Chapter 1

Structural Health Monitoring—An Introduction and Definitions

Christian Boller

Saarland University & Fraunhofer Institute for Non-Destructive Testing, Saarbrücken, Germany (and formerly of The University of Sheffield, Sheffield, UK)

1 BACKGROUND

Structural health monitoring (SHM) is a combination of words that has emerged around the late 1980s. It is very much associated with what we are accustomed with in the classical medical and health sector, combines it possibly with what we are accustomed to in control (monitoring) and links it to what we are truly considering here: engineering structures. SHM (although not expressed like this at the time) possibly dates back much earlier than the 1980s when the combination of words was created. Indeed, it may date back to the origins of structural engineering.

Engineering structures are designed to be safe. The difficulty one trading in this regard is the desire to construct something for a specific purpose out of a material of which one can never know enough in terms of the material's properties as well as the environment the structure is going to operate in. We are happy with the knowledge we can gain in this regard and all we do not know needs to be covered by a safety factor that we have to assume at best guess. The less we know about the operational conditions of a structure and the performance of materials and structures, the higher the safety factor will have to be. This is the risk and dilemma structural engineering is in.

Engineering structures are designed to withstand loads. These loads can be mechanical loads of a static and/or dynamic nature. Loads can, however, also be of an environmental nature such as temperature, humidity or chemical, and again the structure can be exposed to these loads in either a static or even very short term and thus dynamic condition such as a thermo shock. Knowledge of loads applied to a structure has to come from experience. This experience has been either gathered on similar structures in the past or from assumptions. The safest

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

way to design a structure is to design it against an ultimate design limit load, which is the maximum load ever experienced with such a structure added by a safety margin. Designing a structure against this load, however, makes the structure heavy. Often, the maximum load of a structure may just occur once in the structure's life, if ever at all. In that case one may start to question the extreme safety built in, specifically if the maximum load applied would not result in any observable damage.

Loads applied to a structure are the reason for structural deterioration and hence resulting damage. This damage may be generated at a microscopic level and may gradually progress until it becomes observable and critical. Trading with this observability and criticality is the art of damage tolerant design, which has allowed structures to become lighter weight. The way the damage accumulates is of a fairly random nature, and scatter of a factor of 2 in operational life is absolutely normal. This requires a careful means and procedure of inspection at well-defined intervals.

The booming development of sensing technology in terms of sensors decreasing in size and cost, and the combination with microprocessors with increasing power and enhanced materials design and manufacturing in terms of functional materials or even electronic textiles have opened avenues in merging structural design and maintenance with those advanced sensing, signal processing, and materials manufacturing technologies. Taking advantage of this lateral integration is what SHM is about. The central set of questions in this field is therefore

Without compromising safety, could we make our structures

- better available
- lighter weight
- more cost efficient and
- more reliable

by making sensors (and possibly also actuators) to become an integral part of the structure?

What about

- looking at advanced cheap sensors, which are continuously becoming
  - smaller
  - lighter and
  - cheaper?
- making the sensors an integral part of the structural component?
- combining the sensors through
  - advanced microelectronics and possibly
  - wireless technology with
  - advanced microprocessors and
  - advanced signal processing?

If one would try to give an answer to all these, the answer could somehow result in SHM, and a definition for SHM could possibly be the following.

SHM is the integration of sensing and possibly also actuation devices to allow the loading and damaging conditions of a structure to be recorded, analyzed, localized, and predicted in a way that nondestructive testing (NDT) becomes an integral part of the structure and a material.

As a consequence, SHM requires to look at loads as well as damage monitoring with respect to their sensing and assessment algorithms and needs to get those merged in a holistic process such that the health of a structure can be accompanied during the complete life cycle process of the structure considered. How this has emerged and could be further achieved is described as the holistic process in the following articles with further details to be found throughout the wide range of articles being provided throughout this encyclopedia.

2 LOADS MONITORING

Design of engineering structures is at least based on static strength. This is the way structures have been designed since the past mainly by applying a test load that exceeds the maximum operational load by a safety factor to be defined or by designing the structure against the expected maximum load times the safety factor, or possibly both. The problem of fatigue in structures became apparent with the upcoming railway industry in the second half of the nineteenth century. August Wöhler [1] was possibly the first to determine that components—and in his case railway axles—would fracture at loads much lower than the ultimate tensile load if the components were exposed to a repetitive loading. Wöhler furthermore determined that the number of cycles to fracture is related to the level of the repetitive load being applied. This resulted in the effect of materials fatigue to be established and has been mainly described in the form of a fatigue–life curve also often denominated as S–N or Wöhler curve. S–N curves were

and are therefore used to determine the allowable loads of railway axles compared to static loads, which were now named fatigue loads. This principle is well applicable because fatigue loads on railway axles are fairly constant over the life cycle and hence can be considered constant amplitude. Wöhler's principle was shortly expanded to other railway applications such as frames or boogies of railway engines and carriages as well as even railway steel bridges.

Constant amplitude loading is not the way structures are conventionally loaded while operating in service. Most of them are more loaded in accordance to a randomized nature where small loads follow a high load or vice versa such as the different time domain signals shown in Figure 1. Ernst Gaßner [2] was possibly the first to recognize in the 1930s that service loading of structures had to be treated different from constant amplitude loading or in other words constant amplitude loading to be a specific condition within the frame of in-service loading. Fatigue–life curves for in-service loading were therefore defined in terms of the maximum load applied in the in-service spectrum versus the number of loading cycles sustained until fracture. Palmgren [3] from Sweden published his damage accumulation rule (published later in English by Miner [4]) seeming initially trivial for constant amplitude loading, which however became not only more attractive to be applied for randomized in-service loading but also an issue for wide scientific discussion until even today.

Parallel to fatigue analysis, the other wide area of fracture mechanics was initiated during the early twentieth century, mainly driven by people such as Griffith [5] initially. Fracture mechanics was further developed since that time and specifically during WW2, which led to an enhanced understanding of the fatigue process. Furthermore, the combination of fatigue loading and fracture mechanics allowed the new principle of damage tolerant design to be established. Damage tolerant design as opposed to safe life design, which is shown in Figure 2, is a principle that allows damages such as cracks to be present in a structure as long as the overall integrity of the structure is not compromised. This is achieved by either monitoring slow crack growth through well-defined inspection intervals or building in load redundancy in a way such that when a component is due to fracture another component is still able to take over the load without compromising the overall structure itself. Considering damage tolerance in structures has allowed those structures to become much more lighter weight when compared to the traditional safe life design. Damage tolerant design has therefore become mainly the standard for designing civil aviation metallic structures nowadays, with military and even jet engine structures to gradually move towards those design principles as well. However, there are also other areas that have taken advantage of the damage tolerance principle such as marine and offshore structures.

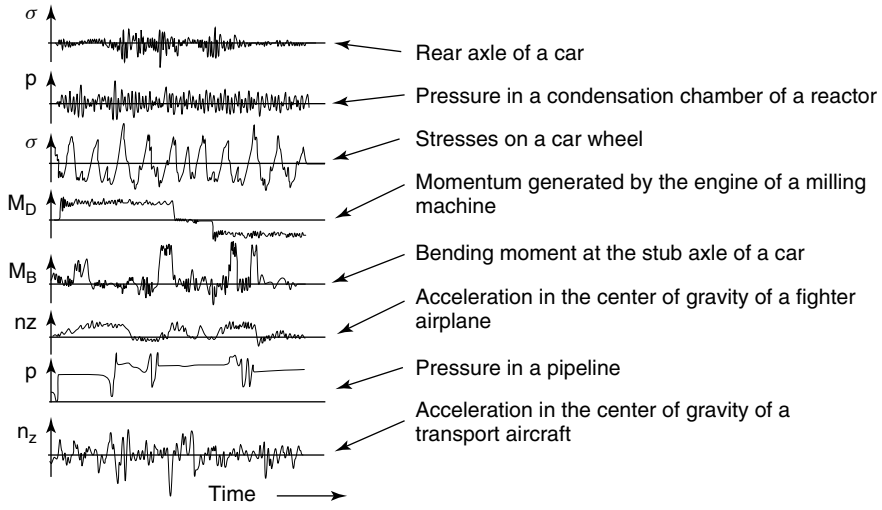


Figure 1. Time domain signals for different in-service random loads.

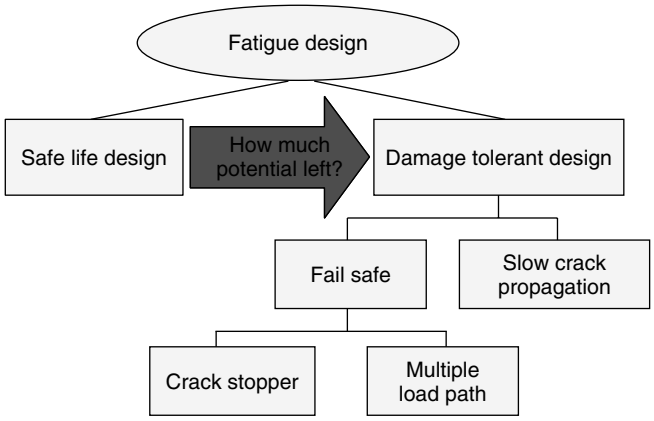


Figure 2. Principles in engineering structural design.

Aircraft structures are hence designed to be either safe life or damage tolerant for a service loading spectrum to be defined during the design of the aircraft. Safe life means that the structure is not allowed to fracture for the loading spectrum defined and up to a given service life. Irrespective of its condition, the component will have to be replaced by a new component after its service life achievement. Damage tolerance is however the alternative to safe life design where a crack is allowed to grow in the structure over a significantly long period of time such that it can be detected during that period safely and does not lead to a critical hazard.

The first aircraft to possibly ever seriously consider damage tolerant design was the De Havilland Comet designed and built in the early 1950s, which was first flown in 1952. This aircraft was possibly the ever most challenging aircraft built in civil aviation history. The design did not just include damage tolerance but also a pressurized fuselage as well as jet engines integrated into the wing. This extreme innovation push however led to some serious crashes in 1953/1954, which resulted from a crack having generated from one of the corner of the windows where the crack had initiated due to underestimation of the window corners' notch combined with the continuous pressurization cycles during each flight. The consequences of this accident were immediate and led to

- the shape of the windows being near to rectangular at the time to be changed to the more ovoid shape we have nowadays;
- the major airframe fatigue test (MAFT) to be introduced where a full-scale aircraft structure is tested under in-service loading conditions on the ground and this being ahead of the so-called fleet leader (the aircraft in a fleet having the most flight cycles accumulated) in terms of flight cycles; and
- operational loads monitoring devices to be considered for aircraft in general.

While the former two consequences can be attributed to structural design, the latter can be clearly attributed to SHM. Although nobody was using the expression of SHM at the time, it was clearly shown that a more precise knowledge of operational loads would be highly essential. This led to the UK Royal Air Force to design a mechanical device that was integrated around the center of gravity of fighter airplanes in the late 1950s and was based on a set of accelerometers being set at different acceleration thresholds (Figure 3). Whenever the aircraft would exceed one of the thresholds, the counter would count those as a measure of exceedances. Hence the numbers being recorded by the different accelerometers would represent the actual load spectrum of the aircraft at the time of recording. The following formula for calculating a flight (or better fatigue) index was introduced

$$\begin{aligned}\text{Flight index } (FI) = K_2 \times S_1 \times (&2.31A + 0.03B \\ &+ 0.001C + 0.001D + 0.28E \\ &+ 3.43F + 10.36G + 18.63H \\ &+ 1.16WL) \end{aligned} \quad (1)$$

Figure 3. Acceleration exceedance monitoring system introduced in the Royal Air Force in the 1950s.

where K_2, S_1, and WL denote the mission coefficient, stores configuration coefficient and landing coefficient, respectively, while A to H represents the readings from the different accelerometers.

Gaßner, who had introduced the eight-step block program sequence as a first step in in-service load sequence analysis [6], continued to determine in-service load sequences for automobiles where a load cycle counting system was developed as shown in Figure 4. Tests performed with this system on various types of roads resulted in a variety of load spectra as shown in Figure 5. This further led to the development of a variety of standard load sequences such as the Gauß and linear distribution spectrum for random loading [7].

Further development was ongoing into a variety of random load sequences such as for transport aircraft (Transport WIng STandard (TWIST)) [9], fighter aircraft (FALSTAFF) [10], helicopters (HELIX FELIX) [11], wind energy structures (WISPER) [12], and possibly much more. Figure 6 shows the TWIST random load sequence for the wing attachment of a transport aircraft as an example. It can be seen that this load sequence is not just composed of a randomization of loads but also of a randomization of different flight types.

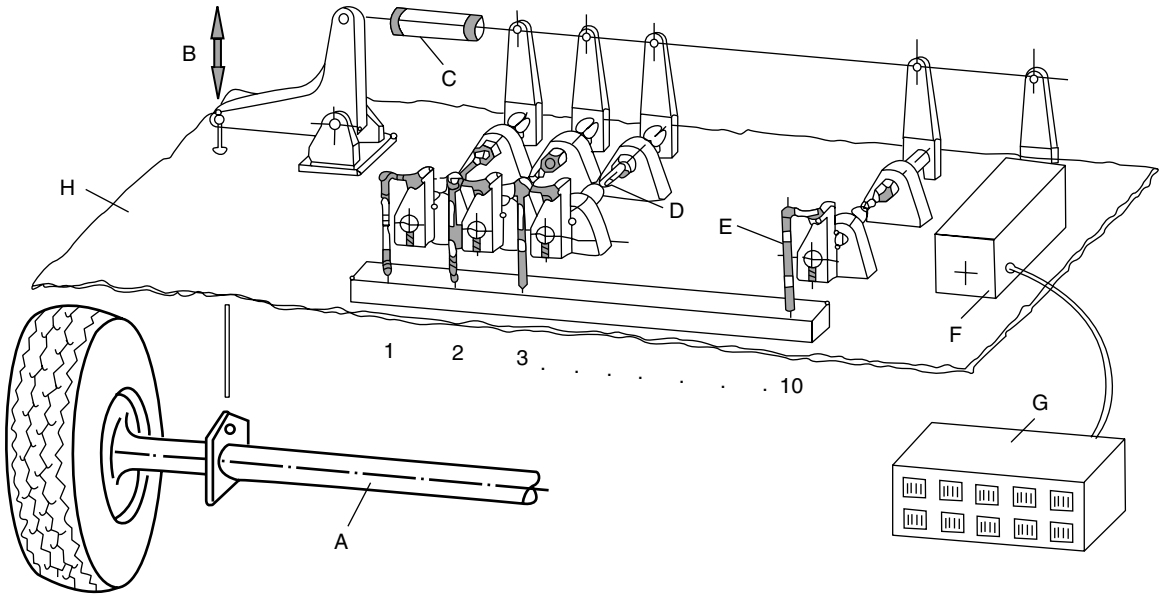


Figure 4. Gaßner's loads monitoring system used in the late 1950s and early 1960s for monitoring road transport load sequences [8].

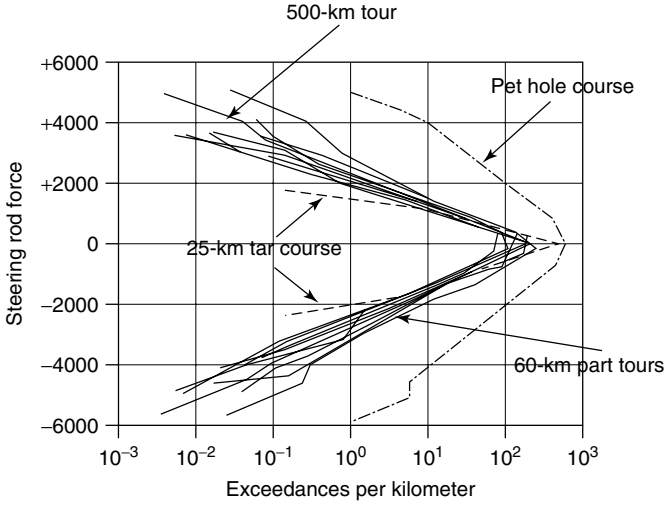
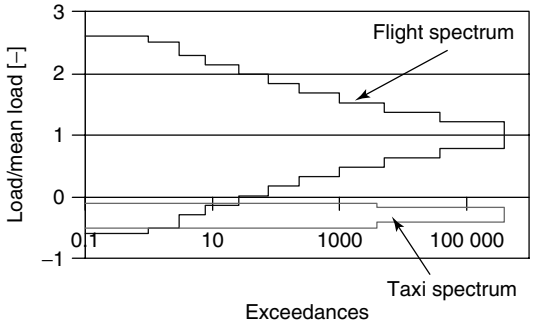


Figure 5. Different road transport load sequences for different road conditions [7].



Flight type	Number of flights	Cycles per step										Cycles per flight
		I	II	III	IV	V	VI	VII	VIII	IX	X	
A	1	1	1	1	4	8	18	64	112	391	900	1500
B	1		1	1	2	5	11	39	76	366	899	1400
C	3			1	1	2	7	22	61	277	879	1250
D	9				1	1	2	14	44	208	680	950
E	24					1	1	6	24	165	603	800
F	60						1	3	19	115	512	650
G	181							1	7	70	412	490
H	420								1	16	233	250
I	1090									1	69	70
J	2211										25	25
Total cycles in all exceedances		1 1	2 3	5 8	18 26	52 78	152 230	800 1030	4170 5200	34 800 40 000	358 665 398 665	

Figure 6. TWIST random load spectrum [9].

What can be concluded from that development in the 1930s to 1980s is that the need for better understanding the true operational loads has been essential. Besides developing standard load sequences, this has been gradually achieved through implementing load monitoring sensors into structures. Aviation has been again a driver in getting those loads monitoring systems integrated. Besides the Royal Air Force (RAF) accelerometer system mentioned above, the Panavia Tornado fighter airplane was possibly one of the first ones getting an Operational Loads Monitoring System called *OLMOS* onboard [13] (*see also*

Usage Management of Military Aircraft Structures). To minimize the number of additional sensors to be implemented, this system was mainly based on monitoring flight parameters only and constructing the operational loads sequence from these with a few strain gauges to be implemented for verification. The system allows overloads to be well detected as well as the usage of the different aircraft in a fleet in general. However, it lacks in allowing the load sequence detected to be used for any further more detailed fatigue analysis. A lot of discussion centered around the question if either flight parameters or strain gauge based monitoring would be the better solution for operational loads monitoring. The Canadian Air Force introduced a strain gauge based system in some of their CF-18 [14], which was then explored by others as well. This discussion culminated in the operational loads monitoring system used on the Eurofighter Typhoon where a flight parameter as well as a strain gauge based operational loads monitoring system has been implemented [15]. This is currently possibly the most advanced operational loads monitoring system being available on an aircraft in general. Similar activities were ongoing on the civil aviation side with an Airbus A320 [16], which never matured due to lack of clarity regarding the operational benefits to be achieved. However, with aircraft nowadays increasingly flying into smaller airports with partially severe weather conditions and possibly associated with fairly short runways and full payload has led to an enhanced probability of hard landings, which resulted in the development of hard landing monitoring devices as described in **Video Landing Parameter Surveys**; **Landing Gear**.

What can be summarized from all this is that all aircraft structures are designed such that they are able to withstand damage resulting from fatigue and static loads as long as the loads do not exceed the design loads. This has been and is validated by numeric analysis and possibly also by destructive materials and component testing, the latter even exceeding to a full-scale fatigue test. As for mechanical loading, a similar design philosophy also applies with regard to impact or environmental damage (i.e., corrosion). In this case, the operational loads would have to be determined in terms of distributions of possible impact loads or sequences of temperature or humidity at different locations around the structure, which is indeed done in the context of monitoring civil engineering structures (*see* **The Influence of Environmental Factors**), electronic systems (*see* **Health Monitoring, Diagnostics, and Prognostics of Avionic Systems**), or mechanical systems such as jet engines (*see* **Monitoring of Aircraft Engines**).

The big advantage of monitoring mechanical loads or even loads in general is not just limited to detecting loads exceeding the design limits but also to feed the information recorded back into the different models of structural assessment being used. Since most of the recent structures are designed on the basis of a digital model using Computer Aided Design (CAD), Finite Element Method (FEM), and Computer Aided Manufacturing (CAM) or Computer Aided Engineering (CAE) in general, where a wide gallery of software tools is provided, it becomes more and more easy to simulate the behavior of structures mainly online. This simulation allows to predict where damage might be most likely to accumulate and finally to occur in form of a crack, corrosion, a delamination or any other sort of deterioration. This is therefore the transition from loads to damage or from new to aged as one may express this in structural terms. A new structure is by definition designed to be damage free as long as no unpredictable accident occurs and will have to face damage once the structure ages. Since structures are designed for a finite life damage is likely to occur beyond that life, which is a consequence of the randomness of damage initiation and the resulting scatter, and can lead to damage to occur twice as late than initially predicted. This is the reason why a large number of structures is still used far beyond their design life.

A good example with regard to aging structures is aircraft. Aircraft are normally considered as aging when they are 15 years old or more. Damage resulting from aging of an aircraft does not necessarily have to result at the locations where this has been expected to happen, which is a consequence of a certain randomness of the damaging process and the scatter in materials' properties. This randomness is also the reason why structural inspections are scheduled at least half of an expected residual life or even less.

Many of the aircraft designed and built in the past or today are conceived to be still used for many decades in the future. Although their design life is definite, they are often used much longer and possibly even more extensive than this was initially planned. Figure 7 gives an overview of some

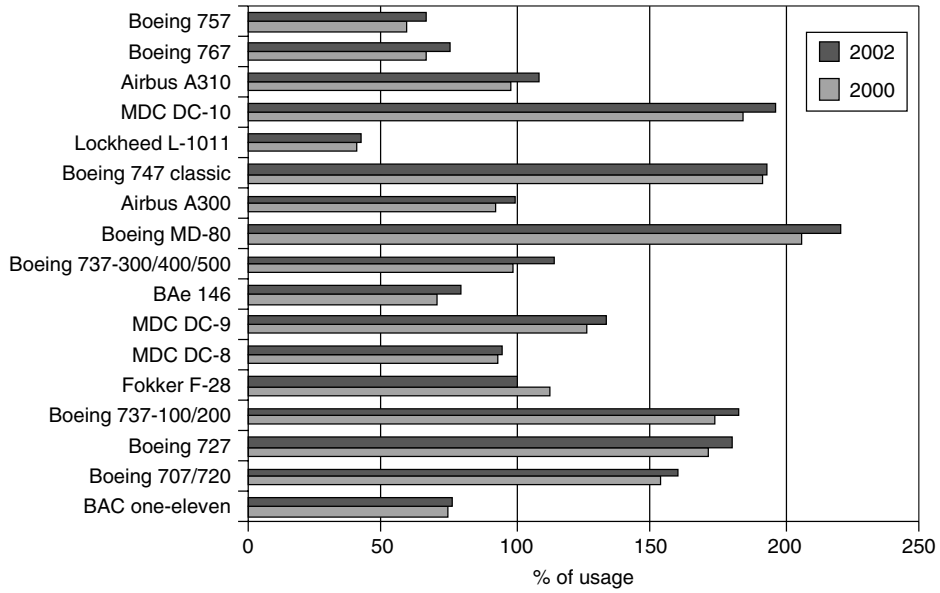


Figure 7. Change of usage of fleet leaders between 2000 and 2002 for different western built aircraft.

typical aircraft types built in the western world and the usage of their fleet leader (the aircraft having been flown most within the fleet) recorded in two different years. It can be observed that for many of the aircraft types the original design life has been exceeded and that there seems to be a trend in continuously expanding the usage of these aircraft in case the demand for aircraft and, thus, air transportation increases. Another extreme example is the US Air Force B-52 bomber, which was designed and built just after WW2 and of which the US Air Force does not plan to take out of service before the year 2045. A set of further examples from the defense environment can be found in Figure 8.

Aircraft operational lives of around a century can hardly avoid a structure to be deteriorated by environmental hazards or threats, nor to be used in a way different from the one it was originally designed for. Examples include conversions from passenger to cargo aircraft or different weapon loads due to different flight missions of a fighter airplane. In any of the cases, additional information is required that will allow the operational conditions such as the loads that the structure has been exposed to and the resulting damage to be known. Since the structure may have been designed quite far in the past where design tools were significantly different from those used today, a conversion into design tools of the twenty-first century may be required as well, such that the structure's integrity can be managed. Looking therefore at the management of the structural integrity of old engineering structures in terms of keeping them "ageless" poses a variety of questions as follows:

- How to handle structures designed in the past in terms of digital modeling?
- What are the load spectra the structure is currently exposed to?
- How to know the degree of damage of an aging structure?
- Where is damage likely to occur on the structure?
- When and where will it be useful to implement sensors on the structure?

An answer to these different questions is given in a first approach below and in further details within various articles of this encyclopedia.

2.1 Digital modeling

In the light of a younger generation taking over in designing engineering structures, the availability of CAD and finite element (FE) models becomes a must. This infrastructure does not exist per se for aircraft, trains, ships, automobiles, buildings, or

Aircraft type	First service flight	Total in service	Average age
Boeing F/A-18 A/B/C/D hornet/super hornet	1978/1983/1987	1203	13.6
Boeing B-52 A/B/H stratofortress	1952/1955/1962	94	46.6
Boeing KC-135 E/R/T stratotanker	1956	568	49/47 (E/R)
Dassault mirage F1 /C/CR/CT/D/E	1973–1992	309	
Dassault/Dornier alpha jet	1978	348	
Fairchild A-10 A/C thunderbolt	1997/2007	364	
Lockheed-martin C-130 A/B/E hercules	1956/1959/1962	1338	44 (E)
Lockheed-martin C-5 A/B galaxy	1969/1980	111	37/20 (A/B)
Lockheed-martin F-16 A/C/D falcon	1979/1981/1989/1994	2982	16.7
Lockheed-martin P-3 orion	1959	368	28.5 (US Navy)
Lockheed-martin U-2 dragonlady	1956	33	
McDonnel douglas F-15 eagle	1974	968	25.5
McDonnel douglas F-4 phantom	1958	725	
MiG-21 (many variants)	1958 – (chinese version)	1528	
MiG-29 (many variants)	1985 –	1047	
Northrop T-38 /A/C	1959/1961	636	
Panavia tornado IDS	1974 –	505	

Figure 8. Aging military aircraft overview.

bridges having been designed 20 years or more in the past. This infrastructure therefore needs to be generated. Means in that regard have been provided in doing such a conversion from solid-state hardware to a digital model by using a laser scanning device. A laser-based projector produces a light stripe on an object where the light stripe is then recorded by a camera and converted into a digital point cloud that can be recorded as an initial digital information. This point cloud is then further converted into a polygon mesh, from there into Non-Uniform Rational B-Spline (NURBS) patches and finally into the desired CAD model. Scanning can be either done manually or automatically by using a robot arm. The digital model up to a coarse-meshed FE model is thus obtained.

2.2 Operational loads

Monitoring operational loads can become a highly complex subject. It results from the multitude of different loads that can apply and the reactions they can have with respect to their variation in height and occurrence. An additional problem in simulation exists with respect to the correct simulation of the reaction forces, i.e., can the structure be rigidly clamped or will a spring and possibly even damper reaction be more appropriate? To find this out requires an iteration process between loads simulation and experimental verification to be initiated. Analysis of the loading influence on the structure can now be done for each of the loads individually with the resulting stress distribution for the load case being obtained from superposition and the resulting figure showing the distribution of stresses and strains. The stress and strain distribution calculated will allow in a first instance to identify the locations where sensors may be placed best to identify loads applied to the structure. Since loading of a structure is however often not solely composed of a single set of loads only but is rather the result of an accumulation of different loads, the analysis done for a single load needs to be performed for all the loads applied on the structure followed by a superposition of the stress analysis results of all these single loading cases, which finally results in the complete picture of the stress distribution for one of the structure's specific loading combinations. This procedure demonstrates the complexity of the loads monitoring and simulation problem and requires a clear strategic approach to be determined when considering the application of loads monitoring for structures of higher complexity such as aircraft.

To experimentally verify the stress distribution simulated requires a loads monitoring system that can be equipped to the structure considered. The

Figure 9. 4-channel loads monitoring system MATCH-II-4 by SWIFT GmbH based on electrical strain gauge monitoring.

system must be relatively small, light weight, robust, and service proven. Examples of such systems being commercially available could be the MATCH-II-4 box from SWIFT GmbH in Germany [17], which is a 4-channel digital data recorder shown in Figure 9. It is designed as a small solid box of 400 g in mass that can be attached to a structure easily and contains a power supply, a CPU with external digital I/O, and a communication interface. The CPU has a 16-bit microprocessor and a Flash-ROM 256 kB memory, which can be reprogramed and stores data up to 864 kB. The I/O can communicate with a standard PC via a RS232C interface. Different types of these boxes can be linked to a network that allows for a multichannel recording of structures of higher complexity. A remote link is further provided as an option. There are larger systems of this kind with airworthiness certification made available for helicopter monitoring where further details can be found in [17].

Different options of strain measurement exist where electrical strain gauges are devices that are very well established for loads measurement. However, each of the sensors has to be connected by individual wires, which make the situation highly complex in case a variety of sensors is involved. A much more elegant and thus smarter solution would exist with the application of optical fiber Bragg grating sensors where virtually tens and hundreds of sensors could be attached to in accordance to the locations worth to be monitored on the structure considered. All sensors could be virtually connected with one single fiber and could be recorded individually by one single piece of equipment. However, such a type of loads monitoring equipment of adequate size and of required cost currently does not exist but may be certainly considerable for larger structures with a high asset value such as aircraft where either cost and/or size of the SHM system may not have such a dominant impact.

2.3 Damage accumulation

The ability to monitor loads also allows load sequences to be monitored, which is the sequence of loads over time and is mainly characterized by peaks and troughs. As shown in Figure 10, each loading cycle can be characterized in terms of load amplitude and mean load in accordance to the rainflow cycle counting method, and the number of cycles with a specific mean load versus load amplitude characteristic is stored in a rainflow matrix. A rainflow matrix can therefore be used to characterize a specific maneuver or mission such as the flight mission of a specific aircraft. The rainflow matrix shown in the lower left part of Figure 10 is an output from the software provided with the SWIFT MATCH-II-4 box.

The availability of an FE loads model and a load spectrum in terms of a rainflow matrix now allows a fatigue–life evaluation analysis to be performed. Knowing the stress distribution through the FE analysis enables fatigue damage accumulation to be calculated at virtually any location of the structure considered. FE fatigue from nCode [18] is a tool that reads the model and results from an FE code, and writes the fatigue results to a postprocessor. Measured or synthesized load data from a rainflow matrix is imported. A database of around 200 standard material fatigue properties or the provision of own fatigue data allows a fatigue analysis to be performed for each loading cycle of the spectrum and for each single element of the structure. Either stress–life or strain–life methods can be used. Damage for each loading cycle and element is accumulated over the load spectrum and finally leads to a picture shown on the lower right-hand side of Figure 10.

The damage distribution figure can so far only be considered as a qualitative one should the component to be analyzed be an aging component where only limited knowledge of the past would be available. The information being available includes the material type, the material properties, the current load distribution, and the current load spectrum. From this knowledge, an experimental quantification could now

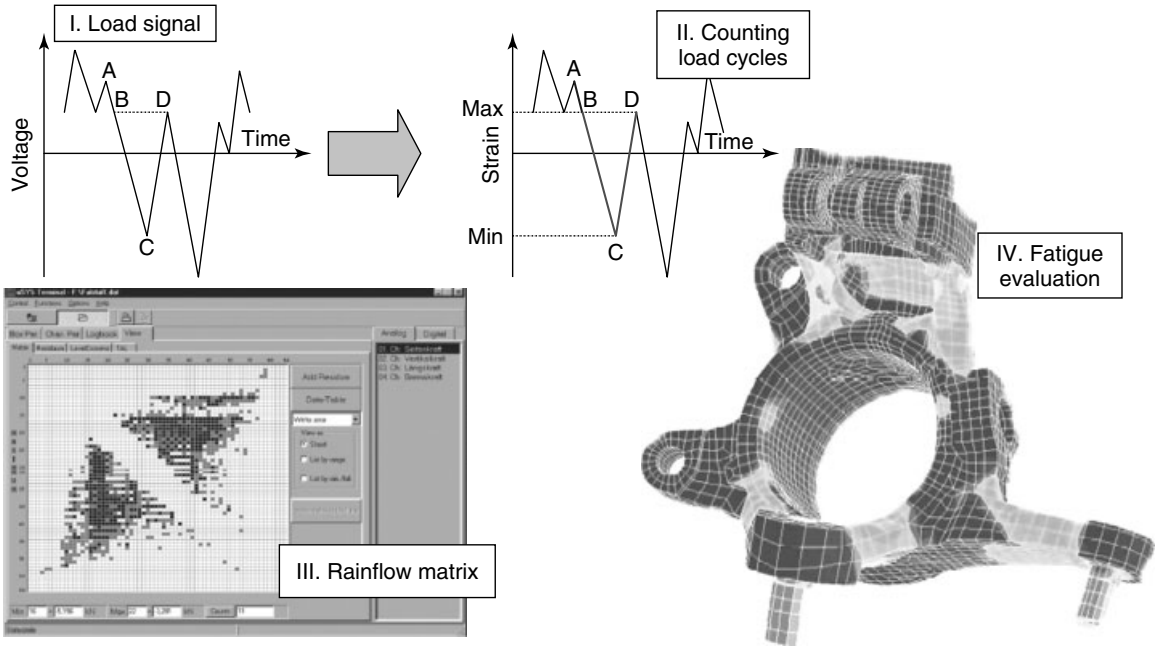


Figure 10. Damage accumulation calculation based on loads monitoring and an FE model.

be made in accordance to the procedure schematically sketched in Figure 11 by cutting off material from the component considered and exposing this material to a fatigue test. The fatigue–life obtained from these tests is the residual fatigue–life of the component at the location where the material was taken from. The amount of damage could therefore be determined as the ratio $(N_f - N_r)N_f$, where N_f is the number of cycles to failure for the pristine material that can be obtained from handbooks and N_r is the number of cycles determined in the experiment. To further prove this concept, more material specimens could be taken for fatigue testing from other locations of the component where the amount of damage has been estimated to be higher.

3 DAMAGE MONITORING

All what has been discussed so far could be solved with loads monitoring only. The other major element of SHM is damage monitoring where different other disciplines come into play. One of them is structural dynamics where the area of condition monitoring has

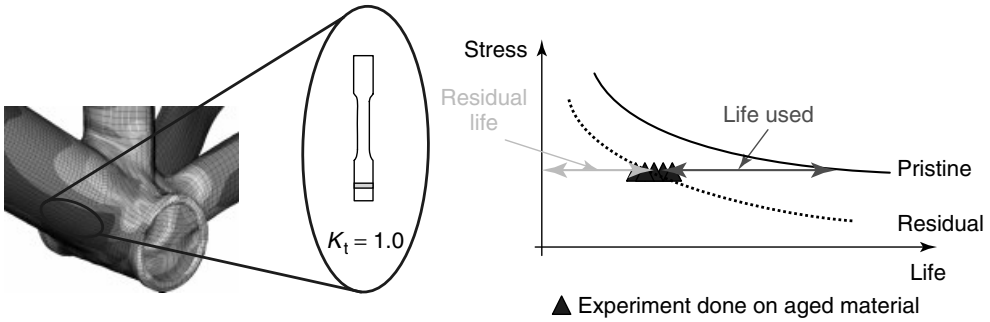


Figure 11. Experimental estimation of aged material's residual life.

a long tradition. Dynamically loaded structures such as rotating machinery are usually loaded at relatively high frequencies with a loading spectrum that can often be determined to be constant amplitude. Hence, if a frequency such as an eigenfrequency (a mode) has turned out to be relevant, it has to be filtered out of a highly randomized and thus possibly noisy signal. The characteristics of these signals may therefore change upon the change of a structure's condition. Hence a crack in a rotating shaft or a gear or a change in a gear's or bearing's surface condition might be easily recognized through the change in the frequency characteristics. Owing to the complexity of the signals, structural dynamics has provided a multitude of signal processing algorithms of which a large number is presented in Section 3 of this encyclopedia.

Besides using vibrations for monitoring damage in materials and structures, there is a multitude of other physical parameters and principles for monitoring damage as well. Those principles include acoustics, temperature, electrical or magnetic fields to just name a few and they are all covered in the wide area of NDT. NDT has provided not only the physical damage monitoring principles discussed in Section 2 but also the variety of sensors to be applied, which are widely described in Section 5. NDT is therefore one of the other major elements contributing to SHM.

Coming back to the structural assessment made so far in terms of loads monitoring and fatigue–life evaluation, the damage accumulation plot shown in the lower right of Figure 10 is an indication regarding where damage accumulates most and is thus most likely to occur. It is related to the loading configuration set and may change if the loading configuration changes as well. However a structure usually follows a specific loading pattern (spectrum) that can be described analytically.

A next question of relevance with respect to damage assessment is related to the incident when damage is likely to occur. As mentioned before fatigue–life can easily vary by a factor of 2 or even more. Hence, the incident of crack initiation can only be determined by continuous inspection where NDT and specific sensing technologies come into play. This can be done, in terms of SHM, by integrating sensing elements around the areas of the structure where damage is most likely to occur first.

Figure 12 shows how the principle can be applied by using acousto-ultrasonics as the monitoring technique. In this case a burst acoustic signal is sent under a Hanning window through some piezoelectric element actuators into the structure and the signals

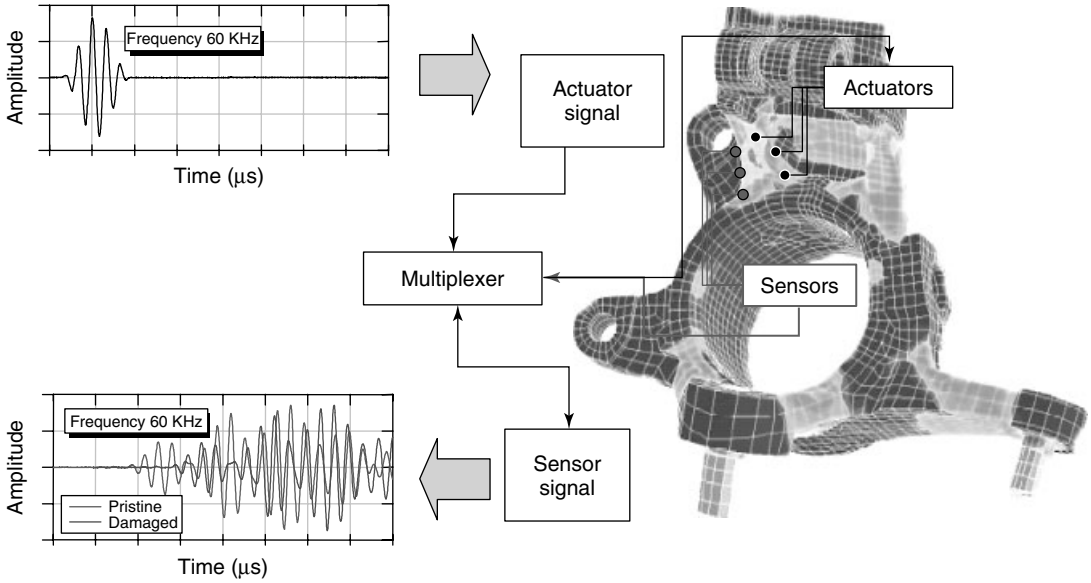


Figure 12. Acousto-ultrasonics damage monitoring principle applied damage critical areas of components.

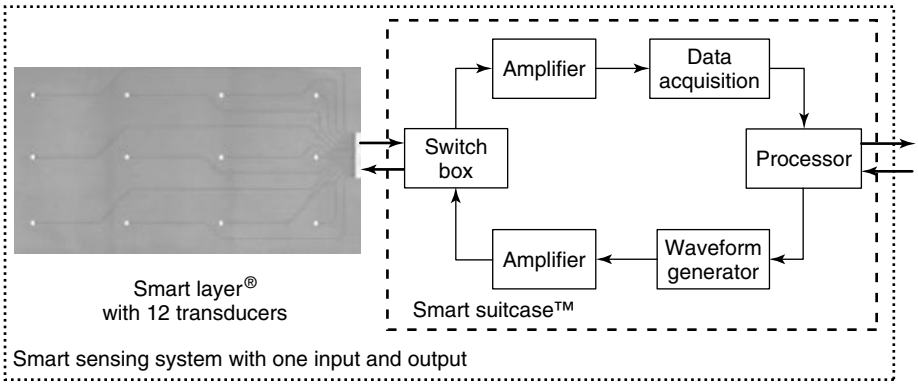


Figure 13. Logic for a sensing system within SHM.

are recorded again by another set of piezoelectric elements acting as sensors. Differences in the signal received will then indicate when damage has been initiated and repair or replacement of the component is required.

4 SENSOR SYSTEM IMPLEMENTATION STRATEGIES

4.1 Sensor or sensor system?

There is a general reluctance in implementing sensors into engineering structures. The reasons for this are understandable since there is no sensor where 100% reliability can be guaranteed for. Hence, the overall reliability of a system may decrease with an increasing number of sensors if the sensors do not provide a significantly additional amount of reliability for the overall engineering structure considered. Consequently, copying the principle of nature such as the human body, where billions of sensors continuously monitor the body itself, may be worth revisiting. As an overriding principle, the number of sensors in an engineering structure is therefore mainly minimized. Minimization of sensors also means that sensors should only be used—and thus possibly also just be implemented—when they are truly required.

However, the number of transducers becomes irrelevant when the complete signal generation and processing unit becomes a part of a sensor system, in practical terms when becoming a sensing system as shown in Figure 13. This system now only has a single interface between the remaining engineering system, such as an aircraft, for which the requirements in terms of input and output as well as reliability can be clearly defined and controlled. Hence, whatever is inside the sensing system is not relevant as long as the system meets the requirements set at the interface to the engineering structure or system. Problems of reliability and redundancy are therefore transferred to the SHM system supplier similar to the approach applied for any avionic or engine system in aviation. Virtually, the biological analogy would be indeed applicable under these conditions, provided the sensing system would meet the requirements accordingly.

Besides the system shown in Figure 13, there is a multitude of other systems being proposed, which are widely described in Section 6. What a holistic SHM system might have idealistically to consist of would be loads monitoring on the one side and damage monitoring on the other. How this might be achievable would require some further strategic considerations, which are subject of the following sections in this article.

4.2 Structural damage classification by MSG-3 process

There is a variety of ways SHM might be implemented in an engineering structure. This very much depends on the way the structure has been designed, the way it can be inspected, and a variety of other

regulations being associated to the structure's operational environment and life cycle. Each area of application has its specifics. One of them to be described here again comes from civil aviation where more examples and specifics are provided in Section 7.

As regards the structural integrity of aircraft structures, a clear route has been identified through the Maintenance Steering Group (MSG), which consists of representatives from the aircraft manufacturers, operators, maintenance, repair and overhaul (MRO) organizations, and the airworthiness authorities. A procedure known as *MSG-3* [18] is now widely applied, whose 'backbone' is shown in Figure 14.

The MSG-3 procedure mainly tries to find out if a component has to be considered as a structure significant item (SSI) or not. If this is the case, then damage can be either introduced to the component through accident, environment, or fatigue. Figure 15 shows the logic of this process combined with the steps where SHM could virtually play a role.

4.3 SHM implementation decision

Accidental damage is something that is unpredictable and can thus occur from day 1 of an aircraft's life. Environmental damage is partially accounted for by design but difficult to handle. Hence environmental damage may be considered from day 1 of an aircraft's life in cases where a large variation in damage initiation can be expected and a corrosion prevention and protection plan may not exist. Fatigue damage is however covered through damage tolerance and is not due to occur before half of the aircraft's design

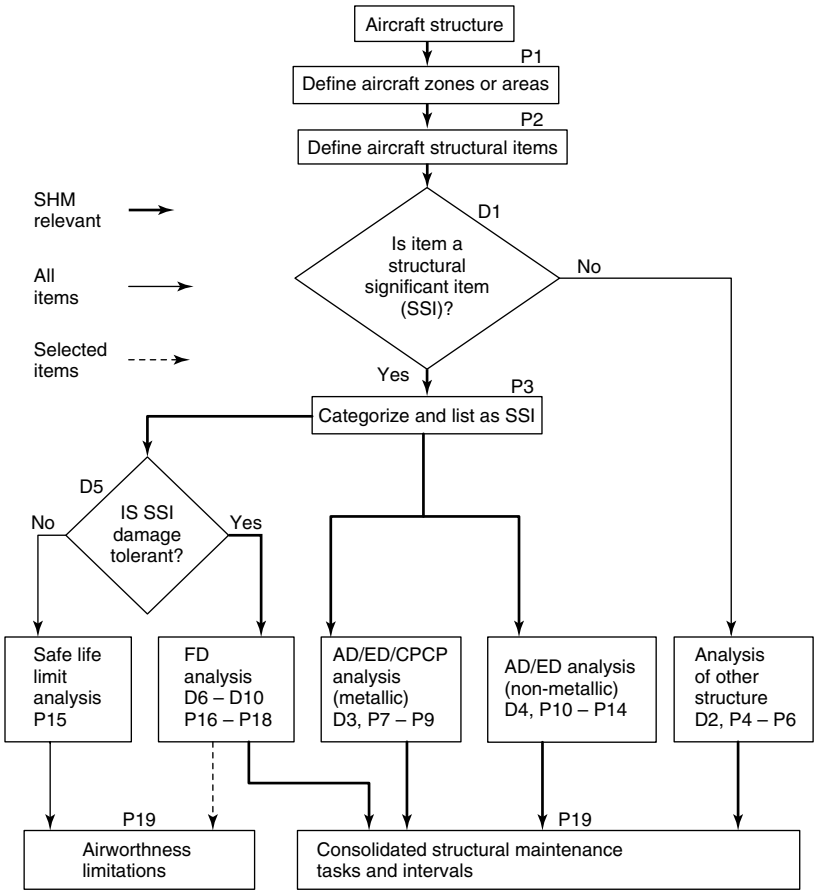


Figure 14. Selection of damage critical aircraft components in accordance to MSG-3 [18].

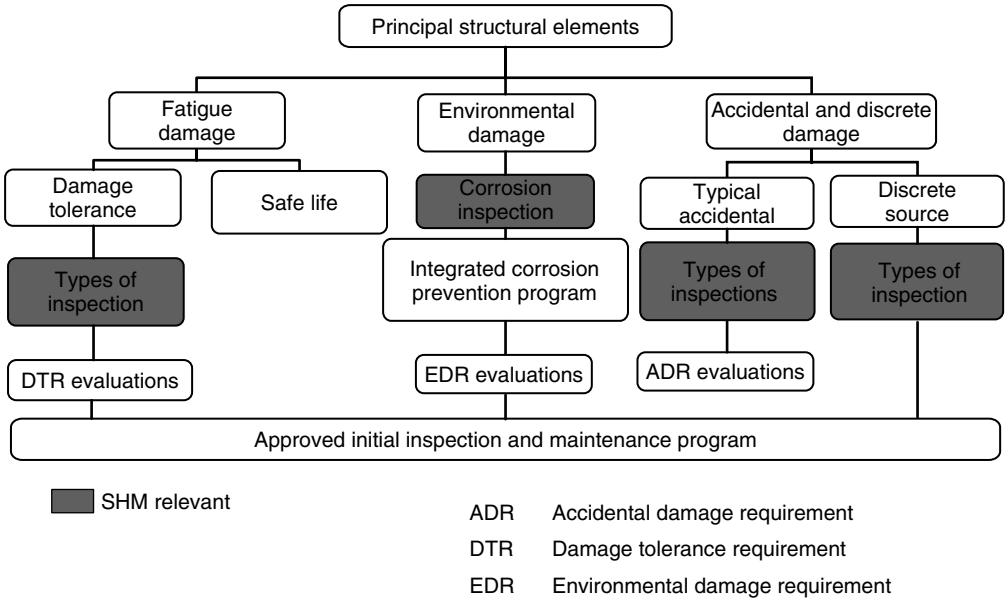


Figure 15. MSG-3 classification of structural damage [18].

life. Fatigue damage monitoring sensors are therefore not due to be implemented before midlife of the aircraft. Consequently, the aircraft manufacturers are following a strategy by looking at sensing accidental and critical environmental damage first which might result in damage monitoring locations as shown in Figure 16. Here, specifically, the locations around passenger and cargo doors are critical where uncontrolled collisions with ground vehicles are likely. Other environmentally critical areas are floor beams and frames close to galleys and lavatories where water spillage combined with corrosion is most likely.

As discussed earlier, loads monitoring using advanced sensing and combining this with a digital model for assessing damage accumulation is the other useful SHM option to be implemented into an aircraft from day 1.

All components being prone to fatigue damage are most likely not to be equipped with SHM systems from the very beginning of an aircraft's operational life since current design rules have been established such that any critical fatigue damage is avoided until the end of the next structural inspection interval defined. Introducing an SHM system would only make sense in case an inspection could be postponed beyond the end of the next inspection interval and this would lead to significant cost savings in terms of the aircraft's life cycle cost (LCC). This however needs to be proven before the SHM system is implemented.

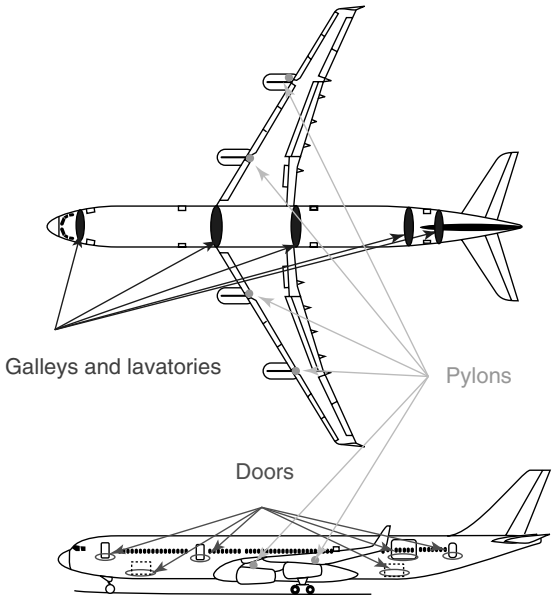


Figure 16. Potential locations for initial damage monitoring on aircraft.

The randomness of fatigue damage otherwise will not make prediction of the damage locations and

specifically their incident easy. Determining these locations and incidents therefore requires a close link to analyzing inspection protocols and here specifically those of the fleet leaders, being the aircraft having accumulated the highest number of flight cycles in a specific fleet. Once these locations and components have been determined, the economic potentials of implementing SHM have to be evaluated for which a procedure is described further below.

5 SHM POTENTIAL DETERMINATION

5.1 Background

A major issue of SHM for aircraft applications is the monitoring of a structure's condition by automation. SHM systems have progressively matured but the question of where to implement those onboard an aircraft to harvest the economic benefits resulting from automated inspection has still to be addressed. Schmidt *et al.* [19] proposed four different areas where SHM could be implemented based on potential fatigue-prone areas over a typical airframe fuselage. Areas identified were stringers, frames, and skins to identify cracks. Schmidt *et al.* also assessed the trade-off relationship that exists between structural weight and frequency of inspections. With increased inspection opportunities gained with SHM, additional damage characteristics could be incorporated into the aircraft design stage, increasing allowable design stresses on components that could result in a reduction in weight. Alternatively, the increased information gained could provide for an extension in the inspection interval, i.e., how frequent the component is inspected.

Schmidt's approach is definitely well applicable with regard to the design of future aircraft. However, with the fleet of existing aircraft, the benefits gained from SHM have to be seen within the frame of automating the inspection process. Inspection automation is, however, only beneficial if it shortens the time of inspection and enhances the availability of the aircraft to be inspected. Hence, what needs to be identified are the structural components along the critical path of the maintenance process and what implications the implementation of SHM would have in shortening the inspection and thus the maintenance process.

A major source along the MSG-3 process where the maintenance process is documented for each aircraft type and where structural inspection tasks are found is the Maintenance Planning Document (MPD). Assigned to each task is a description of the structural location and the nature of inspection (a special detailed inspection (SDI) or a general visual inspection (GVI)), the threshold of the inspection (when the initial inspection could be undertaken) along with the inspection interval (frequency of the inspection), and the estimated time to carry out the task. Optimization and simulation of maintenance processes based on the MPD allowing time critical tasks to be highlighted that may have the potential of being replaced by SHM in terms of maintenance time reduction is therefore what is addressed here. This is demonstrated along with a generic example.

5.2 Structure of maintenance process and problem definition

A maintenance process can be described by depicting the operational life of the aircraft as shown in Figure 17. An aircraft can be in service (white area) or out of service within a maintenance phase (shaded area). The maintenance process can be considered as the entire maintenance undertaken over the aircraft's life and is segmented into maintenance phases, tasks, and jobs, respectively.

Two types of maintenance phases exist: scheduled and unscheduled maintenance. Tasks that belong to scheduled maintenance phases are specified within the MPD and are those to be analyzed. A maintenance task is an action that specifies the action in general providing a description of the area involved, i.e., "inspection of frame 47 for cracks". Maintenance tasks can be segmented further into jobs. Maintenance jobs are more specific actions that make up a maintenance task. However, they have to be undertaken in a particular order. Maintenance jobs include the following: preparation, access gain, removal, inspection, reporting, and reassembly.

With the MPD only specifying the maintenance interval and threshold inspection, commercial airline operators have the freedom to allocate maintenance

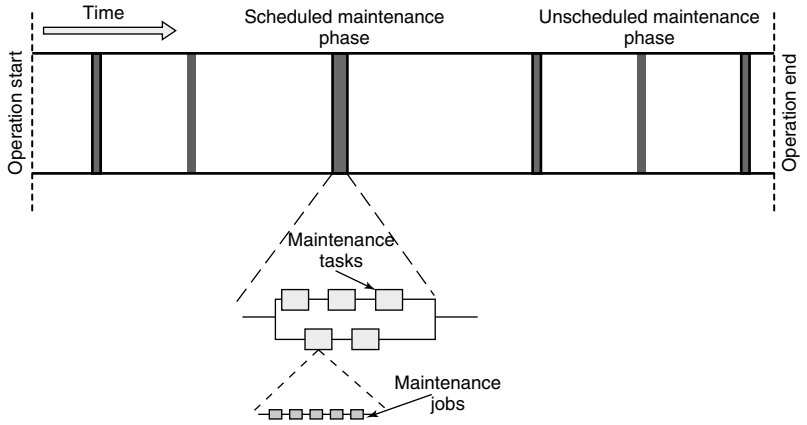


Figure 17. Maintenance process in its entirety and its components.

tasks to a phase in accordance to the allowable maintenance intervals. This results in different inspection tasks in different maintenance phases, which ultimately influences the critical path.

Also with changes to how maintenance is executed (i.e., number of laborers on site and the sequence of tasks executed) changes in the critical path is imminent. Additionally, critical path changes can occur due to changes in times that exist for a maintenance job. In reality, the duration time of a job is even often unlikely to be deterministic. With more complex maintenance phases, there may be a tendency for changes in the critical path with changes in job times. This nondeterministic situation requires optimization procedures that allow for effective maintenance planning as well as reliable identification of structural components being those most likely to positively affect the inspection and maintenance process through integration of SHM. It has to be noted in that regard that any component being identified for SHM may lead to a change in the critical path of the maintenance process, hence also leading to a sequential iteration process. How such an optimization could be achieved is therefore explained in the following paragraphs.

5.3 An integrated methodology for effective maintenance planning

For simulation of the maintenance processes, a commercial software named Arena© has been used [20], which has been successfully applied in post-depot maintenance of helicopters before [21]. Arena is a discrete event process simulation software that has the ability to analyze and predict process performance. A process is modeled in blocks in accordance to a flow chart. Each of the blocks contains the necessary data that describes the discrete event in terms of a process element, a decision, or something else. A process element is analogous to a maintenance action along the process.

Once the process is defined, described, and validated, a simulation run can be initiated. A simulation is based on entities that pass through discrete events (i.e., process blocks, etc.). For the purpose of the study, an entity is allocated as the aircraft to be maintained. The aircraft to which attributes are assigned, virtually moves along the flow chart. The number of entities, i.e., aircraft, can be increased, which allows maintenance in an aircraft hangar to be simulated and thus the maximum of aircraft maintained to be determined under the constraints of the maintenance process. These constraints mainly include the number of workers, tools and maintenance space, and the organization of the maintenance process itself. The constraints such as the number of workers and tools required can be optimized as well. Finally, the process is optimized in its sequence such that a maximum of aircraft is put through the maintenance process at a minimum of time. This simulation will therefore help to determine the maximum of aircraft operability gained from enhancing the aircraft maintenance process just from a logistical perspective. To obtain the complete picture simulation, runs are done over the complete life cycle of an aircraft.

For optimizing a maintenance process in accordance to the objectives set before, the following five steps can be defined:

1. Formation of inspection phases and consolidation of tasks from MPD based on airliner's operation
Following the inspection phases defined by the operator in accordance to the operational conditions, the different maintenance phases mentioned in the MPD are allocated with regard to MSG-3 [18], which is different when compared to the traditional letter checks [22].

2. Optimization (minimization) of maintenance tasks at each inspection phase
To provide a consistent benchmark against SHM and avoid changes in the critical path via different approaches, maintenance tasks in a phase are rearranged in an optimal fashion to a set number of maintenance staff. Scheduling jobs to laborers to minimize maintenance phase times allows time critical tasks to be determined with fixed times (i.e., those that lie on the critical path). The need to schedule inspection tasks to a set number of laborers fits very closely to the flexible job shop problem (FJSP), which is of current interest in the manufacturing sector. A number of approaches have been applied to the FJSP. Gao *et al.* [23] have developed a genetic algorithm (GA) approach that has been used to successfully solve the FJSP in the manufacturing industry, which has been remodified here for optimizing maintenance downtime in a maintenance phase.

3. Identification of critical maintenance tasks
Once each maintenance phase has been optimized in terms of downtime, a critical path analysis is undertaken at each maintenance phase. Uncertainties in duration of each task are addressed through statistical variations in job times and the effect it has on the critical path, using the discrete event simulation option with Arena©.

4. Analysis to suitability of SHM
Highlight tasks being related to the integrity of structural components and that lie on a maintenance phase's critical path, which may be suitable for SHM. Assess if a suitable SHM solution can be achieved for the component.

5. Remodeling of maintenance phases with the inclusion of SHM
Remodel the complete maintenance process with the SHM solutions included. This requires going back to step 2 (or possibly even step 1) and redoing the complete process again. It may lead to new critical paths with new components to be identified where SHM might be applied. This remodeling is repeated until no further potentials can be identified. Comparison of the existing maintenance method to a SHM scenario will finally demonstrate the potential gained.

The method has also been summarized in Figure 18.

5.4 Case study

To demonstrate the method described, assume the maintenance phase to consist of inspection of Gantries (task 1) and Frame 47 (task 2) both depicted in Figure 19. Gantries and Frame 47 are highly loaded structures in an aircraft's center fuselage section that serve as load carriers for fuselage bending and twisting forces mainly, partially superimposed by passenger payload and cabin pressure. Their loading is therefore complex and they are prone to cracking over lifetime. Hence they require inspections at certain time intervals specified in the maintenance document such as an MPD.

To highlight optimization of the maintenance phase, an initial inspection phase interval of 5000 FH was taken, which only consisted of the two maintenance tasks mentioned in more detail below. The objective was to arrange maintenance jobs with a given number of laborers to minimize the makespan of the maintenance phase (the makespan is defined as the difference between end time of the last remaining job and start time of the first job, i.e., the maintenance phase duration). For each task, the three maintenance jobs were removal, inspection, and reinstallation. It was assumed that four labor pairs existed L_1, L_2, L_3, and L_4. L_1, and L_2 were labor pairs that could only undertake access and removal jobs. Labor pairs L_3 and L_4 were only designated to undertake inspections. Figure 20 highlights the problem domain.

Job times (p_{tjl}) related to inspections were gathered from the MPD. Times associated to access were

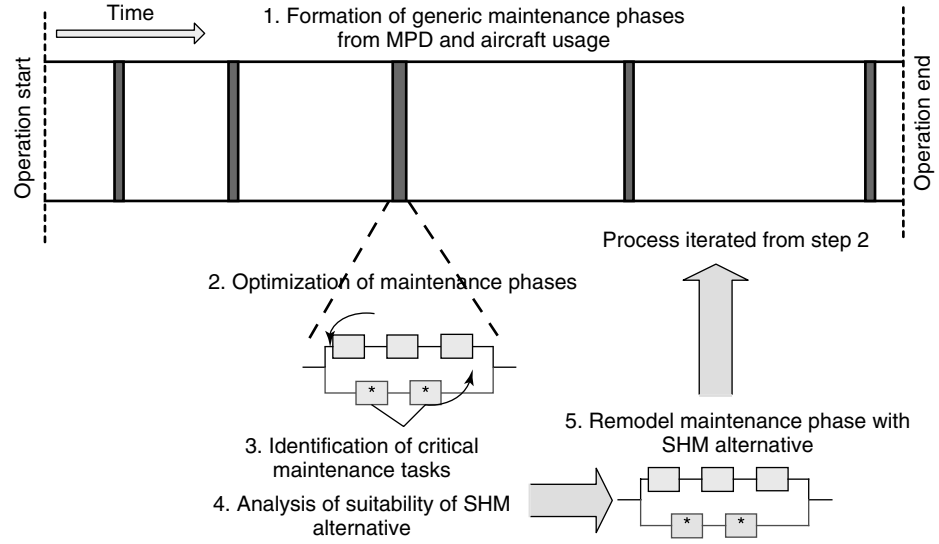


Figure 18. Overview of method for the determination of SHM potentials.

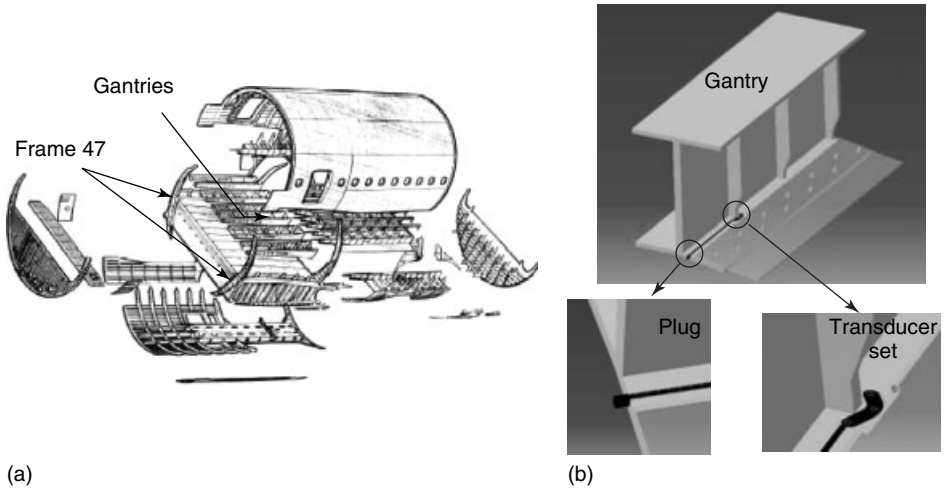


Figure 19. Components considered for case study (a) and proposed SHM solution (b).

assumed based on indications given by maintainers and documentation. Table 1 shows the assumed labor times for each job.

The GA procedure was then used to determine the optimal sequence of maintenance jobs to laborers that provides the minimum makespan, i.e., downtime of the maintenance phase. GAs are an optimization method that are based on an initial random population, which is manipulated through evolutionary mechanisms (selection, crossover, mutation, etc.) to provide better solutions after each generation governed by an objective function. The problem required a suitable permutation representation, which was adopted from Gao *et al.* [23] who were successful in solving similar type problems in manufacturing. The GA developed in Matlab was run and each solution was decoded to a schedule where it was evaluated by the objective function (total duration time). The GA converged to a minimum makespan of 6.3 h for a solution shown in Figure 21, which

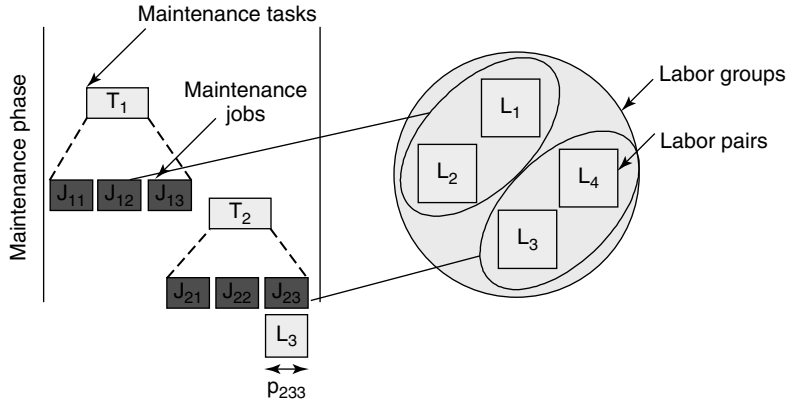


Figure 20. Illustration of the problem domain for 5000 FH maintenance phase.

resulted to around 120 h for the aircraft's complete life cycle.

The critical path could be established from the minimum schedule, which is the one related to task 1. Uncertainties of a standard deviation of 10% of the mean duration times could further be considered in the Arena simulation. However, this did not make a change in the path criticality of the case considered here. The model in Arena© is shown in Figure 22.

A SHM solution was developed as shown in Figure 19(b). Feeding this one back into the optimization process allowed task 1 duration to shrink by roughly a factor of 6 over the life cycle, leading to a saving of around 100 h of aircraft availability and let task 2 now to be on the critical path with around 36 h over the life cycle. Implementing an SHM solution to Frame 47 allowed the total inspection time to be reduced to 22 h, hence enhancement of 100 h in aircraft availability could be made by implementing SHM on both the Gantries and Frame 47.

Table 1. Maintenance tasks in 5000 FH phase

Description	Job	L1	L2	L3	L4
Access task A	J_{11}	2.1	2.1	–	–
Inspection task A	J_{12}	–	–	2.1	2.1
Reinstallation task A	J_{13}	2.1	2.1	–	–
Access task B	J_{21}	0.72	0.72	–	–
Inspection task B	J_{22}	–	–	0.72	0.72
Reinstallation task B	J_{23}	0.72	0.72	–	–

L1, L2: two access staff; L3, L4: two NDT inspectors. All job times specified in unit hours.

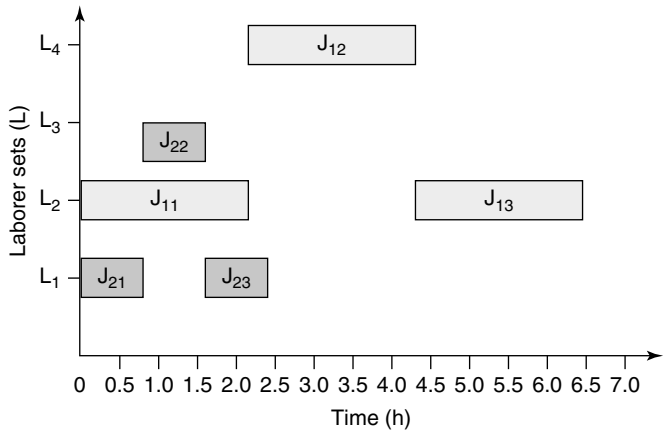


Figure 21. Optimized labor schedule determined by GA.

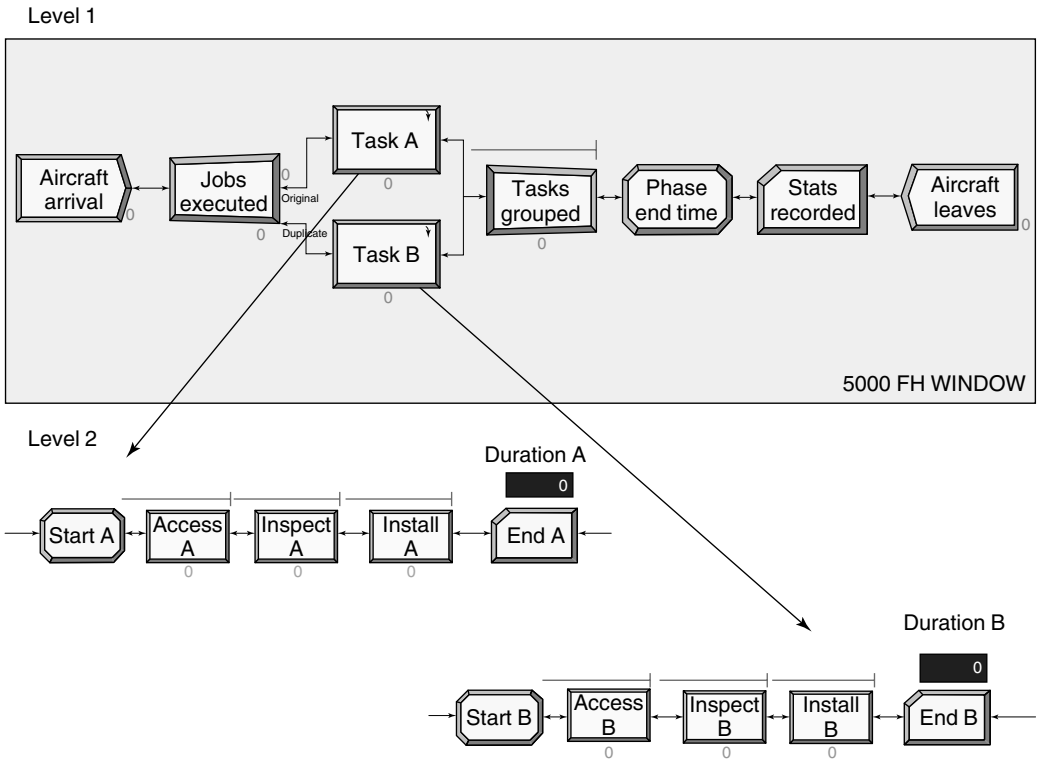


Figure 22. Simulated model of generic 5000 FH maintenance window.

6 CONCLUSIONS

SHM is a topic that has gained a significant high attention in the meantime. The reason is that it integrates a variety of classical disciplines on a lateral basis. This does not only include materials science, fracture, fatigue, strength, NDT, structural dynamics, or design but also include areas such as maintenance logistics, economics, and possibly others. SHM as a result is therefore not just sensors, electronics, or signal processing but rather a system. Such a system must result into something that is shown for a rail axle system in Figure 23. It must consist of a sensor or sensor network, some onboard electronics and a data transmitter possibly in the form of an antenna that does allow to link into the maintenance logistics.

SHM is not just limited to monitoring damage only. It also needs to encompass monitoring of operational loads, which in combination with the various prognostic simulation tools being available today will allow the locations of damage to be predicted in accordance to the operational needs. This will only allow current design and maintenance principles to be included, which is a major necessity should SHM find some promising application. SHM might further change design principles in a way that it could extend the damage tolerance design principle, so far applied with success in civil aviation and also in other areas as well. This however requires SHM systems operating at highest reliability, which is an area where still a lot of effort needs to be invested.

Although the examples being described here have been mainly related to aviation, SHM is by far being limited to this only. The same logic can be principally applied into any other field in engineering such as automotive, naval, railway, civil engineering, the processing, or the heavy machinery industry. Many of those are addressed in the various articles of this encyclopedia.

A big need with SHM is also that whatever will be developed is well integrated into the maintenance concept of the engineering system considered. With

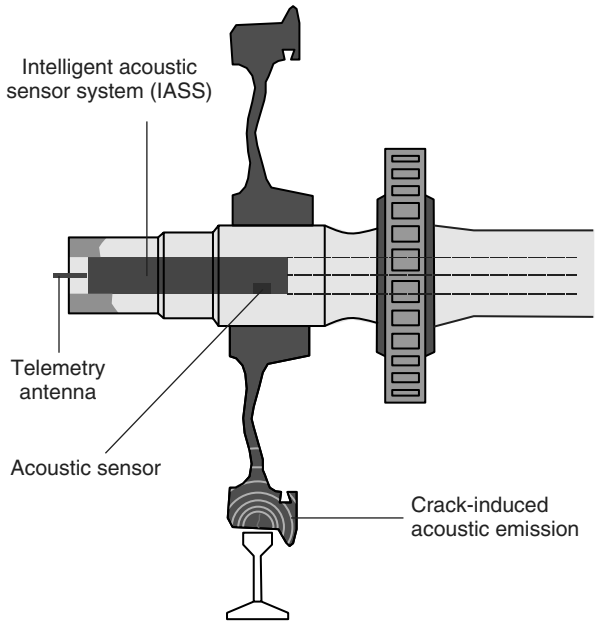


Figure 23. SHM system for train wheel monitoring [24].

regard to aircraft, this requires to follow maintenance principles used for avionics or engines since those are possibly further ahead when compared to structures so far. This also includes the way prognostic maintenance will be done and presented to the operator and maintainer specifically.

A major discussion centers around the question of when SHM should be implemented best along the life cycle of an engineering structure. Implementation at Day 1 makes sense for monitoring operational loads in general and components being prone to accidental damage. Monitoring environmental or fatigue damage from Day 1 will only make sense if an inspection interval can be significantly postponed and this results in a remarkable LCC benefit. These LCC benefits have to be determined in advance before those SHM systems are implemented into an engineering system such as an aircraft. A procedure on how to determine those potentials has been described. For fatigue and environmental damage, monitoring locations have to be determined from ongoing inspection records and here specifically from the fleet leaders of a fleet of engineering structures. Learning from experience will still be the major postulation. The question is only how this experience will be gathered and stored in the future when SHM will be in application.

ACKNOWLEDGMENTS

The author wants to acknowledge the contributions made by Mr. Hrshi Kapoor and Dr. Marilyu Goh to the case study while his stay at the University of Sheffield.

REFERENCES

[1] Wöhler A. Resultate der in der Central-Werkstatt der Niederschlesisch- märkischen Eisenbahn Frankfurt an der Oder angestellte Versuche über die relative Festigkeit von Eisen, Stahl und Kupfer. *Zeitschrift für Bauwesen* 1866 **16**:67–84 (in German).

[2] Gaßner E. *Auswirkung betriebsähnlicher Belastungsfolgen auf die Festigkeit von Flugzeugbauteilen*, Doctoral Thesis, TH Darmstadt and partially in: Jahrbuch der Deutschen Luftfahrtforschung: München /Berlin, 1941; pp. 472–483 (in German).

[3] Palmgren A. Die Lebensdauer von Kugellagern. *VDI-Z* 1924 **58**:339–341 (in German).

[4] Miner MA. Cumulative damage in fatigue. *Journal of Applied Mechanics* 1945 **12**:159–164.

[5] Griffith AA. The phenomenon of rupture and flow in solids. *Philosophical Transactions of the Royal Society, London, Series A*, 1920 **221**:163–197.

[6] Gaßner E. Festigkeitsversuche mit wiederholter Beanspruchung im Flugzeugbau. *Luftwissen* 1939 **6**(2):61–64 (in German).

[7] Svenson O. Beanspruchungskollektiv—Betriebsfestigkeit—Leichtbau. *Leichtbau der Verkehrsfahrzeuge* 1970 **14**(11):178–184 (in German).

[8] Gaßner E. Betriebsfestigkeit gekerbter Stahl- und Aluminiumstäbe unter betriebsähnlichen und betriebsgleichen Belastungsfolgen. *Materialprüfung* 1969 **11**:373–378 (in German).

[9] de Jonge B, Schütz D, Lowak H, Schijve J. *A Standardised Load Sequence for Flight Simulation Tests on Transport Aircraft Wing Structures*, NLR TR 73029 U, 1973. 1973 Also published as LBF Report FB-106.

[10] van Dijk GM, de Jonge JB. *Introduction to a Fighter Aircraft Loading STAndard For Fatigue Evaluations—FALSTAFF*, NLR MP 75017 U, May 1975.

[11] Edwards PR, Darts J (eds). *Standardised Fatigue Loading Sequences for Helicopter rotors—Helix and Felix—Part 1: Background and fatigue evaluation Part 2: Final Definition of Helix and Felix*, NLR TR 84043 U, 1984.

[12] ten Have AA. *WISPER and WISPERX—Final Definition of Two Standardised Fatigue Loading Sequences for Wind Turbine Blades*, NLR TP 91476 U, 1991.

[13] Krauß A. Betriebslastenermittlung für Flugzeugentwurf und—entwicklung. *Proceedings of the 14th Meeting of DVM AK Betriebsfestigkeit*. Rüsselsheim /Germany, 1988 (in German).

[14] Zgela MB, Madley WB. *Durability and Damage Tolerance Testing and Fatigue Life Management: A CF18 Experience*, AGARD CP-506, 1991.

[15] Hunt SG, Hebden IG. Validation of the Eurofighter Typhoon structural health and usage monitoring system. *Smart Materials and Structures* 2001 **10**:497–503.

[16] Ladda V and Meyer H-J. *The Operational Loads Monitoring System OLMS*, NATO AGARD-CP-506, 1991 Paper 15.

[17] www.swift-online.de, 2008.

[18] ATA MSG-3, *Operator/Manufacturer Scheduled Maintenance Development*; Rev. 2005.1; Air Transport Association of America, 2005.

[19] Schmidt, HJ, Telgkamp J, Schmidt-Brandecker B. Application of structural health monitoring to improve the efficiency of an aircraft structure. *2nd European Workshop on SHM*. Munich, Germany, July 2004, pp. 11–18.

[20] Kelton DW, *et al. Simulation with ARENA*, *Second Edition*. McGraw Hill, New York, 2002.

[21] Vigus SE. *A Simulation-based Analysis of the Impact of in-sourcing a Major Process Element on the Coast Guard HH-60J Depot Maintenance Process*, MSc thesis AFIT/GAQ/ENS/03-03, Air Force Institute of Technology, Wright-Patterson AFB, 2003.

[22] Kinnison HA. *Aviation Maintenance Management*, McGraw Hill, 2004.

[23] Gao J, *et al.* A hybrid of genetic algorithm and bottleneck shifting for mutiobjective flexible Job shop scheduling problems. *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*, ACM Publishers, 2007; pp. 1157–1164.

[24] Hentschel D, Frankenstein B, Pridöhl E, Schubert F. Hollow shaft integrated monitoring system for railroad wheels. *10th SPIE Annual International Symposium on Nondestructive Evaluation for Health Monitoring and Diagnostics*: San Diego, CA, 2005; Vol. 5770, pp. 46–55.

FURTHER READING

www.ncode.com, 2008.

Fischer R, Hück M, Köbler H-G, Schütz W. Ein dem stationären Gaußprozess verwandte Beanspruchungs-Zeit-Funktion für Betriebsfestigkeitsversuche. *Fortschritte Ber. VDI-Z Reihe*, 1970; Vol. 5(30) (in German).

Chapter 2
Free and Forced Vibration Models

Muhammad Haroon
Purdue University, West Lafayette, IN, USA

1 INTRODUCTION

Vibrations-based structural health monitoring (SHM) is the process of identifying the health of mechanical systems and structures based on changes in the dynamic behavior caused by damage. The process includes loads identification, damage detection, location, characterization, quantification, and life prediction. Vibrations-based methods are a natural choice because the vibration signature of mechanical systems is a known function of the system's physical parameters. In addition, the data-acquisition process and associated technology are well developed for system identification and relatively easy to apply. The global nature of vibration data also offers advantages in structural damage identification, as demonstrated later in this section. Modal information (natural frequencies and mode shapes), including properties derived from modal information (modal curvature, model strain energy, etc.), has long been used for damage identification. Changes in natural frequencies were among the earliest avenues of investigation and considerable work has been done in this regard; Salawu [1] provides a review of work in this area.

In studying the vibrations-based dynamic behavior of systems, modal properties are the obvious choice for system health analysis because these properties are directly affected by changes in a system's physical parameters (mass, stiffness, and damping) and are straightforward to estimate once models have been developed. Thus, model identification or system identification plays a role in SHM, especially when modal properties are used. A model for the system has to be assumed along with the assumptions about structural behavior, e.g., linear, time invariant, etc. Mechanical damage (e.g., cracks, delamination, and loose bolts) can be related to changes in stiffness and/or damping, which directly affect the system modal properties. Modal parameter–based methods, such as those using natural frequency shifts, are sometimes insensitive to damage and often corrupted by measurement variability, which hinder their general applicability. Many other categories of features can be estimated from vibration data that cover the whole

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

range of possible damage scenarios, operating conditions, structure types, and applications. The subcategories of vibrations-based SHM techniques include those exploiting linear or nonlinear changes and methods utilizing features in the time or frequency domain. The diversity in the available data, information, and features from vibrations-based system interrogation points to the power and versatility of this class of SHM methodologies. Doebling *et al.* [2] and Sohn *et al.* [3] have provided thorough reviews of vibrations-based SHM research.

The aim of vibrations-based SHM is to combine experimental vibration data, e.g., acceleration, with vibration models for damage identification (detection, location, characterization, and quantification) and ultimately develop damage models for prediction. The models can range from simplistic lumped parameter models to complex finite element (FE) or numerical models. Therefore, model development, validation, updating, and uncertainty quantification are important steps in the process.

In the next section, the framework for general vibrations-based SHM is provided followed by a discussion of a number of vibrations-based data interrogation techniques applied to numerical and experimental data. A number of practical examples are presented that step through the four stages of SHM as outlined by Rytter [4]: detection, location, quantification, and damage models for prediction.

2 FRAMEWORK OF VIBRATIONS-BASED SHM (NEWTON'S LAW)

Vibrations-based SHM has its basis in Newton's second law of motion for a constant mass, $\Sigma F = ma$. This simple equation provides the framework for damage identification, including parameters, loads, damage indicators, and various types of vibrations-based SHM techniques. It is used to derive the equations of motion for a lumped parameter system,

$$[M]\underline{\ddot{X}} + [C]\underline{\dot{X}} + [K]\underline{X} + N[\underline{X}(t), \underline{\dot{X}}(t)] = \underline{F} \quad (1)$$

where M is the mass, C is the damping, K is the stiffness, $\underline{\ddot{X}}$, $\underline{\dot{X}}$, $\underline{X}$, and $\underline{F}$ are the acceleration, velocity, displacement, and force vectors respectively, and $N[\underline{x}(t), \underline{\dot{x}}(t)]$ represents nonlinearities in the structural system.

This single equation, or set of equations, provides all of the information for the numerous diverse techniques and approaches for vibrations-based SHM.

Equation (1) contains the kinematic quantities that can be measured in order to interrogate the health of a mechanical system.

Acceleration measurements are the most convenient to make and commercial piezoelectric accelerometers are commonly used to measure dynamic response. Acceleration can be integrated to obtain velocity and displacement, as long as care is exercised in the integration, and then used directly for various SHM techniques [5–7].

The relative motion between structural elements changes with damage and load redistribution accompanying damage and, hence, relative displacements can be monitored to examine the integrity of structures (e.g., this is the primary damage-sensitive feature of interest for buildings subjected to earthquake damage referred to as *interstory drift* [8]). Displacement can be obtained from acceleration measurements via time- and frequency-domain integration but there are drawbacks to this approach [9] including loss of the dc component, which can contain important information for SHM; for example, quadratic nonlinearities cause dc components to appear in the system response. Laser Doppler vibrometry (LDV), a noncontact method, has been used to directly measure displacements [10, 11], including continuously scanning laser Doppler vibrometers (CSLDVs) [12]. Lead zirconate titanate (PZT) transducers are extensively used to measure structural displacement. Martin *et al.* [13] provide a comparison of different types of piezoelectric transducers for measuring structural displacement.

Strain measurements are also often used for SHM, especially in civil applications such as bridge and building health monitoring. The strains can be correlated with deformation and loads on structures in operation [14]. Fiber-optic strain gauges [15, 16], fiber Bragg grating (FBG) sensors [17, 18], and even electromagnetic sensors [19] have been used to monitor strain on structures. Because strain is a localized measurement, a large number of sensors need to be mounted to monitor possible defect locations on structures. Low-cost thin-film piezoelectric

(polyvinyldene fluoride (PVDF)) patches have been proposed to measure strain [20].

External loads need to be measured or estimated for many vibrations-based SHM techniques, for example, frequency response function (FRF)-based damage-identification algorithms [21–23]. In many realistic mechanical systems, inputs are not readily measured, as in automotive road vehicle data, for which the input to the tire patch of the tires from the road is neither measurable nor easy to estimate. For example, consider the equation of motion of a single degree-of-freedom (SDOF) system with base excitation,

$$M\ddot{x} + C\dot{x} + Kx = C\dot{x}_{\mathrm{b}} + Kx_{\mathrm{b}} \tag{2}$$

where x_{b} is the base excitation. In this case, the input is a function of the base motions and system parameters. In such cases, passive methods, which are those relying on output measurements only, need to be utilized for SHM.

The unmeasured external inputs to a system can be modeled, for example, as a stationary stochastic process [24]. Considerable work has been done on indirect estimation of the input forces, including the use of Markov parameters [25, 26], the sum of weighted acceleration technique [27], and the inverse-FRF approach [28, 29]. The inverse-FRF approach suffers from the problem of ill-conditioning. Stites [30] used an overdetermined inverse problem to overcome the ill-conditioning issue for estimating impact loads on mechanical structures. Seydel and Chang [31, 32] developed a real-time technique for determination of force location and time history using piezoceramic sensors and modeling of the system response. Wheel force transducers have been used for wheel force estimation [33]. Considerable work has been done on estimating moving vehicle loads on structures such as bridges. A number of approaches have been taken that involve some form of modeling of the bridge deck [34–36]. Stites [30] provides a survey of the latest developments in indirect force estimation.

Time- and frequency-domain analyses of equation (1) define the modal properties that can be used for damage identification such as natural frequencies, FRFs, antiresonances, modal vectors, dynamic flexibility, Ritz vectors, etc. [2, 3]. It is also the basis of most other vibrations-based SHM techniques. This equation can be rearranged as follows to more clearly illuminate the classes of vibration-based SHM approaches that are possible:

$$[M]\underline{\ddot{X}} = -[C]\underline{\dot{X}} - [K]\underline{X} - N\left[\underline{X}(t), \underline{\dot{X}}(t)\right] + \underline{F} \tag{3}$$

Convenient measurement; Linear or nonlinear; Active or passive; Internal loads

2.1 Internal loads

The displacement- and velocity-dependent terms in equation (3) represent the internal loads with which a mechanical system or component resists the inertial motion caused by excitation forces. The damage causes changes in the system parameters (linear and nonlinear), and these changes alter the internal loading of the components of mechanical systems. It is important to track such changes for reliable diagnosis and prognosis in these structural systems. The magnitude and directionality of the internal loads dictates the rate at which the damage progresses, and if and when the structure will fail. Consequently, the damage and internal loads together determine how damage progresses in mechanical systems (Figure 1).

Although every SHM method works to identify features that are influenced by changes in the system dynamics, including internal loads, a limited amount of work has been done on identifying the actual internal loads in mechanical systems. Neural networks and regression models have been used to identify operating loads on helicopter components [37–40]. Uhl and Pieczara [41] presented a method for identification of operational loading forces for mechanical structures. The method has two major requirements that impose limitations on its practical applicability in an industrial setting:

1. A known and verified finite element model of the structure is available.
2. Responses of the structure to any load vector can be simulated.

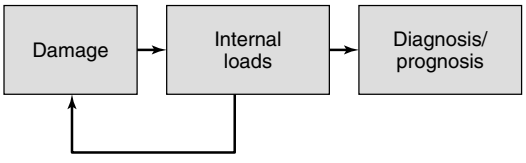


Figure 1. Diagnostic/prognostic methodology.

Haroon and Adams [42, 43] presented a time-domain technique for loads identification that uses system acceleration operating response data. Force-state maps are used to characterize the internal loading (linear and nonlinear) on components of vehicle suspension systems and to track changes in the loads with damage for damage identification. Force-state maps, which represent the displacement- and velocity-dependent internal forces with which a system resists inertial motions, have their origin in the work of Masri *et al.* [44–47] on nonparametric identification of nonlinear systems. Restoring force curves are projections of the force-state maps onto the force-velocity or the force-displacement plane and represent the load characteristics of a system/component as a function of its mechanical state [48]. The force–velocity curves represent the damping loads and the force–displacement curves represent the stiffness loads of a system. These curves allow visualization of the internal loads. An example of a restoring force curve is shown in Figure 2. It is a plot of the displacement versus the acceleration (which is the force scaled by the mass, $F = ma$) of an SDOF system with linear and nonlinear cubic stiffness. The nonlinearity can be characterized by its distinct restoring force curve. Individual nonlinearities have particular restoring forces; therefore, the nonlinear nature of the internal forces can also be characterized by the restoring forces within a system. The area entrained by a restoring force curve is proportional to the magnitude of the internal load and can be estimated to quantify the loads.

2.2 Linear or nonlinear

Damage often causes changes in the nonlinear characteristics of mechanical systems or causes a structure that responds in a linear manner to exhibit nonlinear response characteristics (e.g., loose connections and cracks). Consequently, nonlinear identification can provide useful information for SHM (Brandon [49, 50]). In the past few years, considerable work has been done to identify damage by identifying the nonlinear effects accompanying damage, such as cracks, using frequency harmonics [51] and nonlinear output frequency response functions (NOFRFs) [52]. Nonlinear frequency correlations are also useful for detecting the appearance and propagation of damage. Higher order spectra, specifically bispectra, are useful for crack detection and machine condition monitoring because the nonlinear changes can lead to changes in nonlinear frequency correlations [53–56]. Adams and Farrar [57] developed frequency-domain autoregressive exogenous (ARX) models by considering nonlinearities as internal feedback forces, which contribute to the forced

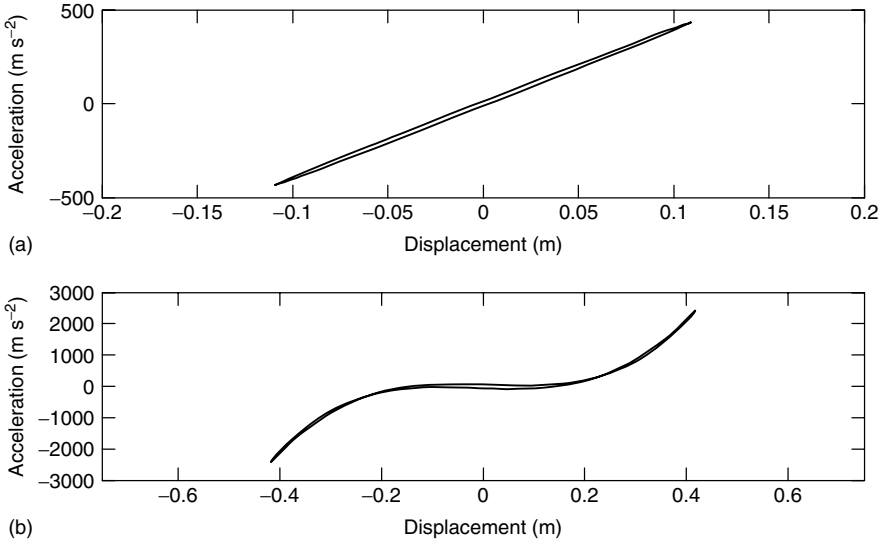


Figure 2. Single degree-of-freedom displacement-restoring force curve; (a) linear stiffness and (b) cubic stiffness.

response of the system. The nonlinear components change with damage and can be used for damage identification. The nonlinear changes in restoring forces can also be used for nonlinear damage detection [58]. Along the same lines of regression of one state variable in terms of another state variable, global phase-space representations of the dynamics of systems have also been used for damage identification. Nonlinear time-series analysis was used to analyze chaotic attractor properties and damage metrics were developed [59–61], which are sensitive to subtle changes in systems and robust to operational variability. In an example of linear phase-plane analysis, Stites *et al.* [62] used phase-plane maps to detect and quantify damage in gas turbine wire harness connectors under harmonic excitation. Other linear health-monitoring techniques include methods based on modal properties [1–3] and methods based on linear FRFs [21–23], including experimental sensitivity functions [21], which detect and quantify damage by comparing the sensitivity of FRFs to system parameters to measured changes in the FRFs. Ackers *et al.* [63] used linear FRF analysis to detect cracks in a metal spindle housed deep within a vehicle end assembly. It is prudent to mention here that the sensitivity of damage features, such as those based on FRFs, to environmental and operational variability is an important consideration for SHM. Cornwell *et al.* [64] discussed the variability in modal properties of a bridge with environmental and operational conditions. Kess and Adams [65] presented a study of the effects of environmental and operational variability on FRFs and transmissibility functions, and showed that frequency bands can be identified where variability is minimum and the possibility of getting false-positive/-negative indications is minimized.

2.3 Active or passive

In cases where the external inputs cannot be measured or estimated accurately, passive methods, which use only measurements of the structural response to external (operational) forces for damage identification, are employed for health monitoring. Transmissibility functions [66, 67], restoring forces [43], phase-plane methods [59–62], and output-only modal parameter estimation techniques [68] are examples of methods that utilize only response measurements to identify damage. Such methods work well in cases where the inputs are difficult to measure, for example in bridge health monitoring, or where restrictions on space and cost hinder the use of actuators to excite structures, but passive methods have the following limitations:

1. There is no guarantee that the excitations are persistent enough or at the frequency required to accentuate the damage, and as such small defects may be difficult to detect.
2. Variability in the unmeasured excitation forces makes the interpretation of the damage features more challenging.
3. Damage cannot be quantified in a physically meaningful way.

A variety of methods are used to excite the structure in active SHM, including shakers, modal impacts, and piezoelectric transducers. Specifically, piezoelectric transducers (Figure 3) are part of a rich and highly active research area for active SHM. Piezoactuators are used to excite structures at high frequencies, where the dynamic response is often more sensitive to local changes within a structure. Electromechanical impedance (EMI) techniques rely on the electromechanical coupling between the electric impedance of piezoelectric materials and the local mechanical impedance of the structure and use the PZT materials to both actuate and sense [69–71]. A number of techniques have been developed that use piezoelectric actuators/sensors to interrogate the health of structures [72–74], with the view to develop embedded sensor technologies. White *et al.* [75, 76] have used piezoactuator/sensor pairs (Figure 4) in the development of the virtual forces technique for SHM.

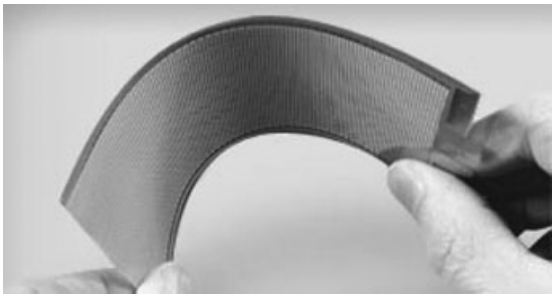


Figure 3. Piezoelectric transducer.

Figure 4. Piezoelectric sensor/actuator pair for SHM. [Reproduced with permission from Ref. 72 & 73, © SPIE, 2000.]

2.4 Time- and frequency-domain analysis

Vibration-based SHM techniques can also be classified according to the domain of analysis, time or frequency. Restoring force curves and autoregressive moving average (ARMA) models are examples of time-domain techniques. In ARMA modeling, response time histories of structural response are fit and the coefficients and residual errors are the damage-sensitive features [77–79]. In addition, time–frequency analysis can be used to analyze any nonstationary events localized in the time domain [3]. Spectrograms (short-term Fourier transform), wavelet analysis, and the Hilbert–Huang transform [80] are examples of time-series analysis techniques. Wavelet transforms have been extensively used for damage identification, especially in rotating structures [81–85].

2.5 Excitation frequency content

Vibrations-based SHM techniques can also be classified on the basis of frequency content of the excitation. Broadband techniques require frequency content over a broad range. Modal parameter estimation techniques [1, 2] and methods requiring spectra of measurements, for example, those based on FRFs [21–23], bispectra [53–56], and frequency-domain ARX models [57], come under this classification. On the flip side, methods that require a narrow frequency range in the input include restoring forces [42, 43] and phase-plane methods [62], which require harmonic excitation.

3 DAMAGE IDENTIFICATION

In this section, numerical and experimental examples of the process of damage identification are presented. Internal loads identification, damage detection, location, and quantification using the classes of vibrations-based damage identification discussed in the previous section are covered, leading up to the development of vibration-based damage prognosis models for prediction of growth of fatigue damage to failure. It is shown that vibration loads and damage information can be combined to develop damage prognosis models.

3.1 Internal loads identification

The internal loads, the forces with which a system resists inertial motion, are a function of the linear and nonlinear parameters of the system. Consequently, the internal loads change with the onset and progression of damage and can be estimated for damage identification. Restoring force curves can be used to identify internal system loads [42, 43]. Response acceleration measurements from two response locations are required to estimate the internal loads between the two locations. A two degree-of-freedom quarter car model (Figure 5) has been used to identify the internal loads in the strut of a passenger vehicle [86]. Acceleration measurements at the spindle and strut-body mount from full-vehicle two-post shaker tests (Figure 6) were used to estimate the restoring forces in the strut. The equation for the sprung mass, M_2, can be written using Newton's Law as

$$M_2\ddot{x}_2 = -C_2(\dot{x}_2 - \dot{x}_1) - K_2(x_2 - x_1) - K_3x_2 + N_1[x_1(t), x_2(t), \dot{x}_1(t), \dot{x}_2(t)] \quad (4)$$

where $x_k(t)$ are the displacements of the unsprung and sprung masses, M_k, C_2 is the suspension viscous damping coefficient, K_k are the stiffness elements in

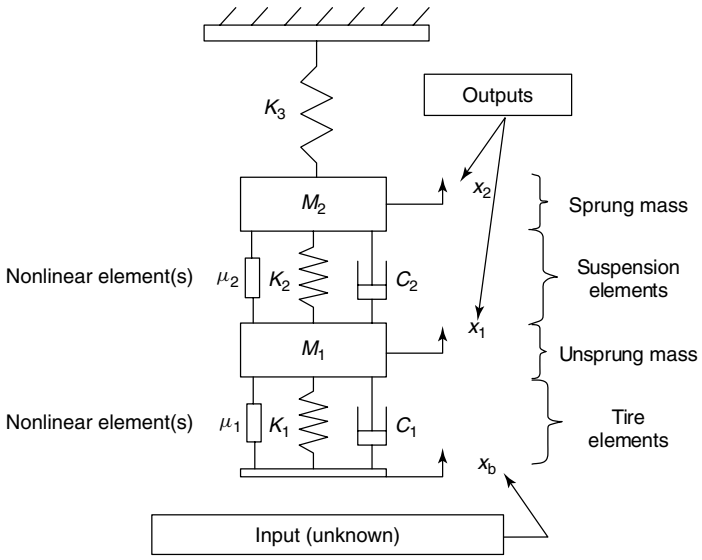


Figure 5. Two degree-of-freedom quarter car model.

the suspension and vehicle body, and $N_1[x_1(t), x_2(t), \dot{x}_1(t)\dot{x}_2(t)]$ denotes the nonlinear forces in the suspension. Because restoring force curves need to be generated at discrete frequencies, slow sine sweep (narrow-band) tests were run and the acceleration signals were numerically integrated to obtain the velocities and displacements. The relative velocity and relative displacement across the suspension were plotted against the acceleration of the sprung mass to generate the restoring force curves and identify the internal loads. Figure 7 shows the frequency characteristic of the velocity-dependent internal loads and Figure 8 shows the same result for the displacement-dependent internal loads.

The velocity-dependent force (Figure 7) initially exhibits primarily hysteresis, but as the frequency increases the force takes on the shape of a Coulomb friction curve, then a piecewise linear characteristic (typical of vehicle shocks), and finally it becomes primarily linear with some hysteresis. The displacement-dependent force (Figure 8) shows primarily hysteresis with backlash at certain frequencies. As an example of comparison of models with experiments, the same restoring forces were also generated from simulations of a full-vehicle numerical model. The ADAMS model with linear stiffness and nonlinear shock damping is shown in Figure 9. As with the experimental results, the displacement-dependent curves show hysteresis with backlash and the velocity-dependent curves show a piecewise linear characteristic.

The internal loads of a sway bar link while subjected to tension–tension fatigue tests were also estimated by measuring the axial acceleration at the two ends (Figure 10). The measurements at the two ends were used to estimate the internal loads. The velocity-dependent internal load has hysteresis and the displacement-dependent load is linear (Figure 11). This example is used throughout the following section to demonstrate the process of damage identification, leading up to the development of a prognostic model.

3.2 Damage detection

A wide variety of vibrations-based methods can be used for damage identification, depending on the available data, the system, and the damage [2, 3]. Shifts in resonant frequencies are the most basic means of detecting damage. Consider a four degree-of-freedom lumped parameter model (Figure 12) simulated in Matlab Simulink®. Damage is introduced in the system in the form of a reduction in stiffness between degrees of freedom 2 and 3. The estimated FRFs for the damaged and undamaged cases are shown in Figure 13. There are clear shifts downward in the second and fourth resonance

Figure 6. Two-post shaker setup and accelerometer locations.

peaks, which represent the change in system modal properties with damage (- - - -). Similarly, damage in realistic structural systems affects the system parameters, which can be detected with the changes in structural natural frequencies [1].

The antiresonances of an FRF are more sensitive to local dynamics and, hence, can provide a better indication of damage in certain cases. For example, modal impact testing of a helicopter blade (Figure 14) with localized damage (change in stiffness) in the form of a gradually tightened strap, which modified the chord stiffness, showed that the antiresonances are more sensitive to a change in system dynamic properties in this case because the damage is more localized (Figure 15).

Damage can also cause nonlinear changes in mechanical systems, in which case nonlinear methods can be used to detect damage. Bolt loosening leads to nonlinear changes associated with Coulomb friction. This damage mechanism was simulated in a building model [57] and identified using frequency-domain ARX models. Consider the two degree-of-freedom quarter car model (Figure 5) with cubic stiffness nonlinearity between the sprung and unsprung mass. A first-order linear ARX model can be used,

$$Y(k) = B(k)U(k) + \sum_{j=-1\neq 0}^{1} A_j(k)Y(k-j) \quad (5)$$

where the response is considered to be caused by the input at the driving frequency and feedback correlation with one frequency above and one frequency below the input frequency. A damage index can be

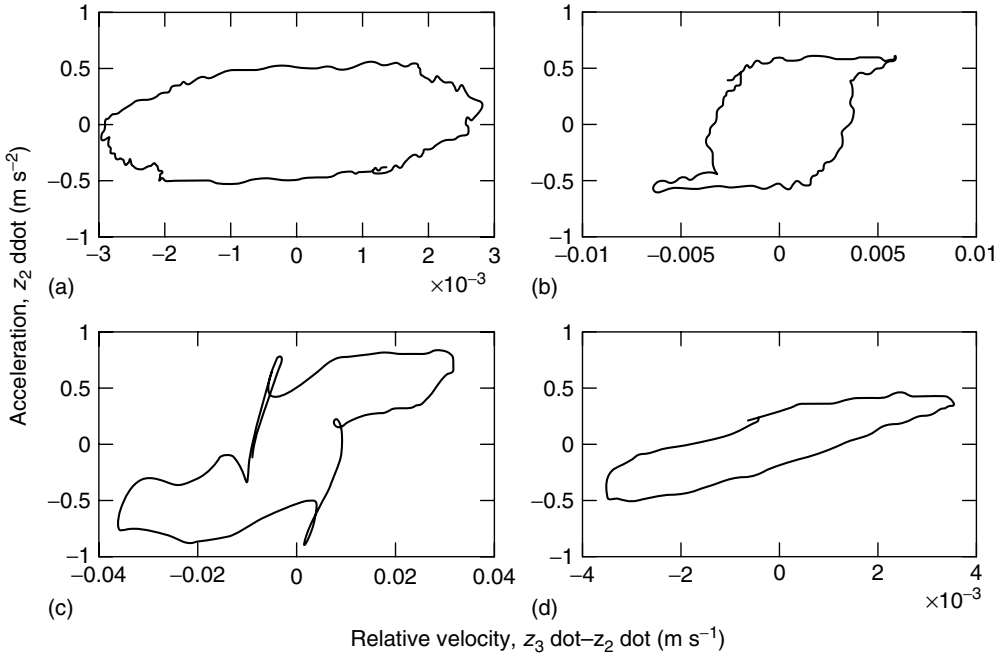


Figure 7. Frequency characteristic of vertical velocity-dependent internal force in the strut: (a) 4.05 Hz, (b) 4.17 Hz, (c) 6.67 Hz, and (d) 12.5 Hz.

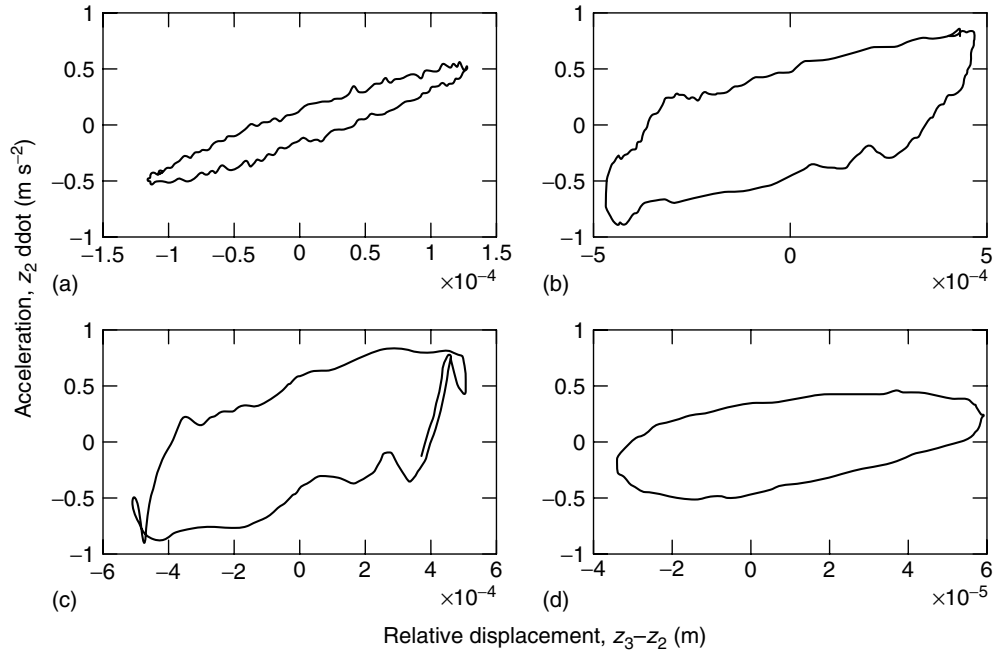


Figure 8. Frequency characteristic of vertical displacement-dependent internal force in the strut: (a) 4.05 Hz, (b) 4.5 Hz, (c) 6.67 Hz, and (d) 12.5 Hz.

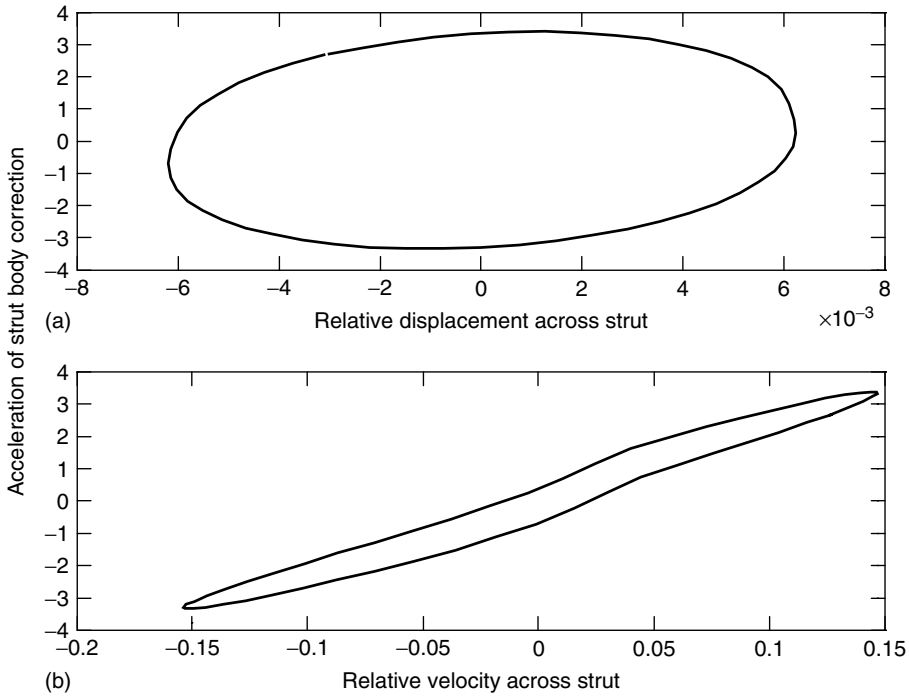


Figure 9. Simulated strut-restoring forces (ADAMS numerical model): (a) displacement-dependent force at 3.7 Hz and (b) velocity-dependent force at 3.7 Hz.

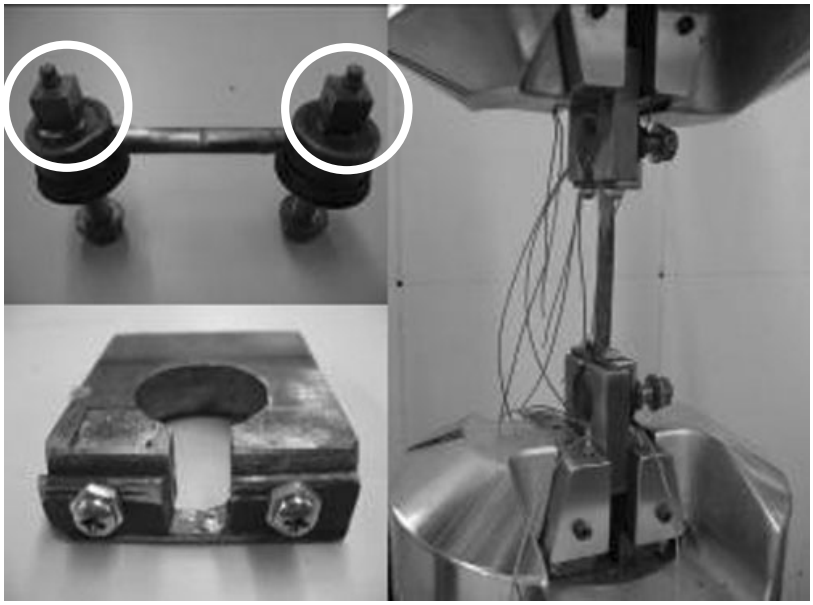


Figure 10. Fatigue testing of sway bar link.

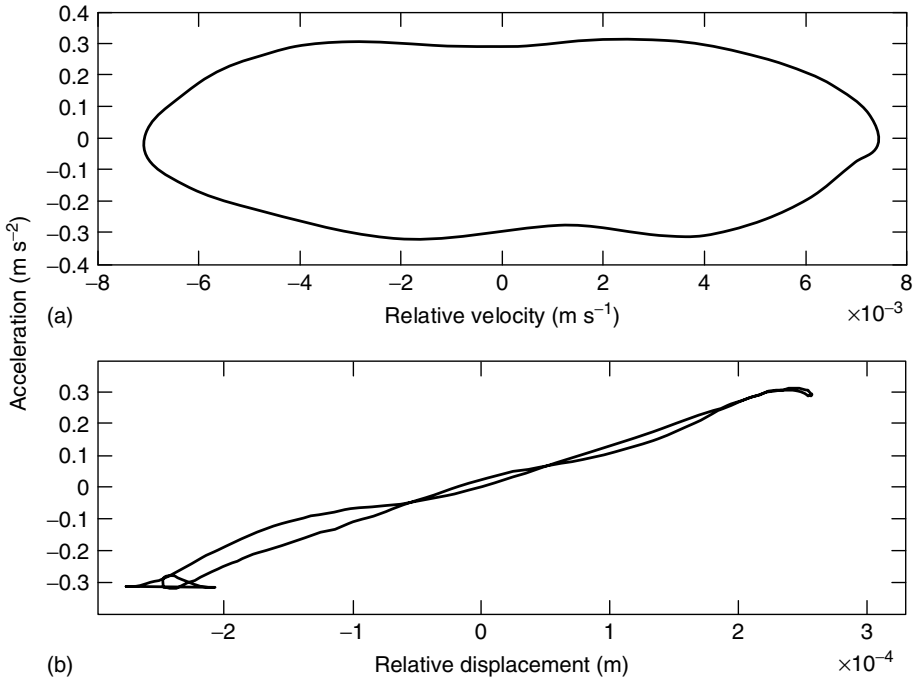


Figure 11. (a) Velocity- and (b) displacement-restoring force in the sway bar link under a tensile load of 7000 N; 14.5 Hz.

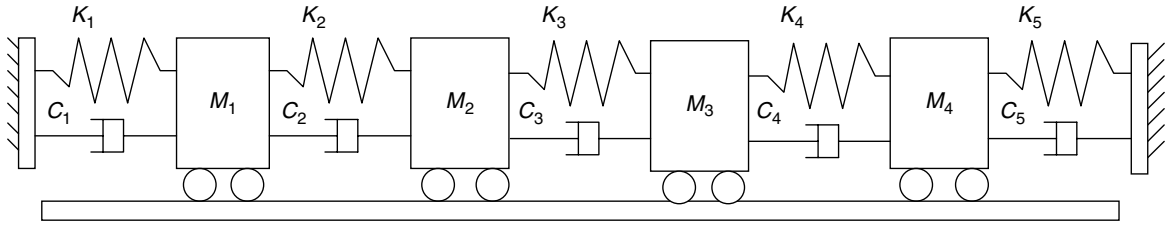


Figure 12. Four degree-of-freedom system.

developed using the nonlinear autoregressive coefficients, $A(k)$. If the damage is assumed to cause a change in the nonlinear parameter, the damage indicator can be used to detect that change. Figure 16 shows that the indicator captures the simulated increase in the nonlinearity with damage.

Similarly, damage simulated by gradual loosening of the lower ball joint in a vehicle suspension (Figure 17) was identified using these damage models [7, 86]. Figure 18 shows that as the bolt is loosened the nonlinear correlations increase owing to the increased friction between the bolt and the ball joint caused by relative motion (undamaged torque was 400 lb-in). The nonlinearity decreases when the bolt is removed because the source of the increased friction is no longer present.

Farrar *et al.* [87] provide a substantial set of examples of nonlinear feature extraction for damage detection, including linearity checks on bridge FRF measurements [88], waveform distortion in rotating machinery [86, 87], time-series analysis [89–93], and higher order spectra [94, 95].

Because the internal loads change with damage, they can also be used to detect damage. The internal loads across the path containing the bolt damage in the vehicle suspension were also estimated as the damage progressed (bolt loosening) (Figure 19). The damage causes a change in frequency characteristics

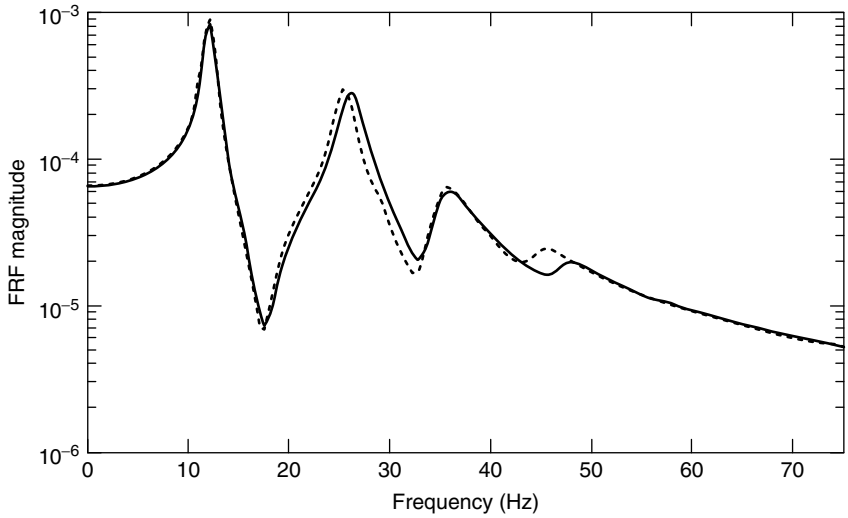


Figure 13. Change in resonant frequencies with damage represented by reduction in stiffness between degrees of freedom 2 and 3; undamaged (———) and damaged (- - - -).

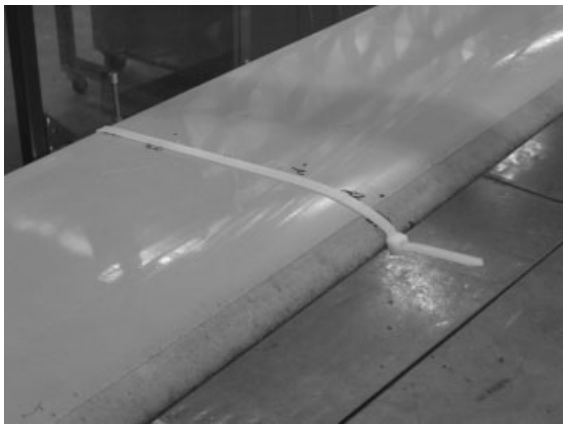


Figure 14. Strap to simulate damage, causing change in stiffness of helicopter blade.

of the velocity-dependent restoring force with the progression of damage. The system moves out of the hysteresis loop at different frequencies because of the damage. The intermediate cases (loosened bolts, 250 in-lb (· · ·) and 100 in-lb (- - - -)) change at higher frequencies (the frequency at which the restoring force corresponding to 100 in-lb bolt torque changes is the highest) compared to the undamaged case, and the most severe damage case (bolt removed) changes at a lower frequency compared to the undamaged case. This same behavior is observed for different input amplitudes. Thus, changes in the characteristic of the nonlinear internal loads can be used to detect damage.

The fatigue testing of the sway bar link in Figure 10 led to the appearance of a circumferential fatigue crack at the lower weld location (Figure 20). The transmissibility across the link and the internal load were used to detect the damage. The transmissibility shows a shift in the level and peak (Figure 21) and the internal loads show an increase in magnitude as the area of the velocity-dependent curve increases (Figure 22). The displacement-dependent internal load shows a decrease in slope, which represents a decrease in stiffness with damage as the link weakens.

3.3 Damage location

Damage localization requires the use of techniques that are sensitive to local changes in a system dynamic's properties, such as the transmissibility functions. Transmissibility functions are like FRFs in many ways, but they are ratios of the same variable (acceleration to acceleration, etc.). This nondimensional property tends to suppress global effects of damage and, thus, enhances the local effects because transmissibilities only contain the zeros, and not the

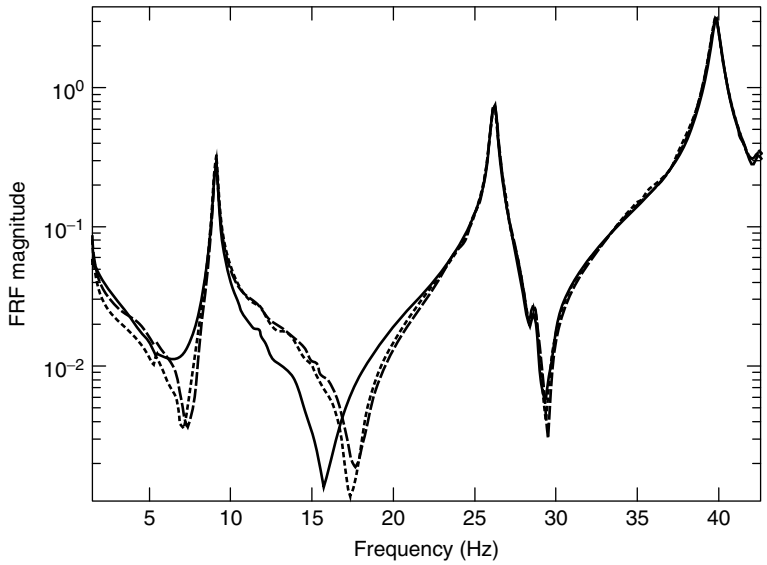


Figure 15. FRFs for helicopter blade damage showing greater sensitivity of antiresonances to localized damage; undamaged (———), damaged level 1 (- - - -), and damaged level 2 (· · · ·).

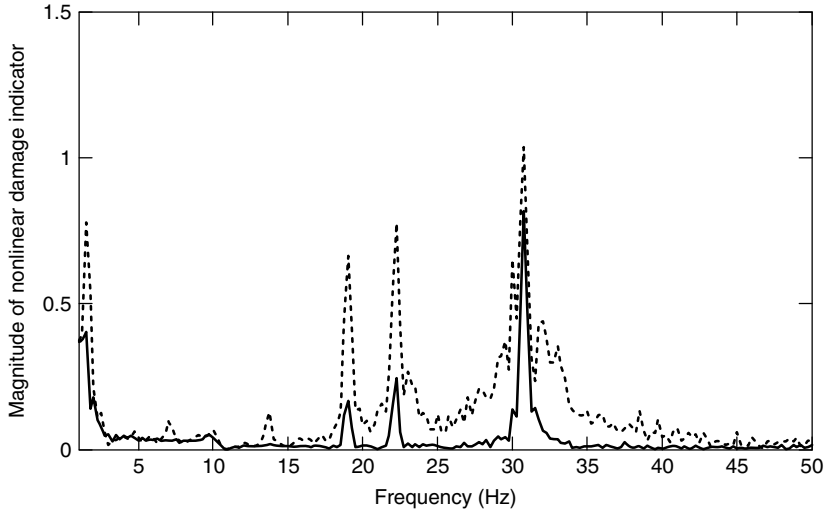


Figure 16. Simulated nonlinear indicator from first-order linear ARX model showing increase in cubic stiffness nonlinearity in the two degree-of-freedom quarter car model; undamaged (———) and damaged (- - - -).

poles, of discrete FRFs, and therefore contain information about more localized regions of structures [67]. Consider once again the four degree-of-freedom system in Figure 12. Suppose damage is introduced in the form of a reduction in stiffness between masses 1 and 2. Transmissibility functions between different response pairs can be used to locate the damage (Figure 23). The transmissibility functions for the response pairs, which have the damage in the path between them (Figure 23a and d), clearly indicate the change in system dynamics with damage, while the response pairs that do not have the damage in the path between them (Figure 23b and c) show no significant change with damage.

Figure 17. Ball joint connecting lower control arm to the steering knuckle.

The bolt damage in Figure 16 can be located in a similar manner. The frequency-domain ARX model in equation (5) is an output-only formulation and, hence, is similar to transmissibility functions. Consequently, it is more sensitive to local changes in dynamic properties. The model in equation (5) was applied to data from locations that did not have the damage in their path. Figure 24 shows the estimated indicator $1 - |A_{j_d}/A_{j_{un}}|$ for the data from points that did not have the damage location in the path between them. It is clear that there are no significant changes in the nonlinear correlations in the path between these two points. This analysis located the damage at the lower ball joint.

A number of other techniques have been developed for damage location. These include methods based on active piezoactuation and sensing, which use traveling, guided waves to sense and locate damage in structures [74, 96], wavelets [97, 98], and strain energy–based methods [98, 99].

3.4 Damage quantification

Quantification is important for developing prognosis models to predict damage growth and failure. Damage indices based on the various damage-identification techniques can be used to track the growth of damage and thresholds can be defined to quantify damage. For example, the level of the index based on the frequency-domain ARX model (Figure 18) can be used, and similarly an index based on transmissibility functions can be developed [67]. However, quantification of damage in a manner that has physical engineering significance is more useful. For example, if it can be stated that a crack has caused a certain change in the stiffness of a structure, the affect of damage on the performance of the structure will be clearer.

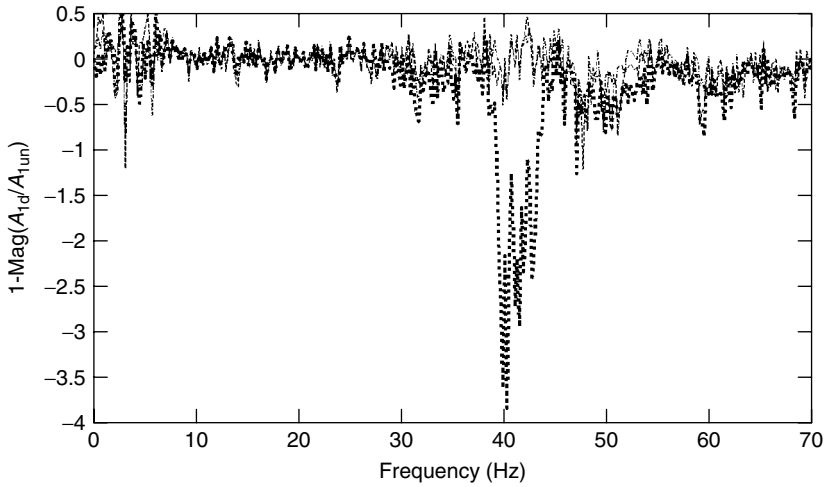


Figure 18. Nonlinear indicator 1-Mag$(A_{1d}(\omega)/A_{1un}(\omega))$ for bolt loosening damage; 250 in-lb torque (- - - -), 100 in-lb torque ($\cdot\cdot\cdot\cdot$) and no bolt (–.–.–).

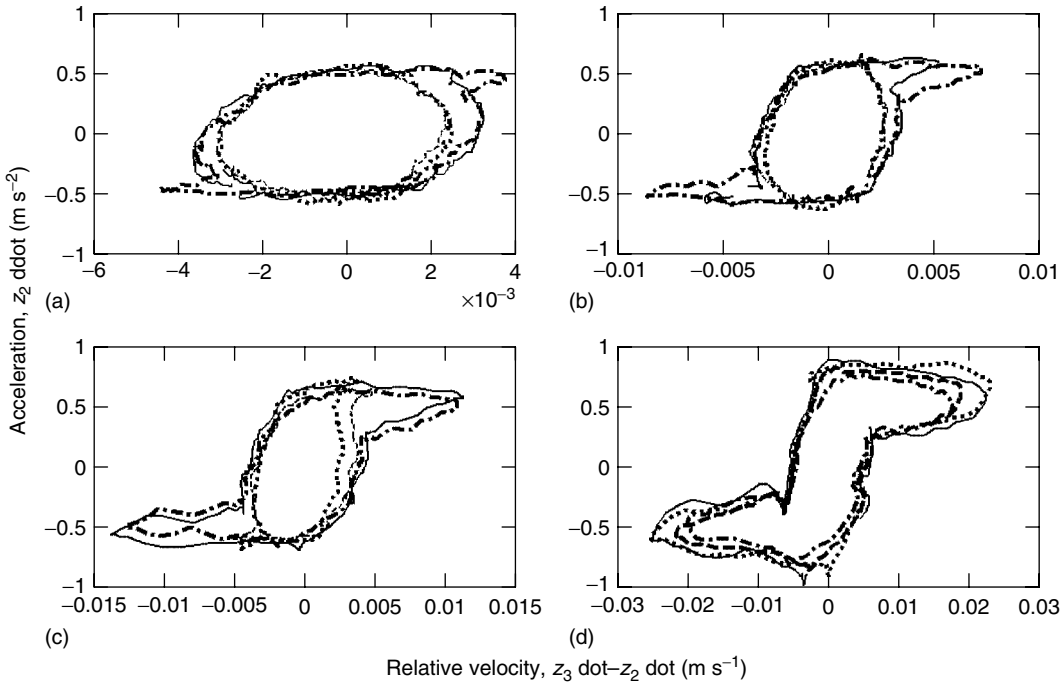


Figure 19. Change in frequency characteristic of vertical damping internal force in the strut with damage (0.5 mm). Bolt torques: undamaged −400 in-lb (- - - -), 250 in-lb (· · · ·), 100 in-lb (- - - -), and no bolt (–.–.–). Frequency: (a) 4.09 Hz, (b) 4.17 Hz, (c) 4.26 Hz, and (d) 4.67 Hz.

Figure 20. Initial circumferential crack in sway bar link under cyclic loading.

Experimental sensitivity functions [21] provide a method that can be used to quantify damage in terms of discrete changes in mass, stiffness, or damping. This method requires a measure of the input force. Consider the two degree-of-freedom quarter car model in Figure 5. Suppose that damage causes a change in the stiffness, K_2, between the sprung and unsprung masses. Estimated or, in the experimental case, measured FRFs can be used to quantify the change in stiffness caused by the damage as

$$\Delta K_{12} \approx \frac{\Delta H_{11}(\omega)}{-[H_{11}(\omega) - H_{12}(\omega)][H_{11}(\omega) - H_{12}(\omega)]} \tag{6}$$

The subscript "12" describes the location of the damaged element. The denominator is called the sensitivity of the FRF to a change in the stiffness K_{12}, $\partial H_{11}(\omega)/\partial K_{12}$. Let us take an example of damage causing a 10% (2500 N m^{-1}) decrease in the stiffness. Using the estimated FRFs in equation (6) to estimate the change in stiffness gives the result in Figure 25. The frequency region between the two resonant frequencies provides an estimated change in stiffness of 2476 N m^{-1}, which is approximately the same as the actual change. Similarly, mechanical system damage can be related to a parametric

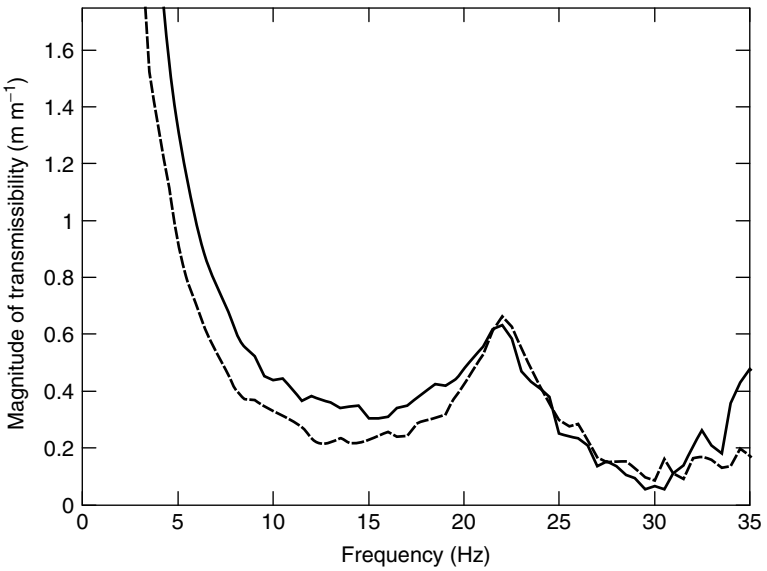


Figure 21. Change in transmissibility with the appearance of the initial circumferential crack in link; undamaged (———) and initial crack (- - - -).

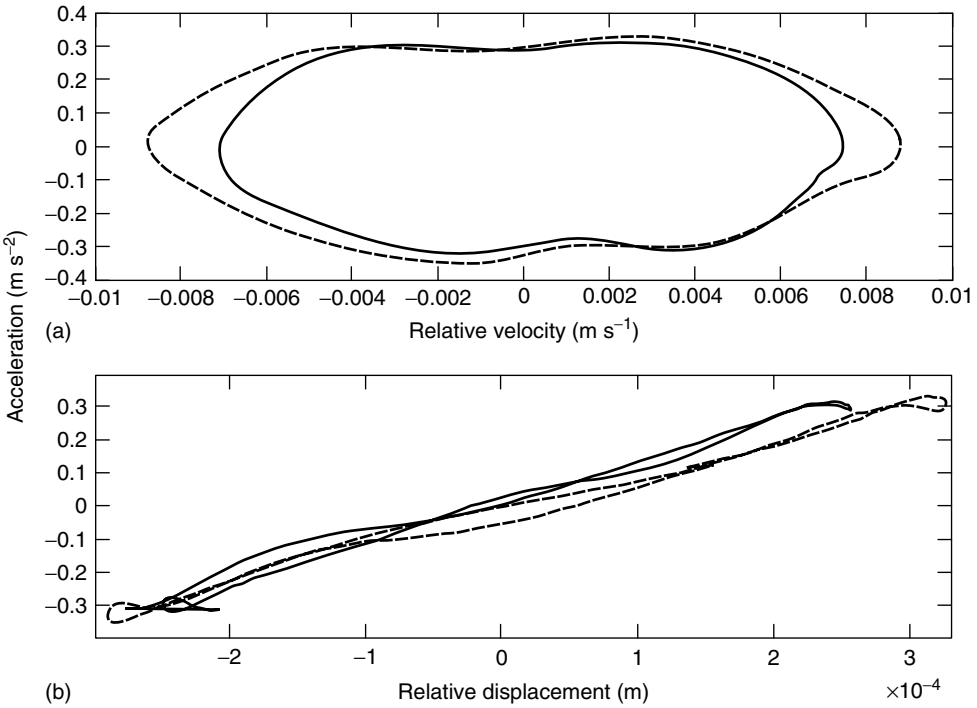


Figure 22. Change in restoring forces with the appearance of the initial circumferential crack in link; undamaged (———) and initial crack (- - - -).

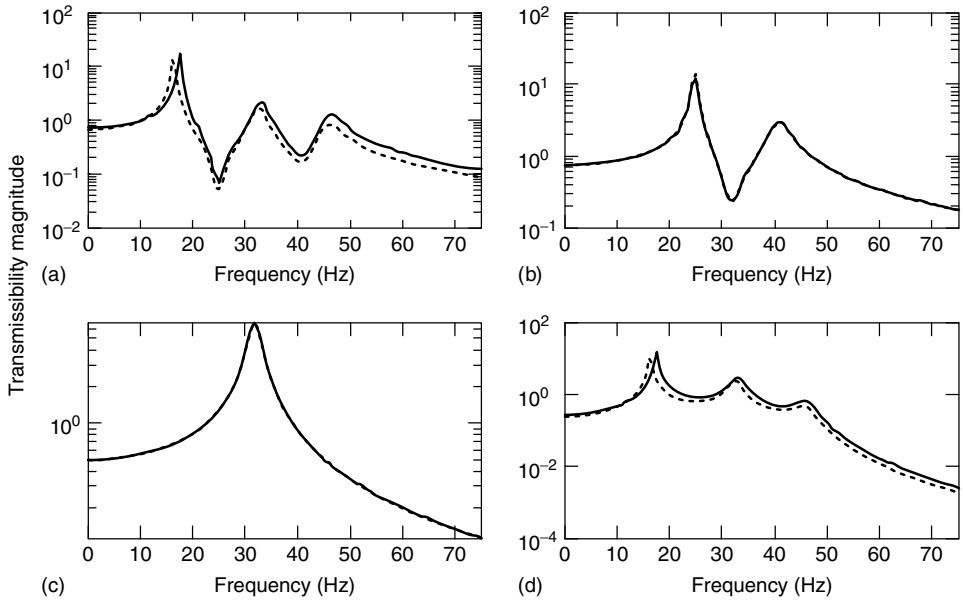


Figure 23. Transmissibility functions between different response location pairs for undamaged (———) and stiffness damage (- - - -) cases in four degree-of-freedom system: (a) x_1 and x_2, (b) x_2 and x_3, (c) x_3 and x_4, and (d) x_1 and x_4.

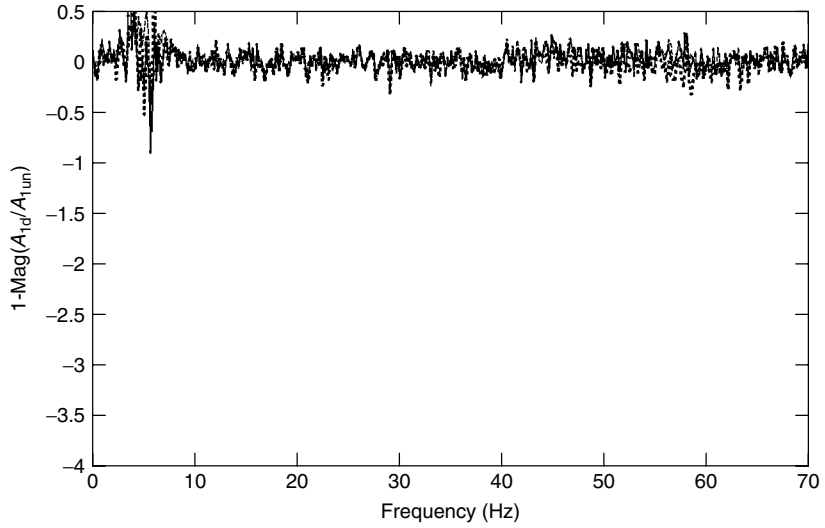


Figure 24. Nonlinear indicator 1-Mag($A_{1d}(\omega)/A_{1un}(\omega)$) for first-order linear ARX model applied to points with no damage in their path; 250 in-lb torque (- - - -), 100 in-lb torque (· · · ·), and no bolt (–.–.–).

change in the structure properties to attach a physical significance to the observed damage in the system.

Changes in the restoring force area, which is proportional to internal system load magnitude, can also be used to quantify damage. The progressive change in area of the velocity-dependent internal load of the sway bar link (Figure 10) as the circumferential crack grew to failure was estimated to quantify the change in internal load with damage (Figure 26 and Table 1). It is clear that the area of the curves, and

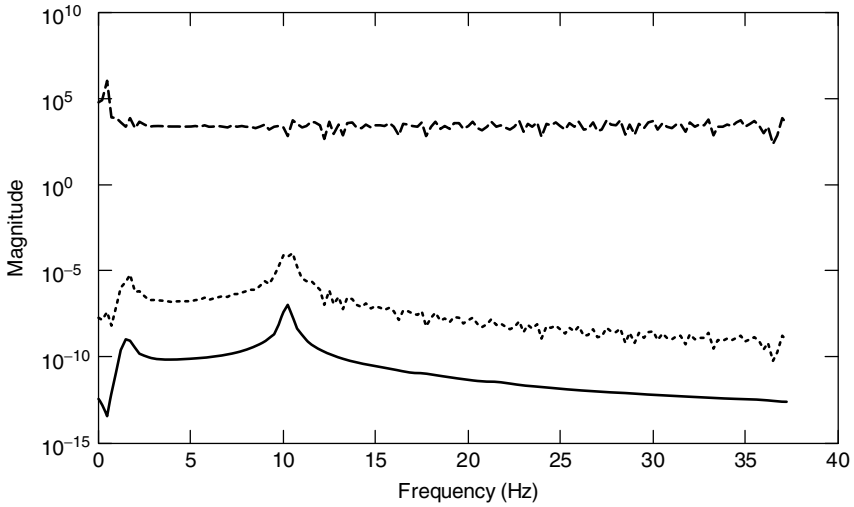


Figure 25. Estimated change in stiffness of the two degree-of-freedom system with damage; stiffness sensitivity $\partial H_{11}(\omega)/\partial K_{12}$ (———), change in FRF, ΔH_{11} (· · · ·), and change in stiffness, ΔK_{12} (- - - -).

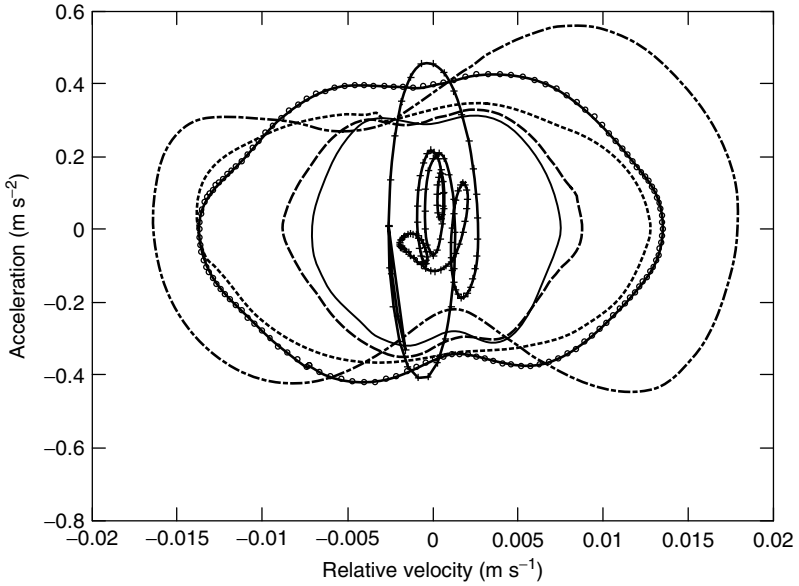


Figure 26. Change in velocity-dependent restoring force with appearance and progressive growth of circumferential crack to failure in a sway bar link under tension–tension fatigue loading; undamaged (———), initial crack (- - - -), progression 1 (· · · ·), progression 2 (–.–.–), just before failure (–+–+–+), and after failure (– ° – ° –).

the load, increases progressively with the damage. The velocity-dependent restoring force also shows an indication of the approaching failure. Immediately before complete failure, the restoring force becomes multiply connected (Figure 26 (–+–+–+)). The internal loads provide a quantifiable change with damage and an indication of impending failure.

Some of the other techniques developed for damage quantification include those based on elemental modal strain energy [100] and methods

Table 1. Changing velocity-dependent restoring force areas with the appearance and progress of circumferential crack in a sway bar link

	Undamaged	Undamaged 2	Crack visible 5000 cycles	Progression 1 10 000 cycles	Progression 2 12 500 cycles	Just before failure 15 000 cycles
Restoring force area ($mm^3\ s^{-2}$)	0.0074	0.0073	0.0088	0.0149	0.0246	0.0028

based on finite element model updating [101], which iteratively minimize the differences between measured and predicted parameters.

3.5 Prognosis

The ultimate aim of SHM is to combine load and damage models that can be used to predict the growth of damage and ultimately failure. And loads identification is an important component of the prognosis process (Figure 1). Information about the current status of damage (health monitoring) alone is not sufficient for accurate prediction of future trends; the nature of system loads (usage monitoring) is also needed for prognosis [102]. A vast amount of work has been done on fatigue and fracture analysis for prognosis, including extensive development of crack propagation laws. But these laws require localized information like geometry of the component, orientation of damage (cracks), and local loading. One of the goals of vibrations-based SHM is to utilize global vibration measurements to identify damage to the system without the need for localized information, which would make the data-acquisition process prohibitively complex, time consuming, and expensive. Prognosis of structural damage at the system level is not a well-developed area and basic research is being conducted. Having said that, damage-based prognostics in the helicopter industry is quite well developed with the advent of the health and usage monitoring system (HUMS) [103]. Risk assessment is another form of damage prognosis. Seismic probabilistic risk assessment (PRA) has been applied to commercial power plants. The annual probability of damage as a result of some future earthquake is obtained [104]. Preliminary work on the use of hidden Markov models for damage prognosis has also been presented [105]. Another example of initial work on damage prognosis is fatigue crack growth prediction using an energy method [106]. Inman *et al.* [107] provide an overview of the technology of prognosis.

Haroon [86] developed empirical mechanical damage growth models, which combine internal loads identification and damage diagnosis. Crack growth laws, which predict the growth of cracks based on the current length of crack and the stress distribution around it, are used as motivation. Damage information from a transmissibility function–based damage index and internal loads quantification from restoring force curve areas are combined to develop prognosis models of the form,

$$\frac{\mathrm{d}(D)}{\mathrm{d}N} = C|D|^m \times |\Delta L|^n \tag{7}$$

where D represents the damage, ΔL is the changing internal component load, and C, m, and n are empirical constants that depend on material properties, material geometry, boundary conditions, external loading, and damage mechanism. The models in equation (7) are used to predict the growth of fatigue cracks in the front sway bar link of a passenger vehicle under tension–tension cyclic fatigue loading (Figure 10). Loads (Figure 26) and damage (Figure 27) information are combined in this model to predict the growth of damage in terms of change in the transmissibility-based index, $D = \Delta T$ (Figure 28).

$$DI_k = \frac{\sum_{i=a}^{b} |\mathrm{Re}(\ln(T_k(\omega_i))) - \mathrm{Re}(\ln(T_{k-1}(\omega_i)))|}{N_f} \tag{8}$$

where T_k is the transmissibility at the kth measurement, T_{k-1} is the previous measurement, ln is the natural log, ω is the frequency, and i is the index that determines the range of frequencies (a to b) over

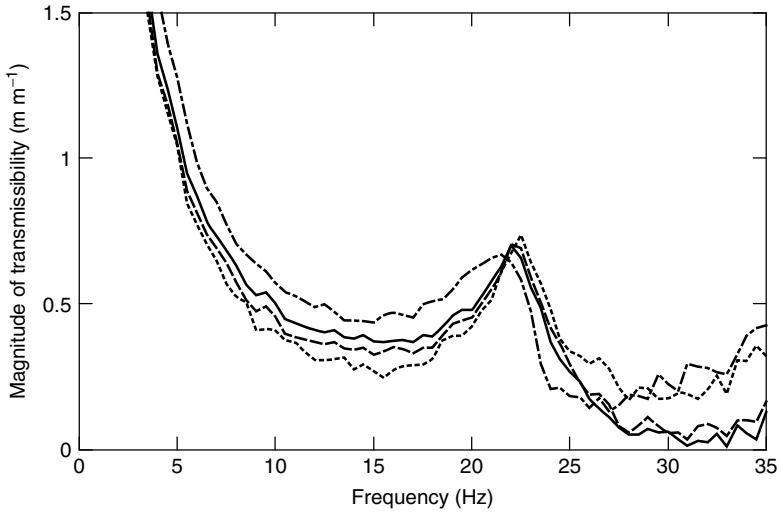


Figure 27. Change in transmissibility with progressive growth of fatigue crack to failure in sway bar link under tension–tension fatigue loading; initial crack (———), progression 1 (- - - -), progression 2 (· · · ·), and just before failure (–.–.–).

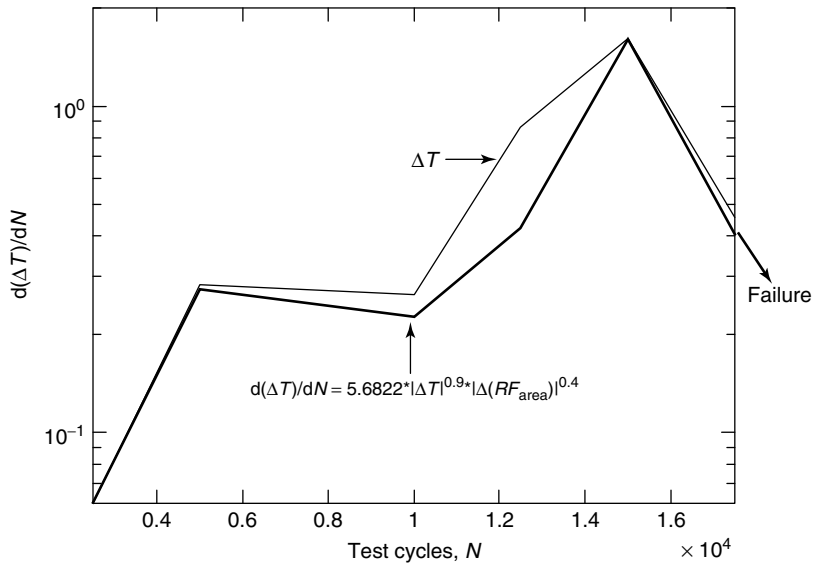


Figure 28. Empirical regression model relating estimated change in transmissibility to change in restoring force area for experimental circumferential crack damage in a sway bar link.

which the change is summed and N_f is the total number of frequencies.

In this case, $m = 0.9$, $n = 0.4$, and $C = 5.6822$. As Figure 28 shows, this model produces the same rate of change of the damage indicator (darker line) as the actual change in the damage indicator (lighter line). The empirical model follows the same trend as the damage indicator; hence, it predicts the growth of damage and the associated damage indicator.

In order to demonstrate that the models actually predict the growth of damage, tests were run on three links with the same material properties, geometries,

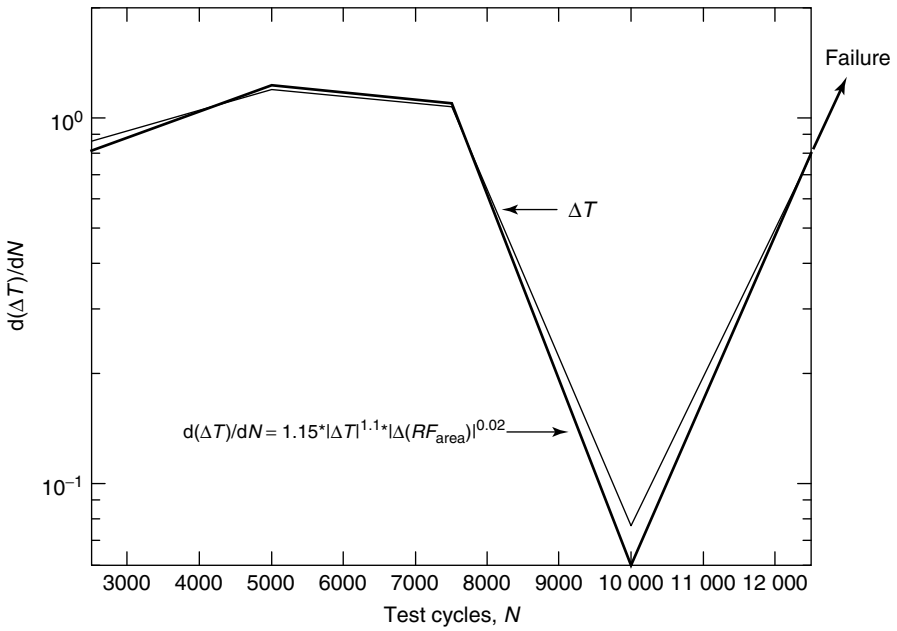


Figure 29. Empirical damage prognosis model for experimental fatigue crack damage in a sway bar link; $m = 1.1$, $n = 0.02$, and $C = 1.15$.

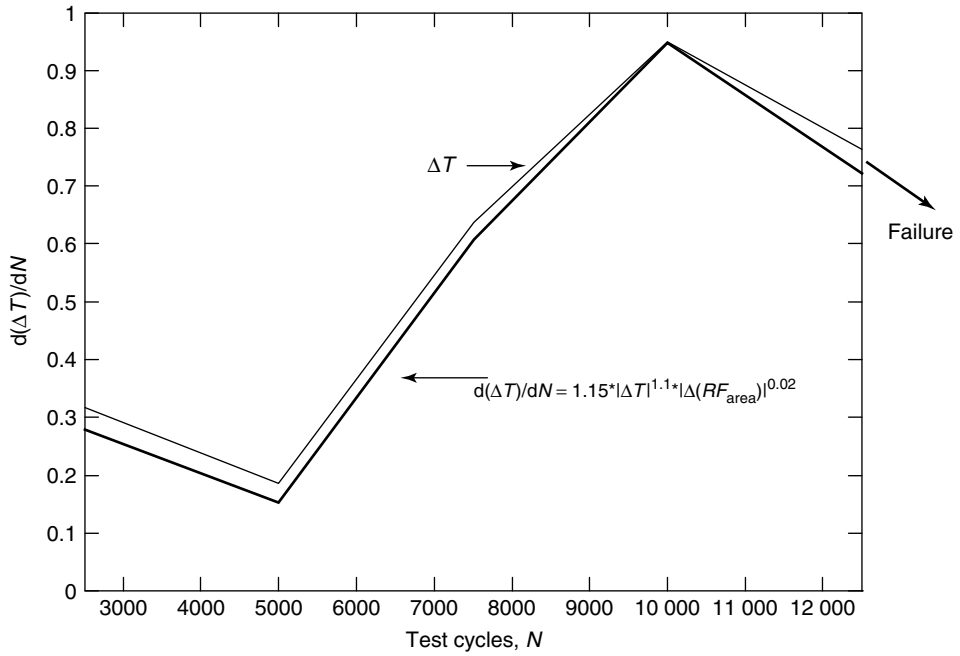


Figure 30. Prediction of growth of experimental fatigue crack damage in sway bar link to failure; Link 1.

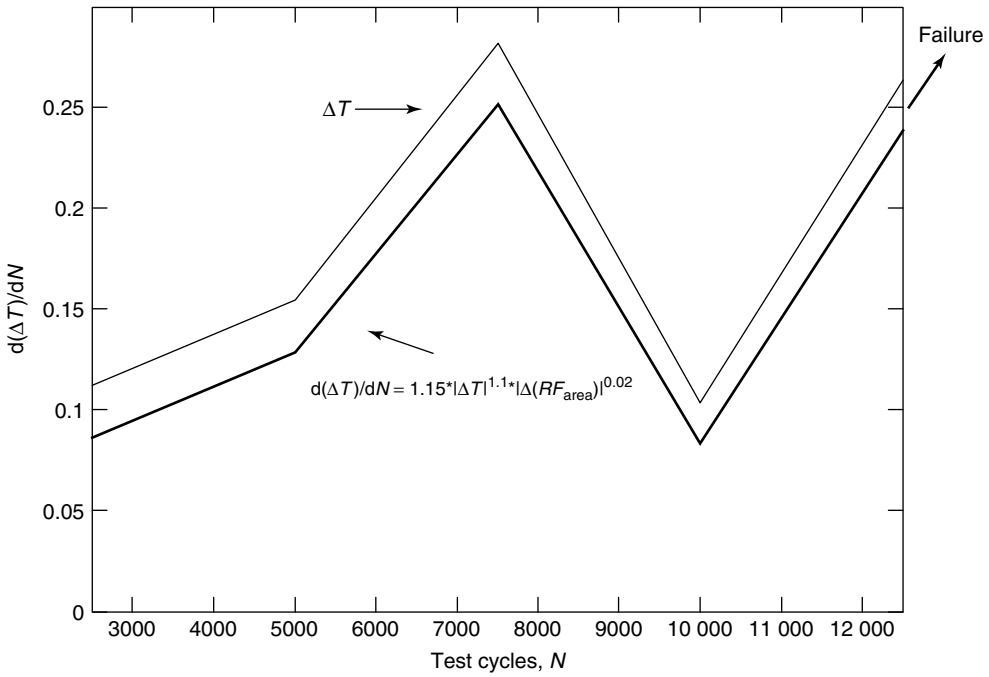


Figure 31. Prediction of growth of experimental fatigue crack damage in sway bar link to failure; Link 2.

loading, and crack mechanism. The model developed using the first link (Figure 29) was used to predict the growth of damage in the other two links (Figures 30 and 31). It is clear that for the same component, material, boundary conditions, loading, and damage mechanism, there exists an empirical relationship between load and damage that predicts the growth of damage to failure. The only error in the prognoses is offsets, which can be attributed to the variability in the placement of the links in the test fixture, variability in the loading, and the fact that all three links had different usage levels; the first two links were 14 years old and had been extensively tested in full-vehicle tests and the last link was new and had never been tested prior to the fatigue testing. This difference in usage is also the primary reason that the damage indicator (or damage) grows differently for the three links; but the same empirical model predicts the growth of damage once it has been initiated. It should be noted that the reason the damage indicator does not have a growing trend is that the change in the transmissibility has been considered rather than the absolute value. The important factor is the empirical model that fits the damage indicator behavior.

It is important to mention that the crack laws require localized information about the loading (stress) and damage (crack length), and it is not feasible to instrument every potential crack initiation region to collect this localized information. On the other hand, the vibrations-based damage prognosis models in [86] only require component-level vibration measurements and this global nature make them more practical for SHM.

4 SUMMARY

Vibrations-based SHM is a class of techniques that have their basis in Newton's Law. Global vibration measurements at the system or component level can be used to interrogate the health of structures. Quite diverse and wide-ranging techniques come under the umbrella of vibrations-based SHM based on the type of data, analysis domain, nature of systems, etc. This collection of techniques points to the power and versatility of this class of SHM methodologies. The available data ranges from acceleration and force to strain, with acceleration being the most convenient measurement, and the techniques can be

further classified according to linear or nonlinear nature of analysis, time-domain or frequency-domain techniques, and availability of input measurements and excitement of structures (active or passive).

Internal loads identification is an important part of the vibrations-based SHM process. The internal loads are a function of system parameters, change with damage, and affect the further growth of damage. As such they can be used to detect damage and develop prognosis models to predict the remaining useful life of structures. Vibrations-based techniques are available for all aspects and stages of SHM: loads ID, damage detection, location, quantification, and life prediction.

Damage prognosis is an area of vibrations-based SHM that is still in its infancy. The aim is to develop damage models to predict growth of damage and remaining useful life. Loads identification can be combined with damage information to develop empirical models to predict damage growth.

REFERENCES

[1] Salawu OS. Detection of structural damage through changes in frequency: a review. *Engineering Structures* 1997 **19**(9):718–723.

[2] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics: A Literature Review*, Los Alamos National Laboratory Report LA-13070-MS. Los Alamos National Laboratory, 1996.

[3] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates SW, Nadler BR, Czarnecki JJ. *A Review of Structural Health Monitoring Literature from 1996–2001,* Los Alamos National Laboratory Report LA-13976-MS. Los Alamos National Laboratory, 2004.

[4] Rytter A. *Vibration Based Inspection of Civil Engineering Structures*, Ph.D. Dissertation. Aalborg University, 1993.

[5] Doebling SW, Farrar CR. Using statistical analysis to enhance modal-based damage identification in structural damage assessment using advanced signal processing procedures. *Proceedings of DAMAS '97*. University of Sheffield, 1997; pp. 199–210.

[6] Ruotolo R, Surace C. Damage assessment of multiple cracked beams: results and experimental validation. *Journal of Sound and Vibration* 1997 **206**(4):567–588.

[7] Haroon M, Adams DE. Nonlinear fault identification methods for ground vehicle suspension systems. *IMAC-XXIV*, Paper No. 44. St. Louis, MO, 2006.

[8] Celebi M, Sanli A, Sinclair M, Gallant S, Radulescu D. Real-time seismic monitoring needs of a building owner—and the solution: a cooperative effort. *Earthquake Spectra* 2004 **20**:333–346.

[9] Han S, Lee JB. Analysis of errors in the conversion of acceleration into displacement. *IMAC XIX*. Kissimmee, FL, 2001; pp. 1408–1413.

[10] Zak A, Krawczuk M, Ostachowicz W. Vibration of a laminated composite plate with closing delamination. *Structural Damage Assessment Using Advanced Signal Processing Procedures Proceedings of DAMAS '99*. University College, Dublin, 1999; pp. 17–26.

[11] Zimmerman DC. Looking into the crystal ball: the continued need for multiple viewpoints in damage detection. *Structural Damage Assessment Using Advanced Signal Processing Procedures Proceedings of DAMAS '99*. University College, Dublin, 1999; pp. 76–90.

[12] Stanbridge AB, Khan AZ, Ewins DJ. Fault identification in vibrating structures using a scanning D vibrometer. *Proceedings 1st International Workshop on Structural Health Monitoring, Current Status and Perspectives*. Stanford University, Palo Alto, CA, 1997; pp. 56–65.

[13] Martin A, Hudd J, Wells P, Tunnicliffe D, Das-Gupta D. Development and comparison of low profile piezoelectric sensors for impact and acoustic emission (AE) detection in CFRP structures. *Structural Damage Assessment Using Advanced Signal Processing Procedures Proceedings of DAMAS '99*. University College, Dublin, 1999; pp. 102–111.

[14] Hillary B, Ewins DJ. The use of strain gauges in force determination and frequency response function measurements. *Proceedings of the International Modal Analysis Conference*. Orlando, FL, 1984; Vol. 2, pp. 627–634.

[15] Rizkalla S, Benmokrane B, Tadros G. Structural health monitoring bridges with fiber optic sensors. *European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000; pp. 501–510.

[16] Kim K, Paik S-H. Optical fiber monitoring system of bridges in Korea. *Proceedings of the 1st International Workshop on Structural Health Monitoring, Current Status and Perspectives*. Stanford University, Palo Alto, CA, 1997; pp. 555–563.

[17] Todd MD, Johnson G, Vohra S, Chen-Chang C, Danver B, Malsawma L. Civil Infrastructure monitoring with fiber optic Bragg grating sensor arrays. *Proceedings of the 2nd International Workshop on Structural Health Monitoring 2000*, Stanford University, Palo Alto, CA, 1999; pp. 359–368.

[18] Todd M, Johnson G, Vohra S. Progress towards deployment of Bragg grating-based fiber optic systems in structural monitoring applications. *European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000; pp. 521–530.

[19] Lhermet N, Claeyssen F, Bouchilloux P. Electromagnetic stress sensor and its applications: monitoring bridge cables and prestressed concrete structures smart systems for bridges, structures, and highways. *Proceedings of SPIE*, San Diego, CA, 1998; Vol. 3325, pp. 46–52.

[20] Satpathi D, Victor J, Wang M, Yang H, Shih C. Development of a PVDF film sensor for infrastructure monitoring. *Smart Structures and Materials 1999: Smart Systems for Bridges, Structures, and Highways, Proceedings of SPIE*. Newport Beach, CA, 1999; Vol. 3671, pp. 90–99.

[21] Johnson TJ, Yang C, Adams DE, Ciray S. Embedded sensitivity functions for characterizing structural damage. *Smart Materials and Structures* 2005 **14**:155–169.

[22] Schulz MJ, Naser AS, Thyagarajan SK, Mickens T, Pai PF. Structural health monitoring using frequency response functions and sparse measurements. *Proceedings of the International Modal Analysis Conference*. Santa Barbara, CA, 1998; Vol. 16, pp. 760–766.

[23] Trendafilova I. Damage detection in structures from dynamic response measurements, an inverse problem perspective. *Modeling and Simulation Based Engineering*. Technical Science Press, 1998, pp. 515–520.

[24] Roberts JB, Dunne JF, Debunos A. A spectral method for estimation of non-linear system parameters from measured response. *Probabilistic Engineering Mechanics* 1995 **10**:199–207.

[25] Kammer DC. Input force reconstruction using a time domain technique. *American Institute of Aeronautics and Astronautics (AIAA) Dynamics Specialists Conference*. Salt Lake City, UT, 1996; Technical Papers A96-27111 06-39.

[26] Steltzner AD, Kammer DC. Input force estimation using an inverse structural filter. *Proceedings of the IMAC XVII*. Kissimmee, FL, 1999; pp. 954–960.

[27] Carne TG, Mayes RL, Batemen VI. Force reconstruction using the sum of weighted accelerations technique—max-flat procedure. *Proceedings of the IMAC XII*. Honolulu, HI, 1994; pp. 1054–1062.

[28] Hundhausen RJ, Adams DE, Derriso MM. Impact loads identification in standoff metallic thermal protection system panels. *Proceedings of SPIE—The International Society for Optical Engineering*, San Diego, CA, 2005; Vol. 5768, pp. 145–156.

[29] Roggenkamp T. *An Investigation of the Indirect Measurement of Broadband Force Spectra*, Ph.D. Dissertation. Purdue University: West Lafayette, IN, 1992.

[30] Stites NA. *Minima-Sensing, Passive and Semi-Active Load and Damage Identification Techniques for Structural Components*, MS Thesis. Purdue University: West Lafayette, IN, 2007.

[31] Seydel R, Chang FK. Impact identification of stiffened composite panels: I. system development. *Smart Materials and Structures* 2001 **10**:354–369.

[32] Seydel R, Chang FK. Impact identification of stiffened composite panels: II. implementation studies. *Smart Materials and Structures* 2001 **10**:370–379.

[33] Decker M, Savaidis G. Measurement and analysis of wheel loads for design and fatigue evaluation of vehicle chassis components. *Fatigue and Fracture of Engineering Materials and Structures* 2002 **25**(12):1103.

[34] O'Connor C, Chan THT. Dynamic wheel loads from bridge strains. *Journal of the Structural Division-ASCE* 1998 **114**(8):1703–1723.

[35] Chan THT, Law SS, Yung TH, Yuan XR. An interpretive method for moving force identification. *Journal of Sound and Vibration* 1999 **219**(3):503–524.

[36] Zhu XQ, Law SS. Identification of vehicle axle loads from bridge responses. *Journal of Sound and Vibration* 2000 **236**(4):705–724.

[37] Haas DJ, Milano J, Flitter L. Prediction of helicopter component loads using neural networks. *Journal of the American Helicopter Society* 1995 **1**:72–82.

[38] Teal RS, Evernham JT, Larchuk TJ, Miller DG, Marquith DE, White F, Deibler DT. Regime recognition for mh47e structural usage monitoring. *Proceedings of the American Helicopter Society 35th Annual Forum*. Virginia Beach, VA, 1997; pp. 1267–1284.

[39] Uhl T. Identification of loads in mechanical structures—helicopter case study. *Computer Assisted Mechanics and Engineering Sciences* 2002 **9**: 151–160.

[40] Zion L. Predicting fatigue loads using regression diagnostics. *Proceedings of the American Helicopter Society 50 Annual Forum*. Washington, DC, 1994; pp. 1337–1358.

[41] Uhl T, Pieczara J. Identification of operational loading forces for mechanical structures. *The Archives of Transport* 2003 **16**(2):109–126.

[42] Haroon M, Adams DE. Active and event-driven passive mechanical fault identification in ground vehicle suspension systems. *Proceedings of IMECE: ASME International Mechanical Engineering Congress and Exposition, Dynamic Systems and Control Division*. Orlando, FL, 2005; pp. 1113–1120.

[43] Haroon M, Adams DE. Component restoring force identification for damage identification in vehicle suspension systems. *International Journal of Vehicle System Modeling and Testing* 2007 (accepted for publication).

[44] Masri SF, Caughey TK. A nonparametric identification technique for nonlinear dynamic problems. *ASME Journal of Applied Mechanics* 1979 **46**:433–447.

[45] Masri SF, Sassi H, Caughey TK. Nonparametric identification of nearly arbitrary nonlinear systems. *ASME Journal of Applied Mechanics* 1982 **49**: 619–628.

[46] Masri SF, Miller RK, Saud AF, Caughey TK. Identification of nonlinear vibrating structures: part I—formulation. *ASME Journal of Applied Mechanics* 1987 **54**:918–922.

[47] Masri SF, Miller RK, Saud AF, Caughey TK. Identification of nonlinear vibrating structures: part II—applications. *ASME Journal of Applied Mechanics* 1987 **54**:923–930.

[48] Crawley EF, Aubert AC. Identification of nonlinear structural elements by force-state mapping. *AIAA Journal* 1986 **24**(1):155–162.

[49] Brandon JA. Structural damage identification of systems with strong nonlinearities: a qualitative identification methodology. Structural damage assessment using advanced signal processing procedures. *Proceedings of DAMAS '97*. University of Sheffield, 1997; pp. 287–298.

[50] Brandon JA. Towards a nonlinear identification methodology for mechanical signature analysis. *Damage Assessment of Structures, Proceedings of the International Conference on damage Assessment of Structures (DAMAS '99)*. Dublin, 1999; pp. 265–272.

[51] Tsyfansky SL, Beresnevich VI. Vibrodiagnosis of fatigue cracks in geometrically nonlinear beams. *Structural Damage Assessment Using Advanced Signal Processing Procedures, Proceedings of DAMAS '97*. University of Sheffield, 1997; pp. 299–311.

[52] Peng ZK, Lang ZQ, Billings SA. Crack detection using nonlinear output frequency response functions. *Journal of Sound and Vibration* 2007 **301**(3–5):777–788.

[53] Haroon M, Adams DE. Identification of damage in a suspension component using narrowband and broadband nonlinear signal processing techniques. Health monitoring of structural and biological systems. *Proceedings of the SPIE* 2007 **6532**:65320Y1–65320Y12.

[54] Zhang G, Chen J, Li F, Li W. Extracting gear fault features using maximal bispectrum. *Key Engineering Materials* 2005 **293–294**:167–174.

[55] McCormick AC, Nandi AK. Bispectral and trispectral features for machine condition diagnosis. *IEE Proceedings—Vision Image and Signal Processing* 1999 **146**(5):229–234.

[56] Hillis AJ, Neild SA, Drinkwater BW, Wilcox PD. Global crack detection using bispectral analysis. *Proceedings of the Royal Society A* 2006 **462**(2069):1515–1530.

[57] Adams DE, Farrar CR. Application of frequency domain arx features for linear and nonlinear structural damage identification. *Proceedings of SPIE's 9th Annual International Symposium on Smart Structures and Materials*. San Diego, CA, 2002; Vol. 4702, pp. 134–147.

[58] Haroon M, Adams DE. Time and frequency domain nonlinear system characterization for mechanical fault identification. *Nonlinear Dynamics—Special Issue on Discontinuous Dynamical Systems* 2007 **50**(3):387–408.

[59] Todd MD, Nichols JM, Pecora LM, Virgin LN. Vibration-based damage assessment utilizing state

space geometry changes: local attractor variance ratio. *Smart Materials and Structures* 2001 **10**:1000–1008.

[60] Moniz L, Nichols JM, Nichols CJ, Seaver M, Trickey ST, Todd MD, Pecora LM, Virgin LN. A Multivariate, attractor-based approach to structural health monitoring. *Journal of Sound and Vibration* 2005 **283**:295–310.

[61] Moniz L, Pecora L, Nichols J, Todd M, Wait JR. Dynamical assessment of structural damage using the continuity statistic. *Structural Health Monitoring* 2004 **3**(3):199–212.

[62] Stites N, Adams DE, Sterkenburg R, Ryan T. Integrated health monitoring of gas turbine engine wire harnesses and connectors. *Proceedings of the 3rd European Workshop on Structural Health Monitoring*. Granada, 5–7 July 2006; pp. 996–1003.

[63] Ackers S, Evans R, Johnson T, Kess H, White J, Adams DE. Crack detection in a wheel end spindle using wave propagation via modal impacts and piezo actuation. *Health Monitoring and Smart Nondestructive Evaluation of Structural and Biological Systems V*, *Proceedings of the SPIE*. SPIE, 2006; Vol. 6177, pp. 104–116.

[64] Cornwell P, Farrar CR, Doebling SW, Sohn H. Environmental variability of modal properties. *Experimental Techniques* 1999 **23**(6):45–48.

[65] Kess HR, Adams DE. A sensitivity method for modeling the effects of operational and environmental variability in structural damage detection. *Proceedings of the IMAC XXIV*, Paper No. 17. St. Louis, MO, 2006.

[66] Zhang H, Schulz MJ, Naser A, Ferguson F, Pai PF. Structural health monitoring using transmittance functions. *Mechanical Systems and Signal Processing* 1999 **13**(5):765–787.

[67] Johnson TJ, Adams DE. Transmissibility as a differential indicator of structural damage. *ASME Journal of Vibration and Acoustics* 2002 **124**(4): 634–641.

[68] Peeters JMB, De Roeck D. Damage identification on the Z24-bridge using vibration monitoring. *European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000; pp. 233–242.

[69] Ayers JW, Rogers CA, Chaudhry ZA. Qualitative health monitoring of a steel bridge joint via piezoelectric actuator/sensor patches. *Proceedings of SPIE* 1996 **719**(2):123–131.

[70] Pardo de Vera C, Guemes J. Embedded self-sensing piezoelectric for damage detection. *Journal of Intelligent Material Systems and Structures* 1998 **9**(11):876–882.

[71] Giurgiutiu V, Zagrai A. Damage detection in thin plates and aerospace structures with electro-mechanical impedance method. *Structural Health Monitoring* 2005 **4**(2):99–118.

[72] Wang CS, Chang FK. Diagnosis of impact damage in composite structures with built-in piezoelectrics network. *Smart Structures and Materials 2000: Smart electronics and MEMS, Proceedings of SPIE*. Newport Beach, CA, 2000; Vol. 3990, pp. 13–19.

[73] Giurgiutiu V. Tuned lamb-wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. *Journal of Intelligent Material Systems and Structures* 2005 **16**(4):291–306.

[74] Sundararaman S, Adams DE, Rigas E. Structural damage identification in homogeneous and heterogeneous structures using beamforming. *Structural Health Monitoring* 2005 **4**(2):171–190.

[75] White J, Adams DE, Jata K. Damage identification of a sandwich plate using the method of virtual forces in semi-realistic conditions. *Proceedings of the European Workshop on Structural Health Monitoring*. Granada, 2006.

[76] White J, Adams D, Jata K. Actuator-sensor pair excitation tuning and self diagnostics for damage identification of a sandwich plate. *Proceedings of the International Workshop on Structural Health Monitoring*. Stanford, CA, 2007; Vol. 1, pp. 669–676.

[77] Heyns PS. Structural damage assessment using response-only measurements. *Structural Damage Assessment Using Advanced Signal Processing Procedures, Proceedings of DAMAS '97*. University of Sheffield, 1997; pp. 213–223.

[78] Sohn H, Farrar CR. Statistical process control and projection techniques for damage detection. *European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2005; pp. 105–114.

[79] Bodeaux JB, Golinval JC. ARMAV model technique for system identification and damage detection. *European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000; pp. 303–312.

[80] Koh SJA, Maalej M, Quek ST. Damage quantification of flexurally loaded RC slab using frequency response data. *Structural Health Monitoring* 2004 **3**(4):293–311.

[81] Wang WJ, McFadden PD. Application of wavelets to gearbox vibration signals for fault detection. *Journal of Sound and Vibration* 1996 **192**(5):927–939.

[82] Sun Z, Chang CC. Structural damage assessment based on wavelet packet transform. *Journal of Structural Engineering* 2002 **128**(10):1354–1361.

[83] Hou Z, Noori M, St. Amand R. Wavelet-based approach for structural damage detection. *Journal of Engineering Mechanics* 2000 **126**(7):677–683.

[84] Staszewski WJ, Tomlinson GR. Application of the wavelet transform to fault detection in a spur gear. *Mechanical System and Signal Processing* 1994 **8**(3):289–307.

[85] Johnson TJ, Adams DE. Rolling tire diagnostic experiments for identifying incipient bead damage using time, frequency, and phase-plane analysis (invited paper). *Proceedings of the Society of Automotive Engineering World Congress, SAE*, Detroit, MI, Paper No. 2006-01-1621, ISBN No. 0-7680-1768-8, 2006.

[86] Haroon M. *A Methodology for Mechanical Diagnostics and Prognostics to Assess Durability of Ground Vehicle Suspension Systems*, Ph.D. Dissertation. Department of Mechanical Engineering, Purdue University: West Lafayette, IN, 2007.

[87] Farrar CR, Worden K, Todd MC, Park G, Nichols J, Adams DE, Bement MT, Farinholt K. *Nonlinear System Identification for Damage Identification*, Los Alamos National Laboratory Report LA-14353-MS. Los Alamos National Laboratory, November 2007.

[88] Farrar CR, Cornwell PJ, Doebling SW, Prime MB. *Structural Health Monitoring Studies of the Alamosa Canyon and I-40 Bridges*, Los Alamos National Laboratory Report LA-13635-MS. Los Alamos National Laboratory, 2000.

[89] Mitchell JS. *Introduction to Machinery Analysis and Monitoring*. PenWel Books: Tulsa, OK, 1993.

[90] Rao JS. *Vibratory Condition Monitoring of Machines*. CRC Press, 2000.

[91] Fugate ML, Sohn H, Farrar CR. Vibration-based damage detection using statistical process control. *Mechanical Systems and Signal Processing* 2001 **15**(4):707–721.

[92] Thouverez F, Jezequel L. Identification of NARMAX models on a modal base. *Journal of Sound and Vibration* 1996 **189**(2):193–213.

[93] Billings SA, Chen S, Backhouse RJ. Identification of linear and nonlinear models of a turbocharged automotive diesel engine. *Journal of Mechanical Systems and Signal Processing* 1989 **3**:123–142.

[94] Rivola A, White PR. Bispectral *Analysis of the Bilinear Oscillator with Application to the Detection of Fatigue Cracks*. *Journal of Sound and Vibration* 1998 **216**:889–910.

[95] George D, Norman H, Farrar CR, Deen R Identifying damage sensitive features using nonlinear time-series and bispectral analysis. *Proceedings of the IMAC XVIII*. San Antonio, TX, 2000; pp. 1796–1802.

[96] Giurgiutiu V, Cuc A. Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention. *The Shock and Vibration Digest* 2005 **37**(2):83–105.

[97] Peng G, Yuan S-F, Xu Y. Damage location of two-dimensional structure based on wavelet transform and active monitoring technology of lamb wave. *Proceedings of SPIE—Nondestructive Evaluation and Health Monitoring of Aerospace Materials and Composites III*, San Diego, CA, 2004; Vol. 5393, pp. 151–160.

[98] Doebling SW, Hemez FM, Peterson LD, Farhat C. Improved damage location accuracy using strain energy based on mode selection criteria. *AIAA Journal* 1997 **35**(4):693–699.

[99] Shi ZY, Law SS, Zhang LM. Structural damage localization from modal strain energy change. *Journal of Sound and Vibration* 1998 **218**(5): 825–844.

[100] Shi ZY, Law SS, Zhang LM. Improved damage quantification from elemental modal strain energy change. *Journal of Engineering and Mechanics* 2002 **128**(5):521–529.

[101] Teughels A, Maeck J, Roeck GD. Damage assessment by FE model updating using damage functions. *Computers and Structures* 2002 **80**(25): 1869–1879.

[102] Farrar CR, Lieven NAJ. Damage prognosis: the future of structural health monitoring. *Philosophical Transactions of the Royal Society A* 2007 **365**:623–632.

[103] Silverman H. T-HUMS—AH64 lead the fleet (LTF) summary and glimpse at hermes 450 MT-HUMS. *AIAC Conference*, Paper 151. Melbourne, 2005.

[104] Ellingwood B. *Validation of Seismic Probabilistic Risk Assessment of Nuclear Power Plants*. Division of Engineering, Office of Nuclear Regulatory

Research, US Nuclear Regulatory Commission, 1994.

[105] Rammohan R, Taha MR. Exploratory investigations for intelligent damage prognosis using hidden Markov models. *IEEE International Conference on Systems, Man and Cybernetics*, Waikoloa, 2005; Vol. 2, pp. 1524–1529.

[106] Liu KC, Wang JA. An energy method for predicting fatigue life, crack orientation, and crack growth under multiaxial loading conditions. *International Journal of Fatigue* 2001 **23**:129–134.

[107] Inman J, Farrar CR, Lopes V, Steffen V. *Damage Prognosis for Aerospace, Civil and Mechanical Systems*. John Wiley & Sons: New York, 2005.

FURTHER READING

Rucka M, Wilde K. Damage location in beam and plate structures by wavelet analysis of experimentally determined mode shapes. *Key Engineering Materials* 2005 **293–294**:313–320.

Chapter 3

Fundamentals of Guided Elastic Waves in Solids

Carlos E. S. Cesnik and Ajay Raghavan

Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA

1 INTRODUCTION

Unlike passive structural health monitoring (SHM) methods, active schemes are capable of exciting the structure, and in a repeatable, prescribed manner, they can quickly examine it for damage, where and when required. Guided-wave (GW) testing has emerged as a promising option among active schemes. It can offer an effective method to estimate the location, severity, and type of damage, and it is a well-established practice in the nondestructive evaluation and testing (NDE/NDT) industry. There, GWs are excited and received in a structure using hand-held transducers for scheduled, offline maintenance. They have also demonstrated suitability for SHM applications having an onboard, preferably built-in, sensor and actuator network to assess the state of a structure during operation. The actuator–sensor pair in GW testing has a large coverage area, resulting in fewer units distributed over the structure.

GWs can be defined as stress waves forced to follow a path defined by the material boundaries of the structure. For example, when a beam is excited at high frequency, stress waves travel in the beam along its axis away from the excitation source, i.e., the beam "guides" the waves along its axis. Similarly, in a plate, the two free surfaces of the plate "guide" the waves in the plane of the plate. In GW SHM, an actuator generating GWs is excited by some high-frequency pulse signal (typically a modulated sinusoidal toneburst of some limited number of cycles). In general, when a GW field is incident on a structural discontinuity, which has a size comparable to the GW wavelength, the wave scatters GWs in all directions. The structural discontinuity could be damage in the structure such as a crack or delamination; a structural feature (such as a stiffener); or a structural boundary. Therefore, to be able to distinguish between damage and structural features, one needs prior information about the structure in its undamaged state. This prior information is typically in the form of a baseline signal obtained for the "healthy state"

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

and is used as reference for comparison with the test case. There are two approaches commonly used in GW SHM: pulse echo and pitch–catch. In the former, after exciting the structure with a narrow bandwidth pulse, a sensor collocated with the actuator is used to "listen" for echoes of the pulse coming from discontinuities. Since the boundaries and the wave speed for a given center actuation frequency of the toneburst are known, the signals from the structural features can be filtered out (alternatively, one could subtract the test signal from the baseline signal). One is then left with signals from the defects (if present). From these signals, defects can be located using the wave speed. In the pitch–catch approach, a pulse signal is sent across the specimen under interrogation and a sensor at the other end of the specimen receives the signal. From various characteristics of the received signal, such as delay in time of transit, amplitude, frequency content, etc., information about the damage can be obtained. Thus, the pitch–catch approach cannot be used to locate the defect unless a dense network of transducers is used. In either approach, damage-sensitive features are extracted from the signal using some signal-processing algorithm (*see* **Signal Processing for Damage Detection**; **Wavelet Analysis**; **Damage Detection Using the Hilbert–Huang Transform**), and then a pattern-recognition technique (*see* **Statistical Pattern Recognition**; **Machine Learning Techniques**; **Artificial Neural Networks**) is required to classify the damage and estimate its severity. These steps involved in GW

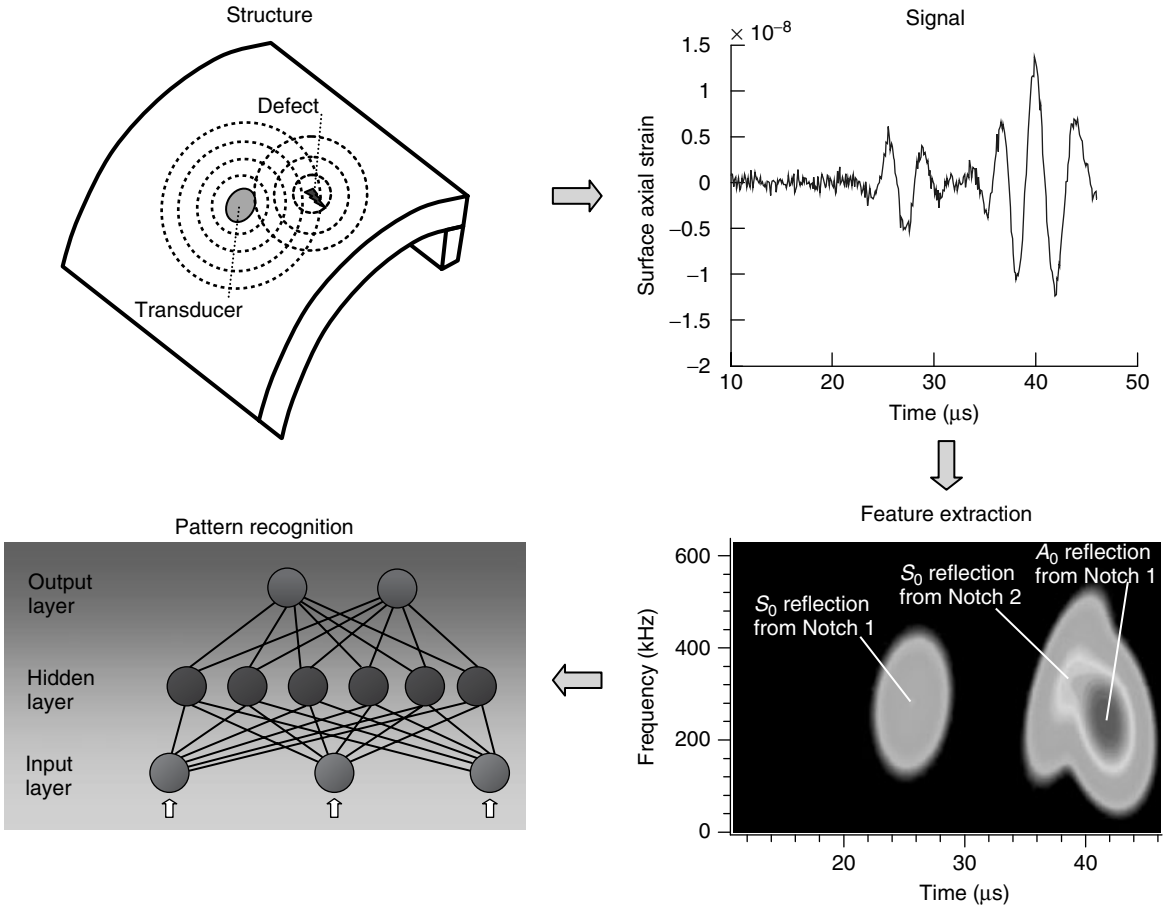


Figure 1. The four essential steps in GW SHM.

SHM are illustrated in Figure 1. Another crucial point to note is that GW SHM always involves the use of some threshold value to decide whether damage is present in the structure. The choice of the threshold is usually application dependent and typically relies on some false-positive probability estimation. A detailed survey paper on this technology, outlining efforts by various researchers in different aspects of this technology and application areas, is presented in [1].

There are several application areas for GWs in solids such as seismology, inspection, material characterization, electronic delay lines, etc., and consequently, they have been a subject of much study [2–5]. A very important class among these areas is that of Lamb waves, which can propagate in a solid plate (or shell) with free surfaces. Because of the abundance of plate- and shell-like structural configurations, this class of GWs has been the subject of much scrutiny. Another class of GW modes is also possible in plates, i.e., the horizontally polarized shear or SH modes. Other classes of GWs have also been examined in the literature. Among them is that of Rayleigh waves [6], which propagate close to the free surface of elastic solids. Other examples are Love [7], Stoneley [8], and Scholte [9] waves that travel at material interfaces. Lamb waves were first predicted mathematically and described by Horace Lamb [10]. Gazis [11] developed and analyzed the dispersion equations for GWs in hollow cylinders. However, neither was able to produce GWs experimentally. This was first done by Worlton [12], who was probably also the first person to recognize the potential of GWs for NDE.

GW testing can become complex owing to the multimodal and dispersive (dependence of wave speed on frequency) nature of these waves. In anisotropic structures, additionally, there can be significant dependence of wave speed on the direction of propagation and "steering" (the tendency of waves to move in a direction different from the launched direction). Hence, a fundamental understanding of wave propagation is essential for the successful application of this method. This understanding can also be very useful for better implementation of acoustic emission (which is a passive SHM method *see* **Acoustic Emission**), where sensors on the structure are tuned to sense GWs excited by damage initiation or growth.

2 FUNDAMENTALS OF ELASTIC WAVE PROPAGATION

The problem of wave propagation is presented below in three steps. First, one-dimensional (1D) wave propagation in a string under tension is introduced to establish some of the basic principles and nomenclature used in the field. This analysis is followed by a general 3D representation of an isotropic solid and analysis of free GWs in an infinite plate. GW excitation by finite-dimensional piezos in isotropic plates is covered next. Finally, wave propagation in transversely isotropic solids is analyzed and wave solutions are found for the GW modes in a multilayered laminated plate.

2.1 1D wave propagation

2.1.1 Introduction

To understand some basic concepts on wave propagation, consider the 1D wave propagation problem along the x direction. The problem of waves in a taut string serves as a good model problem for 1D wave propagation, as illustrated in Figure 2. Consider the following assumptions:

1. The restoring force in the string is due to the tension T.
2. The deflections w are small as are the angles of the string with respect to the horizontal at various points.
3. T is much larger in magnitude than the dynamic strain fluctuations.

From the equilibrium of an infinitesimal piece of the string, one obtains

$$T\frac{\partial^2 w}{\partial x^2} + q = \rho_L \frac{\partial^2 w}{\partial t^2} \tag{1}$$

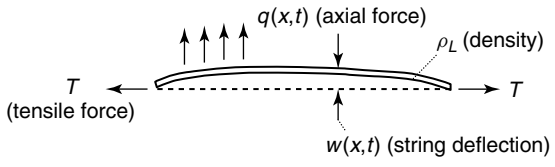


Figure 2. String under tension.

For the free vibrations case, $q = 0$ and equation (1) simplifies to

$$\frac{\partial^2 w}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 w}{\partial t^2} = 0; \quad v^2 = \frac{T}{\rho_L} \tag{2}$$

where v represents the square of the wave speed.

2.1.2 D'Alembert's solution for an infinite string

One can show that

$$w(x,t) = f(x - vt) + g(x + vt) \tag{3}$$

is a solution of equation (2), where f and g are arbitrary functions. This solution can be verified by direct substitution:

$$\frac{\partial^2 f(x - vt)}{\partial x^2} = f''; \quad \frac{\partial^2 f(x - vt)}{\partial t^2} = v^2 f'' \tag{4}$$

Therefore,

$$\text{LHS} = \frac{1}{v^2}(v^2 f'' + v^2 g'') = f'' + g'' = \text{RHS} \tag{5}$$

This solution for the 1D wave equation, known as *D'Alembert's solution*, can thus be used to solve initial value problems on infinite strings. For example, consider the following initial conditions:

$$w(x,0) = U(x); \quad \dot{w}(x,0) = V(x) \tag{6}$$

After inserting these conditions into D'Alembert's solution, equation (3) yields

$$w(x,t)|_{t=0} = f(x) + g(x) = U(x) \tag{7}$$

$$\dot{w}(x,t)|_{t=0} = v(-f'(x) + g'(x)) = V(x) \tag{8}$$

where $()'$ represents the derivative with respect to x. These two equations yield

$$w(x,t) = \frac{1}{2}[U(x - vt) + U(x + vt)] + \frac{1}{2v}\int_{x-vt}^{x+vt} V(s)\,\mathrm{d}s \tag{9}$$

To aid visualization, consider the simple case $V(x) = 0, U(x) = He(x - a) - He(x + a)$, where $He()$ is the Heaviside function. This case is illustrated in Figure 3.

2.1.3 Reflection from a fixed boundary

Consider a string fixed at one end and a wave pulse $f(x - vt)$ approaching the fixed end from $-\infty$. One can solve for the reflected wave from the fixed end at $x = 0$, which imposes the condition $w(0,t) = 0$, by creating an imaginary wave in the $+x$ axis. For this condition, as illustrated in Figure 4, imagine the string to be extended symmetrically to $+\infty$ and further consider a negative (of equal magnitude as the original pulse, but opposite in sign) wave pulse

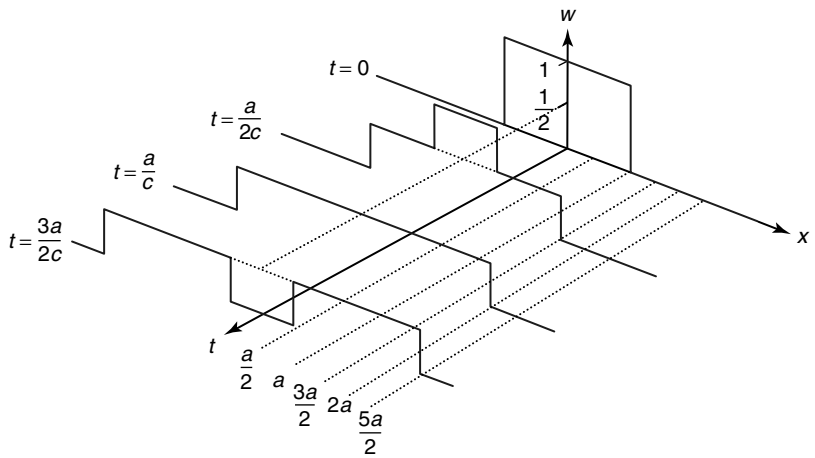


Figure 3. Simple example of transient wave propagation in a 1D string.

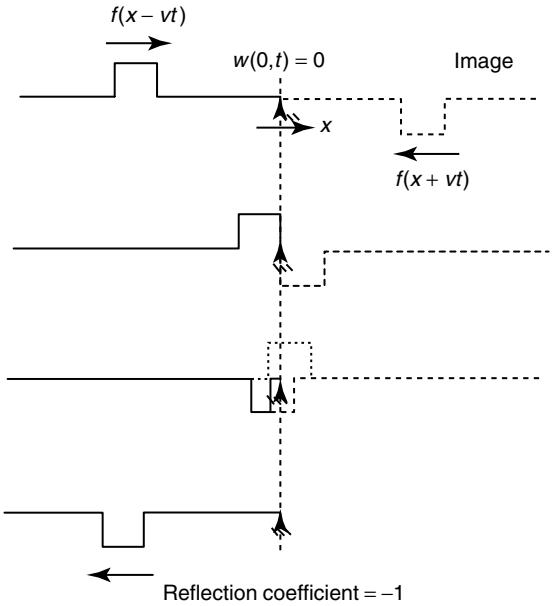


Figure 4. Reflection of a wave pulse in a string from a fixed boundary.

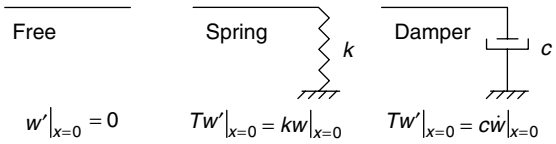


Figure 5. Other boundary conditions for the string problem.

approaching the fixed end simultaneous to the original pulse. As the two waves approach the fixed end, their superposition will ensure the satisfaction of the fixed boundary condition at that end. Since the superposition of the two waves is a feasible solution, by the uniqueness condition, it is indeed the solution. This solution is illustrated in Figure 4. As shown there, for a fixed end, the incident wave packet simply bounces off with its sign reversed. Other boundary conditions are described in Figure 5.

2.1.4 Reflection and buildup of standing waves in finite domains

Consider a finite string of length L with one end fixed and the other being harmonically excited, as shown in Figure 6. Thus, its boundary conditions are

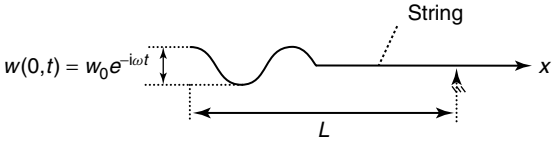


Figure 6. Buildup of waves in a finite string.

$$w(0,t) = w_0 \mathrm{e}^{-\mathrm{i}\omega t}; \quad w(L,t) = 0 \tag{10}$$

Before the rightward wave arrives at the fixed end, it propagates as if it were in an infinite string. Therefore, for the transient wave, D'Alembert's solution for an infinite string applies for this case for $t < L/v$.

The reflection coefficients γ_s at $x = 0$ and γ_e at $x = L$ are defined as follows:

$$\gamma_\mathrm{s} = \left.\frac{w_\mathrm{reflected}}{w_\mathrm{incident}}\right|_{x=0}; \quad \gamma_\mathrm{e} = \left.\frac{w_\mathrm{reflected}}{w_\mathrm{incident}}\right|_{x=L} \tag{11}$$

Since the boundary at $x = L$ is fixed, $\gamma_\mathrm{e} = -1$. Further, since only rightward propagating waves are sought,

$$w(x,t) = f(x - vt) \tag{12}$$

Now using the boundary condition at $x = 0$,

$$w(0,t) = f(-vt) = w_0 \mathrm{e}^{-\mathrm{i}\omega t} \tag{13}$$

One can identify the unknown function f as

$$f(\,) = w_0 \mathrm{e}^{\mathrm{i}\frac{\omega}{v}(\,)} \tag{14}$$

and, therefore, until $vt < L$,

$$\begin{aligned} w(x,t) &= w_0 \mathrm{e}^{\mathrm{i}(\xi x - \omega t)}; \quad \xi = \frac{\omega}{v} \quad \text{for } x < vt \\ &= 0 \qquad\qquad\qquad\quad \text{for } x > vt \end{aligned} \tag{15}$$

The symbols ω, ξ, and v introduced here are commonly encountered when analyzing harmonic waves in a medium. ω is called the *angular frequency*, and is related to the frequency of harmonic oscillations f by the relation

$$\omega = 2\pi f \tag{16}$$

ξ is the wavenumber and is related to the spatial wavelength λ_w of the wave by the following relation:

$$\xi = \frac{2\pi}{\lambda_w} \tag{17}$$

As can be seen from equation (17), the wavenumber is proportional to the number of wavelengths per unit distance. v (also denoted by v_{ph}) is called the *phase speed* and is the speed at which the harmonic wave propagates through the medium. Note that here the wave speed is independent of frequency. It is related to ω and ξ by the equation

$$\frac{\omega}{v} = \xi \tag{18}$$

When the disturbance reaches the fixed end at $x = L$ (at $t = L/v$) for the first time, the solution must be augmented by a leftward wave according to what was described in Section 2.1.4. Therefore, the total solution for $t > L/v$ is

$$w^1{}_{\text{total}} = (w_0 e^{i\xi x}) + (w_0 \gamma_e e^{i\xi L}\ e^{-i\xi(x-L)})e^{-i\omega t} \tag{19}$$

Rightward wave — Leftward wave

The first factor $e^{i\xi L}$ of the new second term appearing in the expression for w is needed to cancel out the first term (i.e., maintain $w = 0$ at $x = L$ as imposed by the boundary condition). The second factor of the second term $e^{-i\xi(x-L)}$ is the resultant leftward wave that has just started.

Once the wave travels all the way back to the left (length L), it will arrive at $x = 0$ at $t = 2L/v$. Again, to satisfy the boundary condition at this point, yet another wave is introduced, so that the new resultant wave is

$$w^2_{\text{total}} = w^1_{\text{total}} + \gamma_s \gamma_e w_0 e^{2i\xi L} e^{i(\xi x - \omega t)} \tag{20}$$

This goes on, and each time the wave hits one of the string boundaries, a new term is added. The terms of this expression are in geometric progression and can be written as

$$\begin{aligned} w(x,t) = w_0 \big(&e^{i\xi x} + \gamma_e e^{i\xi(2L-x)} + \gamma_s \gamma_e e^{i\xi(2L+x)} \\ &+ \gamma_s \gamma_e^2 e^{i\xi(4L-x)} \\ &+ \gamma_s^2 \gamma_e^2 e^{i\xi(4L+x)} + \cdots\big) e^{-i\omega t} \end{aligned} \tag{21}$$

For the steady-state solution, the sum of an infinite number of these terms is needed, and this can be expressed using the formula for geometric series yielding

$$w(x,t) = w_0 \left[\frac{e^{i\xi x} + \gamma_e e^{i\xi(2L-x)}}{1 - \gamma_e \gamma_s e^{2i\xi L}}\right] e^{-i\omega t} \tag{22}$$

Note that the same *steady-state* solution could have been reached by a much simpler approach. Assume a solution of the form

$$w(x,t) = \left(A e^{i\xi x} + B e^{-i\xi x}\right) e^{-i\omega t} \tag{23}$$

and impose $w(0,t) = w_0 e^{-i\omega t}$, $w(L,t) = 0$ to find the constants A and B. However, the earlier approach was adopted to provide an insight into how transients build up into the steady-state solution. A key lesson to learn from the above analysis is that *the solution for the infinite string holds for the finite string for analyzing the transient waves*. This will eventually prove useful in the analysis for GW testing, where essentially transient waves are used, particularly the transmitted pulse and first few echoes. Hence, an infinite plate wave solution can be used as part of the transient wave analysis in finite plates.

2.2 Wave propagation in isotropic solids

The equations governing wave propagation in isotropic plates are presented next. The 1D equations discussed above are first generalized to a 3D solid and then particularized for an infinite plate (or semi-infinite solid).

2.2.1 Isotropic linear elastic 3D solid

The general 3D equilibrium equations of motion for elastic solids are given by

$$\nabla \cdot \boldsymbol{\sigma} + \boldsymbol{f} = \rho \ddot{\boldsymbol{u}} \tag{24}$$

or equivalently by

$$\sigma_{ij,j} + f_i = \rho \ddot{u}_i \tag{25}$$

where the indices i, j, k can assume values 1–3, $\boldsymbol{\sigma}$ is the stress tensor, ρ is the density of the solid, $\boldsymbol{f}$ is the body force per unit volume and $\ddot{\boldsymbol{u}}$ is the acceleration vector. For simplicity, consider the case of a linear elastic isotropic medium. For this medium, the constitutive relations are given by

$$\sigma_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2\mu \varepsilon_{ij} \tag{26}$$

where λ and μ are Lamé's constants for the isotropic medium, δ_{ij} is the Kronecker delta, and ε_{ij} are the components of the strain tensor. To complete the set of elasticity equations, the kinematical relations for the case of geometrically linear deformations are

$$\varepsilon_{ij} = \tfrac{1}{2}\left(u_{i,j} + u_{j,i}\right) \tag{27}$$

Combining these three sets of equations, one can reach the Navier's equation for a 3D isotropic body:

$$(\lambda + \mu)\nabla\nabla\cdot\boldsymbol{u} + \mu\nabla^2\boldsymbol{u} + \boldsymbol{f} = \rho\ddot{\boldsymbol{u}} \tag{28}$$

2.2.2 The Helmholtz decomposition for isotropic media

Consider a decomposition for the displacement vector $\boldsymbol{u}$ into a scalar potential ϕ and a vector $\mathbf{H}$. The physical meaning of these will become evident in due course. This decomposition eases the solution of the problem of wave propagation in a 3D elastic linear isotropic medium. Thus, let the displacement vector be expressed as follows:

$$\boldsymbol{u} = \nabla\phi + \nabla \times \mathbf{H} \tag{29}$$

Furthermore, for uniqueness of the solution, impose the divergence-free condition on $\mathbf{H}$:

$$\nabla\cdot\mathbf{H} = 0 \tag{30}$$

Such a decomposition is always possible and unique, since there are four equations and four unknowns (ϕ and the three components of $\mathbf{H}$) for a given displacement field. Again recalling Navier's elastodynamic equation of motion for zero body force, equation (28), we have

$$(\lambda + \mu)\nabla\nabla\cdot\boldsymbol{u} + \mu\nabla^2\boldsymbol{u} = \rho\ddot{\boldsymbol{u}} \tag{31}$$

The substitution of equations (29) and (30) into equation (31) yields

$$\begin{aligned}(\lambda + \mu)\nabla\left[\nabla\cdot(\nabla\phi + \nabla \times \mathbf{H})\right] \\ + \mu\nabla^2(\nabla\phi + \nabla \times \mathbf{H}) \\ = \rho\left(\nabla\ddot{\phi} + \nabla \times \ddot{\mathbf{H}}\right)\end{aligned} \tag{32}$$

This can be rewritten as

$$\nabla\left\{(\lambda + 2\mu)\nabla^2\phi - \rho\ddot{\phi}\right\} + \nabla \times \left\{\mu\nabla^2\mathbf{H} - \rho\ddot{\mathbf{H}}\right\} = 0 \tag{33}$$

Applying the operators $\nabla\cdot(\,)$ and $\nabla \times (\,)$ on equation (33), one gets

$$(\lambda + 2\mu)\nabla^2\phi - \rho\ddot{\phi} = 0; \quad \mu\nabla^2\mathbf{H} - \rho\ddot{\mathbf{H}} = 0 \tag{34}$$

or

$$\nabla^2\phi = \frac{1}{v_{\mathrm{p}}^2}\ddot{\phi}; \quad \nabla^2\mathbf{H} = \frac{1}{v_{\mathrm{s}}^2}\ddot{\mathbf{H}} \tag{35}$$

Therefore, the original elastodynamic vector equation, equation (31), is decomposed into a scalar wave equation for ϕ and a vector wave equation for $\mathbf{H}$. v_{p} and v_{s} are the wave speeds of the two possible bulk waves in an isotropic unbounded solid, viz., the longitudinal wave (a.k.a. "P" or "dilatational" wave; these are characterized by displacements normal to the propagation direction) and shear wave (a.k.a. "S" or "distortional" wave; these are characterized by displacements normal to the propagation direction).

2.2.3 Rayleigh–Lamb wave equation for an infinite plate

Consider an infinite plate of thickness $2b$, i.e., the domain D: $\{(x_1, x_2, x_3) = (-\infty, \infty) \times (-\infty, \infty) \times (-b, b)\}$ with free surfaces (Figure 7). Waves of interest are in the $x_1 - x_3$ plane, and there are no variations along x_2, i.e., $(\partial/\partial x_2) = 0$.

For this case, Helmholtz's decomposition yields

$$u_1 = \frac{\partial\phi}{\partial x_1} + \frac{\partial H_2}{\partial x_3}; \quad u_3 = \frac{\partial\phi}{\partial x_3} - \frac{\partial H_2}{\partial x_1} \tag{36}$$

$$u_2 = -\frac{\partial H_1}{\partial x_3} + \frac{\partial H_3}{\partial x_1}; \quad \frac{\partial H_1}{\partial x_1} + \frac{\partial H_3}{\partial x_3} = 0 \tag{37}$$

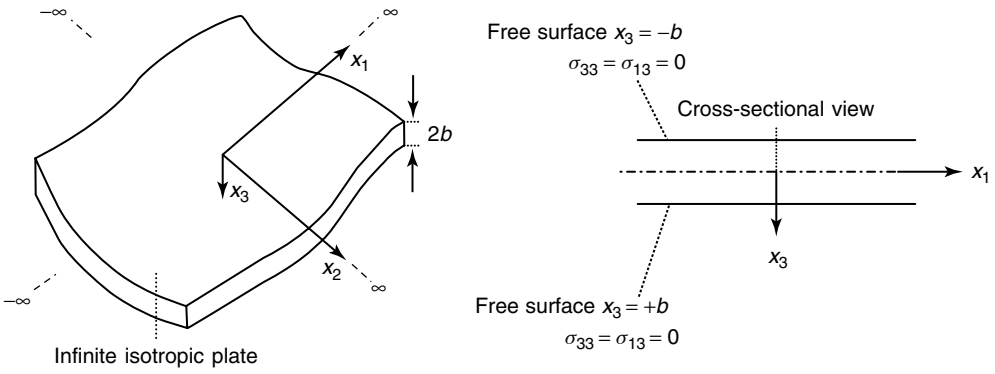


Figure 7. Infinite plate with free surfaces.

Since the boundaries $x_3 = +b$ and $x_3 = -b$ are stress-free, one has

$$\sigma_{33} = 0; \quad \sigma_{13} = \sigma_{31} = 0;$$
$$\sigma_{23} = \sigma_{32} = 0 \text{ at } x_3 = \pm b \tag{38}$$

Using equations (26), (27), and (36), these stress components can be written as

$$\sigma_{33} = (\lambda + 2\mu)\nabla^2\phi - 2\mu\left(\frac{\partial^2\phi}{\partial x_1^2} + \frac{\partial^2 H_2}{\partial x_1 \partial x_3}\right) \tag{39}$$

$$\sigma_{31} = \mu\left(2\frac{\partial^2\phi}{\partial x_1 \partial x_3} + \frac{\partial^2 H_2}{\partial x_3^2} - \frac{\partial^2 H_2}{\partial x_1^2}\right) \tag{40}$$

$$\sigma_{32} = \mu\left(-\frac{\partial^2 H_1}{\partial x_3^2} + \frac{\partial^2 H_3}{\partial x_1 \partial x_3}\right) \tag{41}$$

which will eventually be equated to zero at the plate's free surfaces to find the governing equations. It is interesting to note here that u_1, u_3, σ_{33}, and σ_{31} depend only on ϕ and H_2, while u_2 and σ_{32} depend only on H_1 and H_3. Thus, it suffices to consider the following two cases separately:

Case I

In this case $u_2 = 0$ and the governing equations are

$$\nabla^2\phi = \frac{1}{v_p^2}\ddot{\phi}; \quad \nabla^2 H_2 = \frac{1}{v_s^2}\ddot{H}_2 \tag{42}$$

The surface conditions are

$$\sigma_{33}|_{x_3=\pm b} = 0; \quad \sigma_{13}|_{x_3=\pm b} = 0 \tag{43}$$

Case II

In this case the governing equation is

$$\nabla^2 u_2 = \frac{1}{v_s^2}\ddot{u}_2 \tag{44}$$

The surface condition is

$$\sigma_{32}|_{x_2=\pm b} = 0 \tag{45}$$

As indicated by the previous equations, while the equations (42) are uncoupled, their solutions are coupled through the surface traction-free conditions. As a result, the combination of waves governed by equations (42) in a plate (known as *Lamb waves*) is always "dispersive," i.e., the wave velocity depends on the frequency of the traveling wave. On the other hand, for the waves governed by equation (44), called *SH* or *horizontal shear waves*, there exists a fundamental dispersionless mode at all frequencies.

To derive the equations for Lamb waves, consider first the governing equations in terms of the Helmholtz scalar and vector potentials, i.e., equations (42). Seeking plane wave solutions in the $x_1 - x_3$ plane for waves propagating along the $+x_1$ direction, assume solutions of the form

$$\phi = f\left(x_3\right)\mathrm{e}^{\mathrm{i}(\xi x_1 - \omega t)}; \quad H_2 = h_2\left(x_3\right)\mathrm{e}^{\mathrm{i}(\xi x_1 - \omega t)} \tag{46}$$

Using equations (46) and (42), one gets

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x_3^2} + \left(\frac{\omega^2}{v_p^2} - \xi^2\right) f = 0;$$

$$\frac{\mathrm{d}^2 h_2}{\mathrm{d}x_3^2} + \left(\frac{\omega^2}{v_\mathrm{s}^2} - \xi^2\right) h_2 = 0 \tag{47}$$

Let

$$\frac{\omega^2}{v_\mathrm{p}^2} - \xi^2 \equiv \alpha^2; \quad \frac{\omega^2}{v_\mathrm{s}^2} - \xi^2 \equiv \beta^2 \tag{48}$$

The solutions to the differential equations in (47) are

$$\begin{aligned} f\left(x_3\right) &= A \sin \alpha x_3 + B \cos \alpha x_3; \\ h_2\left(x_3\right) &= C \sin \beta x_3 + D \cos \beta x_3 \end{aligned} \tag{49}$$

Substituting these solutions into equations (39), (40), and (46) and enforcing the conditions given by equations (43), two possible solutions result:

1. Symmetric modes

$$\begin{bmatrix} -\left(\xi^2 - \beta^2\right)\cos \alpha b & 2i\xi\beta \cos \beta b \\ -2i\xi\alpha \sin \alpha b & \left(\xi^2 - \beta^2\right)\sin \beta b \end{bmatrix} \times \begin{bmatrix} B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{50}$$

Nontrivial solutions for the constants B and C exist if and only if

$$\det \begin{bmatrix} -\left(\xi^2 - \beta^2\right)\cos \alpha b & 2i\xi\beta \cos \beta b \\ -2i\xi\alpha \sin \alpha b & \left(\xi^2 - \beta^2\right)\sin \beta b \end{bmatrix} = 0 \tag{51}$$

where "det[ ]" refers to the determinant of the matrix. This expression gives

$$\frac{\tan \beta b}{\tan \alpha b} = \frac{-4\alpha\beta\xi^2}{\left(\xi^2 - \beta^2\right)^2} \tag{52}$$

2. Antisymmetric modes

$$\begin{bmatrix} -\left(\xi^2 - \beta^2\right)\sin \alpha b & -2i\xi\beta \sin \beta b \\ 2i\xi\alpha \cos \alpha b & \left(\xi^2 - \beta^2\right)\cos \beta b \end{bmatrix} \times \begin{bmatrix} A \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{53}$$

Nontrivial solutions for the constants A and D exist if and only if

$$\det \begin{bmatrix} -\left(\xi^2 - \beta^2\right)\sin \alpha b & -2i\xi\beta \sin \beta b \\ 2i\xi\alpha \cos \alpha b & \left(\xi^2 - \beta^2\right)\cos \beta b \end{bmatrix} = 0 \tag{54}$$

This gives

$$\frac{\tan \beta b}{\tan \alpha b} = \frac{-\left(\xi^2 - \beta^2\right)^2}{4\alpha\beta\xi^2} \tag{55}$$

Therefore, given a certain isotropic material, equations (52) and (55) can be solved numerically to find the relation between the driving angular frequency ω and the wavenumber ξ from which the corresponding phase speed v_{ph} can be found. This relation is plotted in Figure 8(a) for an aluminum alloy (material properties used: Young's modulus $E = 70\,\mathrm{GPa}$; Poisson's ratio $\nu = 0.33$; density $\rho = 2700\,\mathrm{kg\,m^{-3}}$). Consequently, a finite time-span pulse (which would have a Fourier transform with a certain spread in the frequency domain) will be distorted as it propagates owing to the different phase speeds of its individual frequency components. Another important characteristic is the group speed dispersion curve (Figure 8b). The magnitude of the group velocity for an isotropic medium (denoted by $\boldsymbol{v}_{\mathrm{gr}}$) is the derivative of the angular frequency with respect to the wavenumber ξ and its direction is coincident with that of the phase velocity. Its magnitude gives a very good approximation to the speed of the peak of the modulation envelope of a narrow frequency bandwidth pulse. This approximation improves in accuracy as the pulse moves further away from the source or if the GW mode becomes less dispersive. The procedure above, although for a simple structure, can be generalized to obtain dispersion curves for complex structures (*see* e.g., **Noncontact Rail Monitoring by Ultrasonic Guided Waves**; **Multiwire Strands**).

The resulting displacement components are given by

$$\begin{aligned} u_1 &= \left\{i\xi\left(A \sin \alpha x_3 + B \cos \alpha x_3\right)\right. \\ &\quad \left. + \beta\left(C \cos \beta x_3 - D \sin \beta x_3\right)\right\} \mathrm{e}^{\mathrm{i}(\xi x_1 - \omega t)} \\ u_3 &= \left\{\alpha\left(A \cos \alpha x_3 - B \sin \alpha x_3\right)\right. \\ &\quad \left. - i\xi\left(C \sin \beta x_3 + D \cos \beta x_3\right)\right\} \mathrm{e}^{\mathrm{i}(\xi x_1 - \omega t)} \end{aligned} \tag{56}$$

For $A = D = 0; B, C \neq 0$, one has symmetric Lamb modes, i.e., symmetric u_1 and antisymmetric u_3 about $x_3 = 0$. For $A, D \neq 0$; $B = C = 0$, one has antisymmetric Lamb modes, i.e., antisymmetric u_1 and symmetric u_3 about $x_3 = 0$. The discussion

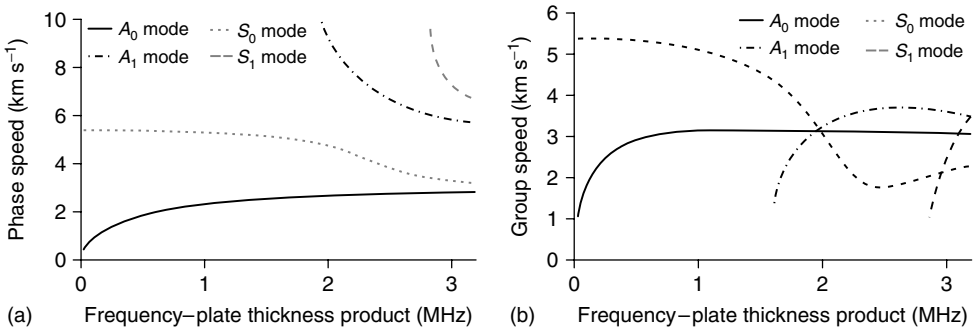


Figure 8. Dispersion curves for Lamb modes in an isotropic aluminum plate structure: (a) phase speed and (b) group speed.

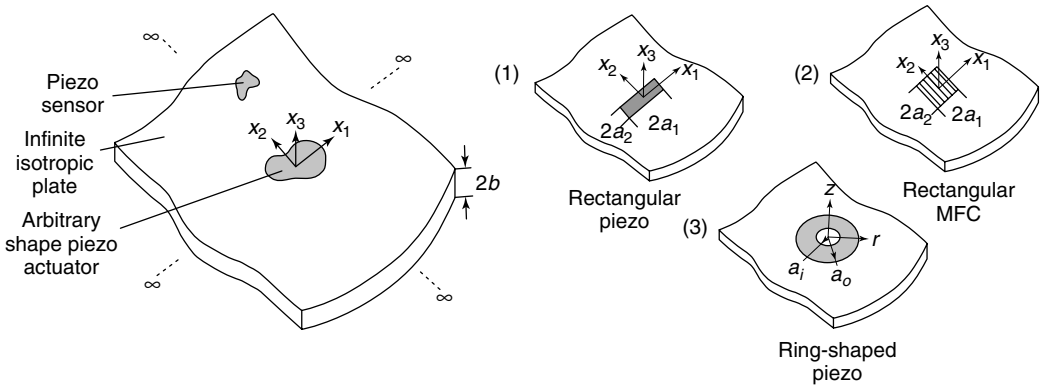


Figure 9. Infinite isotropic plate with arbitrary shape surface-bonded piezo actuator and piezo sensor and the three specific piezo shapes considered.

till this point has only covered free bulk waves and GWs. The next subsection covers GW excitation by piezoelectric wafer transducers (*see* **Piezoceramic Materials—Phenomena and Modeling**; **Piezoelectricity Principles and Materials**; **Piezoelectric Wafer Active Sensors**; **Integrated Sensor Durability and Reliability**), which are among the more commonly used GW SHM transducers.

2.2.4 GW excitation by finite-dimensional piezoelectric wafer transducers in plates

In this section, general expressions for the GW field excited by an arbitrary shape (finite-dimensional) piezoelectric wafer (called *piezo* here) actuator surface bonded on an infinite isotropic plate are presented. Consider an actuator bonded on the surface $x_3 = +b$, as illustrated in Figure 9. The origin is located midway through the plate thickness and the x_3 axis is normal to the plate surface. The starting point is the equations of elasticity in terms of the Helmholtz components, i.e., equations (35). However, for this problem, the 3D elasticity equations of motion cannot be simplified by ignoring variations along one direction in the plane of the plate.

Harmonic excitation at angular frequency ω is assumed. The response to any transient excitation signal can then be found by superimposing the individual responses to the harmonic components of that signal. The 2D spatial Fourier transform is applied to equations (35) in the spatial domain along x_1 and x_2. For a generic variable ψ, it is defined by (denoting the wavenumber components along x_1 and x_2 by ξ_1 and ξ_2, respectively)

$$\tilde{\psi}\left(\xi_1, \xi_2\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi\left(x_1, x_2\right) \mathrm{e}^{\mathrm{i}(\xi_1 x_1 + \xi_2 x_2)}\, \mathrm{d}x_1\, \mathrm{d}x_2 \tag{57}$$

and the inverse transform is given by

$$\psi\left(x_1, x_2\right) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{\psi}\left(\xi_1, \xi_2\right) \times \mathrm{e}^{-i(\xi_1 x_1 + \xi_2 x_2)}\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2 \tag{58}$$

This yields the stresses and displacements in the Fourier wavenumber domain in terms of eight free constants. Similar to the analysis in the previous section, the problem can be decomposed into two independent problems involving four constants each corresponding to symmetric and antisymmetric GW modes. Then, one enforces the forcing condition due to the surface-bonded piezo actuator. This is modeled as causing shear traction normal to its free edge along its boundary on the free surface to which it is bonded. Thus, on the surface $x_3 = b$,

$$\sigma_{31} = \tau_0 \cdot F_1\left(x_1, x_2\right) \quad \sigma_{32} = \tau_0 \cdot F_2\left(x_1, x_2\right)$$
$$\sigma_{33} = 0 \tag{59}$$

where F_1 and F_2 are functions that are zero everywhere except around the edge of the piezo actuator and τ_0 is the magnitude of the shear traction (linearly proportional to the product of the piezo's piezoelectric constant d_{31} or d_{33} and in-plane Young's modulus). This assumes uncoupled dynamics between the actuator and plate. For example, for a rectangular actuator of dimensions $2a_1$ and $2a_2$ along the x_1 and x_2 axes, with its center aligned with the in-plane location of the origin of the coordinate system,

$$\begin{aligned} F_1 &= \left[\delta\left(x_1 - a_1\right) - \delta\left(x_1 + a_1\right)\right] \\ &\quad \times \left[He\left(x_2 + a_2\right) - He\left(x_2 - a_2\right)\right] \\ F_2 &= \left[He\left(x_1 + a_1\right) - He\left(x_1 - a_1\right)\right] \\ &\quad \times \left[\delta\left(x_2 - a_2\right) - \delta\left(x_2 + a_2\right)\right] \end{aligned} \tag{60}$$

where $\delta(\,)$ is the Kronecker delta function and $He(\,)$ is the Heaviside function. Equating the external forcing conditions would give three equations in four unknowns. The fourth equation results from the divergence condition on $\mathbf{H}$, applied on $x_3 = b$. Solving for the constants yields the following equation for out-of-plane displacement in the spatial domain:

$$\begin{aligned} u_3 = \frac{\tau_0}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &\frac{-i.\mathrm{e}^{-\mathrm{i}(\xi_1 x_1 + \xi_2 x_2 - \omega t)}}{D_S(\xi)} \\ &\times \left[2\alpha\beta \sin\alpha b \cos\beta b + (\xi^2 - \beta^2)\sin\beta b \cos\alpha b\right] \\ &\times \left(\xi_1 \tilde{F}_1\left(\xi_1, \xi_2\right) + \xi_2 \tilde{F}_2\left(\xi_1, \xi_2\right)\right)\mathrm{d}\xi_1\,\mathrm{d}\xi_2 \end{aligned} \tag{61}$$

where

$$\xi^2 = \xi_1^2 + \xi_2^2; \quad \alpha^2 = \left(-\xi_1^2 - \xi_2^2\right) + \frac{\omega^2}{v_\mathrm{p}^2};$$

$$\beta^2 = \left(-\xi_1^2 - \xi_2^2\right) + \frac{\omega^2}{v_\mathrm{s}^2}$$

$$D_S(\xi) = \left(\xi^2 - \beta^2\right)^2 \cos\alpha b \sin\beta b + 4\xi^2\alpha\beta \sin\alpha b \cos\beta b$$

The integral form results from applying the inverse 2D Fourier transform in the wavenumber domain. This integrand is singular at the points corresponding to the real roots of $D_S = 0$. The integrands for u_1 or u_2 can also be singular at the roots of $\sin\beta b = 0$ depending on the actuator shape. The former corresponds to the wavenumbers, ξ^S, from the solutions of the Rayleigh–Lamb equation for symmetric modes at frequency ω, i.e., equation (52). The latter corresponds to the wavenumbers of SH waves. While in the above derivation only symmetric modes were considered, the contribution from antisymmetric modes can be found analogously and the complete solution would be a superposition of these two modal contributions.

To illustrate the inversion process in the integral expression for u_3, equations (60) and (61) can be rewritten as

$$\begin{aligned} u_3 = \frac{\tau_0}{\pi^2\mu}\mathrm{e}^{\mathrm{i}\omega t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &\frac{T_S(\xi)}{\xi \cdot D_S(\xi)} \\ &\times \frac{\sin\left(\xi a_1 \cos\gamma\right)\sin\left(\xi a_2 \sin\gamma\right)}{\sin\cos\gamma} \\ &\cdot \mathrm{e}^{-i\xi(x_1\cos\gamma + x_2\sin\gamma)}\,\mathrm{d}\gamma\,\mathrm{d}\xi \end{aligned} \tag{62}$$

with $T_S(\xi) = \xi^2(\beta^2 - \xi^2)\cos\alpha b \sin\beta b - 2\alpha\beta\xi^2 \cos\beta b \sin\alpha b$. Here the Cartesian wavenumbers ξ_1 and ξ_2 have been replaced by the polar wavenumber coordinates ξ and γ. To obtain the 2D spatial inverse

Fourier transform, residue calculus is used. For convenience, the integral in equation (62) is rewritten thus:

$$u_3 = \frac{-\tau_0}{4\pi^2\mu} e^{i\omega t} \int_0^{\infty} \int_0^{2\pi} \frac{T_S(\xi)}{\xi \cdot D_S(\xi)} \times \frac{\left(e^{i\xi a_1 \cos\gamma} - e^{-i\xi a_1 \cos\gamma}\right)\left(e^{i\xi a_2 \sin\gamma} - e^{-i\xi a_2 \sin\gamma}\right)}{\sin\gamma \cos\gamma} \times e^{-i\xi(x_1 \cos\gamma + x_2 \sin\gamma)} \, d\gamma \, d\xi = \frac{-\tau_0}{4\pi^2\mu} e^{i\omega t} \left(\int_0^{\infty} \int_0^{2\pi} \frac{T_S(\xi)}{\xi \cdot D_S(\xi)} \times \frac{e^{-i\hat{\xi}^S((x_1-a_1)\cos\gamma + (x_2-a_2)\sin\gamma - \omega t)}}{\sin\gamma \cos\gamma} \, d\gamma \, d\xi + \cdots \right) \tag{63}$$

The "..." indicates that there are three other similar integrals. Consider the first of the four integrals in the second line of equation (63), say I_1. It is further rewritten as follows:

$$I_1 = \int_{\gamma_1 - \pi/2}^{\gamma_1 + \pi/2} \int_{-\infty}^{\infty} \frac{T_S(\xi)}{\xi \cdot D_S(\xi)} \frac{e^{-i(\xi r_1 \cos(\gamma - \gamma_1) - \omega t)}}{\sin\gamma \cos\gamma} \, d\xi \, d\gamma \tag{64}$$

where $\gamma_1 = \tan^{-1}\left(x_2 - a_2 / x_1 - a_1\right)$ and $r_1 = \sqrt{((x_1 - a_1)^2 + (x_2 - a_2)^2)}$. This form ensures that the coefficient of ξ in the exponent remains positive over the domain of integration. The inner integral along the real ξ axis is solved by considering a contour integral in the lower half of the complex ξ plane, the semicircular portion C with radius $|\xi| = R \to \infty$. Using the residue theorem for the inner integral in equation (64) yields in this case (assuming I is the integrand in I_1)

$$\int_{-\infty}^{\infty} I \, d\xi + \int_C I \, d\xi = -\pi i \sum_{\hat{\xi}^S} \mathrm{Res}\left(\mathrm{I}\left(\hat{\xi}^S\right)\right) \tag{65}$$

where $\hat{\xi}^S$ are the roots of the Rayleigh–Lamb equation for symmetric modes ($D_S = 0$). It can be shown that the contribution over the semicircular portion C of the contour vanishes (see [13] for details). Therefore,

$$\int_C I \, d\xi = 0; \quad \int_{-\infty}^{\infty} I \, d\xi = -\pi i \sum_{\hat{\xi}} \mathrm{Res}\left(\mathrm{I}\left(\hat{\xi}^S\right)\right)$$

$$I_1 = \sum_{\hat{\xi}^S} \int_{\gamma_1 - \pi/2}^{\gamma_1 + \pi/2} \frac{T_S\left(\hat{\xi}^S\right)}{\hat{\xi}^S \cdot D_S'\left(\hat{\xi}^S\right)} \times \frac{e^{-i\left(\hat{\xi}^S r_1 \cos(\gamma - \gamma_1) - \omega t\right)}}{\sin\gamma \cos\gamma} \, d\gamma \tag{66}$$

where $(\,)'$ indicates a derivative with respect to ξ. A similar approach can be used to solve the other three integrals in equation (63), to finally yield the expression for u_3. Expressions for ring-shaped piezo actuators have also been derived in [14]. The GW field excited by anisotropic piezocomposites, such as macro fiber composite (MFC) actuators, consisting of piezo fibers in an epoxy matrix, has been derived for isotropic plates, hollow cylinders, and rectangular-sectional beams in [13, 15]. These have also been extensively validated by numerical and experimental results (see, e.g., Figure 10).

2.3 Wave propagation in anisotropic solids

Fiber-reinforced composite materials are becoming increasingly common in aerospace and other structures owing to their superior stiffness-to-mass ratio and corrosion resistance compared to metals (*see* **Lamb Wave-based SHM for Laminated Composite Structures**; **Commercial Fixed-wing Aircraft**). Therefore, it is useful to have a basic understanding of bulk elastic waves in transversely isotropic solids with uniform density, which serve as a useful model for unidirectional fiber-reinforced composites. This model is most effective if the wavelength is large compared to the fiber diameter and interfiber spacing. This is then used to build the solution for GWs in multilayered laminated plates (Figure 11).

2.3.1 Transversely isotropic elastic 3D solid

First, consider the general solution for bulk waves in a generic anisotropic medium with uniform density ρ. The equations of motion for the bulk medium are

$$\nabla \boldsymbol{c} \nabla^T \bar{\boldsymbol{u}} = \rho \ddot{\bar{\boldsymbol{u}}} \tag{67}$$

where $\bar{\boldsymbol{u}}$ is the local displacement vector (later the global displacement vector $\boldsymbol{u}$ will be introduced for

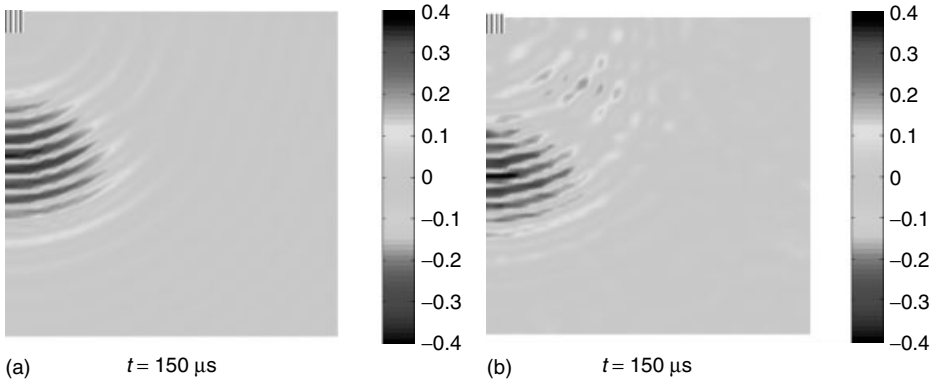


Figure 10. Surface plots showing out-of-plane velocity signals over a $20 \times 20\,\text{cm}^2$ quarter section of an aluminum plate excited by macro fiber composites (the striped portion in the upper left corner) in the A_0 mode using a 30 kHz, 3.5 cycle windowed toneburst: (a) Using the model from [13] and (b) experimental plot obtained using laser vibrometry. Both plots are normalized to the respective peak values over all time instants.

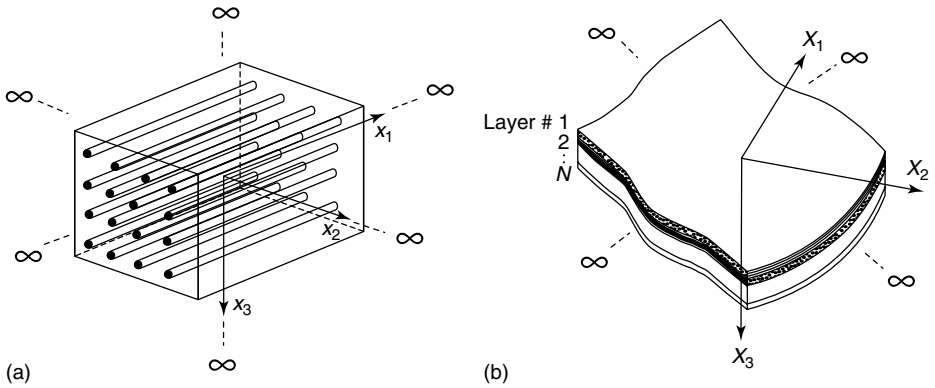


Figure 11. (a) Infinite transversely isotropic solid and (b) multilayered laminated composite plate.

the laminate), $\boldsymbol{c}$ is the stiffness matrix and the operator ∇ is defined as

$$\nabla \equiv \begin{bmatrix} \dfrac{\partial}{\partial x_1} & 0 & 0 & 0 & \dfrac{\partial}{\partial x_3} & \dfrac{\partial}{\partial x_2} \\ 0 & \dfrac{\partial}{\partial x_2} & 0 & \dfrac{\partial}{\partial x_3} & 0 & \dfrac{\partial}{\partial x_1} \\ 0 & 0 & \dfrac{\partial}{\partial x_3} & \dfrac{\partial}{\partial x_2} & \dfrac{\partial}{\partial x_1} & 0 \end{bmatrix} \tag{68}$$

For a unidirectional fiber-reinforced composite, if the fibers are oriented along the 1-direction in the local coordinate system (x_1, x_2, x_3) of the material, the stress–strain relation and the stiffness matrix $\boldsymbol{c}$ for the transversely isotropic material are

$$\begin{bmatrix} \bar{\sigma}_{11} \\ \bar{\sigma}_{22} \\ \bar{\sigma}_{33} \\ \bar{\sigma}_{23} \\ \bar{\sigma}_{31} \\ \bar{\sigma}_{12} \end{bmatrix} = \boldsymbol{c} \begin{bmatrix} \bar{u}_{1,1} \\ \bar{u}_{2,2} \\ \bar{u}_{3,3} \\ \bar{u}_{2,3} + \bar{u}_{3,2} \\ \bar{u}_{1,3} + \bar{u}_{3,1} \\ \bar{u}_{2,1} + \bar{u}_{1,2} \end{bmatrix};$$

$$\boldsymbol{c} = \begin{bmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{22} & c_{23} & 0 & 0 & 0 \\ c_{12} & c_{23} & c_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{55} \end{bmatrix},$$

$$\text{with} \quad c_{44} = \frac{c_{22} - c_{23}}{2} \tag{69}$$

Here $\bar{\sigma}_{ij}$, with i and j taking integer values from 1 to 3, are the local stress components. Next, constants are introduced that correspond to the squares of bulk wave speeds possible along the principal directions:

$$
\begin{aligned}
a_1 &= \frac{c_{22}}{\rho} \text{ (longitudinal wave normal to the fiber direction)} \\
a_2 &= \frac{c_{11}}{\rho} \text{ (longitudinal wave along the fiber direction)} \\
a_3 &= \frac{\left(c_{12}+c_{55}\right)}{\rho} \text{ (shear wave in the plane of isotropy)} \\
a_4 &= \frac{\left(c_{22}-c_{23}\right)}{2\rho} = \frac{c_{44}}{\rho} \text{ (shear wave along the fiber direction)} \\
a_5 &= \frac{c_{55}}{\rho} \text{ (shear wave in the plane of isotropy)}
\end{aligned}
\tag{70}
$$

Viscoelastic damping can be modeled by the use of complex stiffness constants. Suppose the wavenumber components are ξ_1, ξ_2, and ζ along the 1-, 2-, and 3-local directions, respectively. Furthermore, without loss of generality, consider harmonic excitation at angular frequency ω. Then the wave field is of the form

$$
\bar{\boldsymbol{u}} = \bar{\boldsymbol{\Omega}} \mathrm{e}^{-\mathrm{i}\left(\xi_1 x_1 + \xi_2 x_2 + \zeta x_3 - \omega t\right)} \tag{71}
$$

where $\bar{\boldsymbol{\Omega}}$ is a linear superposition of the possible eigenvectors. Then, from equations (67–71), one obtains the Christoffel equation:

$$
\begin{bmatrix}
c_{11}\xi_1^2 + c_{55}\left(\xi_2^2+\zeta^2\right) & \left(c_{12}+c_{55}\right)\xi_1\xi_2 & \left(c_{12}+c_{55}\right)\xi_1\zeta \\
\left(c_{12}+c_{55}\right)\xi_1\xi_2 & c_{55}\xi_1^2 + c_{22}\xi_2^2 + c_{44}\zeta^2 & \left(c_{23}+c_{44}\right)\xi_2\zeta \\
\left(c_{12}+c_{55}\right)\xi_1\zeta & \left(c_{23}+c_{44}\right)\xi_2\zeta & c_{55}\xi_1^2 + c_{44}\xi_2^2 + c_{22}\zeta^2
\end{bmatrix}
\begin{bmatrix} \bar{\Omega}_1 \\ \bar{\Omega}_2 \\ \bar{\Omega}_3 \end{bmatrix}
= \rho\omega^2
\begin{bmatrix} \bar{\Omega}_1 \\ \bar{\Omega}_2 \\ \bar{\Omega}_3 \end{bmatrix}
\tag{72}
$$

For fixed values of ξ_1, ξ_2, and ω, there are six possible roots $\pm\zeta_i$, $i = 1$–3, of this equation. The first two pairs of roots correspond to pairs of "quasi-longitudinal" waves (characterized by displacements dominantly along the wave propagation direction but with small components normal to it) and "quasi-shear" waves (characterized by displacements dominantly normal to the wave propagation direction but with small components along it; see [3]). The wavenumbers in the thickness direction corresponding to these four roots are, respectively

$$
\begin{aligned}
&\zeta_1^2 = -\xi_2^2 + b_1; \quad \zeta_2^2 = -\xi_2^2 + b_2 \\
&b_1 = -\left(\frac{\beta}{2\alpha}\right) - \sqrt{\left(\frac{\beta}{2\alpha}\right)^2 - \frac{\gamma}{\alpha}}; \\
&b_2 = -\left(\frac{\beta}{2\alpha}\right) + \sqrt{\left(\frac{\beta}{2\alpha}\right)^2 - \frac{\gamma}{\alpha}} \\
&\alpha = a_1 a_5; \\
&\beta = \left(a_1 a_2 + a_5^2 - a_3^2\right)\xi_1^2 - \omega^2\left(a_1 + a_5\right); \\
&\gamma = \left(a_2\xi_1^2 - \omega^2\right)\left(a_5\xi_1^2 - \omega^2\right)
\end{aligned}
\tag{73}
$$

The third pair of roots corresponds to quasi-shear waves and their through-thickness wavenumbers are given by

$$
\zeta_3^2 = -\xi_2^2 + \frac{\left(\omega^2 - a_5\xi_1^2\right)}{a_4} \tag{74}
$$

The displacement eigenvectors resulting from equation (72) corresponding to these roots are

$$
\begin{aligned}
\boldsymbol{e}_1 &= \begin{bmatrix} i\xi_1 q_{11} & i\xi_2 q_{21} & i\zeta_1 q_{21} \end{bmatrix}^{\mathrm{T}}; \\
\boldsymbol{e}_2 &= \begin{bmatrix} i\xi_1 q_{12} & i\xi_2 q_{22} & i\zeta_2 q_{22} \end{bmatrix}^{\mathrm{T}}; \\
\boldsymbol{e}_3 &= \begin{bmatrix} 0 & i\zeta_3 & -i\xi_2 \end{bmatrix}^{\mathrm{T}}
\end{aligned}
\tag{75}
$$

where

$$
\begin{aligned}
q_{11} &= a_3 b_1; \quad q_{21} = \omega^2 - a_2\xi_1^2 - a_5 b_1; \\
q_{21} &= a_3 b_2; \quad q_{22} = \omega^2 - a_2\xi_1^2 - a_5 b_2
\end{aligned}
\tag{76}
$$

The other three eigenvectors $\boldsymbol{e}_4$, $\boldsymbol{e}_5$, and $\boldsymbol{e}_6$ are obtained by replacing ζ_i by $-\zeta_i (\zeta_i \geq 0)$. The general

solution for the displacement vector is then given by

$$\begin{aligned}\bar{\boldsymbol{u}} &= \left(C_{1+}\boldsymbol{e}_1 e^{i\zeta_1 x_3} + C_{2+}\boldsymbol{e}_2 e^{i\zeta_2 x_3} + C_{3+}\boldsymbol{e}_3 e^{i\zeta_3 x_3}\right.\\ &\quad + C_{1-}\boldsymbol{e}_4 e^{-i\zeta_1 x_3} + C_{2-}\boldsymbol{e}_5 e^{-i\zeta_2 x_3}\\ &\quad \left. + C_{3-}\boldsymbol{e}_6 e^{-i\zeta_3 x_3}\right) e^{-i(\xi_1 x_1 + \xi_2 x_2 - \omega t)}\\ &= \begin{bmatrix}\boldsymbol{Q}_{11} & \boldsymbol{Q}_{12}\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{E}}_+(x_3) & 0\\ 0 & \bar{\boldsymbol{E}}_-(x_3)\end{bmatrix}\\ &\quad \times \begin{bmatrix}\boldsymbol{C}_+\\ \boldsymbol{C}_-\end{bmatrix} e^{-i(\xi_1 x_1 + \xi_2 x_2 - \omega t)}\end{aligned} \tag{77}$$

where $C_{i\pm}$, $i = 1\text{–}3$, are free constants, and

$$\begin{aligned}&\boldsymbol{Q}_{11} = \begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3\end{bmatrix};\quad \boldsymbol{Q}_{12} = \begin{bmatrix}\boldsymbol{e}_4 & \boldsymbol{e}_5 & \boldsymbol{e}_6\end{bmatrix}\\ &\bar{\boldsymbol{E}}_+(x_3) = \mathrm{Diag}\left[e^{i\zeta_1 x_3}, e^{i\zeta_2 x_3}, e^{i\zeta_3 x_3}\right]\\ &\bar{\boldsymbol{E}}_-(x_3) = \mathrm{Diag}\left[e^{i\zeta_1 x_3}, e^{i\zeta_2 x_3}, e^{i\zeta_3 x_3}\right]\end{aligned} \tag{78}$$

2.3.2 GWs in multilayered laminated plates

With the general solution for the bulk medium in place, one can then solve for free GWs in multilayered laminated plates. The equations in this particular subsection are adapted from [16] and details can be found there. Each ply is a unidirectional fiber-reinforced composite whose fibers are in the plane of the plate.

Owing to the different orientations of the fibers in the different layers, it is useful to work with a global coordinate system (X_1, X_2, X_3) distinct from the local coordinate system, for which the x_1^m axis of the mth ply (m being between 1 and N) is aligned with the fiber direction. However, the X_3 and x_3^m axes are coincident and the two coordinate systems differ only in the plane of the plate. One can relate the displacement vector $\boldsymbol{u}$ in the global system and $\bar{u}^m$ in the local system using the transformation matrix $\boldsymbol{L}^m$ (with ϕ^m being the angle between the X_1 and x_1^m axes):

$$\begin{aligned}&\boldsymbol{u}^m = \boldsymbol{L}^m \bar{\boldsymbol{u}}^m\\ &\text{where}\quad \boldsymbol{L}^m = \begin{bmatrix}\cos\phi^m & -\sin\phi^m & 0\\ \sin\phi^m & \cos\phi^m & 0\\ 0 & 0 & 1\end{bmatrix}\end{aligned} \tag{79}$$

Using "global" wavenumbers K_1 and K_2 along the X_1 and X_2 directions, plane wave solutions of the following form are assumed:

$$\begin{aligned}\boldsymbol{u}^m &= \boldsymbol{U}^m(X_3)\, e^{i(K_1 X_1 + K_2 X_2 - \omega t)};\\ \boldsymbol{\sigma}^m &= \boldsymbol{\Sigma}^m(X_3)\, e^{i(K_1 X_1 + K_2 X_2 - \omega t)}\end{aligned} \tag{80}$$

The global wavenumbers K_1 and K_2 are related to the "local" wavenumbers ξ_1 and ξ_2 in each layer through the following relation:

$$\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} = \begin{bmatrix}\cos(\phi^m) & \sin(\phi^m)\\ -\sin(\phi^m) & \cos(\phi^m)\end{bmatrix}\begin{bmatrix}K_1\\ K_2\end{bmatrix} \tag{81}$$

Traction and displacement continuity must be maintained across the interfaces between the different layers. Therefore, it is convenient to work with a "displacement–stress vector" $\boldsymbol{S}^m$ defined by

$$\boldsymbol{S}^m(X_3) = \left\{\boldsymbol{U}^m(X_3)\ \boldsymbol{\Sigma}_{i3}^m(X_3)\right\}^T \tag{82}$$

Then, from equations (69), (77), and (79)

$$\begin{aligned}\boldsymbol{S}^m(X_3) &= \begin{bmatrix}\boldsymbol{L}^m & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{L}^m\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_{11}^m & \boldsymbol{Q}_{12}^m\\ \boldsymbol{Q}_{21}^m & \boldsymbol{Q}_{22}^m\end{bmatrix}\\ &\quad \times \begin{bmatrix}\boldsymbol{E}_+^m(X_3) & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{E}_-^m(X_3)\end{bmatrix}\begin{Bmatrix}\boldsymbol{C}_+^m\\ \boldsymbol{C}_-^m\end{Bmatrix}\\ &\equiv \boldsymbol{Q}^m(X_3)\,\boldsymbol{C}^m\end{aligned} \tag{83}$$

where

$$\begin{aligned}\boldsymbol{Q}_{21}^m &= \begin{bmatrix}-\rho a_5\xi_1\zeta_1(q_{11}+q_{21}) & -\rho a_5\xi_1\zeta_2(q_{12}+q_{22}) & \rho a_5\xi_1\xi_2\\ -2\rho a_4\xi_2\zeta_1 q_{21} & -2\rho a_4\xi_2\zeta_2 q_{22} & \rho a_4(\xi_2^2-\zeta_3^2)\\ \mu_1 & \mu_2 & 2\rho a_4\xi_2\zeta_3\end{bmatrix}\\ \boldsymbol{Q}_{22}^m &= \begin{bmatrix}\rho a_5\xi_1\zeta_1(q_{11}+q_{21}) & \rho a_5\xi_1\zeta_2(q_{12}+q_{22}) & \rho a_5\xi_1\xi_2\\ 2\rho a_4\xi_2\zeta_1 q_{21} & 2\rho a_4\xi_2\zeta_2 q_{22} & \rho a_4(\xi_2^2-\zeta_3^2)\\ \mu_1 & \mu_2 & -2\rho a_4\xi_2\zeta_3\end{bmatrix}\end{aligned} \tag{84}$$

$$\mu_1 = \rho\left[\left(a_5 - a_3\right)\xi_1^2 q_{11} - \left(a_1 - 2a_4\right) \times \xi_2^2 q_{21} - a_1\zeta_1^2 q_{21}\right]$$

$$\mu_2 = \rho\left[\left(a_5 - a_3\right)\xi_1^2 q_{12} - \left(a_1 - 2a_4\right) \times \xi_2^2 q_{22} - a_1\zeta_1^2 q_{22}\right]$$

$$\boldsymbol{E}_+^m\left(X_3\right) = \mathrm{Diag}\left[\mathrm{e}^{\mathrm{i}\zeta_1\left(X_3 - X_3^{m-1}\right)}, \mathrm{e}^{\mathrm{i}\zeta_2\left(X_3 - X_3^{m-1}\right)}, \mathrm{e}^{\mathrm{i}\zeta_3\left(X_3 - X_3^{m-1}\right)}\right] \tag{85}$$

$$\boldsymbol{E}_-^m\left(X_3\right) = \mathrm{Diag}\left[\mathrm{e}^{\mathrm{i}\zeta_1\left(X_3^m - X_3\right)}, \mathrm{e}^{\mathrm{i}\zeta_2\left(X_3^m - X_3\right)}, \mathrm{e}^{\mathrm{i}\zeta_3\left(X_3^m - X_3\right)}\right]$$

$$\boldsymbol{C}^m = \left\{\boldsymbol{C}_+^m \quad \boldsymbol{C}_-^m\right\}$$

with X_3^m being the X_3 coordinate of the interface between layers m and $(m-1)$. $\boldsymbol{Q}_{21}^m$ and $\boldsymbol{Q}_{22}^m$ are matrices whose columns are the stress eigenvectors for the mth layer corresponding to wavenumbers along the 3-axis, ζ_i and $-\zeta_i$, respectively. These are obtained from the displacement eigenvectors using equation (68). The interface continuity conditions can then be expressed as

$$\boldsymbol{Q}_+^m \boldsymbol{C}^m = -\boldsymbol{Q}_-^{m+1}\boldsymbol{C}^{m+1}$$

where $\boldsymbol{Q}_+^m = \boldsymbol{Q}^m\left(X_3^{m+1}\right)$;

$$\boldsymbol{Q}_-^{m+1} = \boldsymbol{Q}^{m+1}\left(X_3^{m+1}\right) \tag{86}$$

These equations ensure continuity of all displacement and traction components at the interface across two layers. The surface traction-free conditions can be expressed as

$$\hat{\boldsymbol{Q}}_-^1 \boldsymbol{C}_-^1 = \boldsymbol{0}; \qquad \hat{\boldsymbol{Q}}_+^N \boldsymbol{C}_+^N = \boldsymbol{0}$$

where $\hat{\boldsymbol{Q}}_-^1 = \left[-\boldsymbol{L}^1\boldsymbol{Q}_{21}^1 \quad -\boldsymbol{L}^1\boldsymbol{Q}_{22}^1\boldsymbol{E}^1\right]$;

$$\hat{\boldsymbol{Q}}_+^N = \left[\boldsymbol{L}^N\boldsymbol{Q}_{21}^N\boldsymbol{E}^N \quad \boldsymbol{L}^N\boldsymbol{Q}_{22}^N\right] \tag{87}$$

The matrices $\hat{\boldsymbol{Q}}$ correspond to the lower half of $\boldsymbol{Q}$ relating to stress. The system of equations is then solved by assembling equations (86) and (87) together into a $6N \times 6N$ banded matrix (called the *global matrix*, say $\boldsymbol{G}$, which, for a given laminate, is a function of ω, K_1, and K_2):

$$\begin{bmatrix} \hat{\boldsymbol{Q}}_-^1 & \boldsymbol{0} & \cdots & & & & & & \\ \hat{\boldsymbol{Q}}_+^1 & \boldsymbol{Q}_-^2 & \boldsymbol{0} & \cdots & & & & & \\ & & \ddots & & & & & & \\ & \cdots & \boldsymbol{0} & \boldsymbol{Q}_+^{m-1} & \boldsymbol{Q}_-^m & \boldsymbol{0} & \cdots & & \\ & & \cdots & \boldsymbol{0} & \boldsymbol{Q}_+^m & \boldsymbol{Q}_-^{m+1} & \boldsymbol{0} & \cdots & \\ & & & & & & \ddots & & \\ & & & & & \cdots & \boldsymbol{0} & \boldsymbol{Q}_+^{N-1} & \boldsymbol{Q}_-^N \\ & & & & & & \cdots & \boldsymbol{0} & \hat{\boldsymbol{Q}}_+^N \end{bmatrix} \begin{bmatrix} \boldsymbol{C}^1 \\ \boldsymbol{C}^2 \\ \\ \vdots \\ \\ \boldsymbol{C}^N \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \\ \vdots \\ \\ \boldsymbol{0} \end{bmatrix} \tag{88}$$

Alternatively, if the layup is symmetric about the midplane of the plate, then the system can be solved for the symmetric and antisymmetric modes separately, thereby saving some computational time. Then, traction-free surface condition and continuity conditions up to the interface between layers $N/2$ and $(N/2)-1$ are enforced along with conditions of symmetry (u_3, σ_{32}, and σ_{31} being zero at the midplane) or antisymmetry (u_1, u_2, and σ_{33} being zero at the midplane). The problem is thus reduced to two systems, each of complexity $3N \times 3N$. Then, for nontrivial values of the constants,

$$\det \boldsymbol{G}\left(\omega, K_1, K_2\right) = 0 \tag{89}$$

For a given excitation angular frequency ω and angle $\Gamma\left(=\tan^{-1}\left(K_2/K_1\right)\right)$ between the global Cartesian wavenumbers, one can solve equation (89) for the radial wavenumber $K\left(=\left(K_1^2 + K_2^2\right)^{1/2}\right)$, which is related to the phase velocity along that direction ($K = \omega/v_{\mathrm{ph}}$). Thus, the phase velocity also depends on the direction of propagation along with excitation frequency. This directional dependence is usually presented in the form of a "slowness curve", which is a polar plot of the inverse of phase velocity versus propagation angle. This curve is shown in

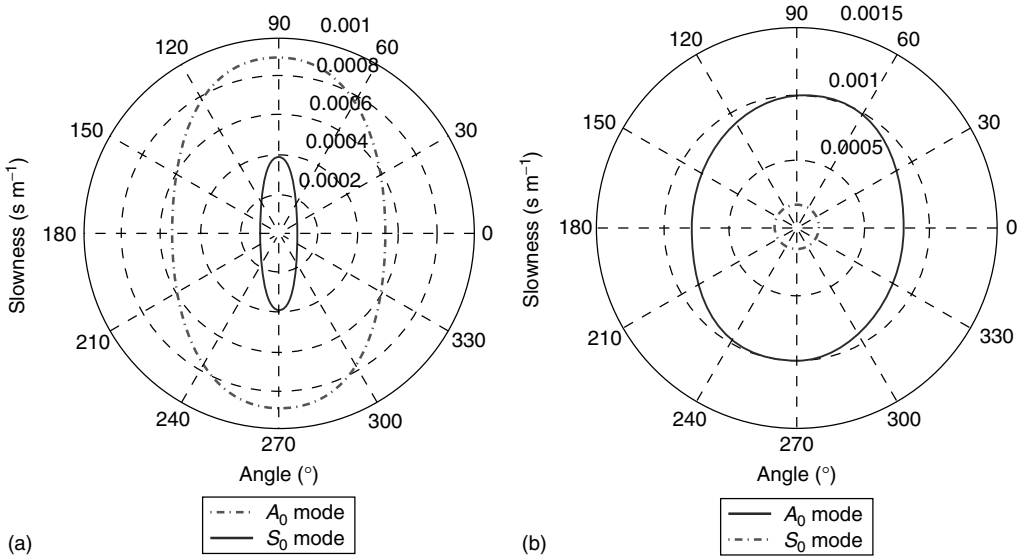


Figure 12. Slowness surfaces (plot of inverse of phase velocity versus direction) in (a) a 1-mm-thick unidirectional graphite–epoxy composite plate at 500 kHz (with graphite fibers along $0^\circ/180^\circ$) and (b) a quasi-isotropic $[0^\circ/45^\circ/-45^\circ/90^\circ]_s$ laminate with each ply being 0.11-mm thick at 200 kHz.

Figure 12(a) for a single-layered unidirectional fiber-reinforced composite and in Figure 12(b) for a quasi-isotropic layup. As seen there, in quasi-isotropic layups, the S_0 mode is approximately isotropic at low frequency-thickness products (up to around 700 kHz for a 1-mm laminate thickness). In composite plates, owing to the anisotropy of phase velocity, the modulation envelope of the GW is 2D and the group velocity is given by the expression

$$\boldsymbol{v}_{\mathrm{gr}} = \frac{\partial \omega}{\partial K_1}\hat{i} + \frac{\partial \omega}{\partial K_2}\hat{j} = -\frac{\nabla_K(\det \boldsymbol{G})}{\partial(\det \boldsymbol{G})/\partial \omega} \qquad (90)$$

where $\hat{i}$ and $\hat{j}$ are unit vectors along the X_1 and X_2 directions. The second form of the formula above is more useful, since the relation between ω, K_1, and K_2 is only available in implicit form of the global matrix determinant. This implies that the group velocity $\mathbf{v}_{\mathrm{gr}}$ is along the normal to the slowness curve (Figure 13a). Thus, the phase velocity and group velocity vectors are, in general, not coincident for anisotropic media. This implies that a plane GW is "steered" in a direction (along the group velocity) different from the direction along which it was launched by a source (which defines the phase velocity), as shown in Figure 13(b). This is explained in Auld [3] in more detail. Semi-analytical expressions for the GW field excited by finite-dimensional piezos in multilayered composite plates are also derived in [17].

3 CONCLUDING REMARKS

A brief introduction to active GW SHM methods was given. The discussion then delved into the rudimentary concepts of elastic waves using the example of 1D wave propagation in taut strings. This example was followed by the governing equations for elastic waves in 3D isotropic elastic solids and a discussion of the GW modes in isotropic plates. The equations for GW excitation by finite-dimensional piezos were then presented. Finally, the more generic governing equations for elastic waves in transversely isotropic solids were covered and this was used to solve for the GW modes in multilayered laminated composite plates. With this basic understanding of GW theory, one is in a much better position to explore other issues in active GW and acoustic emission SHM. In designing effective GW SHM systems, it is important to at least account for the caveats associated with GW propagation discussed, such as the presence

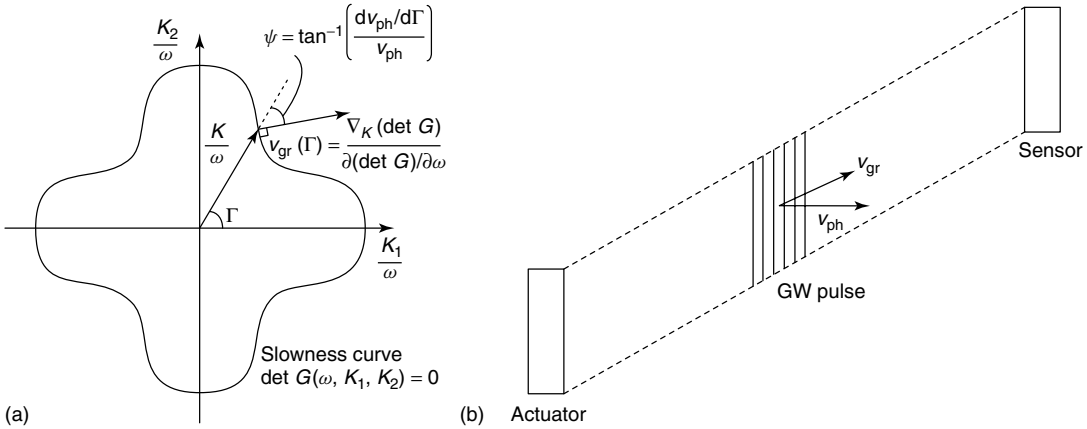


Figure 13. (a) Relation between group velocity and slowness curve and (b) "steering" in anisotropic media.

of multiple modes, dispersiveness, and steering in anisotropic materials.

Beyond these phenomena, depending on the application in question, there may be further challenges to tackle. Apart from the modeling and validation work for GW excitation by finite-dimensional piezos summarized here, the authors' efforts in the past have focused on a variety of research issues in active GW SHM for aerospace structures. An advanced signal processing using chirplet matching pursuits [18] and mode identification has also been developed, which can distinguish between and identify the modes of overlapping, multimodal GW reflections from multiple damage sites (Figure 14). Modeling the effects of elevated temperature (up to 150 °C) on GW SHM has been addressed in [19] (e.g., see Figure 15). This temperature is representative of the maximum limit seen in internal spacecraft structures. Present efforts are directed toward developing/manufacturing of anisotropic piezocomposite transducers and radar-scanning array configurations. As outlined in the review paper by Raghavan and Cesnik [1], much more work is still required before this technology sees widespread field deployment.

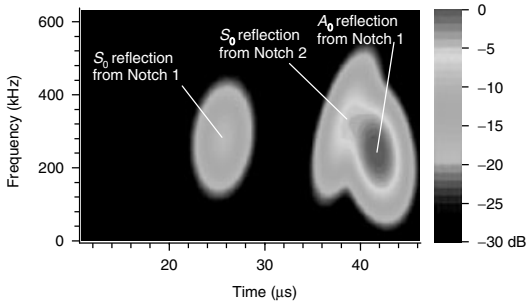


Figure 14. Sample result from the GW signal-processing algorithm in [18] showing its capability to resolve overlapped GW reflections in the time–frequency plane and identify their modes.

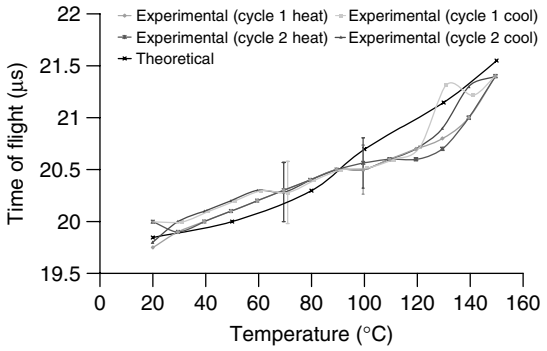


Figure 15. Variation in time of flight with temperature in pitch–catch tests (without damage) using piezoelectric wafer transducers in an aluminum plate specimen from [19].

RELATED ARTICLES

Applications of Acoustic Emission for SHM: A Review

Ultrasonic Methods

Guided-wave Array Methods

Modeling of Lamb Waves in Composite Structures

Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications

Development of an Active Smart Patch for Aircraft Repair

Wind Turbines

REFERENCES

[1] Raghavan A, Cesnik CES. Review of guided-wave structural health monitoring. *The Shock and Vibration Digest* 2007 **39**:91–114.

[2] Rose JL. *Ultrasonic Waves in Solid Media*. Cambridge University Press, 1999.

[3] Auld BA. *Acoustic Fields and Waves in Solids, Second Edition*. R.E. Kreiger Publishing Company: Malabar, FL, 1990; Vols I and II.

[4] Graff KF. *Wave Motion in Elastic Solids*. Dover Publications: New York, 1991.

[5] Achenbach JD. *Wave Propagation in Elastic Solids*. North-Holland: New York, 1984.

[6] Rayleigh JWS. On waves propagated along the plane surface of an elastic solid. *Proceedings of the London Mathematical Society* 1887 **17**:4–11.

[7] Love AEH. *Some Problems of Geodynamics*. Cambridge University Press, 1926.

[8] Stoneley R. Elastic waves at the surface of separation of two solids. *Proceedings of the Royal Society of London, Series A* 1924 **106**:416–428.

[9] Scholte JG. On the Stoneley wave equation. *Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen* 1942 **45**(20–25):159–164.

[10] Lamb H. On waves in an elastic plate. *Proceedings of the Royal Society of London, Series A* 1917 **93**(651):293–312.

[11] Gazis DC. Three-dimensional investigation of the propagation of waves in hollow circular cylinders. *Journal of the Acoustical Society of America* 1959 **31**:568–578.

[12] Worlton DC. Experimental confirmation of Lamb waves at megacycle frequencies. *Journal of Applied Physics* 1961 **32**:967–971.

[13] Raghavan A, Cesnik CES. 3-D elasticity-based modeling of anisotropic piezocomposite transducers for guided-wave structural health monitoring. *Journal of Vibration and Acoustics-Transactions of the ASME* 2007 **129**:739–751. Special issue on damage detection and structural health monitoring.

[14] Raghavan A, Cesnik CES. Finite-dimensional piezoelectric transducer modeling for guided wave based structural health monitoring. *Smart Materials and Structures* 2005 **14**:1448–1461.

[15] Salas KI, Cesnik CES, Raghavan A. Modeling of wedge-shaped anisotropic piezocomposite transducer for guided wave-based structural health monitoring. Presented at the *15th AIAA/ASME/ASC Adaptive Structures Conference*, Paper 2007-1723. Honolulu, HI, 23–26 April 2007.

[16] Lih S-S, Mal AK. On the accuracy of approximate plate theories for wave field calculations in composite laminates. *Wave Motion* 1995 **21**:17–34.

[17] Raghavan A, Cesnik CES. Modeling of guided-wave excitation by finite-dimensional piezoelectric transducers in composite plates. Presented at the *15th AIAA/ASME/ASC Adaptive Structures Conference*, Paper 2007-1725. Honolulu, HI, 23–26 April 2007.

[18] Raghavan A, Cesnik CES. Guided-wave signal processing using chirplet matching pursuits and mode correlation for structural health monitoring. *Smart Materials and Structures* 2007 **16**:355–366.

[19] Raghavan A, Cesnik CES. Studies on effects of elevated temperature for guided-wave structural health monitoring. Presented at the *14th SPIE Symposium on Smart Structures and Materials/NDE*, Paper 6529-9. San Diego, CA, 18–22 March 2008 (to appear in the *Journal of Intelligent Material Systems and Structures*, 2008).

Chapter 5
Electromechanical Impedance Modeling

Andrei N. Zagrai[1] and Victor Giurgiutiu[2]

[1] *New Mexico Institute of Mining and Technology, Socorro, NM, USA*
[2] *Mechanical Engineering Department, University of South Carolina, Columbia, SC, USA*

1 INTRODUCTION

Thin piezoelectric wafers have received considerable attention as active elements of structural health monitoring (SHM) systems [1]. Direct and converse piezoelectric effects (*see* **Piezoelectricity Principles and Materials**) intrinsic to the wafers' material enable a broad spectrum of SHM applications ranging from passive sensing to active generation of elastic waves [2]. To reflect the versatility of embeddable piezoelectrics, in this article, the term *piezoelectric wafer active sensors* (PWAS) is used (*see* **Piezoelectric Wafer Active Sensors**).

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

One distinct SHM methodology facilitated by PWAS is the electromechanical (EM) impedance method [3]. Electromechanical impedance (EMI) is a measure of opposition to the mechanical motion of an electrically excited piezoelectric element. The EMI of the embedded PWAS includes contributions of the piezoelectric element and a host structure. Consequently, the EMI method utilizes local structural dynamic characteristics obtained through impedance measurements for the assessment of structural condition (*see* **Piezoelectric Impedance Methods for Damage Detection and Sensor Validation**) in the immediate vicinity of the sensor [4]. In this respect, [5], the EMI methodology complements embedded ultrasonics, employing the same sensors for long-range inspection. The EMI method assesses the local structural response at relatively high frequencies, ranging from a few kilohertz to hundreds of kilohertz. This is in contrast to vibration methods that explore low-frequency structural responses involving global motion. Monitoring of the structural condition at such high frequencies yields several advantages. The high-frequency response is little affected by global conditions such as flight loads and ambient vibrations; compensation algorithms are employed to address variation of environmental parameters. In addition, because the wavelength of the interrogation signal at high

frequencies is relatively small, the EMI method allows for monitoring small-scale phenomena (i.e., cracks, delamination, disbonds), whose contribution to the global structural dynamics may not be noticeable or detectable by other methods.

It is important to note that the complexity of the EMI spectra depends on structural geometry and constitution. The admittance signatures reflect the resonance behavior and, for structures with simple geometry, yield well separated peaks in the low-frequency range. The impedance is inversely proportional to admittance and therefore indicates a frequency-dependent structural resistance to the applied excitation. An example of the impedance spectrum of a simple structural element, an aluminum circular plate, is shown in Figure 1. The evolution of the spectrum due to a circumferential crack introduced with a jet cutting tool at different distances from the sensor is also presented in the figure. The crack changes the structural condition in the area adjacent to the sensor and this is manifested in an increasing complexity of the impedance spectra. However, the intact (undamaged) structure reveals dynamic characteristics that can be modeled analytically using the principles of structural dynamics. To gain insights into the physical mechanisms governing the EMI method, theoretical underpinnings of the sensor and structural dynamic behavior need to be studied. This is the subject of the article.

2 PIEZOELECTRIC IMPEDANCE SENSORS

Piezoelectric wafer active sensors are typically fabricated from lead zirconate titanate (PZT) ceramic. For small changes in mechanical and electrical parameters, a linear theory of piezoelectricity [6] is applicable and the piezoelectric ceramic is described by the following pair of equations in tensor notation.

$$S_{ij} = s^{E}_{ijkl} T_{kl} + d_{kij} E_k$$
$$D_j = d_{jkl} T_{kl} + \varepsilon^{T}_{jk} E_k \qquad (1)$$

Equations (1) signify a relationship between mechanical strain, S_{ij}, mechanical stress, T_{kl}, electrical field, E_k, and electrical displacement D_j that are coupled through mechanical compliance measured at

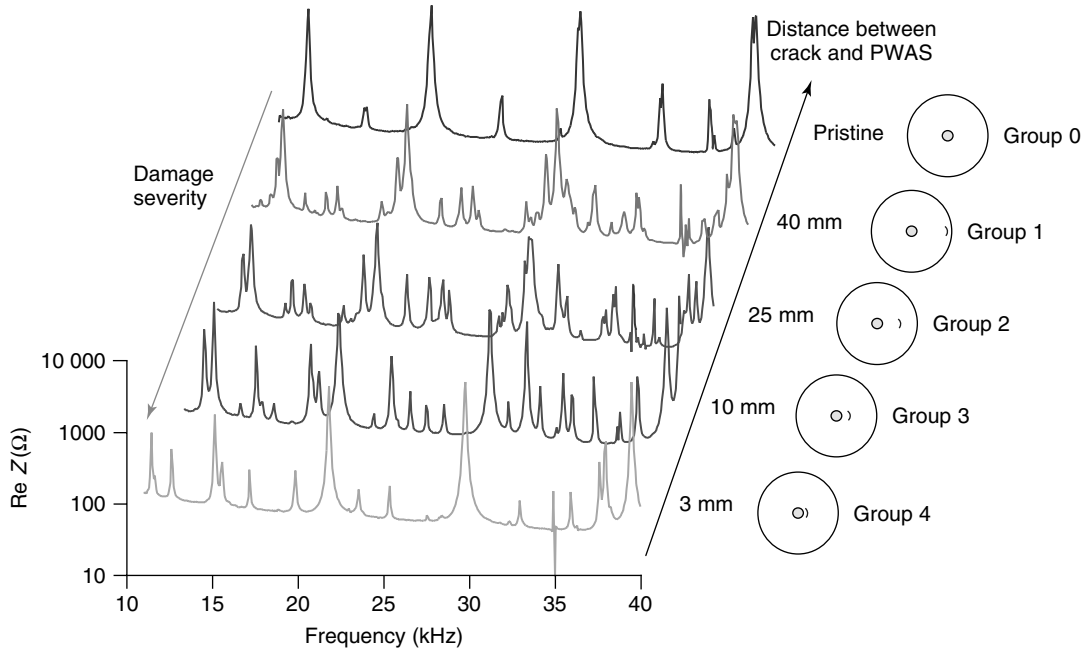


Figure 1. Dependence of the electromechanical impedance spectra on the location of damage.

zero electric field ($E = 0$), s^E_{ijkl}; dielectric permittivity measured at zero mechanical stress ($T = 0$), ε^T_{jk}; and the piezoelectric coefficient, d_{kij}. In general, the tensorial representation in equations (1) would result in nine equations corresponding to respective strain components and polarization directions. Simplification of the constitutive equations is achieved by considering particularities of the piezoelectric sensor configuration and its interaction with the host structure.

2.1 1D PWAS model

One of the most widely used sensor configurations is a thin piezoelectric strip, plate, or disk with the polarization vector parallel to the thickness direction. Effectively, this permits a separation of the sensor geometry into one-dimensional (1D) and two-dimensional (2D) classes. The discussion is provided on an analytical description of the 1D case as the most effective avenue for understanding the modeling aspects of the piezoelectric wafer EMI sensor.

Consider a thin piezoelectric strip with a length l_a, width b_a, and thickness t_a. Figure 2 indicates the sensor geometry, coordinate system, and poling direction E_3. The harmonic voltage $V(t) = \hat{V}e^{i\omega t}$ with amplitude $\hat{V}$ applied across continuous electrodes results in a spatially uniform ($\partial E/\partial x_1 = 0$) electric field $E = V/t_a$ inducing sensor displacements, u_1, u_2, u_3. A one-dimensional case is achieved when thickness and width dimensions are much smaller than the length of the sensor ($t_a \ll b_a \ll l_a$) and corresponding displacement components are practically decoupled. The indicated assumptions allow for reducing the number of equations participating in equations (1) and extensional (longitudinal) motion along the dominant dimension is described by the following system.

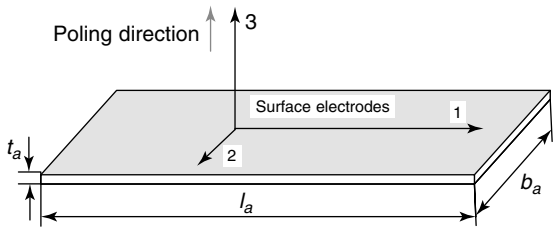


Figure 2. PWAS schematics.

$$S_1 = s^E_{11}T_1 + d_{31}E_3$$
$$D_3 = d_{31}T_1 + \varepsilon^T_{33}E_3 \tag{2}$$

Using Newton's law of motion, $T_1' = \rho\ddot{u}_1$, and the strain–displacement relation, $S_1 = u_1'$, the first of equations (2) yields the classical wave equation:

$$\ddot{u}_1 = c^2 u_1'' \tag{3}$$

where $(\dot{}) = \partial(\,)/\partial t$ and $(\,)' = \partial(\,)/\partial x$, while $c^2 = 1/\rho_a s^E_{11}$ is the wave speed and ρ_a is the density of piezoelectric material. The solution to the equation of motion (3), is separated into temporal and spatial components:

$$u_1(x,t) = \hat{u}_1(x)e^{i\omega t}$$

where

$$\hat{u}_1(x) = \left(C_1 \sin \gamma x + C_2 \cos \gamma x\right) \tag{4}$$

In equation (4), $\hat{u}_1(x)$ is the amplitude of the spatial solution presented in terms of harmonic functions, the wave number $\gamma = \omega/c$, and the constants C_1 and C_2 are dependent on the boundary conditions.

The classical set of boundary conditions includes free and clamped mechanical terminals. A reader may consult the work of Ikeda [7] for the derivation of solutions for classical boundary conditions. Practical applications of piezoelectric impedance sensors imply a more complex form of boundary conditions. The embedded or bonded sensors are essentially constrained by structural elements and, therefore, elastic boundary conditions must be considered. The elastically constrained case opens a path toward the analysis of the complete sensor–structure dynamics because it represents a generic scenario asymptotically approaching free–free and clamped sets as the constraint becomes vanishingly soft, or infinitely stiff [8].

In the 1D case, a bonded piezoelectric sensor is represented as a piezoelectric strip constrained by structural stiffness. The symmetry of the problem yields the constraints of $2k_{\text{str}}$ at each side of the sensor. Figure 3 provides a schematic of the elastically constrained sensor. A spring reaction force at the ends of the sensor is connected with internal stresses according to the following expressions.

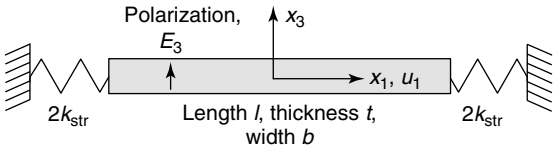


Figure 3. PWAS constrained by the structural stiffness.

$$T_1\left(\frac{l_a}{2}\right) b_a t_a = -2k_{str} u_1\left(\frac{l_a}{2}\right)$$
$$T_1\left(\frac{-l_a}{2}\right) b_a t_a = 2k_{str} u_1\left(\frac{-l_a}{2}\right) \tag{5}$$

Expressing the quasistatic stiffness of the impedance sensor as $k^b_{PWAS} = A_a/s^E_{11} l_a$ and introducing the dynamic stiffness ratio

$$r(\omega) = \frac{k_{str}(\omega)}{k^b_{PWAS}} \tag{6}$$

one obtains

$$u_1'\left(\frac{l_a}{2}\right) + \frac{r(\omega)}{\frac{l_a}{2}} u_1\left(\frac{l_a}{2}\right) = S_{ISA}$$
$$u_1'\left(\frac{-l_a}{2}\right) - \frac{r(\omega)}{\frac{l_a}{2}} u_1\left(\frac{-l_a}{2}\right) = S_{ISA} \tag{7}$$

where $S_{ISA} = d_{31} E_3 = u_{ISA}/l_a$ is the induced strain and u_{ISA} is the induced displacement with the amplitude $\hat{u}_{ISA} = d_{31}\hat{E}_3 \cdot l_a$.

Considering a general solution in the form of equation (4) and denoting $\varphi = \frac{1}{2}\gamma l_a$, one arrives at

$$\begin{aligned}&(\varphi\cos\varphi + r(\omega)\sin\varphi)\, C_1 \\ &\quad - (\varphi\sin\varphi - r(\omega)\cos\varphi)\, C_2 = \tfrac{1}{2}\hat{u}_{ISA} \\ &(\varphi\cos\varphi + r(\omega)\sin\varphi)\, C_1 \\ &\quad + (\varphi\sin\varphi - r(\omega)\cos\varphi)\, C_2 = \tfrac{1}{2}\hat{u}_{ISA}\end{aligned} \tag{8}$$

Constants C_1 and C_2 are obtained by subtracting the first of equations (8) from the second

$$-2\cdot(\varphi\sin\varphi - r(\omega)\cos\varphi) C_2 = 0 \tag{9}$$

which yields $C_2 = 0$, and on adding the two equations

$$2\cdot(\varphi\cos\varphi + r(\omega)\sin\varphi) C_1 = \hat{u}_{ISA} \tag{10}$$

Hence,

$$C_1 = \tfrac{1}{2}\hat{u}_{ISA}/(\varphi\cos\varphi + r(\omega)\sin\varphi) \tag{11}$$

The solution of the wave equation (4) is now expressed as

$$\hat{u}_1(x) = \tfrac{1}{2}\hat{u}_{ISA}\frac{\sin\gamma x}{(\varphi\cos\varphi + r(\omega)\sin\varphi)} \tag{12}$$

In EMI diagnostics, electrical measurements are conducted at the terminals of the sensor. The measured impedance reflects both the electrical and mechanical responses. Modeling of the EM transformation enabled by the sensor is achieved by utilizing the solution (equation 12) in conjunction with the second equation in the system (2). According to the first expression in equations (2), the stress function T_1 depends on the mechanical strain and the electric field:

$$T_1 = \frac{1}{s^E_{11}}\left(S_1 - d_{31}E_3\right) \tag{13}$$

This leads to electrical displacement

$$D_3 = \frac{d_{31}u_1'}{s^E_{11}} - \frac{d^2_{31}E_3}{s^E_{11}} + \varepsilon^T_{33}E_3 \tag{14}$$

When mechanical stress is applied to the piezoelement, it results in an electrical charge that can be obtained by integrating the electrical displacement (equation 14) over the electrode area

$$\begin{aligned}Q(\omega) &= \int_{-\frac{l_a}{2}}^{+\frac{l_a}{2}}\int_{-\frac{b_a}{2}}^{+\frac{b_a}{2}} D_3\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\frac{l_a}{2}}^{+\frac{l_a}{2}}\int_{-\frac{b_a}{2}}^{+\frac{b_a}{2}}\left[\varepsilon^T_{33}E_3 - \frac{d^2_{31}E_3}{s^E_{11}} + \frac{d_{31}u_1'}{s^E_{11}}\right]\mathrm{d}x\,\mathrm{d}y \\ &= \varepsilon^T_{33}E_3 b_a l_a - \frac{d^2_{31}E_3 b_a l_a}{s^E_{11}} + \frac{b_a d_{31}u_1}{s^E_{11}}\Bigg|_{-\frac{l_a}{2}}^{+\frac{l_a}{2}} \\ &= \varepsilon^T_{33}E_3 b_a l_a - \frac{d^2_{31}E_3 b_a l_a}{s^E_{11}} \\ &\quad + \frac{b_a d_{31}}{s^E_{11}}\frac{\hat{u}_{ISA}\cdot\sin\varphi\cdot\mathrm{e}^{i\omega t}}{(\varphi\cos\varphi + r(\omega)\sin\varphi)}\end{aligned} \tag{15}$$

where equation (12) was utilized to express displacement $u_1(x)$ and the argument of the harmonic

functions is $\varphi = \frac{1}{2}\gamma l_a$. Defining the resultant voltage $V(\omega) = \hat{E}_3 \cdot e^{i\omega t}/t_a$, capacitance at zero mechanical stress $C = \varepsilon_{33}^T b_a l_a / t_a$, the EM coupling coefficient $\kappa_{13} = d_{31}/\sqrt{s_{11}^E \varepsilon_{33}^T}$, and $\hat{u}_{\mathrm{ISA}} = d_{31}\hat{E}_3 \cdot l_a$, one arrives at

$$Q(\omega) = V(\omega) \cdot C\left(1 - \kappa_{31}^2\left(1 - \frac{1}{\varphi \cot \varphi + r(\omega)}\right)\right) \tag{16}$$

The time derivative of the charge (equation 16) yields an electric current contributing to expressions for EM admittance and impedance:

$$Y(\omega) = \frac{I(\omega)}{V(\omega)} = i\omega \cdot C\left(1 - \kappa_{31}^2\left(1 - \frac{1}{\varphi \cot \varphi + r(\omega)}\right)\right) \tag{17}$$

$$Z(\omega) = [Y(\omega)]^{-1} \tag{18}$$

Formulations (17) and (18) reveal that the electrical response of the impedance sensor is dominated by the capacitance, C. The mechanical response, reflected through the term containing the EM coupling coefficient κ_{13}, includes the sensor resonance term $\varphi\cot\varphi$ and the contribution of the host structure dynamics participating in the dynamic stiffness ratio $r(\omega)$. Therefore, the ratio $r(\omega)$ allows for the manifestation of structural resonances in the admittance spectrum of the bonded or embedded piezoelectric sensor.

The admittance and impedance formulations in equations (17) and (18) cover a complete frequency range and include the dynamics of both the sensor element and the host structure. Derivation of these expressions has been presented without consideration of the dissipative effects. Dissipation via structural damping can be introduced by utilizing a complex representation of sensor and structural material parameters, e.g., complex mechanical compliance $\bar{s}_{11}^E = s_{11}^E \cdot (1 + i \cdot \eta)$, where $i = \sqrt{-1}$ and η is the damping constant. It follows that

$$\bar{Y}(\omega) = i\omega \cdot \bar{C}\left(1 - \bar{\kappa}_{31}^2\left(1 - \frac{1}{\bar{\varphi} \cot \bar{\varphi} + \bar{r}(\omega)}\right)\right) \tag{19}$$

and

$$\bar{Z}(\omega) = \frac{1}{i\omega \cdot \bar{C}}\left(1 - \bar{\kappa}_{31}^2\left(1 - \frac{1}{\bar{\varphi} \cot \bar{\varphi} + \bar{r}(\omega)}\right)\right)^{-1} \tag{20}$$

To illustrate the application of the presented approach to modeling of the impedance sensors, consider a sensor with free–free boundary conditions. Understanding this case is important in the sensor-characterization process before installation on a host structure. The generic admittance equation (19) is used with a notion that $\bar{r}(\omega) = 0$ for the unconstrained sensor. This formula is used for modeling piezoelectric sensors that measure vibrations along the dominant in-plane dimension. Results for sensor geometries ranging from a rectangular wafer with a 1 : 1 aspect ratio to a piezoelectric strip with the aspect ratio 4 : 1 are presented in Figure 4. Analysis of the high aspect ratio case is important for verifying the validity of the 1D model, while consideration of the low aspect ratio is useful from a practical point of view, as the popular geometry of the impedance sensor is a rectangular piezoelectric wafer. Difficulties in analytical modeling of the unconstrained rectangular piezoelectric wafers arise from the absence of the close-form solution for this essentially 2D case. Applying equation (19) for calculating admittance of the square plate sensor yields the admittance spectrum presented in Figure 4(a). The length, width, and thickness of the sensor are $l_a = 6.99\,\mathrm{mm}$, $b_a = 6.56\,\mathrm{mm}$, and $t_a = 0.215\,\mathrm{mm}$ respectively; the parameters of the piezoelectric ceramic are $\varepsilon_{33}^T = 15.470 \times 10^9\,\mathrm{F\,m^{-1}}$, $s_{11}^E = 15.3 \times 10^{-12}\,\mathrm{Pa^{-1}}$, $d_{31} = -175 \times 10^{-12}\,\mathrm{m\,V^{-1}}$, $\kappa_{31} = 0.36$. Theoretical resonance frequencies identifiable in Figure 4(a) include 208 kHz (1L), 222 kHz (1 W), 621 kHz (2L), 663 kHz (2 W), 1038 kHz (3L), 1108 kHz (3 W), and 1451 kHz (4L). The experimental admittance responses obtained with the impedance analyzer is also presented in the figure. Comparison of theoretical and experimental spectra reveals relatively high errors (19 and 37%) for the first length (1L) and width (1 W) modes. Discrepancy between experimental and calculated responses arises from the 2D stiffening effect typical of low-aspect ratio in-plane vibrations. This effect is not captured by the 1D theory. However, higher vibration modes, which are less affected by the 2D stiffening, show favorable agreement

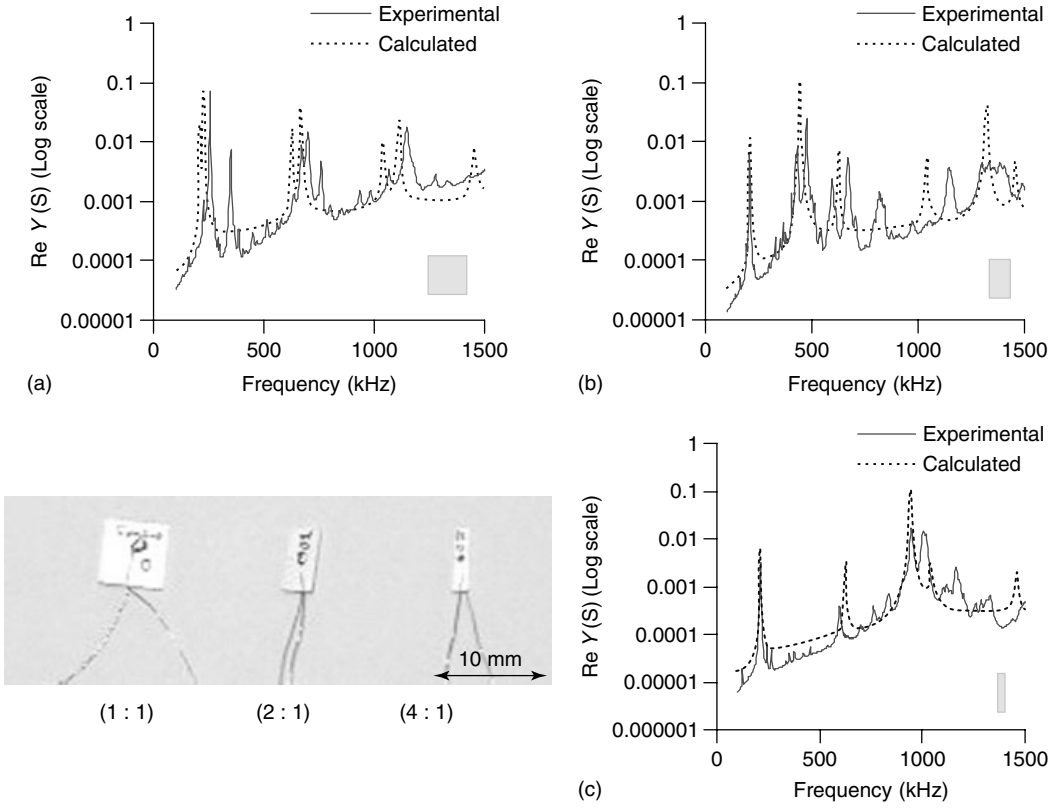


Figure 4. Experimental and calculated admittance spectra of PWAS with the following aspect ratios: (a) 1 : 1 — square wafer, (b) 2 : 1 — half-width wafer, and (c) 4 : 1 — quarter-width wafer.

between experimental and theoretical results. Hence, with the exception of the first modes, a reasonable approximation can be achieved by utilizing a 1D analytical formulation.

Agreement between theoretically calculated and practically measured resonances improves considerably for the piezoelectric wafers with increasing aspect ratio. Figure 4(b,c) provides a pictorial representation of this fact for piezoelectric sensors with aspect ratios of 2 : 1 and 4 : 1. For instance, predictions for the half-width sensor suggest less than 2% error for 1L and 1B resonances. A quarter-width sensor manifests outstanding agreement with the theory for vibration modes below 1 MHz. In addition to clearly identifiable vibration modes, several other peaks are present in the experimental curve. These modes are attributed to the edge roughness producing secondary vibration effects. Analysis of the admittance spectra in Figure 4 suggests a clear trend toward improving the prediction accuracy as the sensor geometry approaches the 1D case.

2.2 2D PWAS model

The one-dimensional modeling approach shows good agreement between the results of analytical calculation and experimental data for sensor geometries with well separated length, width, and thickness dimensions. However, the majority of structural identification and health monitoring applications of the impedance method feature sensors of rectangular or circular configurations. As was indicated in the previous section, no close-form solutions exist for the unconstrained rectangular (e.g., square) sensors and inferring the sensor response using numerical analysis is recommended. For two-dimensional (2D) modeling of the impedance sensors, consider

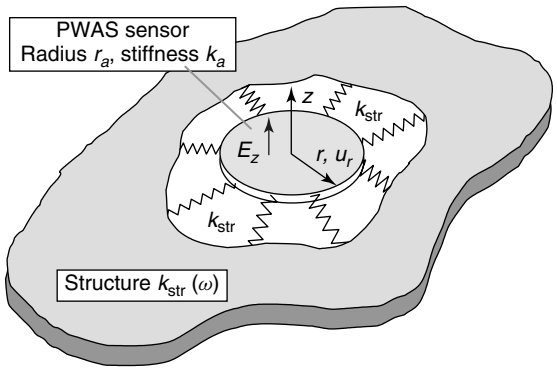


Figure 5. A circular impedance sensor constrained by the structural stiffness, $k_{str}(\omega)$.

circular plate piezoelectric wafers. This sensor geometry allows for the consideration of the axisymmetric problem and, hence, significantly simplifies the analysis procedure. The geometry of the circular PWAS bonded to the host structure is presented in Figure 5. The coordinate system with radial (r), circumferential (θ), and vertical (z) variables is adopted for describing the sensor response. The symmetry of the problem implies that derivatives with respect to the circumferential variable are zero. This results in the reduced set of equations in the system (1). Assuming the in-plane isotropy of the elastic and piezoelectric constants, one arrives at the following [6]:

$$\begin{aligned} S_{rr} &= s_{11}^E T_{rr} + s_{12}^E T_{\theta\theta} + d_{31} E_z \\ S_{\theta\theta} &= s_{12}^E T_{rr} + s_{11}^E T_{\theta\theta} + d_{31} E_z \\ D_z &= d_{31}(T_{rr} + T_{\theta\theta}) + \varepsilon_{33}^T E_z \end{aligned} \tag{21}$$

The analysis of the system (21) follows a procedure outlined for the previously discussed 1D case. It can be shown that the equation of motion for a thin circular plate is

$$\frac{\partial T_{rr}}{\partial r} + \frac{T_{rr} - T_{\theta\theta}}{r} = \rho_a \cdot \frac{\partial^2 u_r}{\partial t^2} \tag{22}$$

Using the definition of radial and circumferential strains,

$$S_{rr} = \frac{\partial u_r}{\partial r}, \quad S_{\theta\theta} = \frac{u_r}{r} \tag{23}$$

and stress relationships inferred from the first two of equations (21),

$$T_{rr} = \frac{1}{s_{11}^E(1-\nu^2)}\left(\frac{\partial u_r}{\partial r} + \nu\frac{u_r}{r}\right) - \frac{d_{31}E_z}{s_{11}^E(1-\nu)} \tag{24}$$

$$T_{\theta\theta} = \frac{1}{s_{11}^E(1-\nu^2)}\left(\nu\frac{\partial u_r}{\partial r} + \frac{u_r}{r}\right) - \frac{d_{31}E_z}{s_{11}^E(1-\nu)} \tag{25}$$

one obtains an equation of motion in terms of the sensor radial displacement, u_r,

$$\frac{\partial^2 u_r}{\partial r^2} + \frac{1}{r}\frac{\partial u_r}{\partial r} - \frac{u_r}{r^2} - \frac{1}{c^2}\frac{\partial^2 u_r}{\partial t^2} = 0 \tag{26}$$

where $c = 1/\sqrt{\rho_a s_{11}^E \cdot (1-\nu_a^2)}$ is the speed of axially symmetric extensional waves. Assuming a harmonic solution of equation (26) in the form

$$u_r(r,t) = \hat{u}_r(r) \cdot \mathrm{e}^{-i\omega t} \tag{27}$$

rearrangement and multiplication by r^2 yields

$$r^2\frac{\partial^2 \hat{u}_r}{\partial r^2} + r\frac{\partial \hat{u}_r}{\partial r} + \hat{u}_r\left(\frac{\omega^2 r^2}{c^2} - 1\right) = 0 \tag{28}$$

It can be shown that equation (28) is analogous to the Bessel differential equation [9] whose solution is readily obtained in terms of a linear combination of Bessel functions of the first and second kinds:

$$\hat{u}_r(r) = A \cdot J_1\left(\frac{\omega r}{c}\right) + B \cdot Y_1\left(\frac{\omega r}{c}\right) \tag{29}$$

Since the displacement of the plate is always finite, even as r approaches zero, one must choose $B = 0$. The final solution is

$$\hat{u}_r(r) = A \cdot J_1\left(\frac{\omega r}{c}\right) \tag{30}$$

The coefficient A is determined from the boundary conditions. For bonded or embedded piezoelectric wafers, these boundary conditions reflect the contribution of the structural stiffness. Figure 5 provides an illustration of the sensor constrained by the adjacent structure. Therefore, the force relationship at $r = r_a$ is

$$N_a(r_a) = -k_{str}(\omega) \cdot u_r(r_a) \tag{31}$$

Equation (31) is used to calculate radial and tangential stress components of the sensor:

$$T_{rr}(r_a) = \frac{N_a(r_a)}{t_a} = \frac{-k_{\text{str}}(\omega)u_{\text{PWAS}}(r_a)}{t_a}$$
$$T_{\theta\theta}(r_a) = \frac{1}{s_{11}^E}\left(\frac{u_r(r_a)}{r_a} - s_{12}^E T_{rr}(r_a) - d_{31}E_z\right) \quad (32)$$

Substituting the stress equations into the first expression in equation (21), one obtains as follows [5]:

$$\frac{\partial u_r(r_a)}{\partial r} = -s_{11}^E \frac{k_{\text{str}}(\omega)\cdot u_r(r_a)}{t_a} + \frac{s_{12}^E}{s_{11}^E} \times \left(\frac{u_r(r_a)}{r_a} - s_{12}^E \frac{k_{\text{str}}(\omega)\cdot u_r(r_a)}{t_a} - d_{31}E_z\right) + d_{31}E_z \quad (33)$$

A dynamic stiffness ratio is defined as

$$\chi(\omega) = \frac{k_{\text{str}}(\omega)}{k_{\text{PWAS}}^d} \quad (34)$$

where the static stiffness of the circular plate impedance sensor is $k_{\text{PWAS}}^d = t_a/[r_a s_{11}^E(1-\nu_a)]$.

Incorporating the dynamic stiffness ratio equation (34) into equation (33) leads to

$$\frac{\partial u_r(r_a)}{\partial r} = -\chi(\omega)\cdot(1+\nu)\frac{u_r(r_a)}{r_a} - \nu\frac{u_r(r_a)}{r_a} + (1+\nu)d_{31}E_z \quad (35)$$

where $\nu_a = -s_{12}^E/s_{11}^E$ is the Poisson ratio. The constant A in the general solution, equation (30), is obtained by the direct substitution of equation (30) into equation (35):

$$A = \frac{(1+\nu_a)\cdot d_{31}E_0}{\frac{\omega}{c}J_0\left(\frac{\omega r_a}{c}\right) - \frac{(1-\nu_a-\chi(\omega)\cdot(1+\nu_a))}{r_a}\cdot J_1\left(\frac{\omega r_a}{c}\right)} \quad (36)$$

The third of equations (21) yields the electrical displacement D_z:

$$D_z = \frac{d_{31}(1+\nu_a)}{s_{11}^E(1-\nu_a^2)}\frac{\omega}{c}AJ_0\left(\frac{\omega r}{c}\right) - \left(\frac{2d_{31}^2}{s_{11}^E(1-\nu_a)} + \varepsilon_{33}^T\right)E_0 \quad (37)$$

Equation (37), the electrical displacement, is integrated to obtain the charge

$$Q = \int_0^{2\pi} \mathrm{d}\theta \int_0^{r_a} D_z r\,\mathrm{d}r = \pi r_a^2 \varepsilon_{33}^T E_0 \mathrm{e}^{i\omega t}\cdot\left[1 - k_p^2 + k_p^2 \times \frac{(1+\nu_a)J_1(\varphi_a)}{\varphi_a J_0(\varphi_a) - (1-\nu_a-\chi(\omega)\cdot(1+\nu_a))\cdot J_1(\varphi_a)}\right] \quad (38)$$

where the planar coupling factor $k_p = \sqrt{2d_{31}^2/[s_{11}^E\cdot(1-\nu_a)\varepsilon_{33}^T]}$ is introduced to account for the EM coupling and $\varphi_a = \omega r_a/c$.

Differentiation of the charge in equation (38) with respect to time results in the electrical current $\hat{I} = i\omega\cdot\hat{Q}$ participating in the final expression for admittance

$$Y(\omega) = i\omega C(1-k_p^2)\left[1 + \frac{k_p^2}{1-k_p^2}\frac{(1+\nu_a)J_1(\varphi_a)}{\varphi_a J_0(\varphi_a) - (1-\nu_a)J_1(\varphi_a) + \chi(\omega)(1+\nu_a)J_1(\varphi_a)}\right] \quad (39)$$

It should be noted that the admittance in equation (39) accounts for the dynamics of the sensor element (through ϕ_a) and the dynamics of the host structure (through $\chi(\omega)$).

The impedance $Z(\omega)$ is defined as

$$Z(\omega) = \left\{i\omega C(1-k_p^2)\left[1 + \frac{k_p^2}{1-k_p^2}\frac{(1+\nu_a)J_1(\varphi_a)}{\varphi_a J_0(\varphi_a) - (1-\nu_a)J_1(\varphi_a) + \chi(\omega)(1+\nu_a)J_1(\varphi_a)}\right]\right\}^{-1} \quad (40)$$

It is worth noting that equation (40) represents a general formulation for vibration of an elastically constrained piezoelectric disk and provides a theoretical description of the impedance response seen by the impedance analyzer at the piezoelectric sensor terminals. Hence, expression (40) facilitates the comparison of the experimental and theoretical impedance signatures of 2D axisymmetric circular structures.

Equation (39), with $\chi(\omega) = 0$, was utilized to simulate an admittance response of an unconstrained piezoelectric circular wafer with diameter $d_a = 6.98\,\text{mm}$ and thickness $t_a = 0.216\,\text{mm}$. The following properties of piezoelectric material were considered: $\varepsilon_{33}^T = 15.470 \times 10^9\,\text{F m}^{-1}$, $s_{11}^E = 18 \times 10^{-12}\,\text{Pa}^{-1}$, $d_{31} = -175 \times 10^{-12}\,\text{m V}^{-1}$, $k_p = 0.63$. No damping was introduced in this simulation. Figure 6 shows experimentally obtained and theoretically calculated admittance. The figure presents the dynamic behavior of the sensor below 1.5 MHz. Within this frequency range, three in-plane radial modes are clearly identifiable: 300, 784, and 1247 kHz. The calculated response closely follows the experimental data. The maximum discrepancy between measured and predicted natural frequencies did not exceed 2.1%. Therefore, the suggested modeling approach can be employed to predict EMI and admittance responses of the 2D axisymmetric structures.

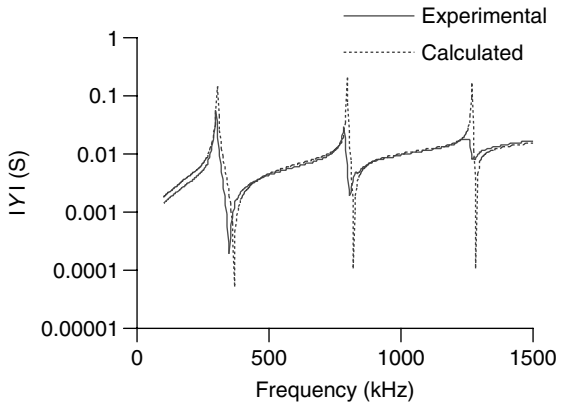


Figure 6. Experimental and calculated admittance responses of an unconstrained circular wafer impedance sensor.

3 ELECTROMECHANICAL IMPEDANCE MODEL FOR 1D STRUCTURES

3.1 Analytical model for 1D sensor/structure interaction

Previous sections discussed modeling of the impedance sensor elastically constrained by the adjacent structure. According to expressions (20) and (40), the sensor impedance contains contributions of both sensor dynamics and structural dynamics. The latter participates in the cumulative impedance response through the dynamic stiffness ratios. Therefore, to model the impedance measured across terminals of the sensor bonded to the structural surface or embedded into the host structure, an expression for the dynamic structural stiffness must first be determined. This can be achieved by considering mechanisms underlying structural excitation and sensing.

When an active sensor is bonded to the host structure as presented in Figure 7, the applied alternating voltage results in expansions and contractions of the sensor, which induce elastic waves traveling in the structure. Reflections of the elastic waves produce standing wave patterns associated with structural vibration modes. Finally, the dynamic deformation of the structural element during vibrations affects the sensor and, hence, is reflected in its impedance signature.

To illustrate the EMI modeling approach, consider a 1D elastic structure (a beam or a rod) with the bonded piezoelectric sensor. In the suggested representation depicted in Figure 7, the PWAS lies between points x_a and $x_a + l_a$ on the structural surface; l_a is the assumed length of the sensor. Expansion and contraction of the sensor produce a reaction

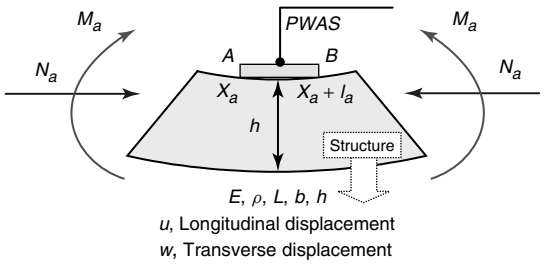


Figure 7. Forces and moments produced as a result of interaction between PWAS and a host structure.

force F_{PWAS} from the beam onto the sensor and an equal and opposite force from the sensor onto the beam. The sign of F_{PWAS} depends on the sign of the voltage applied to the piezoelectric. It is transparent from Figure 7 that the dynamic force $F_{\text{PWAS}} = \hat{F}_{\text{PWAS}} e^{i\omega t}$ induced by the sensor onto the beam results in the following axial force and bending moment [8]:

$$M_a = F_{\text{PWAS}} \frac{h}{2}, \quad N_a = F_{\text{PWAS}} \tag{41}$$

The spatial distribution of this axial force and bending moment is controlled by the sensor and can be expressed in terms of a Heaviside function, $H(x - x_a) = 1$ for $x_a < x$; 0 otherwise.

$$N_a(x,t) = \hat{N}_a[-H(x - x_a) + H(x - x_a - l_a)] \cdot e^{i\omega t} \tag{42}$$

$$M_a(x,t) = -\hat{M}_a[-H(x - x_a) + H(x - x_a - l_a)] \cdot e^{i\omega t} \tag{43}$$

It follows that the force induced by the impedance sensor produces both extensional (longitudinal) and transverse (flexural) vibrations. The initial work of Liang *et al.* [10] discussed flexural vibrations of an elastic beam. Park *et al.* [11] considered longitudinal modes but neglected flexural vibrations. In the following analysis, a unified treatment of longitudinal and flexural behavior is presented.

3.1.1 Extensional vibrations of 1D structure

It can be shown [8] that in the presence of an additional axial force acting on the infinitesimal element of the beam, in-plane displacement $u(x,t)$ is described by the following equation:

$$\rho A \cdot \ddot{u}(x,t) - EA \cdot u''(x,t) = N_a'(x,t) \tag{44}$$

Equation (44) includes parameters of the beam: cross-sectional area A, density ρ, and modulus of elasticity E. The solution is sought in terms of modal expansion [12]

$$u(x,t) = \sum_{p=0}^{\infty} B_p U_p(x) \cdot e^{i\omega t} \tag{45}$$

where $U_p(x)$ are orthonormal mode shapes satisfying $\int U_q(x) U_p(x)\, dx = \delta_{qp}$, with $\delta_{qp} = 1$ for $q = p$, and 0 otherwise. The modal participation factor B_p controls the amplitudes of vibration modes. Following the separation of variables solution procedure and utilizing the expression for natural frequencies, $c^2 \cdot U_p''(x) = -\omega_p^2 \cdot U_p(x)$, with $c = \sqrt{E/\rho}$, resulting from the consideration of the homogeneous equation $\rho A \cdot \ddot{u}(x,t) - EA \cdot u''(x,t) = 0$, upon multiplication by $U_q(x)$ and integration over the length of the beam, L, one obtains

$$B_p = \frac{\int_0^L U_q(x) N_a'(x)\, dx}{\rho A \cdot (\omega_p^2 - \omega^2)} = \frac{\hat{N}_a}{\rho A} \cdot \frac{[U_p(x_a + l_a) - U_p(x_a)]}{(\omega_p^2 - \omega^2)} \tag{46}$$

Expression (46) is derived for undamped vibrations. In practice, damping is always present in any dynamic system. To account for the energy dissipation, viscous damping with the factor ζ_p is introduced. The modal participation factor of the damped beam is

$$B_p = \frac{\hat{N}_a}{\rho A} \cdot \frac{[U_p(x_a + l_a) - U_p(x_a)]}{\omega_p^2 + 2i\zeta_p \omega \omega_p - \omega^2} \tag{47}$$

Substituting equation (47) into the modal expansion (45), the longitudinal displacement can be expressed as

$$u(x,t) = \frac{\hat{N}_a}{\rho A} \sum_{p=0}^{\infty} \frac{[U_p(x_a + l_a) - U_p(x_a)]}{\omega_p^2 + 2i\zeta_p \omega \omega_p - \omega^2} \times U_p(x) \cdot e^{i\omega t} \tag{48}$$

3.1.2 Transverse vibrations of 1D structure

The Euler–Bernoulli model of a beam is considered in this section. Flexural vibrations are excited by the bending moments because of expansion and contraction of the impedance sensor. The resulting equation of motion is

$$\rho A \cdot \ddot{w}(x,t) - EI \cdot w''''(x,t) = -M_a''(x,t) \tag{49}$$

The solution for transverse displacement is the modal expansion

$$w(x,t) = \sum_{s=0}^{\infty} C_s W_s(x) \cdot \mathrm{e}^{i\omega t} \tag{50}$$

where $W_s(x)$ are the orthonormal flexural mode shapes satisfying the homogeneous differential equation $EI \cdot W_s''''(x) = \omega_s^2 \cdot \rho A \cdot W_s(x)$. Substitution of $W_s''''(x)$ in terms of $W_s(x)$, multiplication by $W_v(x)$ and integration over the length of the beam gives the modal participation factor

$$\begin{aligned} C_s &= \frac{-\int_0^L W_s(x)\hat{M}_a''(x)\,\mathrm{d}x}{\rho A \cdot (\omega_s^2 + 2i\zeta_s\omega\omega_s - \omega^2)} \\ &= \frac{\hat{M}_a \cdot [W_s'(x_a) - W_s'(x_a + l_a)]}{\rho A \cdot (\omega_s^2 + 2i\zeta_s\omega\omega_s - \omega^2)} \end{aligned} \tag{51}$$

where a damping ratio ζ_s is introduced to account for energy dissipation.

Therefore, the expression for the flexural displacement is

$$\begin{aligned} w(x,t) = \frac{\hat{M}_a}{\rho A} \sum_{s=0}^{\infty} \frac{[W_s'(x_a) - W_s'(x_a + l_a)]}{\omega_s^2 + 2i\zeta_s\omega\omega_s - \omega^2} \\ \times W_s(x) \cdot \mathrm{e}^{i\omega t} \end{aligned} \tag{52}$$

3.1.3 Cumulative structural response

The dynamic structural stiffness, $k_{\mathrm{str}}(\omega)$, presented by the structure to PWAS can be represented as

$$k_{\mathrm{str}}(\omega) = \frac{\hat{F}_{\mathrm{PWAS}}(\omega)}{\hat{u}_{\mathrm{PWAS}}(\omega)} \tag{53}$$

where $\hat{u}_{\mathrm{PWAS}}(\omega)$ is the displacement amplitude at frequency ω and $\hat{F}_{\mathrm{PWAS}}(\omega)$ is the amplitude of the reaction force. Consider a generic point P on the surface of the beam; the horizontal displacement of this point is determined by contributions of longitudinal and flexural components:

$$u_P(x,t) = u(x,t) - \frac{h}{2}w'(x,t) \tag{54}$$

where $u(x,t)$ and $w(x,t)$ are displacements defined by equations (48) and (52), respectively. Letting P be A and B, as presented in Figure 7, and taking the difference, one determines the elongation

$$\begin{aligned} u_{\mathrm{PZT}}(x,t) &= u_B(x,t) - u_A(x,t) \\ &= u(x_a + l_a, t) - u(x_a, t) \\ &\quad - \frac{h}{2}\left[w'(x_a + l_a, t) - w'(x_a, t)\right] \end{aligned} \tag{55}$$

On substituting equations (48) and (52) into equation (55) and using formulation (41), $\hat{u}_{\mathrm{PWAS}}(\omega)$ becomes

$$\begin{aligned} \hat{u}_{\mathrm{PWAS}}(\omega) = \frac{\hat{F}_{\mathrm{PWAS}}}{\rho A}\Bigg\{ \sum_{p=0}^{\infty} \frac{[U_p(x_a + l_a) - U_p(x_a)]^2}{\omega_p^2 + 2i\zeta_p\omega\omega_p - \omega^2} \\ + \left(\frac{h}{2}\right)^2 \sum_{s=0}^{\infty} \frac{[W_s'(x_a + l_a) - W_s'(x_a)]^2}{\omega_s^2 + 2i\zeta_s\omega\omega_s - \omega^2} \Bigg\} \end{aligned} \tag{56}$$

The representation of the structural response in terms of the frequency response function (FRF) to the single input single output (SISO) excitation applied by PWAS can be achieved by dividing equation (56) by $\hat{F}_{\mathrm{PWAS}}$. Therefore, the cumulative FRF consists of longitudinal and flexural components $H(\omega) = H_u(\omega) + H_w(\omega)$:

$$\begin{aligned} H(\omega) = \frac{1}{\rho A} \cdot \Bigg[\sum_{p=0}^{\infty} \frac{[U_p(x_a + l_a) - U_p(x_a)]^2}{\omega_p^2 + 2i\zeta_p\omega\omega_p - \omega^2} \\ + \left(\frac{h}{2}\right)^2 \sum_{s=0}^{\infty} \frac{[W_s'(x_a + l_a) - W_s'(x_a)]^2}{\omega_s^2 + 2i\zeta_s\omega\omega_s - \omega^2} \Bigg] \end{aligned} \tag{57}$$

The structural response can be presented using the dynamic stiffness

$$k_{\mathrm{str}}(\omega) = [H(\omega)]^{-1} \tag{58}$$

3.1.4 Boundary conditions for the elastic beam

The mode shapes in the definition of FRF and dynamic structural stiffness depend on the boundary conditions of the host structure. Practical implementation of free–free boundary conditions is relatively straightforward and in further development it is attractive to consider an elastic beam with free ends.

The mode shapes of an extensionally vibrating free–free elastic beam are governed by the following

expression [12]:

$$U_p(x) = A_p \cos(\gamma_p x),\ \gamma_p = \frac{p\pi}{L},\ \omega_p = \gamma_p c,$$
$$c = \sqrt{\frac{E}{\rho}},\ p = 1, 2, \ldots \tag{59}$$

where the scale factor $A_p = \sqrt{2/L}$ is determined through a normalization process.

The mode shape functions for the transverse vibrations include harmonic and hyperbolic functions:

$$W_s(x) = A_s\big[\cosh\gamma_s x + \cos\gamma_s x - \sigma_s(\sinh\gamma_s x + \sin\gamma_s x)\big] \tag{60}$$

where wave speed $c_w = \sqrt{EI/\rho A}$, frequency $\omega_s = \gamma_s^2 c_w$, and scale factor $A_s = 1/\sqrt{\int_0^L W_s^2(x)\,\mathrm{d}x}$ participate in the definition. Numerical values of $L \cdot \gamma_s$ and σ_s for $s \leq 5$ can be calculated; for $s > 5$, $\gamma_s = 1/L(2s+1)\pi/2$ and $\sigma_s = 1$. The expressions above allow for modeling a 1D structure (elastic beam) with free–free boundary conditions. However, it needs to be mentioned that equations (57) and (58) are generic and permit incorporation of other sets of boundary conditions.

3.2 Experimental verification of the 1D impedance model

The model of the structural response introduced in the previous section plays an important role in determining the cumulative impedance of the sensor/structure system. In effect, the model suggests that the impedance signature measured by the impedance analyzer at the sensor terminals presents not only the sensor characteristic, but also structural dynamic features reflected in the dynamic stiffness (equation 58) and, hence, in the dynamic stiffness ratio $r(\omega)$. Therefore, expression (58) can be utilized in conjunction with equations (19) and (20) to predict impedance or admittance responses of the sensor/structure system and allow for direct comparison between theoretical and experimental data.

To determine the validity of the presented 1D impedance model, a set of experiments involving small steel beams has been conducted. Simulation parameters were selected to represent experimental specimens depicted in Figure 8. The specimens were

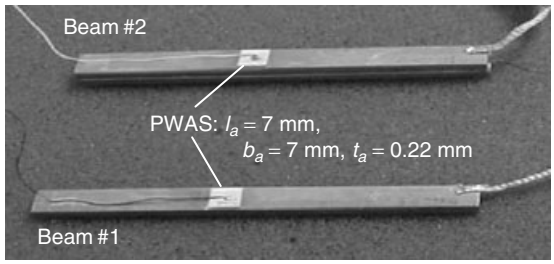


Figure 8. Experimental specimens representing one-dimensional structures.

fabricated from steel with nominal $E = 200\,\mathrm{GPa}$ and $\rho = 7750\,\mathrm{kg\,m^{-3}}$. The beams were of the same length and width, $L = 100\,\mathrm{mm}$ and $b = 8\,\mathrm{mm}$, but of different thicknesses: $h_1 = 2.6\,\mathrm{mm}$ and $h_2 = 5.2\,\mathrm{mm}$. The double-thickness beam was fabricated by gluing two single-thickness specimens back-to-back. This configuration was introduced to simulate the effect of corrosion in metallic structures and disbonding/delamination in composite structures. Noticeable in Figure 8, both specimens were instrumented with the impedance sensors placed at $x_a = 40\,\mathrm{mm}$ from the left end.

The numerical simulation was performed by coding equations (58) and (20) in the mathematical software. The viscous damping of $\zeta_p = \zeta_s = 1\%$ was introduced for steel beams. A frequency range that contains the first set of natural frequencies of the experimental specimens was considered and free–free boundary conditions were simulated using common foam. The real part of the structural system impedance was measured with the HP 4194A Impedance Analyzer. The results of the theoretical calculations and experimental testing are presented in Figure 9. For the single-thickness beam, discrepancy in the position of the experimental and calculated impedance peaks is within 4% for the first five frequencies. Consistent with the experimental observation, the model predicts five flexural peaks. The seventh peak in the impedance spectrum in Figure 9(a) is attributed to the longitudinal vibrations. It is observable at approximately 25 kHz in the single thickness specimen, causing peak splitting due to the close proximity of the flexural and longitudinal frequency contributions. As expected from the theoretical analysis, the longitudinal peak has the same 25 kHz location on the impedance spectrum of the double-thickness specimen presented

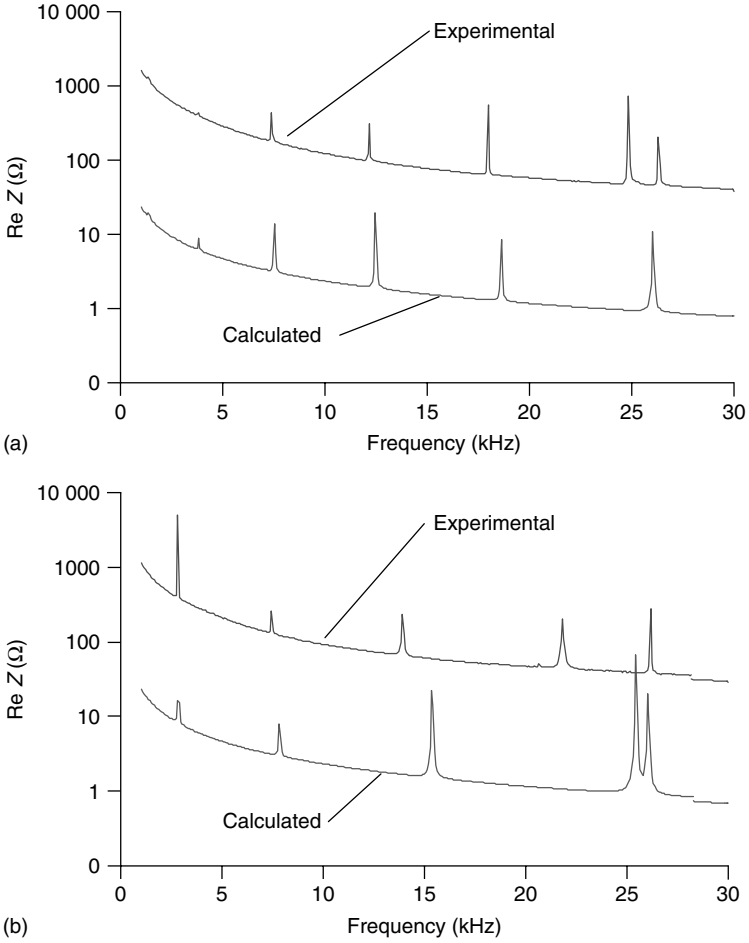


Figure 9. Experimental and calculated impedances of (a) single-thickness beam (beam #1) and (b) double-thickness beam (beam #2).

in Figure 9(b). The figure reveals a doubled period for the impedance peaks, which is in agreement with the theory for transverse vibrations. However, the discrepancy between the experimental and calculated data increases at higher frequencies and reaches almost 17% for the fifth impedance peak. It is likely that this discrepancy is caused by the adhesive layer between the two identical single-thickness beams. Even though the errors seem larger for the double-thickness beam, the confirmation of theoretical predictions by the experimental results is apparent. For the single-thickness beam, the errors between calculated and measured impedance peaks are small, and within the range normally accepted in experimental modal analysis.

4 ELECTROMECHANICAL IMPEDANCE MODEL FOR 2D STRUCTURES

4.1 Analytical model for 2D sensor/structure interaction

An axisymmetric two-dimensional geometry provides a straightforward modeling approach for the impedance sensor of a circular configuration. To take advantage of symmetry, the sensor is assumed to be bonded to the center of the circular plate. In this setting, the center of the circular plate coincides with the center of the sensor as presented in Figure 10.

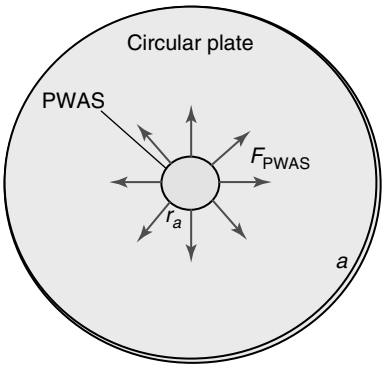


Figure 10. Excitation of a circular plate by PWAS.

Following considerations for the force and moment excitations discussed in the preceding section, the effect of the active sensor on the structure can be presented in terms of the following harmonic loads:

$$N_a(r,t) = \hat{N}_a \cdot [-H(r_a - r) + H(-r)] \cdot \mathrm{e}^{i\omega t} \quad (61)$$

$$M_a(r,t) = \hat{M}_a \cdot [H(r_a - r) - H(-r)] \cdot \mathrm{e}^{i\omega t} \quad (62)$$

Owing to the axially symmetric geometry, no circumferential component is entered into expressions (61) and (62).

4.1.1 Extensional vibrations of 2D structure

Extensional vibrations are generated by a harmonic axial force exerted by a piezoelectric sensor on a structural element. Adding this force to the balance of forces results in the equation of motion with an additional forcing component:

$$\frac{Eh}{1-\nu^2}(u''(r,t) + u'(r,t)/r - u(r,t)/r^2) - \rho h \cdot \ddot{u}(r,t) = -(N_a' + N_a/r) \quad (63)$$

where $(\dot{}) = \partial(\,)/\partial t$, $(\,)' = \partial(\,)/\partial r$ and $u(r,t)$ is the in-plane displacement of a circular plate, E is the modulus of elasticity, ν is the Poisson ratio, ρ is density, and h is the thickness of the plate.

Assuming the solution for the in-plane displacement that satisfies the homogeneous form of equation (63), one expresses $u(r,t)$ in terms of orthonormal mode shape functions $R_k(r)$.

$$u(r,t) = \sum_{k=0}^{\infty} P_k R_k(r) \cdot \mathrm{e}^{i\omega t} \quad (64)$$

where P_k is the modal participation factor for the extensional vibrations. Mode shapes $R_k(r)$ satisfy the orthonormalization condition

$$\rho h \cdot \int_0^{2\pi} \int_0^a R_k(r) R_l(r) r \, \mathrm{d}r \, \mathrm{d}\theta = \rho h \cdot \pi a^2 \cdot \delta_{kl} = \rho h \cdot \pi a^2 \cdot \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases} \quad (65)$$

Separation of the time- and space-dependent variables leads to the equation for natural frequencies

$$\frac{E}{\rho \cdot (1-\nu^2)} \left(R_k''(r) + \frac{1}{r} R_k'(r) - \frac{1}{r^2} R_k(r) \right) \cdot \frac{1}{R_k(r)} = -\omega_k^2 \quad (66)$$

Substituting solution equation (64) into equation (63), taking into consideration equation (66), multiplying by $R_l(r)$ and applying equation (65) gives a formulation for the modal participation factor P_k:

$$P_k = \frac{\int_0^{2\pi} \int_0^a R_k(r) \cdot (N_a' + N_a/r) r \, \mathrm{d}r \, \mathrm{d}\theta}{\pi a^2 \cdot \rho h \cdot (\omega_k^2 - \omega^2)} \quad (67)$$

The contribution of the force N_a is accounted for by substituting the excitation equation (61) into equation (67). Introducing viscous damping via ζ_k, one obtains

$$P_k = \frac{2\pi \hat{N}_a \cdot \left[r_a R_k(r_a) - \int_0^a R_k(r)(H(r_a - r) - H(-r)) \, \mathrm{d}r \right]}{\pi a^2 \cdot \rho h \cdot (\omega_k^2 + 2i\zeta_k \omega \omega_k - \omega^2)} \quad (68)$$

Considering the modal participation factor (68) and expansion (64), one arrives at the expression for the in-plane displacement of the circular plate,

$$u(r,t) = \frac{2\hat{N}_a}{a^2 \cdot \rho h} \cdot \sum_{k=0}^{\infty} \frac{\left[r_a R_k(r_a) - \int_0^a R_k(r)(H(r_a - r) - H(-r))\,\mathrm{d}r \right]}{(\omega_k^2 + 2i\zeta_k \omega\omega_k - \omega^2)} R_k(r) \cdot \mathrm{e}^{i\omega t} \tag{69}$$

4.1.2 Transverse vibrations of 2D structure

The problem of transverse vibrations of circular plates is very popular in engineering analysis owing to the large number of practical applications. In the particular geometry presented in Figure 10, the piezoelectric sensor expands (and contracts) axisymmetrically producing a symmetric bending moment $M_a(r,t)$. Hence, in the following derivation, the symmetric loading and boundary conditions are considered. It follows that for a plate continuous in the circumferential direction (i.e., $0 \le \theta \le 2\pi$), the load is θ-independent, and boundary conditions do not vary around the circumference. Adding moment $M_a(r,t)$ to the equilibrium of bending moments acting on the element of the plate, one arrives at the equation of motion describing transverse vibrations of a circular plate under the action of the piezoelectric impedance sensor

$$D\nabla^4 w(r,t) + \rho h \cdot \ddot{w}(r,t) = M_a'' + 2M_a'/r \tag{70}$$

where $D = Eh^3/12(1-\nu^2)$ is the flexural rigidity of a plate; for the particular axisymmetric case $\nabla^4 w(r,t) = w''''(r,t) + 2 \cdot w'''(r,t)/r - w''(r,t)/r^2 + w'(r,t)/r^3$.

The solution for transverse vibrations is presented in terms of superposition of modes

$$w(r,t) = \sum_{m=0}^{\infty} G_m Y_m(r) \cdot \mathrm{e}^{i\omega t} \tag{71}$$

Axisymmetric mode shape functions $Y_m(r)$ satisfy the orthonormalization condition

$$\rho h \cdot \int_0^{2\pi} \int_0^a Y_n(r) \cdot Y_m(r) r\,\mathrm{d}r\,\mathrm{d}\theta = \rho h \cdot \pi a^2 \cdot \delta_{mn} = \rho h \cdot \pi a^2 \cdot \begin{cases} 1 & m = n \\ 0 & m \ne n \end{cases} \tag{72}$$

and participate in the equation for natural frequencies

$$D \cdot \nabla^4 Y_m(r) = \omega_m^2 \cdot \rho h \cdot Y_m(r) \tag{73}$$

Substitution of the solution given by equation (71) into the equation of motion (70), enables one to apply the separation of variables solution procedure outlined in the preceding sections. Utilizing formulation (73) in conjunction with the effect of damping, multiplying the result by $Y_n(r)$ and applying equation (72), upon integration over the area of the plate, one obtains an expression for the modal participation factor

$$G_m = \frac{\int_0^{2\pi} \int_0^a Y_m(r) \cdot (M_a'' + 2M_a'/r) r\,\mathrm{d}r\,\mathrm{d}\theta}{\rho h \cdot \pi a^2 \cdot \left(\omega_m^2 + 2i\zeta_m \omega\omega_m - \omega^2\right)} \tag{74}$$

Substitution of the modal participation factor (74) into the solution equation (71) and consideration of the expression for moment load (equation 62) yields the final form of the transverse displacement

$$w(r,t) = \frac{-2\hat{M}_a}{\rho h \cdot a^2} \cdot \sum_{m=0}^{\infty} \frac{\left[r_a \cdot Y_m'(r_a) - Y_m(r_a) + Y_m(0) \right]}{\left(\omega_m^2 + 2i\zeta_m \omega\omega_m - \omega^2\right)} \times Y_m(r) \cdot \mathrm{e}^{i\omega t} \tag{75}$$

4.1.3 Structural response of the circular plate

The dynamic structural stiffness in terms of the force and displacement of the piezoelectric impedance sensor is given in expression (53). Since the displacement of the sensor at the center of the plate is zero,

the total displacement containing in-plane and out-of-plane (transverse) components is

$$u_{\mathrm{PWAS}}(r,t) = u(r_a,t) - \frac{h}{2}\cdot w'(r_a,t) \tag{76}$$

Therefore, a frequency-dependent amplitude $\hat{u}_{\mathrm{PZT}}(\omega)$ can be obtained by substituting extensional (equation 69) and transverse (equation 75) displacements of the circular plate into equation (76). Using $k_{\mathrm{str}}(\omega) = \hat{F}_{\mathrm{PWAS}}(\omega)/\hat{u}_{\mathrm{PWAS}}(\omega)$, it follows that the dynamic structural stiffness of a circular plate is

$$k_{\mathrm{str}}(\omega) = a^2\rho\cdot\left[\frac{2}{h}\sum_{k=0}^{\infty}\frac{\left[r_a R_k(r_a) - \int_0^a R_k(r)(H(r_a-r)-H(-r))\,\mathrm{d}r\right]R_k(r_a)}{(\omega_k^2 + 2i\zeta_k\omega\omega_k - \omega^2)} + \frac{h}{2}\sum_{m=0}^{\infty}\frac{[r_a\cdot Y_m'(r_a) - Y_m(r_a) + Y_m(0)]Y_m'(r_a)}{(\omega_m^2 + 2i\zeta_m\omega\omega_m - \omega^2)}\right]^{-1} \tag{77}$$

4.1.4 Boundary conditions for the circular plate

To facilitate comparison of experimental and theoretical results, it is attractive to consider a thin circular plate with free edges. Extensional vibration modes of a circular plate are governed by the following expression involving Bessel functions of the first kind:

$$R_k(r) = A_k\cdot J_1(\lambda_k r) \tag{78}$$

where λ_k is determined from the frequency equation

$$\frac{J_1(\lambda a)}{\lambda a}(1-\nu) = J_0(\lambda a) \tag{79}$$

With the first five roots

$$\lambda_k a = 2.05, 5.39, 8.57, 11.73, 14.88. \quad k = 1, 2, \ldots \tag{80}$$

Using equation (80), one determines natural frequencies corresponding to extensional vibrations of a circular plate with a free edge:

$$\omega_k = c\cdot\lambda_k \tag{81}$$

where $c = \sqrt{E\rho\cdot(1-\nu^2)}$. Implementation of the condition of equation (65) gives an expression for the amplitude constant A_k:

$$A_k = [J_1^2(\lambda_k a) - J_0(\lambda_k a)J_2(\lambda_k a)]^{-\frac{1}{2}} \tag{82}$$

Transverse vibration modes are described by Liessa [13]

$$Y_m(r) = A_m\cdot[J_0(\lambda_m r) + C_m\cdot I_0(\lambda_m r)] \tag{83}$$

Note that expression (83) is valid for the case when the nodal diameters are absent, which arises for axisymmetric vibrations under the applied symmetric load as suggested in Figure 10. In equation (83), m represents a mode shape number and is related to a number of nodal circles. The frequency equation for transverse vibrations of the circular plate with a free edge is relatively complex. Numerical procedures are employed to solve for λ_m, C_m, and A_m. Tabulation of these parameters for a particular case of free edge boundary conditions was presented by Itao and Crandall [14]. The availability of eigenvalues λ_m allows for the calculation of flexural natural frequencies according to the formulation

$$\omega_m = \frac{\lambda_m^2}{a^2}\cdot\sqrt{\frac{D}{\rho h}} \tag{84}$$

where a is a radius of the circular plate.

4.2 Experimental verification of the 2D impedance model

Experiments were conducted to investigate the dynamic behavior of thin circular plates subjected to free boundary conditions and the loading described in Figure 10. The experimental specimen was fabricated from aircraft-grade aluminum with $E = 70.3$ GPa,

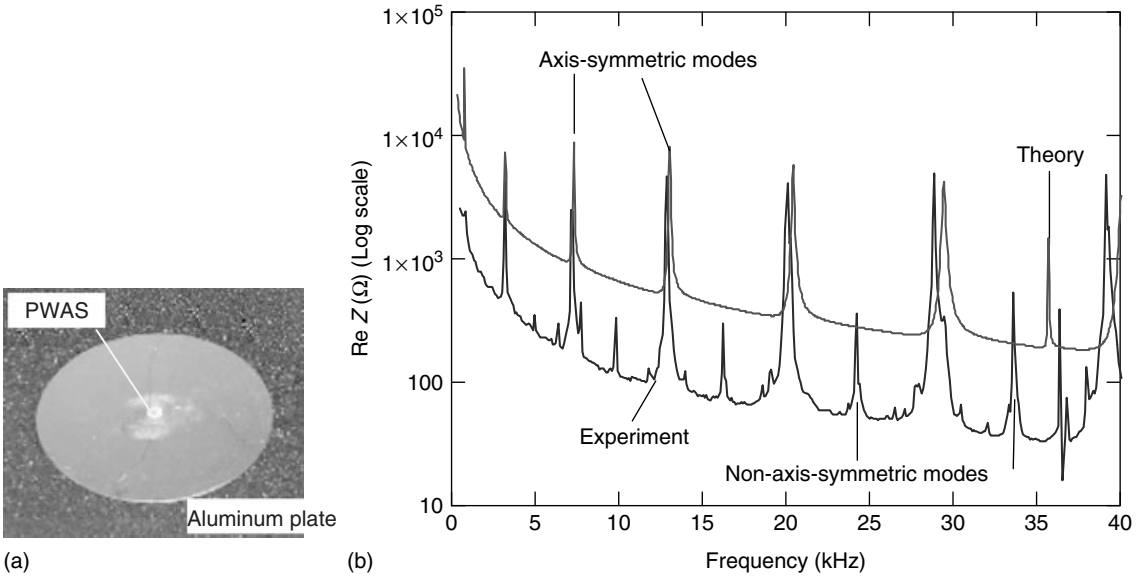


Figure 11. (a) An aluminum circular plate specimen featuring centrally located PWAS; (b) theoretical and experimental electromechanical impedances of a circular plate in (a).

$\rho = 2.8 \times 10^3\,\mathrm{kg\,m^{-3}}$, and $\nu = 0.345$. The thickness of the plate was 0.8 mm and the diameter was 100 mm. A circular PWAS (see Section 2.2 for details on sensor parameters) was installed at the center of the aluminum plate. An attempt was made to match the center of the plate with the center of the sensor as closely as possible to realize the axisymmetric geometry and load conditions considered in the theoretical development. Commercially available foam was utilized to imitate free edge boundary conditions. Figure 11(a) presents the configuration of the experimental specimen and the position of the impedance sensor.

The impedance of the 2D sensor/structure system was measured with an HP 4194A Impedance Analyzer. The experimental impedance is presented in Figure 11(b) along with the impedance calculated using expression (40). This expression describes an impedance of the sensor as measured at the electrical terminals and explicitly incorporates the structural dynamic stiffness from equation (77), reflecting the contribution of free edge boundary condition via equations (78–84). Noticeable from equations (40) and (77), the present model includes dynamics of a 2D impedance sensor and dynamics of the 2D structure with the generic boundary constraints. Theoretical consideration of a free edge boundary condition for the aforementioned plate and sensor geometries resulted in the impedance response featuring seven flexural peaks and one extensional peak. Damping was accounted for by assuming the extensional vibrations damping factor $\zeta_k = 0.07\%$ and the transverse vibrations damping factor $\zeta_m = 0.4\%$. Figure 11(b) indicates a good agreement between theoretically calculated and experimentally determined impedances. Comparison of the experimental and theoretical impedance data suggests validity of the presented modeling approach and advocates the importance of analytical procedures in improving our understanding of the EMI response of complex structures.

5 IMPEDANCE MODELING: CURRENT STATUS AND FUTURE PERSPECTIVES

Over the past decade, EMI modeling has evolved dramatically from the 1D approach pioneered by Liang *et al.* [10] to coupled numerical–analytical

techniques employed for 3D structures by Madhav and Soh [15]. Although it is probably impossible to provide a complete list of contributions, authors of this article would like to highlight some recent modeling efforts that may be of interest to researchers in this area.

The discussed EMI modeling approach incorporates both the dynamics of a piezoelectric impedance sensor and dynamics of the host structure. The latter includes longitudinal (extensional) and flexural vibrations. In this respect, the 1D development presented in this article further extends the work of Liang *et al.* [10]. Noticeably, this and most of the other models assume perfect bonding between the sensor and the host structure. Xu and Liu [16] introduced the impedance of a bonding layer and Bhalla and Soh [17] further advanced the model to include an effect of the shear lag.

Fundamental aspects of 2D impedance modeling were considered by Zhou *et al.* [18], who presented an analytical formulation for the rectangular geometry of the sensor and a thin plate; interaction of a sensor and the host structure via bending moments produced flexural vibrations of a plate. The impedance modeling for axisymmetric 2D structures was presented in this paper and in [5]. Dynamic behaviors of the sensor and a host structure were included, as well as extensional and transverse structural vibrations. Bhalla and Soh [19] introduced the concept of the effective impedance to address difficulties in model validation for complex structures. In a recent paper [15], the author's numerical–analytical procedure was extended into the 3D geometry. Cheng and Lin [20] developed a model that includes interaction of multiple impedance sensors and their respective masses. Addressing integration of the custom-built electronic circuitry into the impedance testing, several EMI models based on electrical circuit analysis [21, 22] have been suggested and validated against experimental data and established analytical formulations. There is no doubt that in future scientists and engineers will witness further improvements in the accuracy of impedance modeling, development of the EMI representations for structures of significant geometrical complexity, and increasing integration or cross-coupling of analytical formulations, numerical simulations, and experimental testing.

REFERENCES

[1] Crawley EF, de Luis J. Use of piezoelectric actuators as elements of intelligent structures. *AIAA Journal* 1987 **25**(10):1373–1385.

[2] Giurgiutiu V, Cuc A. Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention. *The Shock and Vibration Digest* 2005 **37**(2):83–105.

[3] Liang C, Sun FP, Rogers CA. An impedance method for dynamic analysis of active material systems. *Proceedings, 34th AIAA/ASME/ASCE/AHS/ASC SDM Conference*. LaJolla, CA, 1993; pp. 3587–3599.

[4] Park G, Sohn H, Farrar CR, Inman D. Overview of piezoelectric impedance-based health monitoring and path forward. *The Shock and Vibration Digest* 2003 **35**(6):451–463.

[5] Zagrai AN, Giurgiutiu V. Electro-mechanical impedance method for crack detection in thin plates. *Journal of Intelligent Material Systems and Structures* 2001 **12**(10):709–718.

[6] ANSI/IEEE Std 176-1987 IEEE Standard on Piezoelectricity.

[7] Ikeda T. *Fundamentals of Piezoelectricity*. Oxford Science Publications: 1996.

[8] Giurgiutiu V, Zagrai AN. Embedded self-sensing piezoelectric active sensors for on-line structural identification. *Journal of Vibration and Acoustics* 2002 **124**(1):116–125.

[9] Abramowitz M, Stegun I. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications: 1964.

[10] Liang C, Sun FP, Rogers CA. Coupled electromechanical analysis of adaptive material systems—determination of the actuator power consumption and system energy transfer, impedance modeling of active material systems. *Journal of Intelligent Material Systems and Structures* 1994 **5**:12–20.

[11] Park G, Cudney HH, Inman DJ. An integrated health monitoring technique using structural impedance sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**(6):448–455.

[12] Meirovitch L. *Elements of Vibration Analysis*, *Second Edition*, McGraw-Hill: 1986.

[13] Liessa A. *Vibration of Plates*. Acoustical Society of America: 1993.

[14] Itao K, Crandall SH. Natural modes and natural frequencies of uniform, circular, free-edge plates. *Journal of Applied Mechanics* 1979 **46**:448–453.

[15] Madhav AVG, Soh CK. An electromechanical impedance model of a piezoceramic transducer-structure in the presence of thick adhesive bonding. *Smart Materials and Structures* 2007 **16**:673–686.

[16] Xu YG, Liu GR. A modified electromechanical impedance model of piezoelectric actuators-sensors for debonding detection of composite repair patches. *Journal of Intelligent Material Systems and Structures* 2002 **133**(6):389–405.

[17] Bhalla S, Soh CK. Electromechanical impedance modeling for adhesively bonded piezo-transducers. *Journal of Intelligent Material Systems and Structures* 2004 **15**:955–972.

[18] Zhou S, Liang C, Rogers C. An impedance-based system modeling approach for induced strain actuator-driven structures. *Journal of Vibration and Acoustics* 1996 **118**:323–331.

[19] Bhalla S, Soh CK. Structural health monitoring by piezo-impedance transducers. *ASCE Journal of Aerospace Engineering* 2004 **17**(4):154–165.

[20] Cheng CC, Lin CC. An impedance approach for vibration response synthesis using multiple PZT actuators. *Sensors and Actuators, A* 2005 **118**:116–126.

[21] Zagrai A. Electro-mechanical analogies for modeling the structural impedance response. *Proceeding of SPIE* 2007 **6532**:6532OF.

[22] Peairs DM, Inman DJ, Park G. Circuit analysis of impedance-based health monitoring of beams using spectral elements. *International Journal of Structural Health Monitoring* 2007 **6**:81–94.

Chapter 4
Acoustic Emission

Michael R. Gorman
Digital Wave Corporation, Englewood, CO, USA

1 INTRODUCTION

Acoustic emission (AE) is a transient mechanical wave produced by some disturbance, such as a crack that suddenly grows when a material is stressed. The AE technique is one of the nondestructive evaluation (NDE) techniques recognized by American society for nondestructive testing (ASNT). The material needs to be stressed in order to use AE. The stress level should be low enough to be nondestructive; it should not compromise structural use. AE can be applied as part of routine quality assurance procedures. For example, most pressure vessels are routinely proofed as part of the manufacturing process. An advantage of AE is that it can be used to monitor structures in service and serve as an early warning of impending failure. AE signals in engineering structures can reveal the presence of growing flaws, detect impacts, locate leaks, locate frictional interferences, and, with the right setup, provide the precise load at the onset of fracture.

AE as an NDE technique has progressed to modal acoustic emission (MAE), which operates on a good physical basis. MAE is conceptually similar to seismology, SONAR and other branches of mechanics and acoustics. AE waveforms can be predicted by normal mode vibration solutions for finite plates. Wave modes allow identification of sources. This use of normal modes for interpreting AE waveforms significantly reduces the number of false calls that typically occur in amplitude/duration/counts (so-called AE parameters) techniques used in the past.

There were several problems with parameter AE, the major problem being that most of the information in the waves was filtered out by the resonant transducer, 100–300 kHz frequency filter, and "AE analyzer" circuits that reduced the signals to a handful of numbers thought to be representative of the AE wave. Dr Goranson, former Head of Structures at Boeing, in his paper titled

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

"Jet Transport Structures Performance Monitoring," Structural Health Monitoring, Current Status and Perspectives, *Proceedings of the International Workshop on Structural Health Monitoring*, Stanford University, CA, September 18–20, 1997, stated:

> 'Classical' AE data consists of various parameters intended to represent the AE waveform. The parameters characterizing a given waveform are not necessarily unique to that waveform and could be the same for other waves generated by other types of sources.

The MAE approach eliminates the nonuniqueness problem pointed out by Dr Goranson.

Another major problem practitioners soon discovered was that there were often significant errors in the source location as well. Source location is arguably the most critical part of AE. To date, source location is not part of any American society of mechanical engineers (ASME) codes. In contrast, SONAR and seismological methodologies focus on determining source location.

The language of MAE is much different than the old "parameter" or "counts" AE. The method uses familiar engineering terms and theoretically based Newtonian physics. This engineering language makes it easier to explain to personnel charged with deciding the fate of a given structure. The connections between sources and waves make physical sense, just as they do in seismology.

MAE has renewed interest in AE as an NDE technique. It has built confidence in its results. The main ideas of the waves are described in this article. Examples of waves in plates are given, which show the variation in the mode shapes and frequencies encountered as the plate thickness increases. This variation with plate thickness is probably the least understood part of MAE. Actual fracture waves in a steel pressure vessel are compared with theoretically predicted waves, and crack depth is obtained. Also, a new formula is presented for calculating front end frequency. Finally, a summary is provided with some further thoughts about using MAE in structural health monitoring (SHM).

1.1 MAE background

The physical material in which a wave propagates is called the *medium*. The discussion here is confined to guided waves in plates, which provides ample material since aircraft, pipelines, pressure vessels, storage tanks, spacecraft, and so on, are constructed with platelike sections. MAE waves in plates propagate for most practical purposes in just two modes. One mode is called the *extensional*, or E, wave, and the other mode is called the *flexural*, or F, wave. The term *mode* has a meaning similar to vibrational modes but that aspect is not important here.

It was shown in [1] that there is a simple and direct connection between a source and the waves excited in plate. An ordinary pencil lead, broken on the surface of a plate, creates a large F wave and very small E wave. A pencil lead break on the edge results in a much larger E wave. The F wave amplitude may still be quite large, depending on how near to the centroidal plane the lead is broken. A crack is generally an in-plane feature, so this source must produce an E wave. This general characteristic of cracks is the basic principle behind the practical use of MAE.

The theoretical study of elastic waves excited by fracture sources has great importance in the field of AE [2]. Without this understanding, it would not be known how to properly set up a test and the meaning of any waves captured would likely be unintelligible to the observer. Numerous solutions have been proposed to calculate the elastic waves excited in an infinite plate for various AE sources [3–7] using complicated transform and inverse transform integrals. Although those solutions found use in calibrating AE sensors [8], they are difficult to extend for studying AE waveforms in finite plates where arbitrary source and receiver positions and edge reflections must be considered.

If the normal mode solution to the classical plate bending equation could be used to compute the flexural wave excited by a step-function surface load on a finite thin plate, then reflections could be accounted for and, due to the simplicity of the equations, there would be a possibility of extending the solutions to composite materials. This calculation was done in [9]. The solutions obtained were simple explicit forms that showed good agreement with the low-frequency part of an experimentally measured pencil lead break signal, including the reflections. It is well known, however, that the solutions of the classical plate bending equation are invalid at higher frequencies, so the calculations did not give zero displacement

in the received waveform prior to the arrival of the Rayleigh surface wave. In addition, classical plate theory cannot be used to describe the dispersive behavior of the extensional mode. To overcome those limitations, a higher-order Mindlin plate theory [10, 11] was used to compute the elastic wave solutions using the same normal mode expansion approach for a finite plate [12, 13]. These solutions give very good agreement with experimentally measured E and F waves excited by various sources.

Phenomenologically speaking, the E and F waves are similar to the longitudinal (P or L) and shear (S) waves that are familiar in earthquake engineering (P, S) and in ultrasound (L, S), but there is a very important difference: L and S waves do not change shape as they propagate; E and F waves do. L and S are bulk waves; they propagate in media whose dimensions are large compared to the wavelength. Once the wavelength is on the order of one of the dimensions of the medium, like the thickness of a plate, the geometry of the part affects the wave motion considerably.

The E wave in a plate has its largest particle displacement in the plane of the plate, basically a stretching motion, while the F wave is a bending motion with its largest displacement out of plane or perpendicular to the plate's surface. The stretching motion of the E wave in the plane of the plate produces a motion of the surfaces of the plate, that is, the surfaces undergo a Poisson contraction toward each other. Visualizing the stretching of a rubber eraser in its long direction gives the correct image. It also demonstrates the inherent symmetric nature, with respect to the midplane. E and F waves have been observed in impacts on rods and plates (see [14] and [15] and the references therein).

The E wave has a "plate velocity" of around $5080\,\mathrm{ms}^{-1}$ in steel; the L wave has a velocity of $5920\,\mathrm{ms}^{-1}$. As noted above, the Fourier frequencies in an L wave pulse "stay together" as they propagate, that is, they all travel at the same speed and thus an L wave does not change its shape. This property makes the L wave useful for measuring the thickness of a material. A typical L wave frequency is 5 MHz. The wavelength is related to the frequency by

$$\lambda = \frac{c}{f} \tag{1}$$

Given a velocity of $5920\,\mathrm{ms}^{-1}$ in steel, the wavelength is 1.2 mm. This value is much smaller than the dimensions of the steel parts typically measured with an ultrasonic thickness gauge.

The E wave is dispersive, so the "plate velocity" only holds absolutely at zero frequency but is useful for estimating mode presence. At a frequency of 100 kHz, a typical MAE frequency, the wavelength is 5.1 mm. This is larger than the thickness of many structural sections. At 50 kHz, $\lambda = 10.2$ mm in steel.

For completeness of this present discussion, the E and F waves in plates are also known as the *lowest-order Lamb modes*, S_0 and A_0. S stands for symmetric and A stands for antisymmetric. Again, the plane of symmetry is the midplane of the plate. The inherent symmetry of a stretching motion was noted above. The inherent antisymmetry of a bending motion is also straightforward to visualize with the eraser. There are an infinite number of Lamb modes corresponding to the infinite number of sinusoidal standing waves (these are *the modes*) that are possible to imagine existing across the thickness of the plate.

Because the waves expected from crack growth can be predicted theoretically, the use of MAE is not dependent on a database of repeated tests on identical specimens. The need for such a database has often been a barrier to the use of the technique in the past. With the "counting" techniques for example, merely changing the gain changed the results and required a whole new database. Different manufacturers' equipment "counted" differently. This limitation was extremely frustrating, to say the least.

1.2 Practical testing

The main goal of MAE for SHM is to locate and identify indications for follow-up NDE techniques to characterize so that stress and fracture analysts can assess the importance of flaws.

To perform a test, broadband transducers are coupled to a part, the part is stressed appropriately, and the waves are recorded. An AE transducer is about the size and weight of a standard accelerometer and can be mounted on about any structure. An MAE recorder is essentially a multichannel digital oscilloscope with some specialized software. The frequency spectrum of AE waves ranges from the high sonic frequency, about 5 kHz, to the low

ultrasonic frequency, about 2 MHz. Frequencies in large structures rarely exceed a megahertz. Detection equipment is now inexpensive enough to be used for continuous monitoring for fail-safe applications.

Briefly put, the waves expected to propagate in a given test can be predicted in advance with reasonable accuracy. If the measured waves have the right shape and frequencies, the source can be attributed to crack growth (or delaminations, fiber breaks, etc., in composite materials). The tester should be very aware that waves created by noise will also propagate as E and F waves. Noise waves will, in general, look different from crack waves, i.e., their shapes and frequency spectra will be different. There are all sorts of sources of noise that can be introduced by direct contact or through the air.

Once the basic concepts and techniques are mastered for a given type of test, the technique is straightforward to apply and interpret. In fact, it is now regularly and routinely applied by technicians in the United States and Canada. In repetitive type tests, test technicians have been trained to recognize noise waves and crack waves. Handling new test situations requires more extensive training, and final interpretation of test results often requires significant expertise; a radiologist checks human X ray and MRI readings because the human brain is still the best at dealing with the complex mix of factors that go into the correct diagnosis.

2 MODELING MAE WAVES

2.1 Effect of plate thickness

It is well known that there are two parts to the solution to the wave equations governing plates. The homogeneous equation results when the forcing function is set to zero. The solution yields the dispersion equation. A dispersion plot shows how wave velocity varies with frequency for each mode of propagation. The inhomogeneous equation, or forced vibration problem, gives the displacements, i.e., the actual waveforms, for a given source. Just as in seismology and vibrations, closely matching the shape of AE waves to the predicted shape is critical for correct interpretation.

Many examples of MAE waves in very thin plates ($\lambda < h$, h = plate thickness) are now out in the literature. What is not well known, but has been critical in practical testing over the past 10 years, is how shapes of the E and F waves change dramatically when going from 5-mm-thick aircraft skins to 50-mm-thick pipelines and pressure vessels. Consequently, we show higher-order plate theory solutions for several plate thicknesses here. Both surface impulses and edge impulses are calculated. The impulse forcing function is an important case. Solutions for other functions can then be found by convolution. In practice, it has been found that the impulse source produces waveforms that match measured waveforms well enough to guide the data analyst to correct conclusions without knowing the exact source time and spatial functional dependence.

The calculations were performed using the tool WavePredictor™, which contains the higher-order plate theory governing equations described in [12] and [13]. The software tool allows selection of plate geometry, forcing function, source and receiver positions, transducer size, and digital recorder setup. WavePredictor™ is a Windows™-based software tool and a trademark of Digital Wave Corporation. Windows™ is a trademark of Microsoft Corporation.

The source motion and position directly influence the wave motion. Cracks open up in the plane of the plate and create an E wave. If the crack is symmetrically located about the centroidal axis, no bending moment is present and the wave motion is purely extensional, or an E, wave. When the crack tip is above or below the centroidal plane, crack growth generates both E and F waves. The F wave usually appears much larger on the recorder display because its motion is strongly out of plane and waves are usually measured with a normally sensitive transducer that is mounted on the plate surface.

The graph of phase velocity against frequency is a useful way of viewing the dispersion curve because it allows the tester, or the software, to determine the speed for a given frequency component in the FFT of a captured wave. If the arrival times are determined using the wave of that frequency, then the correct velocity can be used to calculate the source location. Source location determined in this manner will be accurate. For composite materials, the velocity varies as a function of the angle with respect to the main fiber axis. This is in addition to the variation of velocity with frequency. Effectively, a new dispersion curve is generated at each angle. A computer program can compute the source position accurately only if it

has the ability to properly account for all of these possibilities.

Dispersion curves are well known, but one is shown here for completeness. It has the typical shape for an E wave. Figure 1 shows a dispersion curve for an E wave in a 0.635 mm-thick aluminum plate. The velocity at zero frequency is called "the" plate velocity. Its value is very close to 5200 $m s^{-1}$. This is to be compared with the bulk longitudinal velocity of 6300 $m s^{-1}$ in aluminum.

There are several interesting points about this figure. First of all, there are two curves. The upper one is the phase velocity and the lower one is the group velocity. At the lower frequencies, both the phase and group velocities vary slowly with frequency and then drop rather quickly. Then they nearly level off once again at the higher frequencies. The intermediate range is said to be highly dispersive. The mildly dispersive range at the lower frequencies is very useful for estimating wavelength. This is the longest wavelength wave because it is the fastest; it will arrive before any other wave. The velocity used in source location is the group velocity because that is the velocity at which the energy of a narrow range, or group, of frequencies propagates. The double-valued nature of the velocities means that those frequency groups will mix. The bottom of the group velocity trough is an important point to keep in mind. These groups are moving more slowly than certain groups in the flexural mode and will, thus, arrive later. This can be confusing, if not understood, because it distorts the "standard look" of the E and F wave arrivals. The effect is particularly severe in thicker plate sections.

2.2 Sources normal to the plane of the plate

Figures 2–4 show waves in aluminum plates of standard US thicknesses of 0.25, 0.5, and 0.75 in. All other dimensions below are in meters. The waves were excited by an impulse source normal to the plate and on the upper surface where the receiver is also mounted. These cases are shown because the

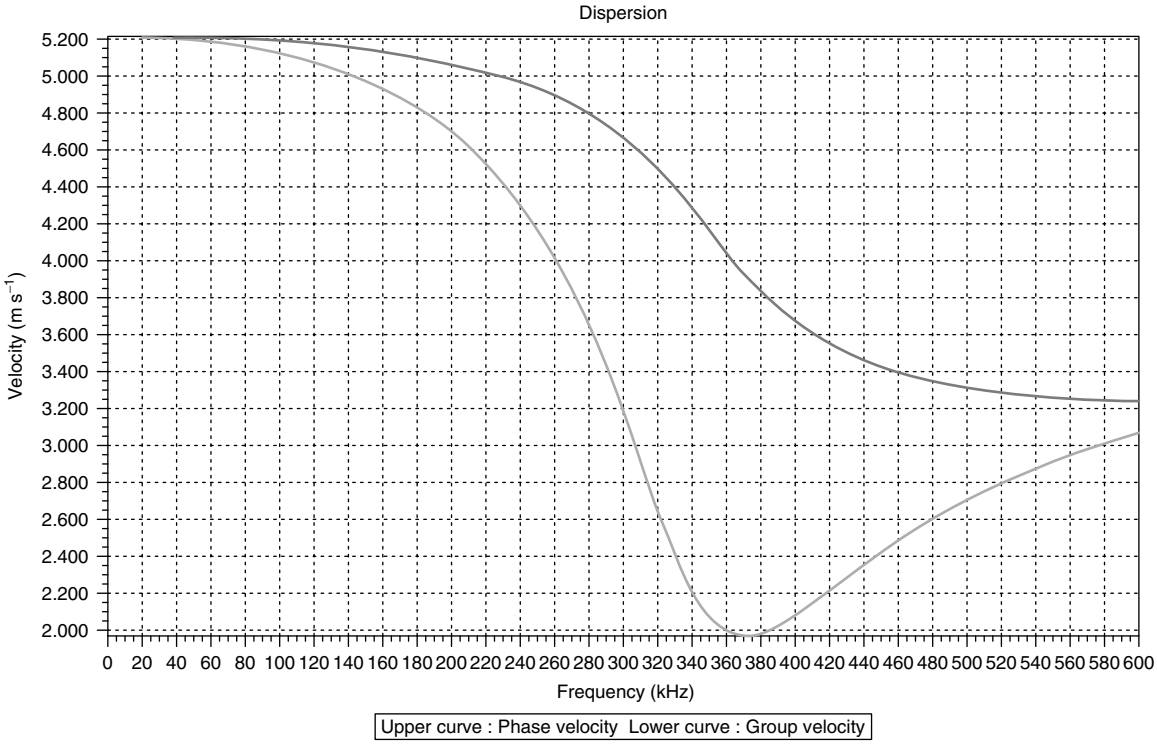


Figure 1. E wave dispersion curve for aluminum plate 0.0635 m (0.25 in.) thick.

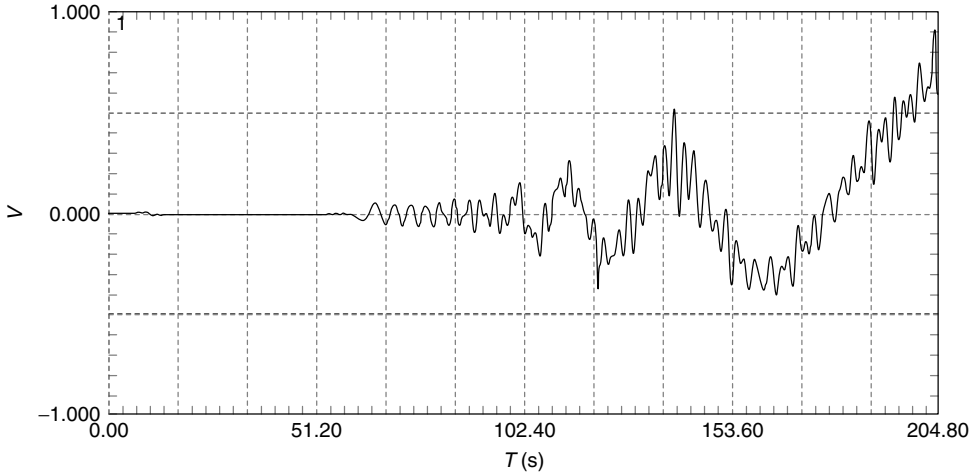


Figure 2. Wave in a 0.25-in.-thick Al plate excited by an impulsive source normal to the upper surface. The front-end frequency is about 180 kHz. Reflections are superposed on the low-frequency portion of the F wave. Small ripples at front are artifacts of calculation.

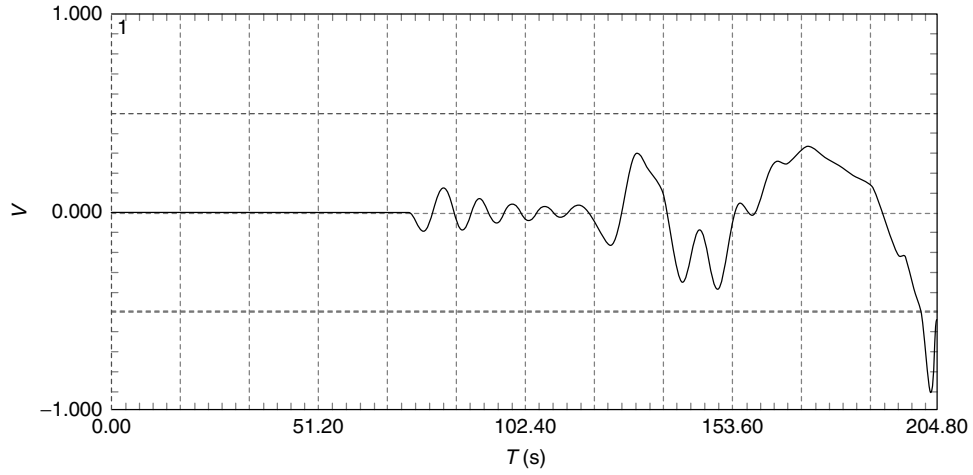


Figure 3. Wave in a 0.50-in.-thick Al plate excited by an impulsive source normal to the upper surface. The front-end frequency is about 100 kHz. Reflections are superposed on the low-frequency portion of the F wave.

pencil lead break source is the most used source and the break is almost always done on the surface of the plate. In contrast, the crack source, which is the source to be detected and located, is usually in-plane and not on the surface. The waves from this source are shown in the next section.

The lateral plate dimensions were chosen to be 0.8 m × 0.6 m in the x and y directions, respectively. The impulsive source was placed at (0.2, 0.2, $h/2$) m, where h is the plate thickness, and the receiver was fixed at (0.4, 0.4, $h/2$) m. The digitization rate was chosen to be 5 MHz. Low-pass filtering was applied so the essential features, particularly the front-end arrival and shape, were more prominent. A 20-kHz high-pass filter was applied. There were 1024 sample points, which gives 204.8 μs in the wave window. Owing to the digital nature of the calculation, small ripples can show up at early times. These ripples should be ignored. Voltage versus time is shown just as would be observed on a digital oscilloscope screen.

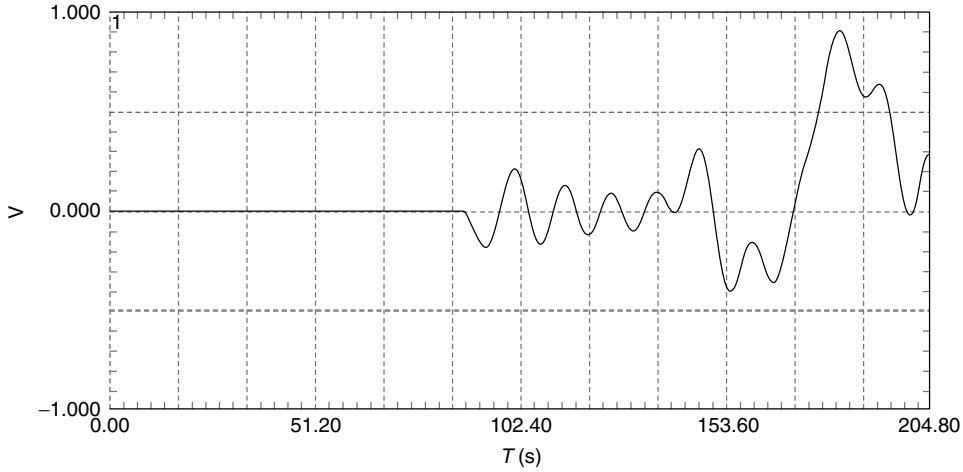


Figure 4. Wave in a 0.75-in.-thick Al plate excited by an impulsive source normal to the upper surface. Reflections are superposed on the low-frequency portion of the F wave. The front-end frequency is about 70 kHz.

The essential features to be observed in Figure 2 are the arrival of the E wave, with its low frequencies merging into higher frequencies, followed by the much larger F wave, with its high frequencies arriving before its lower frequencies. The front end of the E wave is very "clean", which is to say that there are no reflections superposed on it. It gradually merges into the F wave front end, unless the propagation distance is large enough for the two modes to separate, which is not the case here. The F wave invariably has reflections, or in the case of pressure vessels, "wrap" waves, superposed on it. Remember that the front end, or low frequency, part of the E wave travels about twice as fast as the fastest components in the F wave, so these E wave frequency components will speed around a vessel and will interfere with various parts of the direct wave depending on the source receiver distance.

Figures 3 and 4 show similar features for the thicker plates of the same lateral dimensions. The E wave reflection is interfering with the F wave. The reflection is more dispersed than the direct E wave.

For a 1-in.-thick plate, the front-end frequency is about 50 kHz. For a 2-in.-thick plate, the front-end frequency is about 25 kHz.

A formula that provides a straightforward means of calculating the front-end frequency of a pencil lead break on the surface of a material is

$$f = \frac{c}{4h} \tag{2}$$

where c is the plate wave speed for the material and h is the plate thickness [16]. This formula was determined by plotting the measured time between the first two peaks in a well-formed E wave against the thickness of plates of different thicknesses. It gives only an approximate value of the front-end frequency for a certain source to receiver distance at which the modes are well formed, which usually occurs at about $20-40h$ for a wide range of structural materials. This formula can be used to determine quickly, on the spot so to speak, if a tester is looking at plate waves.

2.3 Sources parallel to the plane of the plate

The impulse source on the edge is considered next. The "front-end frequency" is slightly lower for the in-plane source compared to the out-of-plane source. The effects of different source locations on the edge of a 0.25-in.-thick plate, relative to the midplane of the plate, are also studied. The effects are very noticeable and differences in the waveforms due to different source positions can give an idea of the length of the crack. For example, crack growth in a pressure vessel usually starts on the inner diameter (ID) surface and progresses, with pressure cycling, toward the outer diameter (OD). Initially the F wave is excited with very large energy compared to the E wave. As the crack tip progresses toward the midplane of

the plate, the relative energies in the modes changes dramatically. At the midplane the F wave energy vanishes entirely. As a crack progresses to through-wall, the modes change again. This knowledge has been used for some years to determine crack position and length in pressure vessels, and an example is given in Section 4.

For Figures 5–8 the plate dimensions are the same as above for the surface source. Now, however, the source is at (0, 0.2, z), where z is in the thickness direction. $z = 0$ is at the midplane. The receiver is mounted on the upper plate face, as in most AE tests. Again, voltage versus time is shown just as would be observed on an oscilloscope.

Figure 5 shows the waves excited for an impulsive source normal to the edge and at the midplane. The wave is purely extensional. Figures 6 and 7 show waves that result from the same type of impulsive

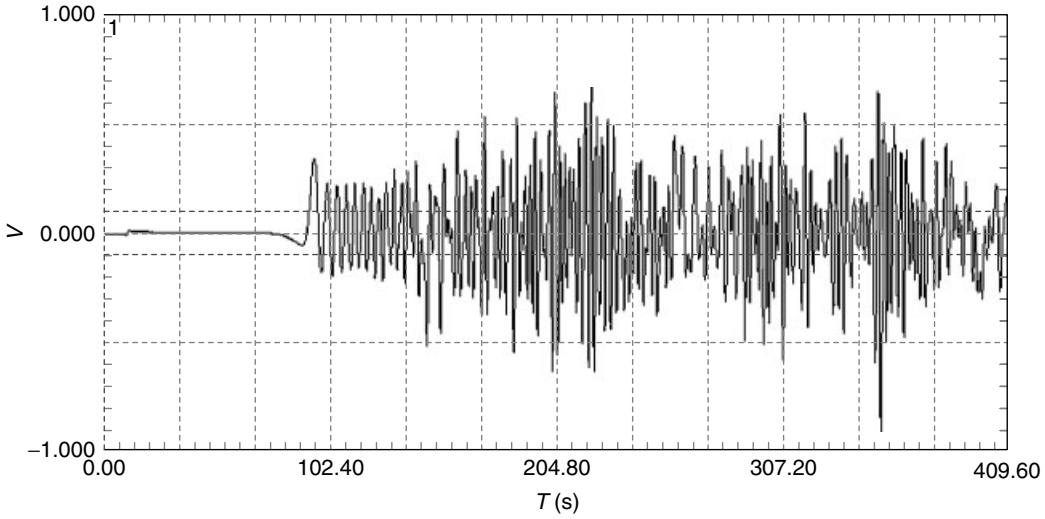


Figure 5. Waves due to impulsive source at midplane on the edge of a 0.25-in.-thick aluminum plate. Front-end frequency is about 100 kHz.

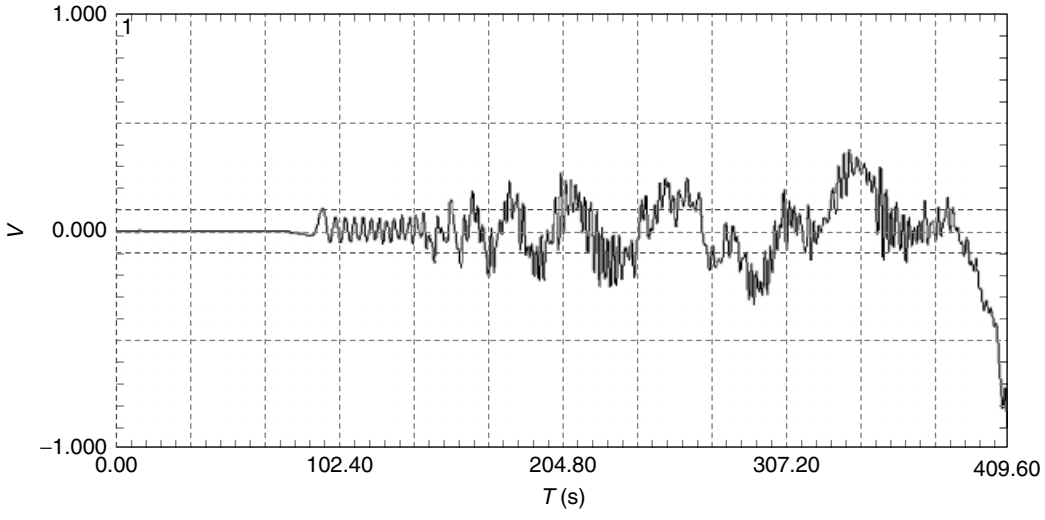


Figure 6. Same plate as Figure 5, only the source is at 0.00127 m up from center edge.

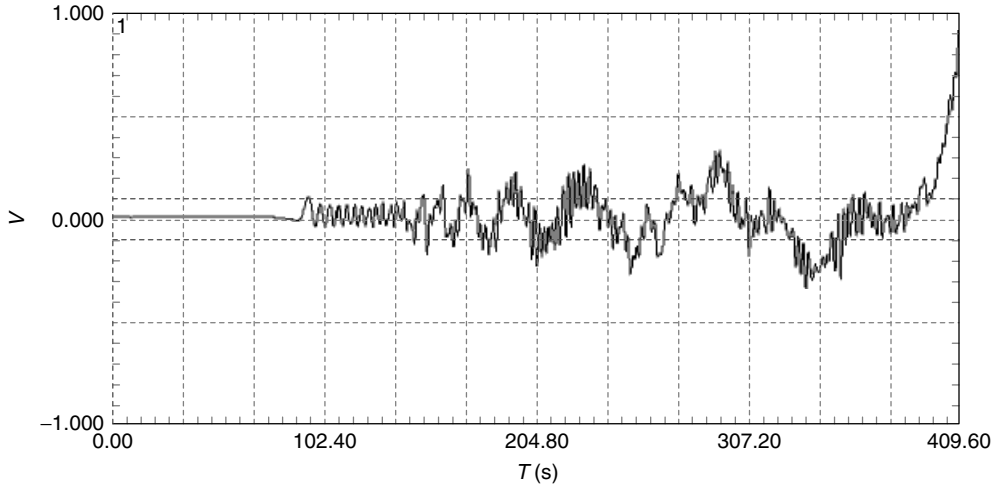


Figure 7. Same plate as Figure 5, but with source at −0.00127 m. Note the inversion of the F wave compared to the one seen in Figure 6.

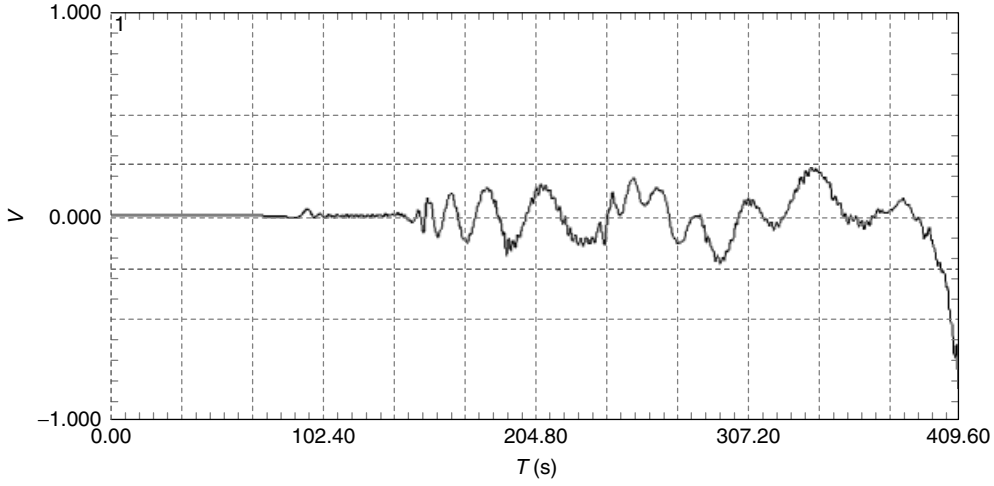


Figure 8. A 0.25-in.-thick aluminum plate. Source at 0.002875 m on edge. SR04. This is the 10% notch case. Note that E wave is far less prominent than when source is deeper but the front-end frequency is still critical and distinguishable.

source at identical positions above and below the midplane. The differences should be carefully noted. The E wave is identical in both figures but the F waves are 180° out of phase as expected. A growing crack would exhibit the same information and the inversion of the F wave would occur as the crack tip passed the midplane, thus revealing its progress from, say, ID to OD if it were in a pressure vessel.

Figure 8 shows the same type of source near the upper surface but still on the edge of the plate. In fact, this case represents the 10% notch situation. Now the E wave energy is a small fraction of the F wave energy. The bending moment is much greater than when the source was near the midplane and thus the F wave is excited to a greater degree. Although, the E wave energy is less, it is nevertheless observable and the front-end frequency, of which little has been said yet in this section, is still distinguishable. It is critical to verify the front-end frequency in order to confirm the crack growth. Noise sources are far

Table 1. Aluminum plate front-end frequencies for an in-plane source

Thickness	Front-end frequency (kHz)
0.006350 m (0.250 in.)	140
0.012700 m (0.500 in.)	80
0.019050 m (0.75 in.)	60
0.02540 m (1.0 in.)	47

less likely to have the same impulsive nature as a crack. The first arrival frequency has been used over and over in MAE testing to distinguish crack growth from noise. It is observed in a later section of this work that, in cyclic work, the presence of a crack can be observed by frictional emission emanating from interference of the crack faces, particularly as they close like teeth. This frictional-type emission is very repeatable and locates at precisely the same position in the specimen until the length of the crack macroscopically increases.

Table 1 gives the front-end E wave frequencies for an in-plane source (simulating a crack) for various thicknesses of 6061 aluminum plate.

These numbers are only approximate and correspond to a given source to receiver distance. They are displayed here to give an idea of what to expect in various thickness plates. The calculation (or measurement) for the actual source to receiver distance should be used for analysis of data. These numbers can be obtained for steel and various composites as well using the program WavePredictor™.

Note that, for greater than 0.25-in. wall, the frequency ranges above are not within the 100–300 kHz range commonly used in AE testing and called for in every american society for testing and materials (ASTM) and ASME standard. This is especially true for the thicker materials typically tested. Also, note that the "front-end frequency" is lower for an in-plane source than for an out-of-plane source. This is very useful for source identification.

3 BRIEF TREATMENTS OF OTHER IMPORTANT TOPICS

3.1 Noise

Noise must be mentioned to complete the sketch of MAE being given here. All sorts of unwanted acoustical events are generated and propagated in structures. These unwanted events are referred to as noise. Of great significance is the ability of MAE to identify noise events. Mechanical noise will propagate as plate modes. A resonant transducer and narrow bandwidth filter combination often make noise difficult to distinguish from fracture events (see the quote by Dr Goranson above).

Noise events can mislead a tester to conclude whether a part is defective or even failing. Wave modes are the key to separating fracture sounds from noise (unwanted sounds). This separation can be difficult; noise can be loud and is produced by a number of sources, particularly frictional rubbing on and in a material. The key to identifying noise is to check for the front-end frequencies and mode shapes calculated using wave mechanics.

3.2 Beneficial frictional emission

Sometimes mechanical rubbing can be useful. For example, the emissions produced by fretting of a crack's surface can be identified in many cases and used to accurately locate a crack in, say, a metallic wing skin. Frictional emission in composite materials can be used to accurately identify the location of delaminations. Relative amounts of frictional emission may be useful for characterizing damage in composites. The advantage of frictional emission is that it is a repeatable source.

3.3 Source location

Source location is extremely important [17–20]. It is also the most definite result that can be obtained from AE methods. Source identification may not always be possible due to distortion or propagation losses but source location can usually be achieved. The critical nature of a source depends on its location on a structure. Source location accuracy can also be limited by distortion, but only severe distortion makes the front end, or some definite phase point, difficult to determine with accuracy at the various arrival positions. The typical cause of severe distortion is an inhomogeneous propagation path. Placing a sensor on the outer bolted cap to a flange rather than on the vessel itself would result in a wave path with multiple

reflections that would result in severe distortion of the wave. However, the severity of these reflections needs to be determined on a case-by-case basis.

Usually the source can be located accurately only if the waves are used. Fixed thresholds often do not lead to accurate source location. With digital wave recorders, this disadvantage of fixed threshold recorders can be eliminated.

Figure 9 shows the same wave as captured at two different positions. It also shows two horizontal dotted red lines in each wave window. These are threshold crossing lines at an arbitrary 40 mV level. It can be readily seen that the waves cross the fixed threshold at very different phase points. The arrival time difference by threshold crossing is 46 μs. The correct time difference calculated from the front ends of the waves is 18 μs. On the basis of a velocity of 5.4 mm μs^{-1}, the error is 151 mm. The distance between the sensors was 152 mm. Adjusting the threshold for this particular case would mean it will be incorrect for another source position. Thus, source location based on fixed threshold crossings is to be avoided. Only source location based on methods of picking similar phase points on the waves themselves is to be considered acceptable.

Another point to mention is the use of pencil lead breaks to set up source location. The pencil lead break is usually done on the surface of a vessel, thus creating a large F wave. Usually, it is this wave that triggers the threshold crossing. However, the E wave is the one to look for when trying to detect crack growth. The only way to get around this discrepancy is to calibrate on the E wave created by a lead break or some source better suited to the vessel under test. The E wave created by a pencil lead break will be very small but still visible 40 ft away on the other end of a trailer tube. For thicker vessels, like metallic structures greater than 2-in. thick, and composite vessels greater than even 0.50-in. thick, the pencil lead break does not excite the appropriate lower frequencies with sufficient energy and another type of source is recommended such as a tap. Various devices that provide a predictable tap with the right frequency content can and should be improvised.

For composite materials, any noticeable anisotropy in the velocity curves must be taken into account.

4 EXAMPLE OF PRESSURE VESSEL TESTING

A DOT 3AA specification tube (steel, wall thicknesses of 0.590 in.) had an internal notch introduced by cutting one into the original seamless pipe and then forging the domes (Figure 10). The tube was then hydrostatically fatigued to grow a crack at the notch location.

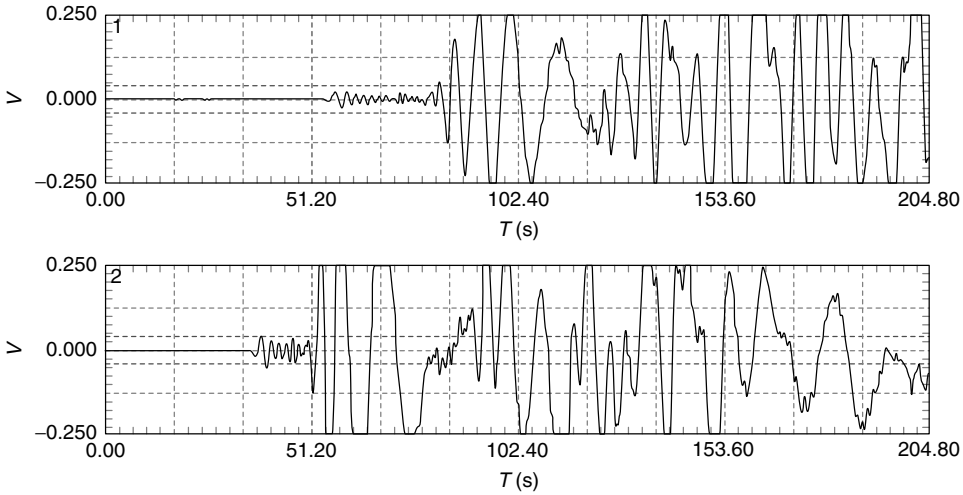


Figure 9. Waves from same source captured at two different positions. Inner most dotted horizontal lines represent positive and negative fixed voltage thresholds, both negative and positive.

Figure 10. DOT spec 3AA tube undergoing cyclic fatigue. Ellipse shows position of crack near the center. MAE transducers are at the ends of small white wires. A 3T tube, seen at the right, was also tested. Good crack growth data were obtained.

A fracture wave detector (Digital Wave Corporation, DWC) was used to record the wave signals from the sensors. The sensors used were DWC B225 sensors. These sensors have broad frequency response over the range of 50–350 kHz with a sensitivity of greater than 1 V μm^{-1}. PA2040 amplifiers and a filter/conditioning module (DWC) provided amplification and filtering. Two sensors were located near the notch, while two other sensors were located near the domes of the flask.

The 3AA cylinder was monitored with MAE equipment for 400 cycles from cycle 9001 to 9400. Figure 6 shows a typical waveform of a signal obtained for crack growth in the cylinder. It was deduced that this event came from the crack because it arrived at sensors 1 and 2 first (there were no other possible sources near these sensors), it occurred near the peak pressure of the cyclic pressure, and the correct wave mode (extensional) was formed. This event occurred near sensor number 2. As the wave propagated out from the crack growth, the plate modes began to form. The modes had not formed yet at sensors 1 and 2, because they were close to the source. It takes approximately 20–40 plate thicknesses for the modes to begin to clearly form [4].

The modes are clearly seen at sensor 3, while at sensor 4, the extensional mode is just beginning to separate from the flexural mode in the displayed A-scan. These points in time are noted in the figure. Figures 11–16 show the theoretically predicted waveforms from crack growth at various depths for a plate of the same thickness as the 3AA flask. It can be seen that the near-surface cracks (Figures 11 and 12) match the signal from the crack growth the closest. As the crack grew deeper (Figures 13–15) the extensional mode became larger, while the flexural mode decreased. Thus, by direct comparison of the crack growth signal to the theoretically predicted signals, the crack growth shown in Figure 11 is only 1–2-mm deep.

Figure 17 shows an event from corrosion on the outside diameter of a flask "popping off" the cylinder. Because there is no depth through the wall from the source, the waveform is purely flexural with no extensional components. Thus, by monitoring the wave modes, source can be identified, and follow-up inspections could be limited to those areas exhibiting flaw signals of interest.

5 EXAMPLE OF COMPOSITE ROCKET MOTOR CASE

Rocket motor cases are simply pressure vessels (Figure 18). They have two domes on the ends connected by a cylindrical section. Skirts cover up the domes, so the entire component looks like a long tube. The pressure vessel is designed to contain the pressure of the gas created when the fuel is ignited. Composite cases are usually filament wound on large mandrels. Some are very large. Others are quite small. The case shown here was made of graphite/epoxy.

The case was impacted by a blunt-nosed tup with a large mass attached just behind the nose. By varying the impact velocity, damaged areas of different sizes were created. Damage area was determined by an ultrasonic mapping that gave the extent of the damage.

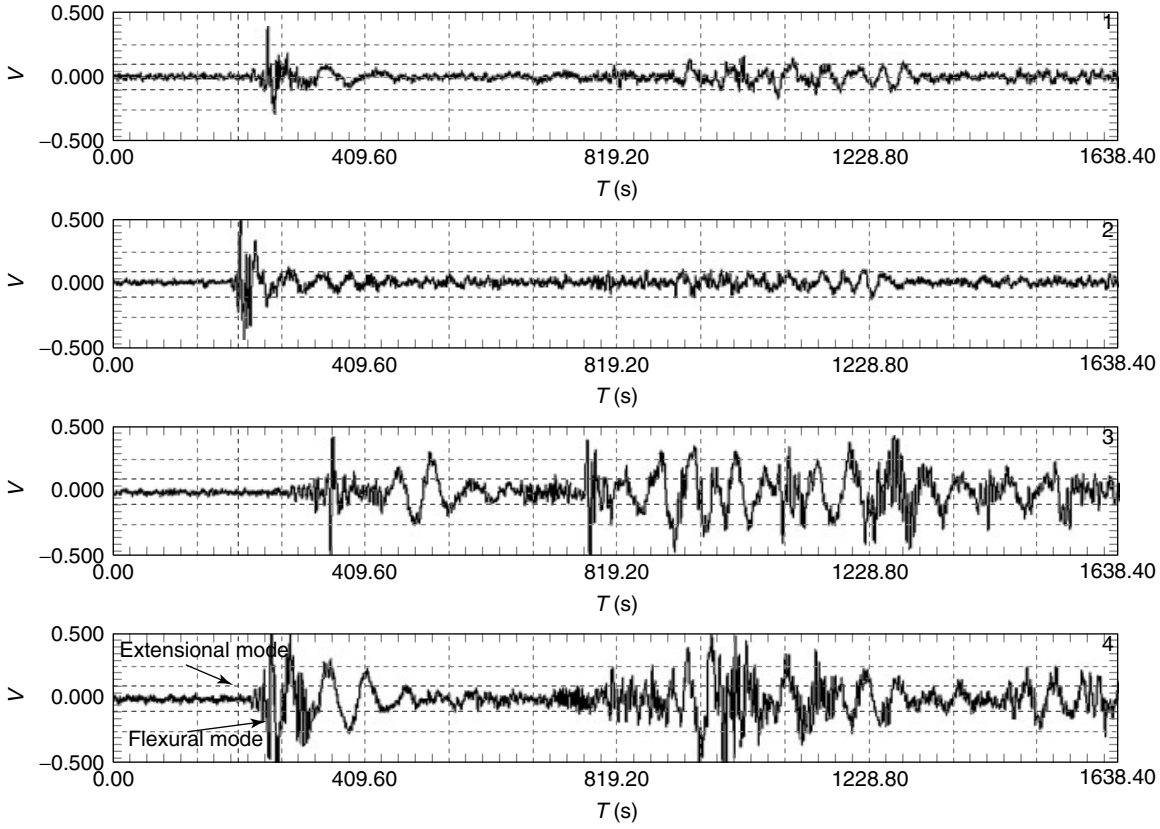


Figure 11. Waveform from actual crack growth in 3AA tube.

After a case is made, it is then moved into a hydrotest bay where it is filled with water and pressurized to ensure that it can withstand the pressure expected during actual firing of the motor, if it were loaded with propellant.

In this instance, the case was a test article. It was being studied to see the effect of impact of various energies, which is described here.

The case had wideband MAE sensors coupled to it and MAE waves were recorded as the case was pressurized. It is not the intent here to document all the details of the test. The intent here is to show some of the waves and describe their characteristics as they relate to the growth of delaminations in the case.

During pressurization, waves with both E and F waves were observed. The pressure was increased to a certain point, held for a specified period and then the case was depressurized. Every time events with large enough energy were observed, delamination growth was confirmed by ultrasonic methods. The delamination waves had very large flexural modes associated with them. The energy of the delamination wave was determined and a correlation with delamination growth area was observed. In fact, once this relationship was developed, the ultrasound confirmed it in every case. It became a dependable diagnostic tool.

The case was hydrotested to observe the effect of impact. A case of such large size could be impacted at several locations without interaction between impact sites. It was observed that increasing impactor energies created increasing damage in the composite. Delamination growth was detected by MAE and confirmed with ultrasound. An example from one of the impact sites is shown in Figure 19.

The delamination growth mapped by posttest ultrasound closely matched the source location map of MAE events. Following the ordinate upwards (increasing pressure), one can see the events begin

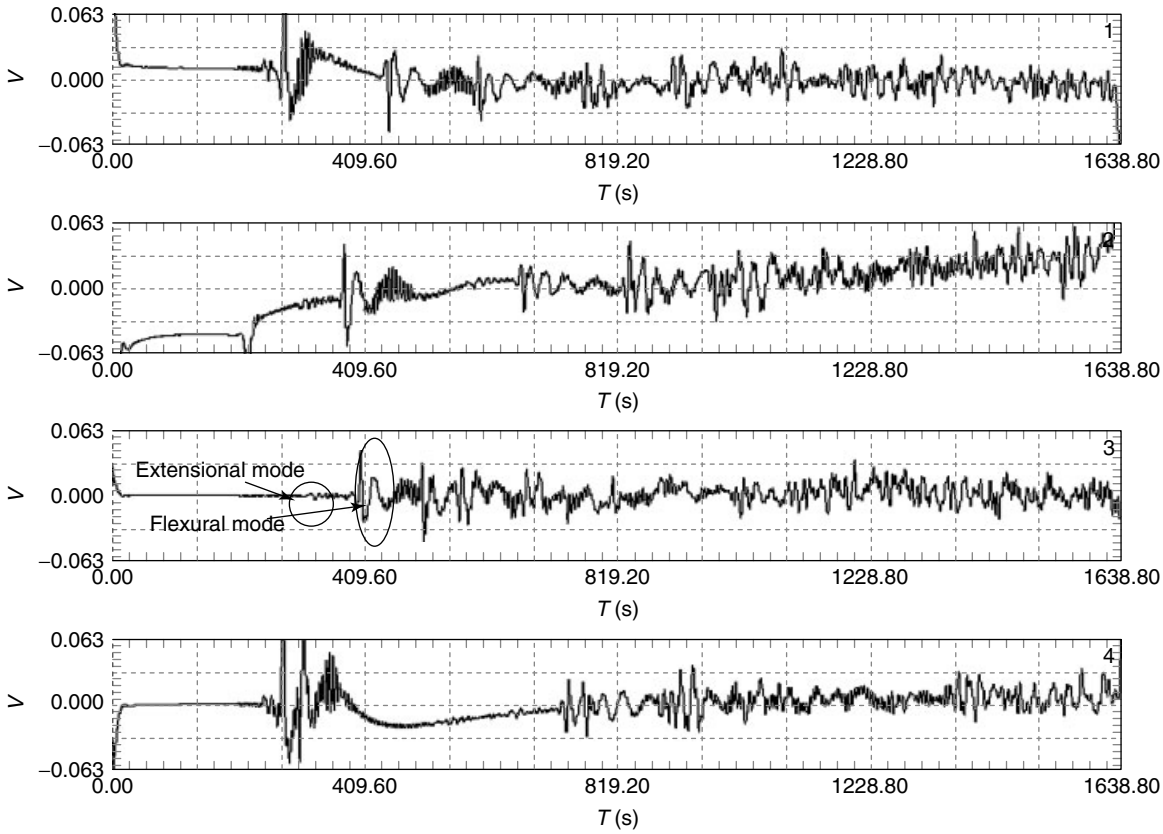


Figure 12. 0.04-in. (1-mm)-deep crack from OD. The modes have been identified on the waveform on channel 3. The waveforms show that the amplitude difference between the extensional mode and flexural mode is quite large. This is expected for near-surface cracks.

to spread outward on the composite laminate. Larger energies are also present.

Waveforms from the delamination growth are shown in Figures 20–24. Large flexural modes are observed with measurable delamination growth. This is to be expected because delamination growth is caused by interlaminar shear stresses between the delaminating layers. These double couples create bending waves, i.e., the flexural waves. These forces can be visualized as two typical pencil lead breaks displaced a small distance apart on opposite surfaces of the plate. Of course, the actual delamination source is somewhere in the plate's interior. This is why very large delaminations actually bulge out and are easily observed visually.

The waves shown in Figure 24 are those associated with damage growth at the impact site. The damage associated with the smaller energy events was too small to be observed ultrasonically. These are thought to be due to crack growth in the matrix that did not lead to observable growth in the delamination boundaries.

6 SUMMARY AND ADDITIONAL MATERIAL

AE as a nondestructive test technique has progressed to MAE, which operates on a good physical basis. It must be understood at the outset that even the language of MAE is much different than the old "parameter" or "counts" AE. This language is very important. The method is easier to explain both to technical and managerial personnel because it

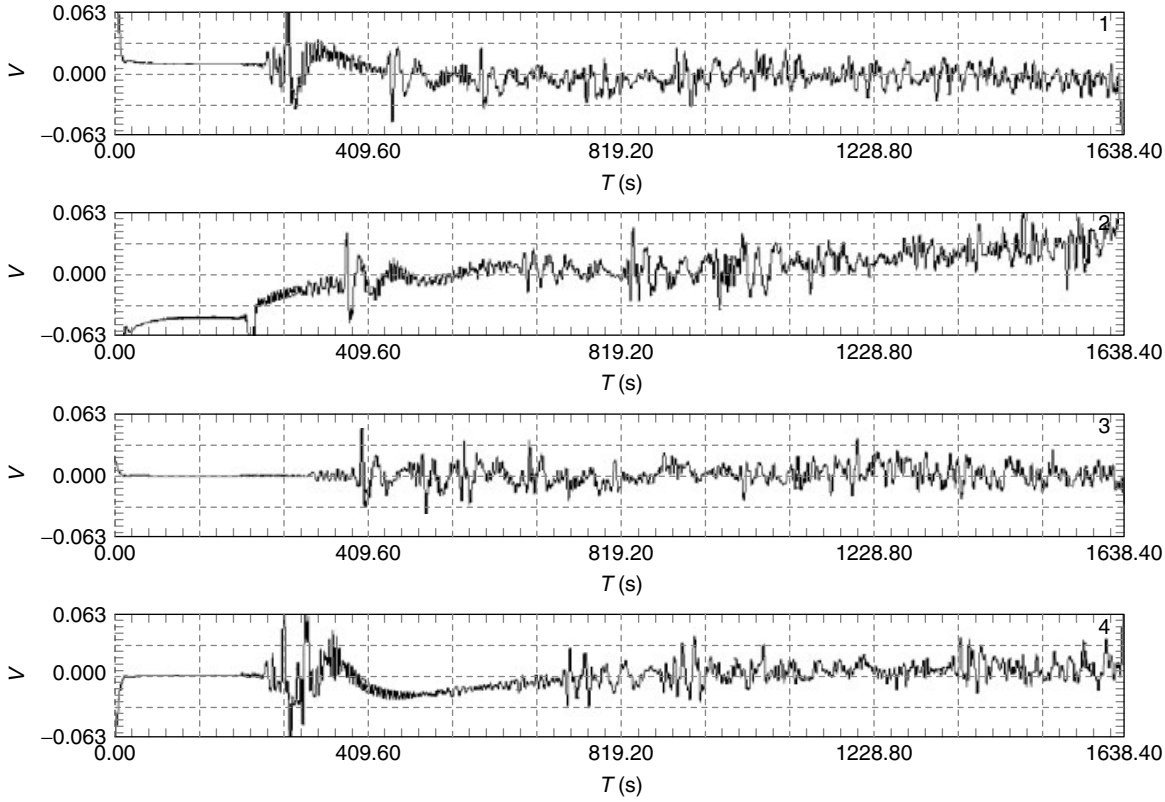


Figure 13. 0.079-in. (2-mm)-deep crack from OD. The extensional mode is now larger. This particular waveform seems to match the crack growth waveform the closest. The large offset in channels 1 and 2 is due to the numerical simulation.

contains familiar engineering terms that make a direct connection with related engineering fields.

Waves are either bulk or guided. Bulk waves are important for MAE tests of thick structures where the source to receiver distance is short and direct. Large engineering structures are built with platelike sections, so most of the practical MAE testing is concerned with plate wave propagation. The source motion and position directly influence the wave motion. An understanding of the material and its fracture behavior is very important.

As a crack propagates, from the ID to the OD of a pressure vessel, it generates both E and F waves. No E wave means the signal is not from a crack and is some form of noise.

MAE waves are broadband, that is, they contain a broad spectrum of frequencies (Fourier frequencies) depending on the source. A very sharp pulse consists of more and higher frequencies than a slower pulse. An understanding of the material and its fracture behavior is very important. Cracks usually generate sharp pulses. Some materials fail by ductile tearing. Stroking the plate's surface with a finger generates lower frequency waves.

The frequencies are created by the source. Those captured are influenced by the material. Higher frequency components present at the source may be rapidly attenuated by the material and not detected far away. This aspect is where experience plays a role in setting up a test. A few wave propagation checks with various sound sources are usually carried out prior to loading a structure. "Coverage" of an entire structure can usually be achieved with a reasonable number of sensors because plate waves are guided waves and travel great distances on the order of several meters depending on the material and geometry, and the presence of reflective surfaces or inhomogeneities that cause scattering.

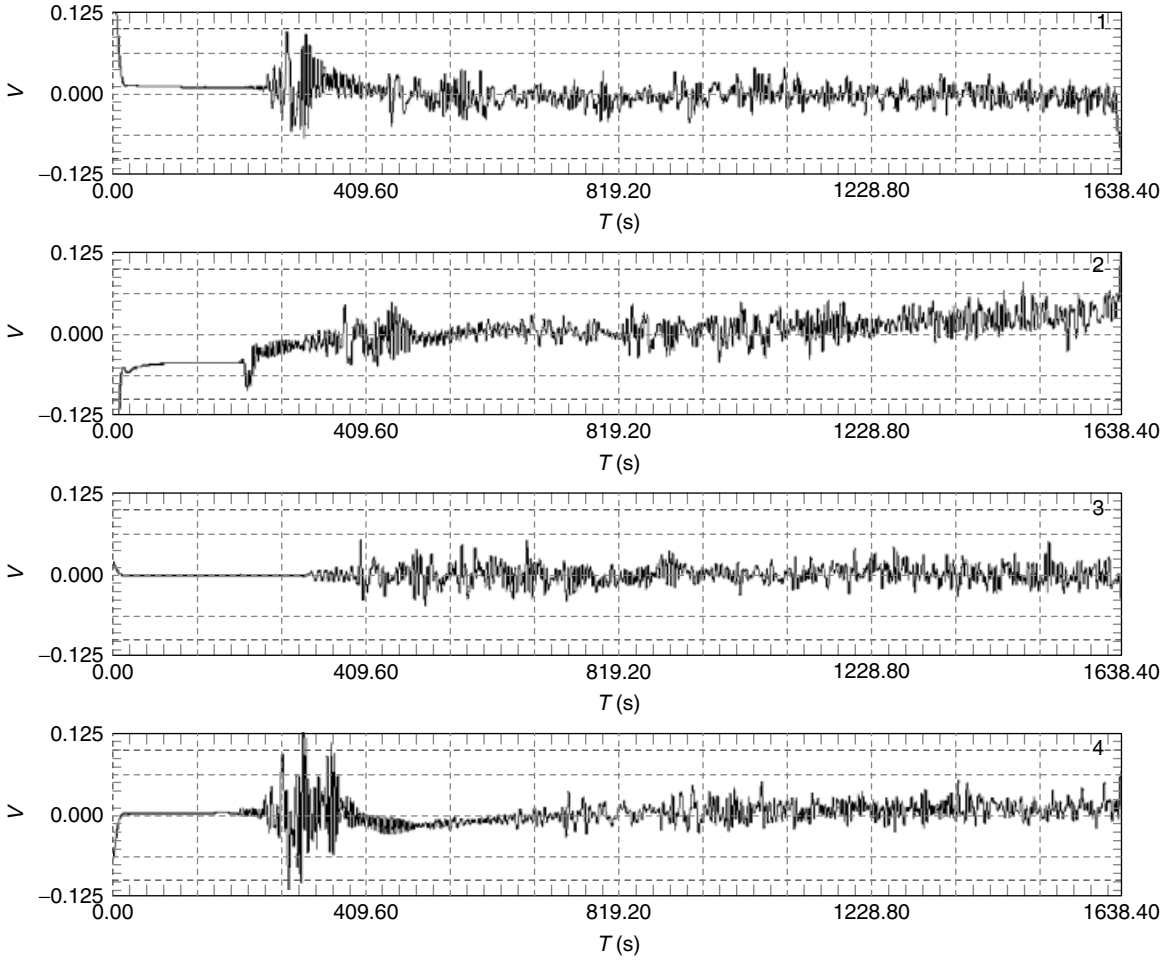


Figure 14. 0.12-in. (3-mm)-deep crack from OD. As the crack grows deeper, the extensional mode to flexural mode ratio gets larger.

Plate waves are dispersive meaning that they change shape as they propagate. This is because theory predicts that each Fourier component has a different velocity. A graph of phase velocity against frequency is called a *dispersion curve*. In practice, the two waves look almost like inversions of each other in that the low-frequency components of the E wave arrive before its high-frequency components, while the high-frequency components of the F wave arrive before its low-frequency components. It is important to use multiple channels in a test so that the dispersion behavior can be confirmed.

MAE is a predictive science that elucidates the direct connection between source and waves produced. Some think that the "digital" aspect is the whole difference between the old style AE and MAE. Nothing could be further from the truth. One method of AE is theoretically based, the other is not. Fewer specimens are needed with MAE and design changes can be handled theoretically, which means that a whole new series of tests can be eliminated.

Waveforms created by crack growth in a pressure vessel and delamination growth in composite rocket motor case were presented and briefly discussed. These examples demonstrate the power of the MAE technique to precisely identify and locate the source of emission. The defect growth was confirmed by ultrasound.

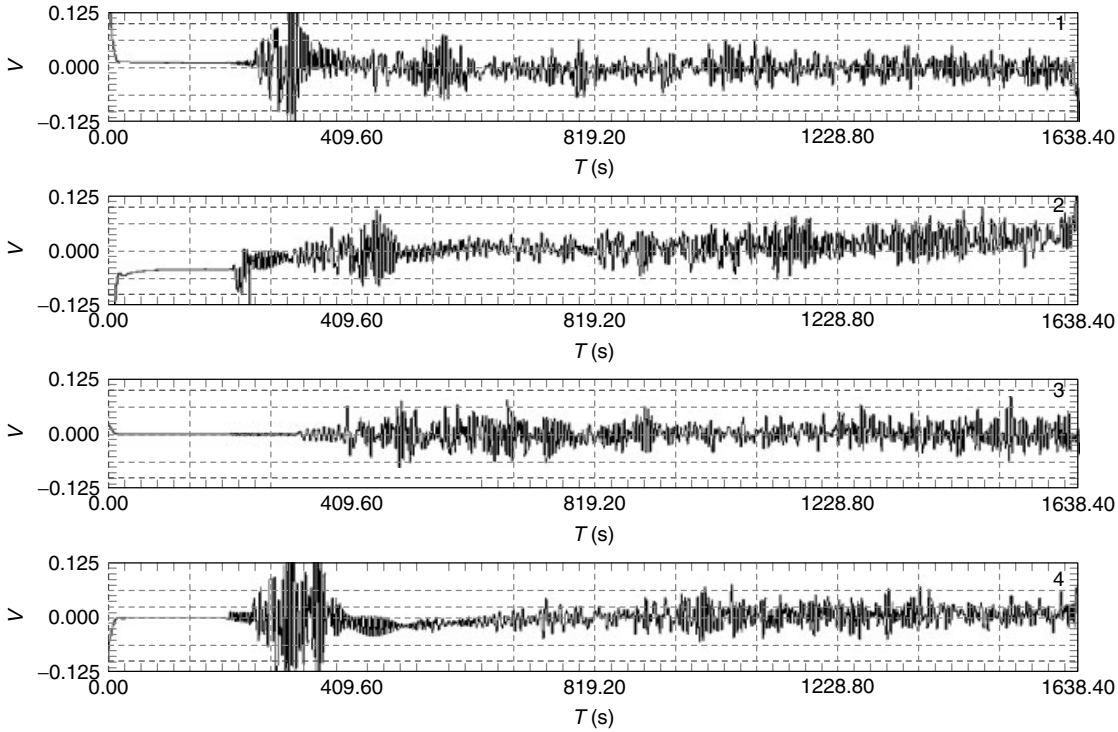


Figure 15. A 0.16-in. (4-mm)-deep crack from OD.

It is important to use multiple channels in a test so that the dispersion behavior can be confirmed.

Examples of noise were given. Noise is understood as sound waves that are undesirable in a given test. Frictional sounds produced by two parts rubbing together are very common. The waves created are usually narrowband and consist of low-frequency components. On the other hand, electromagnetic interference (EMI) is usually very high frequency but even this is not universal. Electromagnetic fields exhibit variations, too. Importantly, EMI does not exhibit modal propagation characteristics so EMI waves do not look like propagating plate waves. This is another reason why it is important to use more than one channel in any test.

It is important to capture the front end of the wave. This means allowing enough time in the pretrigger portion of the digital memory. Location analysis should be done with the front part of the E wave if possible. This is often the only undistorted part of the wave. The "front-end frequency" and its associated velocity give the most accurate location.

Like sonar, seismology, room acoustics, etc., further improvements in technology are still possible. The research areas are divided into three main areas: sources, sensors, and digital signal processing. The front-end frequency can be predicted only approximately for real sources. The exact time function of most real sources is not known especially in new testing situations. The impulse has worked well for fracture sources in steel and graphite/epoxy composites. The spatial part of the driving function depends on the stress state near the crack. This state is known inexactly. Theoretically, it is possible to determine crack length from AE waveforms as described in the text and given in the example on pressure vessel testing. Transducers have not had much fresh insight since the national institute for standards and technology (NIST) work in the early 1980s. Signal processing in MAE work is still in its infancy.

The opportunity is taken here to point out to the skeptics who have had negative experiences with AE in the past, and there are many, that the AE method developed a checkered past as attempts to use it for

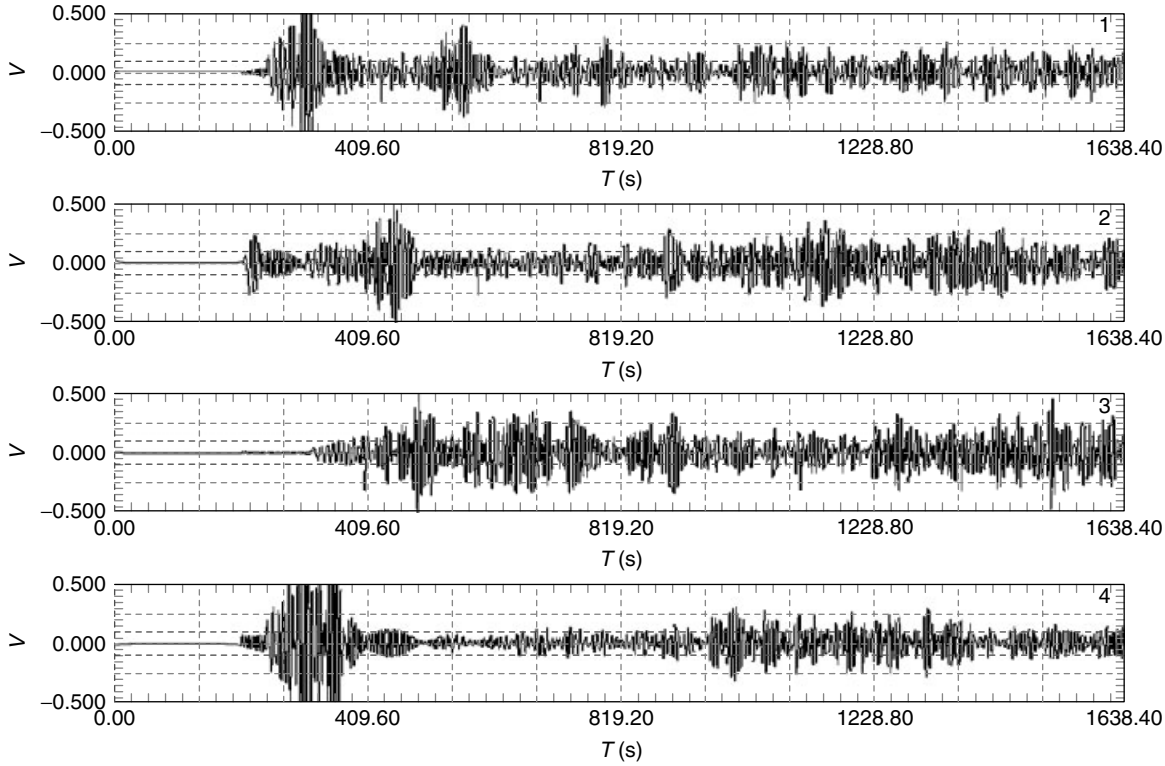


Figure 16. A 0.29-in. (7.5-mm)-deep crack from OD. This waveform is purely extensional, since it is at the midplane of the tube wall.

SHM without fundamental understanding resulted in disappointment, particularly in pressure vessels and aircraft. Practical testing is now on a firm theoretical footing. (Not all have learned from Goranson's experience—even as this paper was being written, a new standard was being balloted with the same old "one size fits all" approach—a resonant transducer and 100–300 kHz bandpass—which is found in nearly every Code and Standard extant regardless of the test article material and geometry. Professor Pao once pointed out with respect to an ASME standard balloted long ago "Were there not serious reservations expressed by concerned engineers, our honored society would have been burdened with a code that had no theoretical foundation" [2].)

To aid the buyer of SHM methods, no AE test should be performed unless the types of waves that will be measured are predicted and justified on fundamental theoretical grounds in advance. How would a good recording of music be made without knowing what type music was to be recorded and in what setting? The test equipment should be digital recorders set so the waves can be captured with clean front ends. Transducers should be of the appropriate size (diameter) and frequency range for the waves expected. (Are all sonar receivers the same?)

All analysis should be done, so the physics of source and wave is clear. The results should be as definitive and quantitative as possible. For source location, experimentally measuring wave velocity with an artificial source like a pencil lead break is acceptable in simple cases like isotropic linear source location on a tube trailer tube, but confirmation of the velocity with frequency on a theoretically predicted dispersion curve will lead to better results with a real crack source because the waves produced are different. This result means that the material properties used were correct and can be used in calculating the wave shapes, which, in turn, enables the tester to show the direct and physical meaning of the results

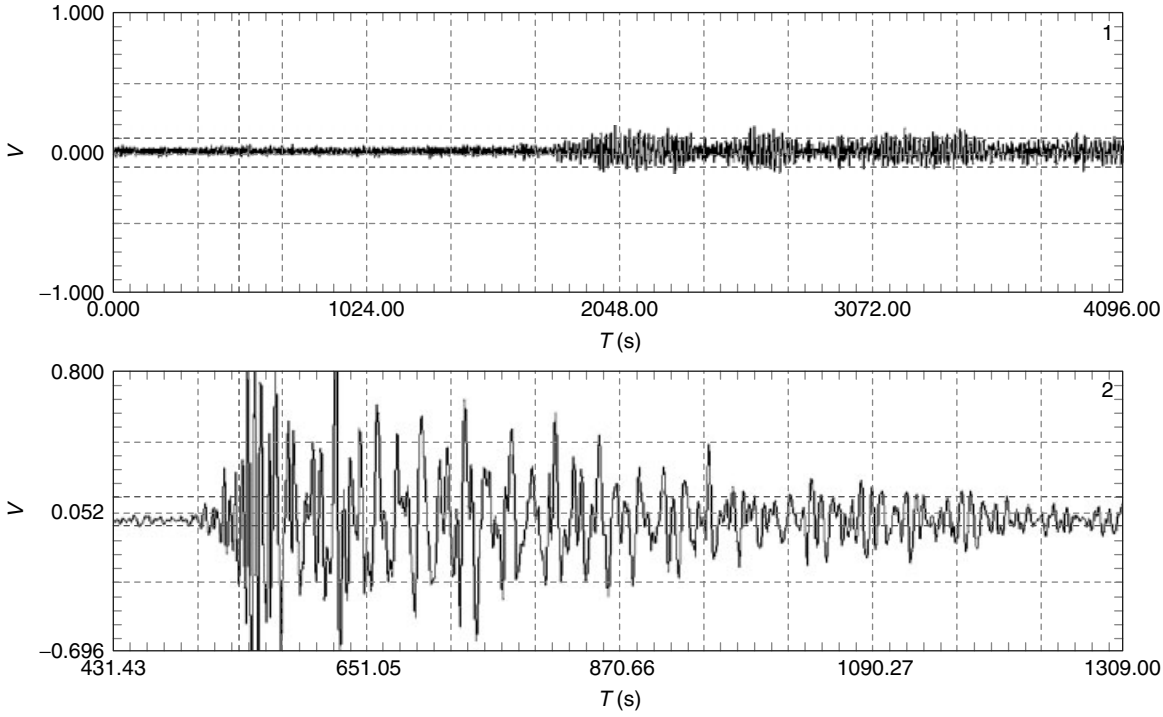


Figure 17. Waveform from corrosion on the outside of the tube. Channel 2 has been expanded to show the details in the waveform. Note the higher-to-lower frequency content as a function of increasing time. This is a characteristic of a flexural mode. Also, no extensional mode component is seen in channel 1. The extensional mode would have separated from the flexural mode component over the propagation distance, if it had been present, and been observed easily.

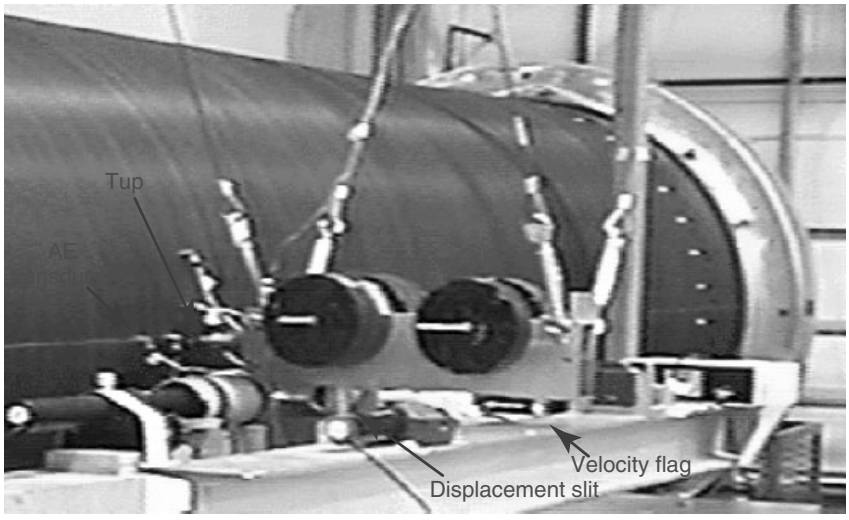
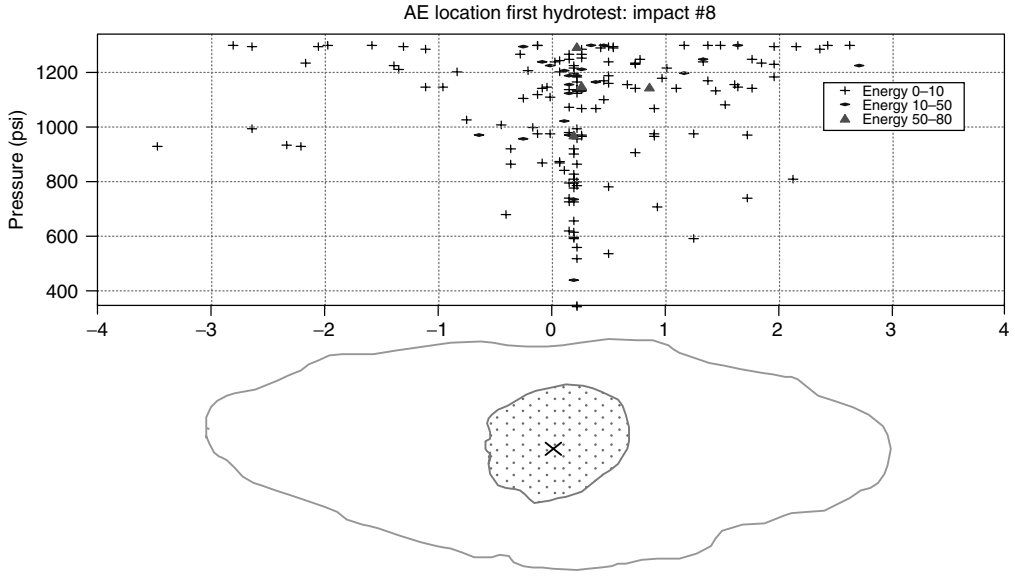


Figure 18. This shows a portion of a composite rocket motor case and how it was impacted. A rocket case is essentially a pressure vessel. Figure kindly supplied by D.S. Gardiner, Alliant Techsystems Inc.



Impact #	Energy (ft-lb)	Max load (lbf)	Displacement (in.)	UT prehydro ($in.^2$) 0.5 MHz	UT prehydro ($in.^2$) 1.5 MHz	UT posthydro ($in.^2$) 1.5 MHz
K522-08	50.4	2028	0.413	1.08	0.89	10.74

Figure 19. A map of the delamination growth at impact site #8 during hydrotest. Figure kindly supplied by D.S. Gardiner, Alliant Techsystems Inc.

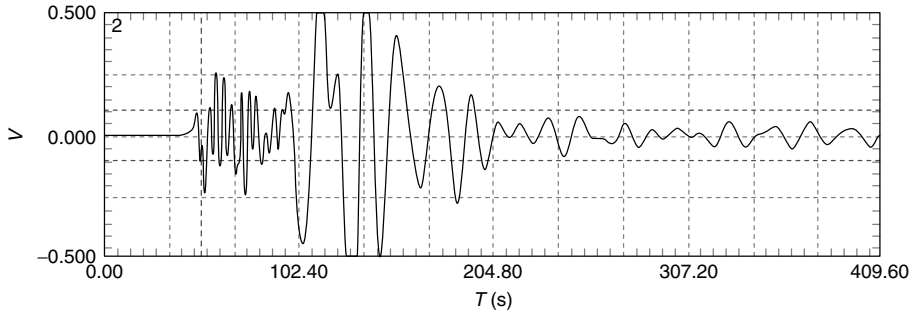


Figure 20. Typical wave from small delamination growth in the composite rocket motor case shown in Figure 18.

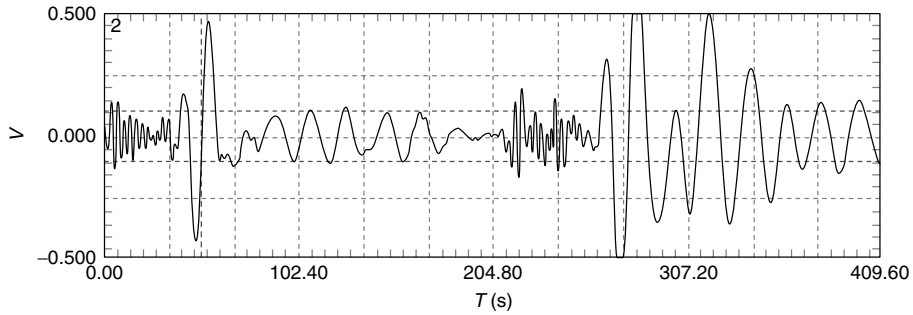


Figure 21. Two small delamination events in the composite rocket motor case shown in Figure 18.

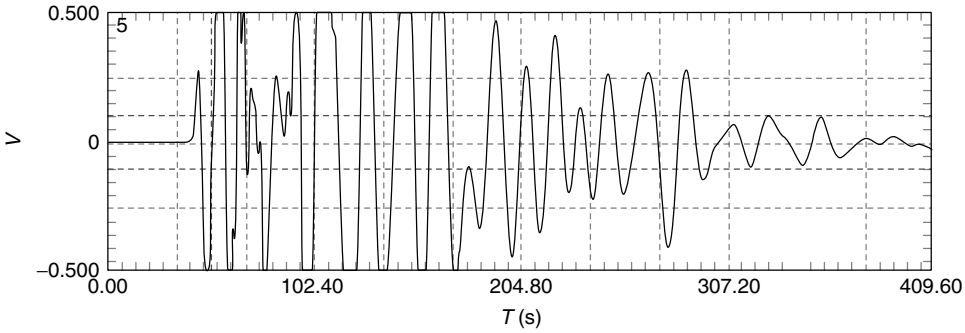


Figure 22. Medium energy delamination event in the composite rocket motor case shown in Figure 18.

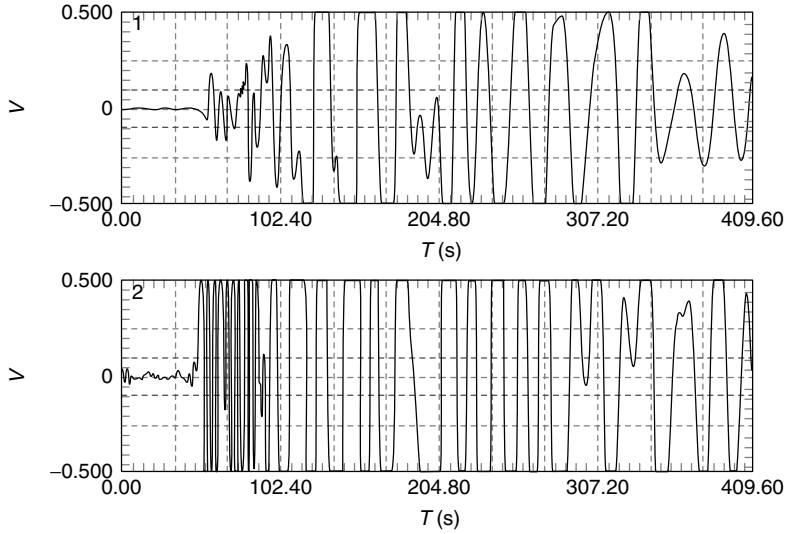


Figure 23. Large delamination event in the composite rocket motor case shown in Figure 18. The waves were so energetic that they saturated the amplifiers.

to the end user of the information. That is the kind of SHM information in which engineers and plant managers can have confidence.

REFERENCES

[1] Gorman MR. Plate wave acoustic emission. *Journal of the Acoustical Society of America* 1991 **90**(1): 358–364.

[2] Pao YH. Theory of acoustic emission. *Elastic Waves and Non-Destructive Testing of Materials*. ASME, 1978; AMD-Vol. 29, p. 107.

[3] Ceranoglu AN, Pao YH. Propagation of elastic pulse and acoustic emission in a plate. *Journal of Applied Mechanics* 1981 **48**:125.

[4] Weaver RL, Pao YH. Axisymmetric elastic waves excited by a point source in a plate. *Journal of Applied Mechanics* 1982 **49**:821.

[5] Vasudevan N, Mal AK. Response of an elastic plate to localized transient sources. *Journal of Applied Mechanics* 1985 **52**:356.

[6] Mindlin, RD. Influence of rotary inertia and shear on flexural motions of isotropic elastic plates. *Journal of Applied Mechanics* 1951 **18**:31.

[7] Lih S-S, Mal AK. On the accuracy of approximation plate theories for wave field calculations in composite laminates. *Wave Motion* 1995 **21**:17.

[8] Proctor TM, Breckenridge FR, Pao YH. Transient waves in an elastic plate: theory and experiment

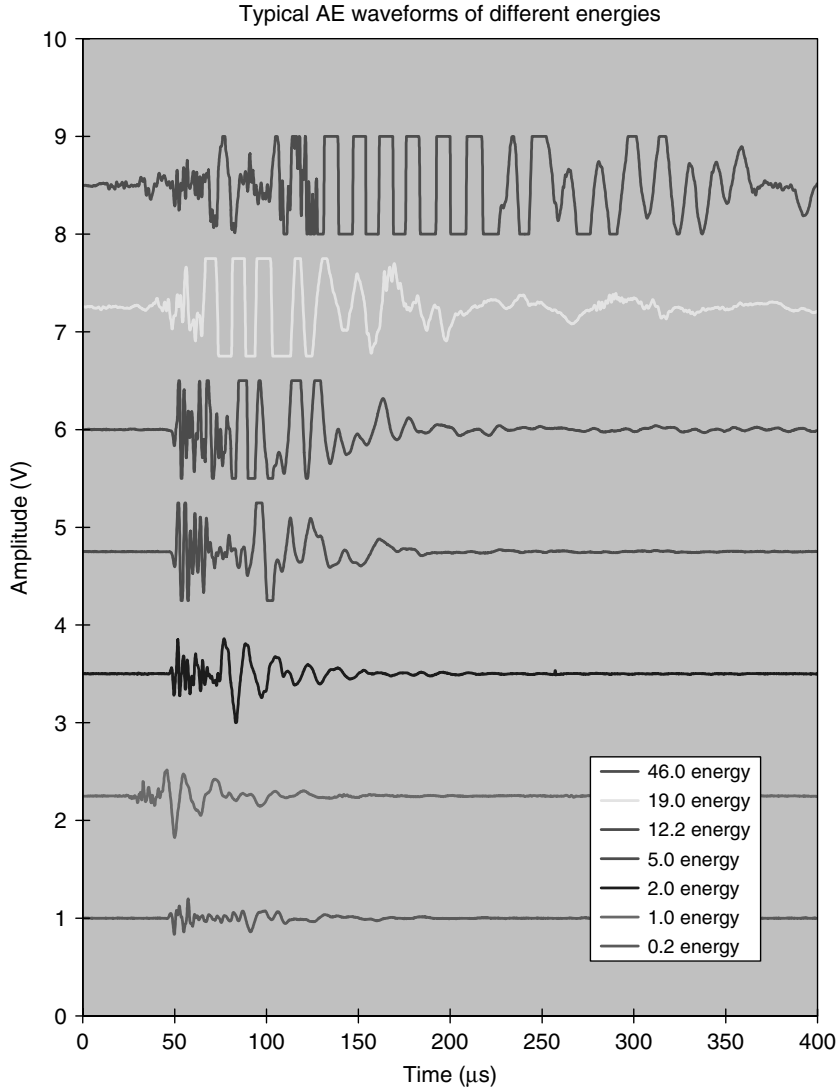


Figure 24. Energies in AE events in the composite rocket motor case shown in Figure 18. Relative energy scale. The two largest energies were from measurable delaminations, confirmed by ultrasound scans. Figure kindly supplied by D.S. Gardiner, Alliant Techsystems Inc.

compared. *Journal of the Acoustical Society of America* 1983 **74**:1905.

[9] Gorman MR, Prosser WH. Application of normal mode expansion to acoustic emission waves in finite plates. *Journal of Applied Mechanics* 1996 **63**:555.

[10] Mindlin RD. Influence of rotatory inertia and shear on flexural motions of isotropic elastic plates. *Journal of Applied Mechanics* 1951 **18**:38.

[11] Mindlin RD, Medick MA. Extensional vibrations of elastic plates. *Journal of Applied Mechanics* 1959 **26**:561.

[12] Huang W. *Review of Progress in Quantitative Nondestructive Evaluation*. Plenum press, New York, 1998; Vol. 17.

[13] Huang W, Ziola SM, Gorman MR. *Review of Progress in Quantitative Nondestructive Evaluation*. Plenum press, New York, 1999; Vol. 18.

[14] Kolsky H. *Stress Waves in Solids*. Dover Publications: New York, 1963.

[15] Graff KF. *Wave Motion in Elastic Solids*. Dover Publications: New York, 1991.

[16] Gorman MR, Ziola SM. *Digital Wave Corporation. Internal Report*. 2002.

[17] Gorman MR, Ziola SM. Plate waves produced by transverse matrix cracking. *Ultrasonics* 1991 **29**:245–251.

[18] Ziola SM, Gorman MR. Plate waves produced by transverse matrix cracking. *Journal of the Acoustical Society of America* 1991 **90**(5):2551–2556.

[19] Searle I, Ziola SM, Rutherford P. *SPIE—The International Society for Optical Engineering*. SPIE: Bellingham, WA, 1995; Vol. 2444, pp. 212–223.

[20] Martin CA, VanWay CB, Lockyer AJ, Kudva JN, Ziola SM. *SPIE—The International Society for Optical Engineering*. SPIE: Bellingham, WA, 1995; Vol. 2444, pp. 204–211.

Chapter 6
Thermomechanical Models

Minh P. Luong
LMS CNRS UMR7649, Ecole Polytechnique, Palaiseau, France

1 INTRODUCTION

Thermomechanical coupling effects in engineering materials and structural components have traditionally been neglected in thermal stress analyses. The temperature field and the deformation induced by thermal dilation and mechanical loads were solved separately; however, this effect could become significant when mass inertia is not negligible, due to the flux of heat generated through the boundary of the body, or if the material is loaded beyond its stable reversible limit. The relevance of coupled thermomechanical analysis has been demonstrated for a variety of problems, such as fault analysis of nuclear reactors, damping of stress wave propagation, deformation localization after bifurcation, and strength softening of material due to the heat generated by repeated plastic deformations.

Internal energy dissipation was recognized by a number of well-known scientists [1–3]. By carrying out experiments on the cyclic twisting of cylindrical bars, Dillon [4] identified the work done to the system by plastic deformation as the major contribution to the heat effect, and proposed an internal dissipation rate **D** related to plastic strain rate. The thermal effect due to thermomechanical coupling at the tip of a moving crack has been investigated [5] within the framework of thermodynamics, taking into account stress and strain singularities. The heat generated owing to plastic deformation causes a large local temperature increase, which is expected to affect the selection of failure modes during dynamic fracture and thus influence the fracture toughness of the material. Well-developed empirical theories of plastic deformation in metals allowed engineers to successfully predict the behavior of a variety of structures and machine elements loaded beyond the elastic limit for purposes of design.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 BACKGROUND OF THERMOMECHANICAL COUPLING IN SOLIDS

Infrared thermography is a convenient technique for producing heat images from the invisible radiant energy emitted from stationary or moving objects at a distance and without surface contact or in any way influencing the actual surface temperature of the objects viewed. The temperature rise ahead of a fatigue crack has been measured using a thermographic camera to demonstrate the local heating at the tip. Attempts have been made to measure and characterize the heat generated during the cyclic straining of composite materials. The scanning infrared camera has been used to visualize the surface-temperature field on steel [6], wood [7], engineering materials [8], and fiberglass–epoxy composite samples during fatigue tests. Recently, this infrared thermographic technique has been applied in sports engineering [9]. A consistent theoretical framework is necessary to correctly interpret the thermal images.

The development of the thermoelastic–viscoplasticity governing equations requires three types of basic assumptions [10–13].

1. The basic thermomechanical quantities completely describing the thermodynamic processes with time: the motion x, the second Piola–Kirchhoff stress tensor $\mathbf{S}$, the body force per unit mass b, the Helmholtz free energy ψ, the specific entropy s, the heat supply r, the mass density ρ in the reference configuration, the absolute temperature T, the heat flux vector per unit area q, the elastic strain tensor $\mathbf{E}^{\mathrm{e}}$, the inelastic strain tensor $\mathbf{E}^{\mathrm{I}}$, and a set of internal state variables $\mathrm{IV} = \alpha^{(\mathrm{I})}$ characterizing the material irreversible behavior. All these quantities are functions of the reference position vector $\mathbf{X}$ and the time t. It is noteworthy that the inelastic strain rate can be omitted from the aforementioned list, to be recovered as an internal state variable. However, the internal state variables (IVs) provide useful information on the microscopic state and on the microstructural defects, while the inelastic strain provides information only on the current geometry.
2. The fundamental equations of mechanics defining balance laws of linear momentum, angular momentum, mass, and energy, as well as the second law of thermodynamics expressed in the above variables.
3. The constitutive relations that describe the material response and assure the compatibility of the constitutive equations with the fundamental equations of mechanics.

When restricting the analysis to perfectly viscoelastic–plastic material and small perturbation assumption, this leads to the following coupled thermomechanical equation:

$$\rho C_v \theta_{,t} = \rho r + \mathrm{div}(k\ \mathrm{grad}\theta) - (\beta : D : \mathbf{E}^{\mathrm{e}}_{,t})\theta + \mathbf{S} : \mathbf{E}^{\mathrm{I}}_{,t} \tag{1}$$

where $\rho\,(\mathrm{kg}^{-1}\,\mathrm{m}^{-3})$ denotes the mass density, $C_v\,(\mathrm{J\,kg}^{-1}\,\mathrm{K}^{-1})$ the specific heat at constant deformation, $\theta_{,t}\,(\mathrm{K\,s}^{-1})$ the time derivative of the absolute temperature, r the heat sources, div the divergence operator, $k\,(\mathrm{W\,m}^{-1}\,\mathrm{K}^{-1})$ the thermal conductivity, grad the gradient operator, $\beta\,(\mathrm{K}^{-1})$ the coefficient of the thermal expansion matrix, : the scalar product operator, D the fourth-order elastic stiffness tensor, $\mathbf{E}^{\mathrm{e}}_{,t}$ the time derivative of the elastic strain tensor, $\mathbf{S}$ the second Piola–Kirchhoff stress tensor, and finally $\mathbf{E}^{\mathrm{I}}$ the inelastic strain tensor. The volumetric heat capacity $C = \rho C_v$ of the material is the energy required to raise the temperature of a unit volume by $1\ ^{\circ}\mathrm{C}$ (or 1 K).

Since the physical process underlying the problem is highly diversified, the modeling is thus approached from a purely phenomenological point of view. Such an approach can be useful in the interpretation of the energetics of the thermoelastic–plastic behavior. The classical theory of rate-independent isotropic or kinematic hardening plasticity is considered to be an adequate basis for such modeling as it offers the simplest constitutive model for elastic behavior of the material while still allowing consistent inclusion of two-way thermomechanical coupling effects.

When using internal state variables that describe structural changes in materials, the right-hand side member will be completed by other terms representing the cross-coupling effects [10]. These effects influence the evolution of temperature through the second-order terms when compared with the internal dissipation term. Their contribution to internal heating during the adiabatic process is small and so they are sometimes neglected.

This coupled thermomechanical equation suggests the potential applications of the infrared scanning technique in diverse engineering domains: detection of fluid leakage [14, 15], nondestructive testing using thermal conduction phenomena, elastic stress measurements [16], and localization of dissipative phenomena [17–19]. Thus the detected temperature change, resulting from four quite distinctive phenomena, must be correctly discriminated by particular test conditions and/or specific data reduction. This analysis is the principal difficulty when interpreting the thermal images obtained from experiments under the usual conditions.

3 THERMAL PHENOMENA

The first term on the right-hand side of equation (1) is related to the existence of sources or heat sinks in the scanning field [20]. The surface heat patterns displayed on the scanned specimen may result from either external heating, referred to in the literature as *passive heating* where local differences in thermal conductivity cause variations on isothermal patterns, or internally generated heat referred to as *active heating* [21].

3.1 Application to pharmacodynamics

This term has been recently exploited in a noninvasive technique to detect variations in skin temperature and thus to analyze the pharmacodynamic properties of topical corticosteroids using vasoconstriction as a marker [22]. Several subjective and objective methods are available to categorize the potency of topical corticosteroids on healthy skin. They are based on analysis of the vasoconstriction caused by corticosteroids. The parameters that have been studied involved changes in skin color. In this work, cutaneous thermal images were recorded in real time. A single application of short-duration test was performed without massage or occlusion. A predetermined 1.2-cm-radius template surface was used for all tests in order to standardize the results with those of the standard skin-blanching test that was performed after occlusion according to McKenzie's protocol (Figure 1).

Several skin disorders have been studied by infrared thermography, including allergies, skin infections, burns, tumors, and microangiopathy [23]. However, when used according to the standard method, this technique has been considered insufficiently sensitive to detect slight changes in superficial blood flow induced by topical corticosteroids relative to heat emitted from underlying vascularization [24]. This work proposed differential imaging that makes the

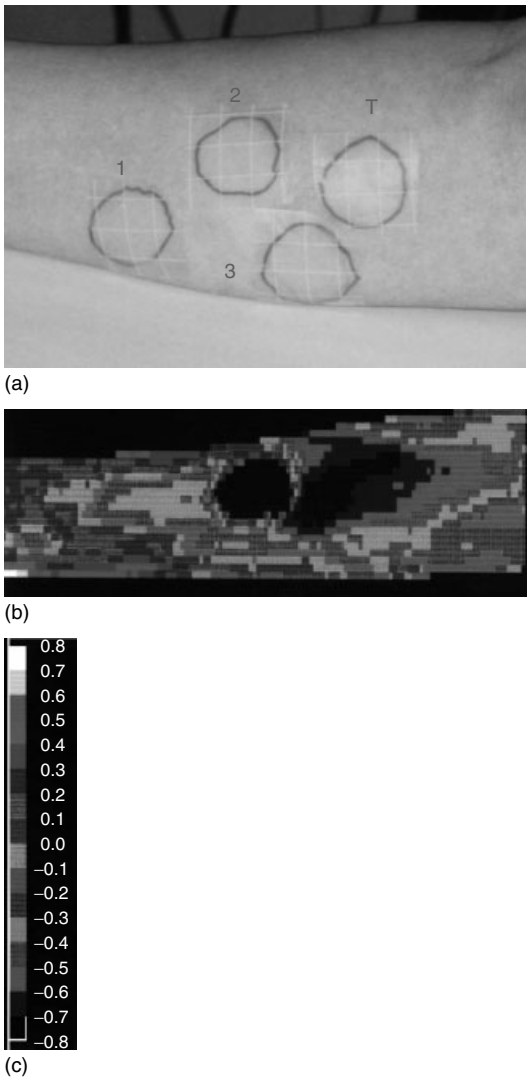


Figure 1. (a) Application of corticosteroids on a forearm to evaluate their bioavailability; (b) change in temperature of a forearm obtained by subtraction between two heat images before and 5 min after application of betamethasone dipropionate cream; and (c) temperature change scale ($\Delta\theta$ °C).

technique more sensitive. The possibility of analyzing thermograms recorded by subtraction of thermal images (differential thermography) allows for the distinction of changes in temperature by subtracting the areas, which do not change, similar to the radiological concept of quantitative or subtractive imaging. Moreover, when subtraction is performed as a function of time, the rate of variations in temperature (from $t = 0$ to t_x) over time, can be assessed, and the rate of change (acceleration between time points: t_x to t_{x+1} to t_{x+2}), can be calculated by subtracting the images from various time points. Differential infrared thermography, therefore, permits continuous quantitative evaluation of cooling in real time, without contact.

Four topical corticosteroids (reference: French classification: Class I (ultrapotent anti-inflammatory activity), Class II (potent anti-inflammatory activity), Class III (moderate anti-inflammatory activity), and Class IV (mild anti-inflammatory activity)) were tested simultaneously at different sites on the forearm to obtain as constant a background microcirculation as possible, as the background circulation depends on ambient skin temperature, resting state, and limb positioning [25]. The initial results reveal temperature changes in the first 3 h for each topical corticosteroid after a single application without occlusion (Figure 1b and c). In contrast, standard testing of vasoconstriction evaluates skin blanching after 6 h [26], as this is the minimum period accepted to date to obtain clear discrimination between topical corticosteroids. An interval of 16 h is often used because the patches can be applied in the evening for evaluation the following day.

The reference methods for evaluation of vasoconstrictive activity are based on the presence or absence of skin blanching, evaluated by attributing an arbitrary score from 0 to 3 or 4 by a trained reader, and it is, therefore, operator dependent [27]. The main value of differential infrared thermography compared to these methods is that of a *measurable objective parameter*.

In conclusion, differential infrared thermography has two notable advantages: (i) it provides a measurable objective parameter and (ii) it allows evaluation of the kinetics of topical corticosteroids shortly after application, well before the appearance of the skin-blanching effect.

3.2 Leakage detection on structural components

Infrared thermography is particularly useful when applied to the detection of thermal phenomena related to gas flow through holes in metallic walls. In these cases, calorific diffusion of the thermal gradient caused by the expansion or the compression of a gas during its passage through the hole is readily detected with an infrared camera. A short-duration test has been performed with success on large components of missile-jet tubes (Figure 2a–c). The results obtained confirmed that the infrared thermography is an excellent means of ready control, without contact, and can be easily set up for the detection of well-localized leakages.

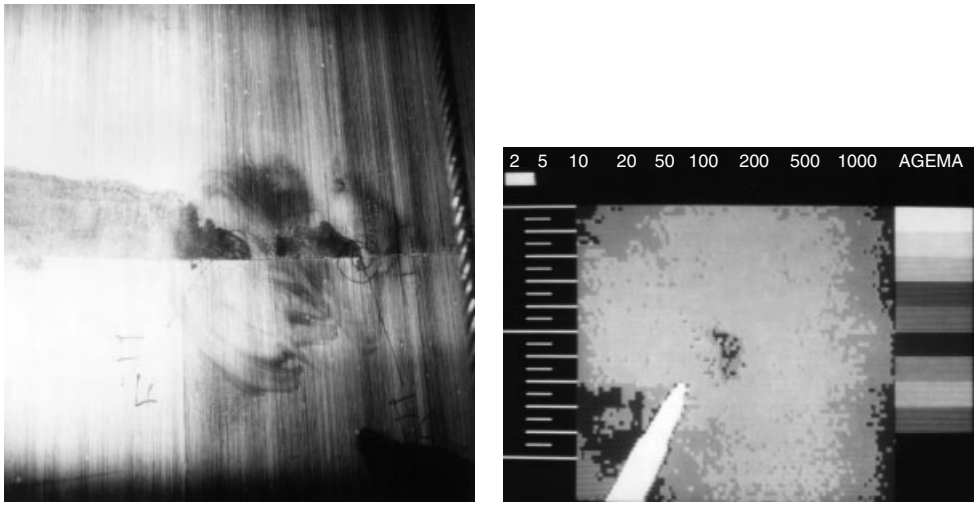
4 CONDUCTION PHENOMENA

The second term on the right-hand side of the thermomechanical equation governs the heat transfer by thermal conduction in which the heat passes through the material leading to a uniform specimen temperature. The second-order tensorial nature of the thermal conductivity, k, may sometimes be used for the detection of anisotropy in heavily loaded materials.

Variations in thermal conductivity may arise because of local inhomogeneities or flaws in the material [28]. Where an unsteady state exists, the thermal behavior is governed not only by its thermal conductivity but also by its heat capacity. The ratio of these two properties is termed the *thermal diffusivity*, $\alpha = k/C\ (\mathrm{m^2\,s^{-1}})$, which becomes the governing parameter in such a state. A high value of the thermal diffusivity implies a capability for rapid and considerable changes in temperature (Figure 3). It is important to bear in mind that two materials may have very dissimilar thermal conductivities but, at the same time, they may have very similar diffusivities. A pulsed heat flux has been used to characterize a delamination within a composite by the discontinuity caused in the temperature time history [29].

5 THERMOELASTIC COUPLING IN SOLIDS

The third term illustrates the thermoelastic coupling effect. The thermoelastic behavior of materials is



(a) (b)

(c)

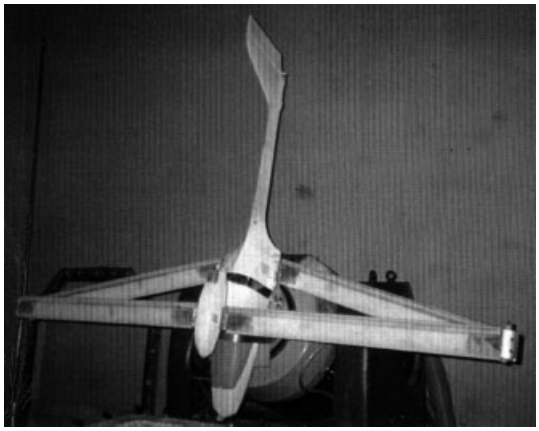
Figure 2. (a) Localization of gas leakage on a large component of missiles-jet tube; (b) leakage location #4 detected with a difference pressure of 2.5 kPa between the two faces (temperature full scale 2 °C); and (c) leakage location #5 detected with a difference pressure of 5 kPa between the two faces (temperature full scale 2 °C).

Figure 3. Thermal diffusivity characteristics of the front face of the royal palace in Fès Marocco.

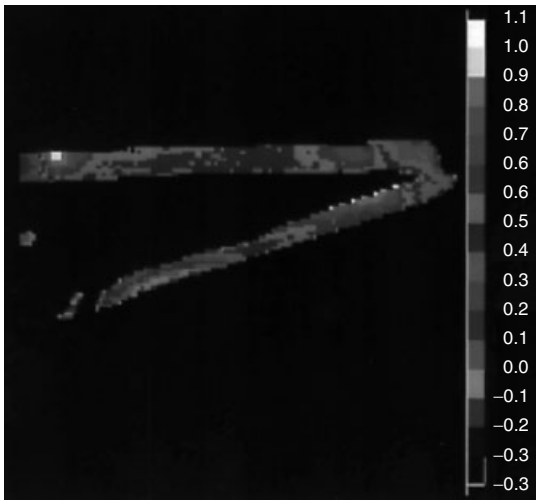
revealed by everyday experience in the form of linear or bulk expansions and contractions under the effect of temperature changes and in the applications of the elastic properties of metals and even polymers in the production of springs, pins, clips, and the like. From a phenomenological point of view, thermoelasticity is thereby linked to a notion of reversibility. The material responds to mechanical or thermal excitation in an instantaneous manner and, when the excitation is removed, returns to its initial state without showing any memory of the recent changes. Within the linear elastic range and when subjected to tensile or compressive stresses, a material experiences a reversible conversion between mechanical and thermal energy causing temperature change. Provided adiabatic conditions are maintained, the relationship between the change in the sum of principal stresses and the corresponding change in temperature is linear and independent of loading frequency [16]. This thermoelastic coupling term may be significant in cases of isentropic loading where no energy is added and no energy losses occur. A stress analysis technique is known as *stress pattern analysis by thermal emissions* (SPATE) that measures the temperature due to the thermoelastic heating and cooling of a body under cyclic loading [30].

Experience shows thermoelasticity to be a common type of behavior. It is characterized by *reversibility* of the response of the material to the excitation it undergoes. Thermoelastic behavior is modeled by assuming that current values of the temperature and strain tensor in the material element are sufficient to define its physical state. The free energy arises as the *thermodynamic potential*, a function of the current values of the temperature and strain tensor. The stress tensor is obtained by differentiating with respect to the strain tensor. In an isotropic material undergoing small perturbations, linear thermoelasticity is characterized by *two* elastic constants and *one* thermal expansion coefficient. Elastic and thermal strains can be uncoupled as follows: total strain = elastic strain + thermal strain. In practice, the thermoelastic equilibrium problem is solved by using the superposition principle. The problem consists of adding together the solutions (i) of the isothermal problem with excitations and (ii) of the purely thermal problem [31].

A vibratory test was performed on a reduced scale model of a double-winged aircraft (Figure 4a)



(a)



(b)

Figure 4. (a) Reduced scale model of a double-winged aircraft and (b) elastic stress concentration recorded on the test specimen (temperature scale is given in degrees Celsius).

in order to evaluate the weakness zones around connections between the two wings themselves and the two wings with the fuselage (Figure 4b). The double-wing design is assumed to lead to greater maneuverability. The big disadvantage of the biplane layout was that the two wings interfered with each other aerodynamically, each reducing the lift produced by the other. This interference meant that for a given wing area the biplane produced more drag and less lift than a monoplane. This mechanical statement is supported by the hot locations on the wings caused by stress concentrations.

6 OCCURRENCE AND DETECTION OF DAMAGE

The last term defines the energy dissipation generated by plasticity and/or viscosity [32]. Internal energy dissipation has been recognized by many scientists [33]. The work done to the system by plastic deformation is identified as the major contribution to the heat effect. In thermoelastic–plasticity, there exists a general acceptance that not all mechanical work produced by the plastic deformation can be converted to the thermal energy in the solid. A significant portion of the work is believed to have been spent in the change of material microscopic structure. The work done in plastic deformation per unit volume can be evaluated by integrating the material stress–strain curve. This internal dissipation term constitutes an important part of the nonlinear coupled thermomechanical analysis.

The quantification of this intrinsic dissipation for engineering materials is an extremely difficult task without infrared thermography [34].

The infrared thermographic technique is mainly concerned with temperature differences (or thermal gradients) that exist in the material rather than with the absolute values of temperature. It conveniently detects the dissipation regime of the material under loading.

Ignoring the significance of the coupled thermomechanical equation, an unsuccessful attempt [35] has been made to monitor thermoemissions during fatigue crack propagation tests. The technique used was unable to quantitatively detect the changes in thermoemission caused by small and slow crack-tip advances.

Experimental evidence shows that only part of the input plastic deformation power is expended to change the material's microstructure; the other part is dissipated in the form of heat (Table 1).

In materials testing in an industrial environment, thermal noise often generated by gripping systems may sometimes obscure the intrinsic dissipation of the tested specimen. This difficulty can be overcome when using thermal image subtraction or *differential thermography* as shown, for instance, in the case of an XC55 steel specimen subjected to rotating–bending fatigue testing. This procedure of thermal image processing provides the fatigue limit of steel materials within a few hours instead of the several months required when using the standard staircase method [40].

Table 1. Thermomechanical coupling models reported in literature

Postulates and internal variable IV	Internal energy dissipation rate
Dillon [4] IV = plastic strain rate	Plastic power
Lee [36] IV = plastic power	90–100% of plastic power
Nied and Battermann [37] IV = a function of dislocation energy per unit length	A part of plastic power
Raniecki and Sawczuk [38] IV = work hardening	A function of IV
Mroz and Raniecki [39] IV = work hardening	A function of IV
Lehmann [3] IV = a part of plastic power	A part of plastic power

6.1 Mechanical evaluation of wood construction materials

Damage and failure behavior of wood in tensile, compressive, or shear loading is an important consideration in connection with designs or regulations of wooden structures subjected to high allowable working stresses.

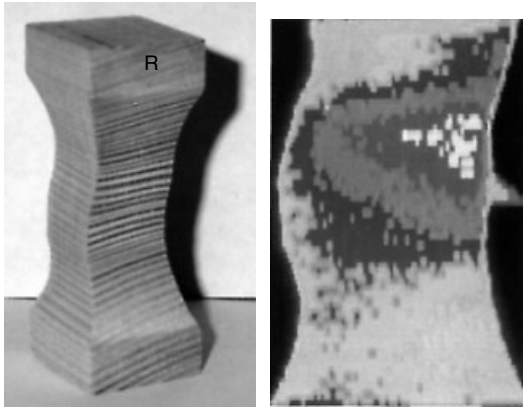


Figure 5. (a) Pine wood specimen sampled in the radial direction and (b) localization of intrinsic dissipation in wood specimen sampled in the radial direction (each color hue corresponds to 0.2 °C) evidencing the failure mechanism.

Several series of monotonic unconfined compression tests have been conducted on square specimens of pinewood, prepared along its anisotropy directions (longitudinal L, radial R, and transverse T). The wood specimens were especially designed (cross section $S_0 = 256\,\mathrm{mm}^2$, $h_0 = 20\,\mathrm{mm}$) with enlarged ends to prevent sliding, bending, or premature buckling, caused by heterogeneity, bad alignment of compression loading, or other significant end effects (Figure 5a). The thermography testing allowed observations of the physical process of failure (Figure 5b).

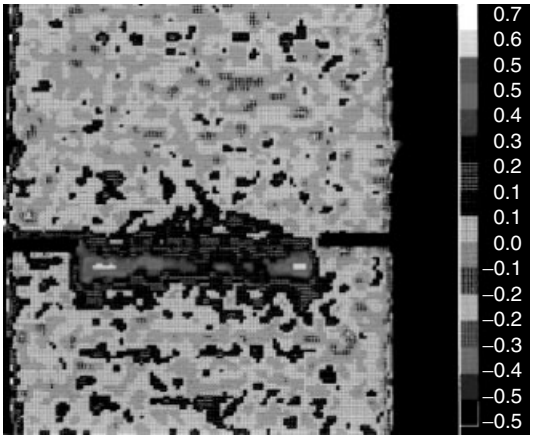
In several architectural applications, there is a need for development of high-ductility connections for braced frame systems and energy dissipating connections for antiseismic applications. Metal-plate-connected wood trusses are widely used in residential constructions (Figure 6a) and are increasingly used in agricultural and other commercial constructions. A reason for their widespread use and continued growth of applications for wood trusses concerns the efficiency and effectiveness of punched-metal-plate connectors. Extensive engineering design services support an almost unlimited variety of components that can be assembled with plates and dimension lumber (Figure 6b).

Failure modes including the teeth pulling out of the wood, failure of the wood member within the plated region, yielding of the plate loading, and buckling of the plate in gaps between wood members are depicted by thermal images (Figure 6c) before failure.

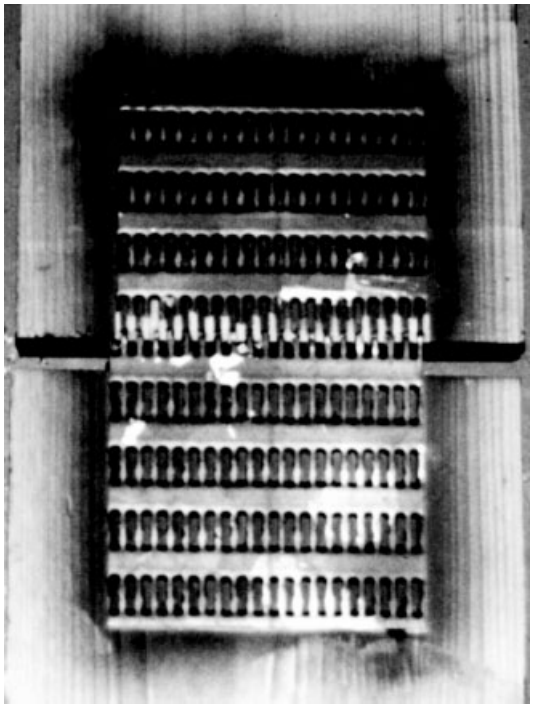
6.2 Application to plain concrete specimens

Concrete materials present a low thermomechanical conversion under monotonic loading. Plastic

(a)



(c)



(b)

Figure 6. (a) Metal-plate-connected wood trusses; (b) splice joint under tension; and (c) intrinsic dissipation on metal-plate connection.

deformation (whereby microcracking and slips occur creating permanent changes globally or locally) is one of the most efficient heat production mechanisms. Most of the energy required to cause such plastic deformations is dissipated as heat. Such heat generation is more easily observed when it is produced in a fixed location by reversed applied loads. These considerations define the use of vibrothermography as a nondestructive and noncontact technique for observing the damage process of concrete materials [41].

In the laboratory, the high-frequency servo-hydraulic test machine provides a means of vibration and dynamic testing of engineering materials. A vibratory loading at 100 Hz, applied to the specimen subjected to a given static compression, exhibits, in a nondestructive manner, the irreversible plastic strain concentrations around gaps or cracks. The contribution of the plasticity term is revealed by the rapid evolution of heat dissipation once the stable reversible stress domain is exceeded, demonstrating the occurrence of an unstable crack propagation or coalescence of flaws existing in the concrete specimen. Experimental results have already shown the following:

1. Under a vibratory excitation, between 25 and 50% of the nominal uniaxial compression $\sigma_N = F/S_0$, the heat dissipation, detected for 2000 load cycles, is small, even at the hottest location.
2. When $0.50 \leq \sigma/\sigma_N \leq 0.75$, stress concentrations around cracks or defects are readily detected at the 1000th load cycle.
3. For $0.63 \leq \sigma/\sigma_N \leq 0.88$, cracking occurs increasingly in the reduced section part of the specimen.

Infrared thermography readily depicts intrinsic dissipation localization, announcing quite different mechanisms of damage preceding concrete failure (Figure 7a). The different phases of heat dissipation, operating during an unstable failure, are readily described by heat patterns. When defects or weakness zones are present on the specimen, infrared observations evidence the progressive mechanism of defect coalescence (Figure 7b). The rate of heat generation at the hottest location is used to detect the threshold of the failure process if compared with the traditional stress–strain curve. These results have been readily extended to rock materials [42].

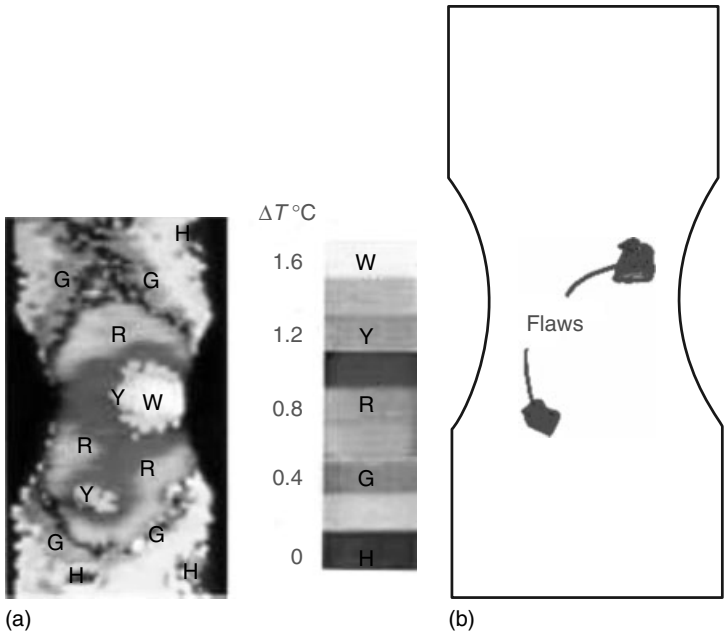


Figure 7. Plain concrete specimen subject to a nominal uniaxial vibratory compression σ_N between 63 and 88% of peak resistance. (a) Thermal image of the specimen and (b) examination of the specimen after test.

6.3 Rapid evaluation of the endurance limit of plain concrete

In accordance with the coupled thermomechanical equation, the analysis of thermal images consists of isolating the intrinsic dissipation from thermal noises by simply subtracting the thermal image at a reference time from the thermal image at 1000 load cycles. Computer-assisted thermography software allows for data reduction of the thermal images using the function subtraction of images. The resulting subtracted image shows the temperature change between two compared images, obtained under nearly identical test conditions. This image processing provides quantitative values of intrinsic dissipation.

This procedure is applied for each load step. The manifestation of the fatigue damage mechanism is revealed by a break in the intrinsic dissipation regime. The starting load level must be chosen below

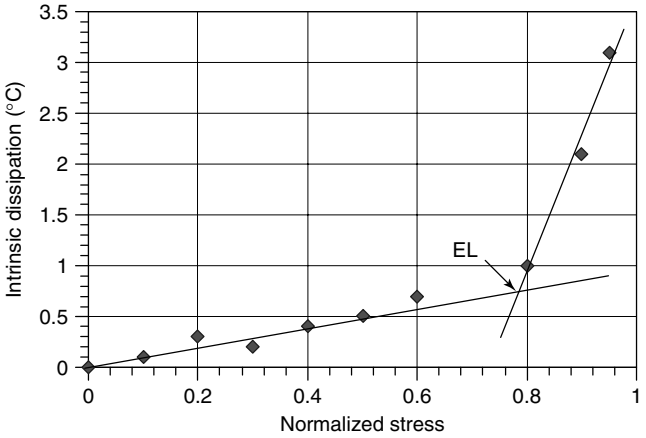
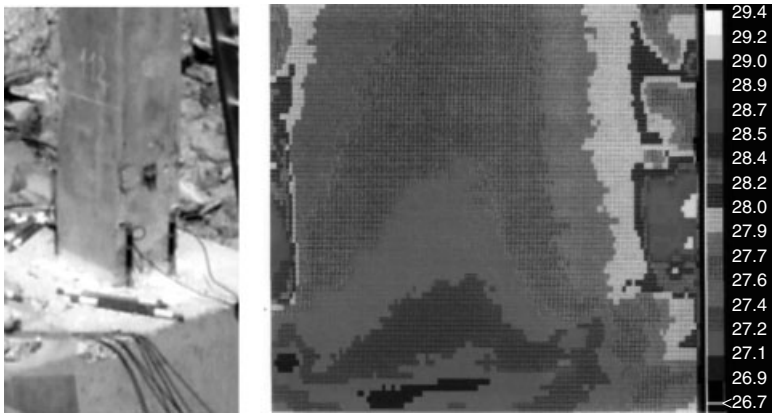
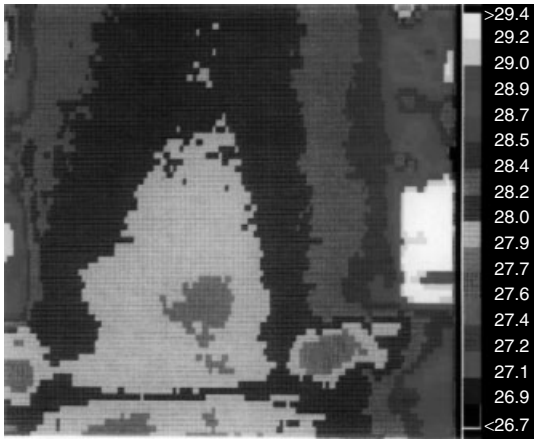


Figure 8. The endurance limit of the concrete specimen under compressive loading is graphically determined by an abrupt change of the dissipative regime.

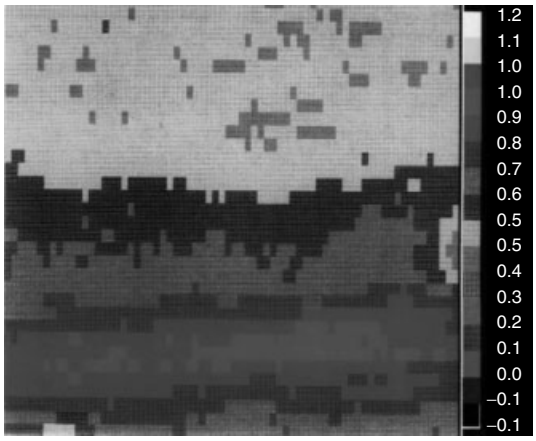
Figure 9. Experimental reinforced concrete structure under seismic-type loading.



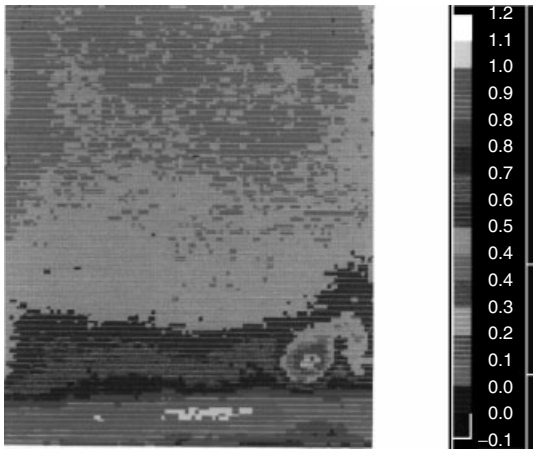
(a)



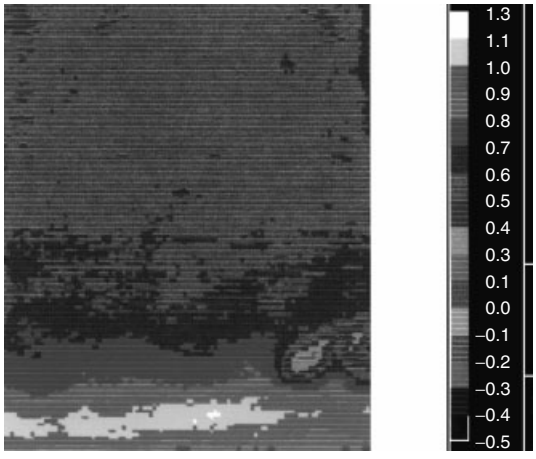
(b)



(c)



(d)



(e)

Figure 10. (a) Thermal image of a column base of the reinforced concrete structure before seismic-type loading; (b) thermal image of a column base of the reinforced concrete structure after N load cycles; (c) intrinsic dissipation located at the column base obtained by thermal images subtraction; and (d) intrinsic dissipation due to $N = 270$ load cycles; (e) intrinsic dissipation due to $N = 360$ load cycles.

the fatigue limit. It depends significantly upon the concrete characteristics. For example, assume that a test is begun at a stress level of about 20% of failure nominal stress, then 30%, 40%, etc. This testing is continued until the temperature rise attains several degrees Celsius. For each load step, an averaging treatment (among 4, 8, 16, or 32 thermograms) provides more stable thermal images.

Experimental results are summarized in Figure 8, which illustrates how the endurance limit (EL) is determined using a graphical procedure. The threshold of critical thermal dissipation is roughly the same for different chosen number of load cycles. The threshold corresponds approximately to the value deduced from standard procedures. These experiments have shown that the infrared thermographic technique can provide the EL of concrete within a few hours instead of several months needed when using the traditional standard staircase method. These results are consistent with those obtained on concrete prisms subjected to compressive fatigue testing [43].

6.4 Infrared thermographic scanning of a full-scale earthquake resistant concrete structure

The damaged areas are located and highlighted by heat patterns. These results support and validate the assumptions to be taken into consideration in numerical procedures for stability and integrity assessment of concrete structures. The phenomenological behavior in consideration is therefore the standard of reference, allowing the use of the methods and results of continuum mechanics for analyzing and modeling their engineering performance. Information about the location and significance of structural defects as a basis for maintenance decisions, including the extreme case of removal from service, can be obtained through inspection and nondestructive evaluation.

The proposed infrared thermographic procedure involves careful examination of those areas where defects are most likely to occur. Analyzing the structure and the service histories of similar structures in similar environments can identify the critical areas. The application of infrared scanning to inspection of concrete structure relies on the fact that the energy is dissipated during the process of accumulative damage when internal cracks or flaws develop. The most likely severe earthquake can be withstood if the members are sufficiently ductile to absorb and dissipate seismic energy by inelastic deformations with little decrease in strength.

Under seismic-like loading, simulated by a rotating mass exciter placed on the top of the building, plastic hinges form progressively at the column bases where heat dissipation mainly occurs in the overstressed steel reinforcements. The heat, flowing out through the crushed concrete, can be observed using infrared thermography as a function of the number of load cycles (Figure 9).

Computer assisted thermography software allows the data reduction of thermal images using the function subtraction of images and shows the progressive evolution of heat dissipation at a column base before crack lines become visible. The resulting image is a subtracted image showing the temperature change between the reference time and after N

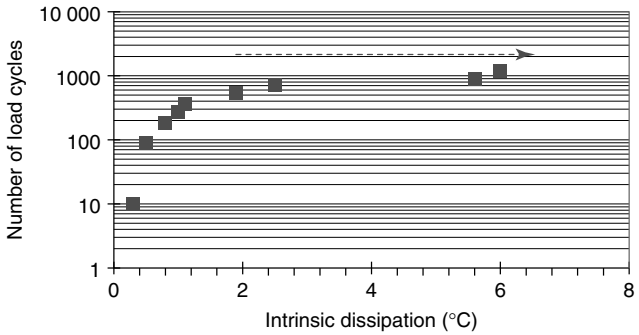


Figure 11. Endurance lifetime evaluation of a column base of the steel reinforced concrete structure subjected to seismic-type loading.

load cycles (Figure 10a–c). This image processing provides quantitative values of intrinsic dissipation corresponding to different numbers N of load cycles.

The results obtained after 270 load cycles (Figure 10d) and after 360 load cycles (Figure 10e) lead to a ($\log N$ – intrinsic dissipation) plot (Figure 11) that allows evaluation of fatigue lifetime of a column base of the experimental reinforced concrete structure subjected to seismic-type loading.

The above data processing (heat image subtraction) applied for the overall structure under solicitations readily provides useful information about localization of dissipation or damageable zones.

6.5 Infrared observations of energy dissipation on concrete structures subject to shaking table loading

Load-bearing walls in reinforced concrete structures are commonly used worldwide. The research described hereafter was done within the framework of the European Consortium of Earthquake Shaking Tables and the Innovative Seismic Design New and Existing Structures, Topic 5 Shear wall structures research project, which is supported by the European Commission. Two large-scale 360 MN specimens (SPECIMEN A and SPECIMEN B) representing one-third scale five-story buildings were tested under dynamic seismic-like loading on the large shaking table Azalée at the Commissariat à l'Energie Atomique Saclay (French Atomic Energy Agency, Saclay, France). The loading input signal is an artificial accelerogram (far-field earthquake) characterized by its peak ground acceleration (PGA) values.

The mock-up tests performed aimed at demonstrating the major influence of boundary conditions at the base of the model, and the feasibility of optimizing low ratio and adequate distribution of reinforcements to obtain multicracking zones (multifuse concept) in opposition with the traditional pseudo-plastic hinge localized at the base of a steel reinforced wall (monofuse concept).

6.5.1 SPECIMEN A highlighting the effects of reinforcement ratios

SPECIMEN A, composed of two lightly reinforced walls anchored to the shaking table, is designed as recommended in Eurocode 8 allowing a plastic hinge at the base. Special attention has been paid to the influence of different reinforcement ratios and boundary conditions. During the tests performed, SPECIMEN A mock-up suffered high damage levels. Its behavior was mostly conditioned by its flexural bending. Examination after tests evidenced failure of steel reinforcements (Figure 12).

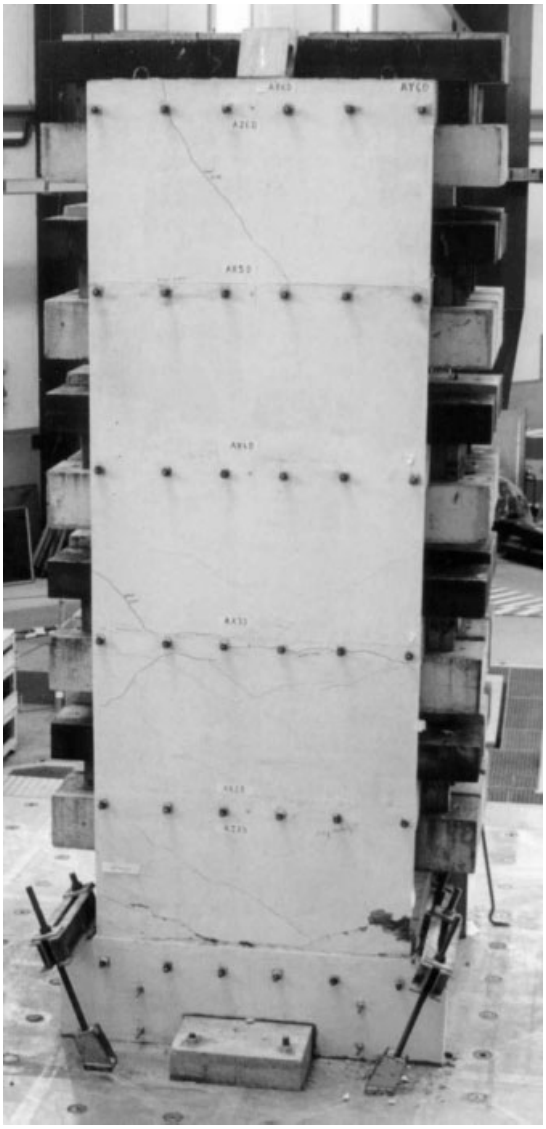
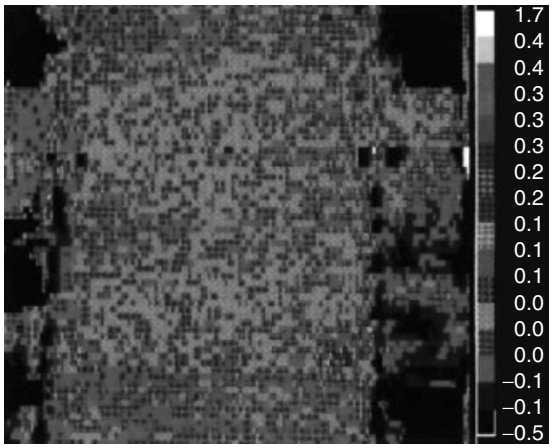
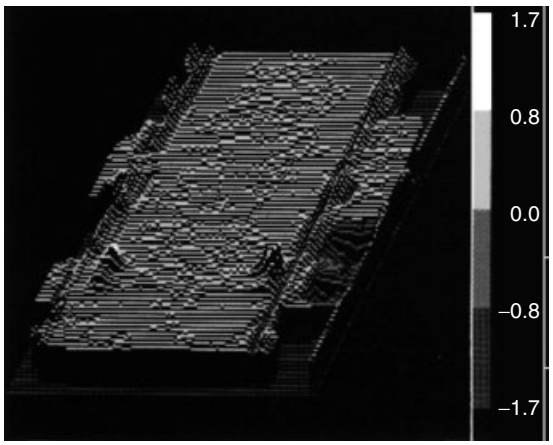


Figure 12. Experimental reduced scale structure in steel reinforced concrete SPECIMEN A tested on the shaking table of CEA Saclay.



(a)



(b)

Figure 13. (a) Intrinsic dissipation caused by plasticity of the steel reinforcements after testing at peak ground acceleration $PGA = 8.0\,\mathrm{m\,s^{-2}}$ (temperature changes ΔT are given in degrees Celsius) and (b) visualization of dissipation mechanism due to plasticity of steel reinforcements.

In this case, the dissipation mechanism caused by plasticity of steel reinforcements can be considered as an internal parameter so that equation (1) has to be completed with two supplementary terms representing the cross-coupling effects, where the former is caused by the dependence of the stress tensor on temperature (reversible) while the latter is induced by the same dependence of the generalized force conjugated to the internal state vector (irreversible). This phenomenon appeared on the concrete surface with a delay depending on the depth of reinforcements. Thus, infrared thermography readily evidenced and localized, on the scanned wall surface, the plasticity of steel reinforcements with a delay due to heat conduction characteristics of concrete (Figure 13a and b).

6.5.2 SPECIMEN B highlighting the effects of boundary conditions

SPECIMEN B, composed of two lightly reinforced walls, was designed according to French seismic code PS92 recommendations and simply rested on a 40-cm-thick sand layer (Figure 14). This test aimed to reproduce the phenomena of uplift and the fact that such a nonlinear phenomenon was capable of isolating the structure from ground-borne excitation. In this case, it is expected that soft boundary conditions will determine the seismic behavior of structural walls.

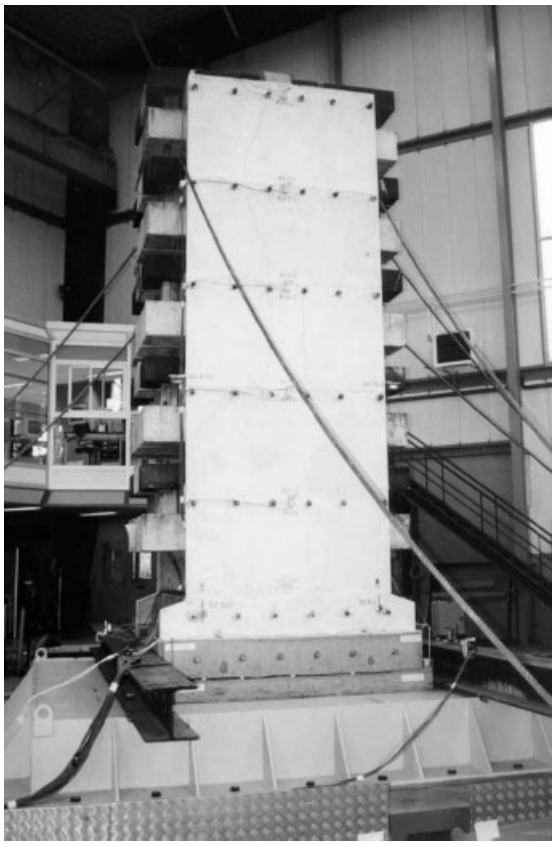


Figure 14. SPECIMEN B resting on a fine sand layer.

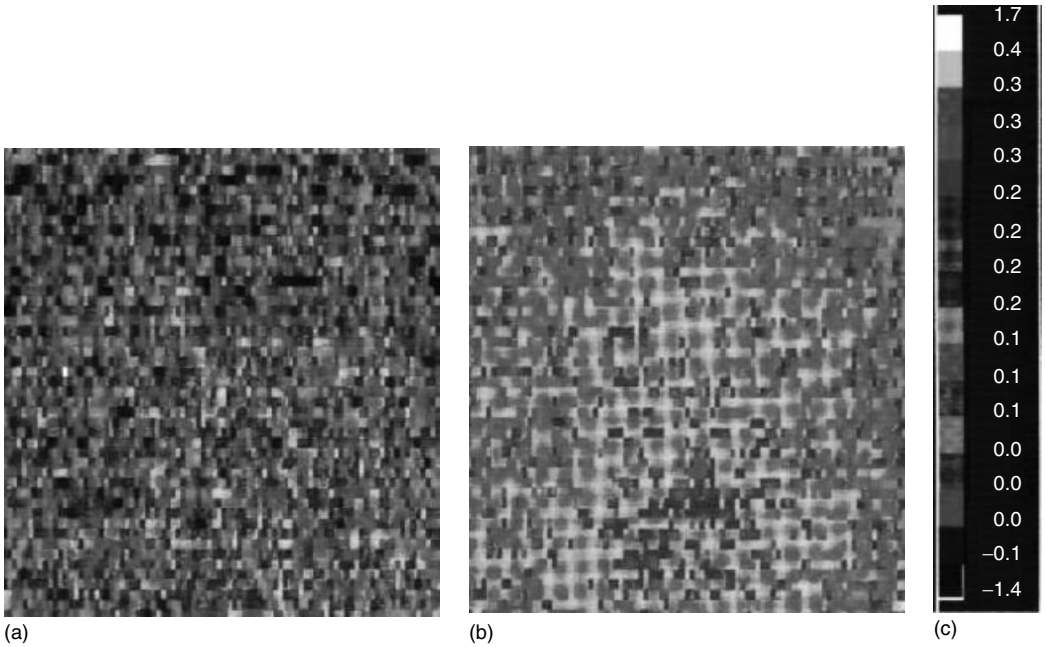


Figure 15. (a) Dissipation due to slippage of steel reinforcements recorded 6 min after seismic test at $PGA = 11\,\mathrm{m\,s^{-2}}$ during 30 s; (b) dissipation due to slippage of steel reinforcements recorded 10 min after seismic test at $PGA = 11\,\mathrm{m\,s^{-2}}$ during 30 s; and (c) heat changes are given in degree Celsius.

Table 2. Magnitude order of temperature change

Dissipative mechanisms	Range of temperature changes	Time delay
Plasticity of concrete under compression	Up to 10 °C	In real time
Plasticity of steel reinforcements	Some degrees Celsius	Several minutes of delay
Slippage between steel reinforcements and concrete matrix	Tenths of degrees Celsius	A delay of tens of minutes

As in the above case, infrared thermography evidenced friction between steel reinforcements and concrete matrix with a delay necessitated by heat conduction through the concrete layer (Figure 15a and b).

6.5.3 Dissipative mechanisms and their range of temperature changes

Experimental results showed that the discrimination of the involved dissipative mechanisms is very delicate. Fortunately, this work, originally intended to validate diverse different dissipative mechanisms, provided the interesting discriminative characteristics of temperature changes as in Table 2.

7 CONCLUDING REMARKS

This work has demonstrated that (i) the *evolution* of heat change facilitates the detection of thermobiological coupling and (ii) the *dissipativity* of engineering materials or structures under solicitations is a highly sensitive and accurate *manifestation of damage*.

Thanks to the thermomechanical coupling, infrared thermography provides a nondestructive, noncontact, and real-time test to observe the physical process of

concrete degradation and to detect the occurrence of its intrinsic dissipation. Thus, it readily provides a measure of the material damage and aids in defining a limit of acceptable damage and endurance limit of concrete under load beyond which the material is susceptible to failure. It should be pointed out that the inelastic strain due to compressive loading provides information only on the current geometry while the internal state variables provide information on the internal state and on the microstructural defects.

The method allows not only qualitative work such as finding flaws, defects, or weakness zones but also quantitative analysis of the effects of flaws and defects on strength and durability of concrete structural components. This useful and promising technique offers an accurate illustration of crack initiation, and readily detects the onset of its unstable propagation through the material and/or flaw coalescence when cyclic loading generates increasing irreversible microcracking. Mechanical test data generated under noncyclic conditions are insufficient to provide a comprehensive insight into the damage development in brittle concrete under cyclic loading. Design procedures ignoring fatigue phenomena may be seriously flawed if the concrete structures concerned are loaded cyclically.

The examples reported in this article and several published results (*see* **Thermal Imaging Methods**) demonstrate the versatility of infrared thermographic technique in various domains of application, provided that the physical phenomenon is correctly interpreted in a consistent theoretical framework.

The main interest of this energy approach is to unify microscopic and macroscopic test data. The parameter *intrinsic dissipation* under consideration is a scalar quantity, which is easy to evaluate accurately. Subsequently, it may suggest multiaxial design criteria, highly relevant for full-scale testing of engineering structures.

REFERENCES

[1] Bui HD, Ehrlacher A, Nguyen QS. Thermomechanical coupling in fracture mechanics. In *Thermomechanical Coupling in Solids*, Bui HD, Nguyen QS (eds). Elsevier Science IUTAM, 1987, pp. 327–341.

[2] Farren WS, Taylor GI. The heat developed during plastic extension of metals. *Proceedings of the Royal Society* 1925 **A/107**:422–428.

[3] Lehmann TH. Coupling phenomena in thermoplasticity. *Structural Mechanics in Reactor Technology (SMiRT5)*. The International Association for Structural Mechanics in Reactor Technology: Berlin, 1979, Paper L1/1.

[4] Dillon Jr OW. Coupled thermoplasticity. *Journal of the Mechanics and Physics of Solids* 1963 **11**:21–33.

[5] Bui HD, Ehrlacher A, Nguyen QS. Etude expérimentale de la dissipation dans la propagation de la fissure par thermographie infrarouge. *Comptes Rendus Academie Sciences* 1981 **293/II**:1015–1017.

[6] Luong MP. Infrared thermographic scanning of fatigue in metals. *Nuclear Engineering and Design* 1995 **158**:363–376.

[7] Luong MP. Infrared scanning of failure processes in wood. In *Selected SPIE Papers*, *CD-ROM Series in PDF Thermal Sensing and Imaging*, Snell JR, Burleigh DD (eds). The International Society for Optical Engineering, April 1999, p. 7.

[8] Luong MP. Infrared thermographic characterization of engineering materials. *SPIE Proceedings Series Infrared Technology XVI*. SPIE, 1990; Vol. 1314, pp. 275–284.

[9] Luong MP. Infrared observations of the mechanical performance of tennis strings. *Thermosense XIII Proceedings of SPIE*. The International Society for Optical Engineering: Orlando, FL, 2001; Vol. 4360, pp. 624–635.

[10] Duszek MK, Perzyna P. The localization of plastic deformation in thermoplastic solids. *International Journal of Solids and Structures* 1991 **27**(11): 1419–1443.

[11] Haupt P. On the thermomechanical modelling of inelastic material behaviour. In *IUTAM Symposium on Micro- and Macrostructural Aspects of Thermoplasticity*, Bruhns OT, Stein E (eds). Kluwer Academic Publishers, 1999, pp. 3–14.

[12] Kratochvil J, Dillon Jr OW. Thermodynamics of elastic-plastic materials as a theory with internal state variables. *Journal of Applied Physics* 1969 **40**:317–325.

[13] Mandel J. Variables cachées, puissance dissipée, dissipativité normale. *Sciences et Techniques de l'Armement* 1980:37–49, Special Issue by Numéro spécial, Janvier 1980, Thermodynamique des comportements rhéologiques.

[14] Luong MP, Martin A. Détection de microfuites par thermographie infrarouge. *Actes ASTELAB 90 Colloque International Essais Industriels Coopération Européenne*. Paris, June 1990.

[15] Weil GJ. Techniques of infrared thermographic leak testing. *Infrared and Thermal Testing. ASNT NDT Handbook*. American Society for Nondestructive Testing: Columbus, OH, 2001; Vol. 3, Chapter 18, Part 1, pp. 602–608.

[16] Beaudoin JL, Bissieux C, Offerman S. Thermoelastic stress analysis. *Infrared and Thermal Testing, ASNT NDT Handbook*. American Society for Nondestructive Testing: Columbus, OH, 2001; Vol. 3, Chapter 11, Part 7, pp. 339–341.

[17] Luong MP. Characteristic threshold and infrared vibrothermography of sand. *Geotechnical Testing Journal ASTM* 1986 **9**(2):80–86.

[18] Luong MP. Thermomechanical couplings in solids. *Infrared and Thermal Testing ASNT NDT Handbook*. American Society for Nondestructive Testing: Columbus, OH 2001; Vol. 3, Chapter 11, Part 8, pp. 342–347.

[19] Luong MP. Introducing infrared thermography in soil dynamics. *Infrared Physics and Technology* 2007 **49**:306–311.

[20] Cielo P, Maldague X, Déom AA, Lewak R. Thermographic nondestructive evaluation of industrial materials and structures. *Materials Evaluation* 1987 **45**:452–460.

[21] Reifsnider KL, Henneke EG, Stinchcomb WW. *The Mechanics of Vibrothermography. Mechanics of Nondestructive Testing*. Plenum Press: New York, 1980, pp. 249–276.

[22] Luong MS, Luong MP, Lok C, Carmi E, Chaby G, Visieux V. Bioavailability of topical corticosteroids evaluated by differential thermography. *Annales de Dermatologie et de Venereologie* 2000 **127**:701–705.

[23] Di Carlo A. Thermography and the possibilities for its applications in clinical and experimental dermatology. *Clinical Dermatology* 1995 **12**:329–336.

[24] Aiache JM, Lafaye C, Bouzat J, Rabier R. Measurements of corticosteroids topical availability by thermography. *Journal de Pharmacie de Belgique* 1980 **35**:187–195.

[25] Henry F, Fumal I, Pierard GE. Postural skin colour changes during the corticosteroid blanching assay. *Skin Pharmacology and Applied Skin Physiology* 1999 **12**:199–210.

[26] Queille-Roussel C. Le test de vasoconstriction en peau saine—techniques et applications. *Annales de Dermatologie et de Venereologie* 1988 **115**:491–503.

[27] Shah VP, Peck CC, Skelly JP. Vasoconstriction skin blanching-assay for glucocorticoids, a critique. *Archives of Dermatology* 1989 **125**:1558–1561.

[28] McLaughlin Jr PW, Mirchandani MG, Ciekurs PV. Infrared thermographic flaw detection in composite laminates. *Journal of Engineering Materials and Technology* 1987 **109**:146–150.

[29] Balageas DL, Déom AA, Boscher DM. Characterization and nondestructive testing of carbon-epoxy composites by pulsed photothermal method. *Materials Evaluation* 1987 **45**:461–465.

[30] Oliver DE. Stress pattern analysis by thermal emission (SPATE). *Dynamic Stress Analyser*. Ometron Limited: UK, 1986; Vol. 14, pp. 1–28.

[31] Salençon J. *Handbook of Continuum Mechanics, General Concepts, Thermoelasticity*, Ecole Polytechnique ISBN 3-540-41443-6. Springer-Verlag: 2000.

[32] Mandel J. Energie élastique et travail dissipé dans les modèles. *Cahiers du Groupe Français de Rhéologie* I/1. SEDOCAR Paris, September 1965; pp. 9–14.

[33] Bui HD. Dissipation d'énergie dans une déformation plastique. *Cahiers du Groupe Français de Rhéologie*. I/1. SEDOCAR Paris, September 1965; pp. 15–19.

[34] Chrysochoos A, Dupré JC. An infrared set-up for continuum thermomechanics. *Quantitative Infrared Thermography QIRT 92*, *Eurotherm*. Editions Europ-Éennes Thermique et Industrie: Paris, 1992; Vol. 27, pp. 129–134.

[35] Leaity GP, Smith RA. The use of SPATE to measure residual stresses and fatigue crack growth. *Fatigue and Fracture of Engineering Materials and Structures* 1989 **12**(4):271–282.

[36] Lee EH. Elastic plastic deformations at finite strains. *Journal of Applied Mechanics* 1969 **36**:1–6.

[37] Nied HA, Battermann SC. On the thermal feedback reduction of latent energy in the heat conduction equation. *Materials Science and Engineering* 1972 **9**:243–245.

[38] Raniecki B, Sawczuk A. Thermal effects in plasticity. *Zeitschrift für Angewandte Mathematik und Mechanik* 1975 **55**:333–341, 363–373.

[39] Mroz Z, Raniecki B. On the uniqueness problem in coupled thermoplasticity. *International Journal of Engineering Science* 1976 **14**:211–221.

[40] Luong MP. Nondestructive evaluation of fatigue limit of metals using infrared thermography. *Material*

Research Society Symposium Proceedings 1998 **503**:275–280.

[41] Luong MP. Infrared thermovision of damage processes in concrete and rock. *Engineering Fracture Mechanics* 1990 **35**(1–3):127–135.

[42] Luong MP. Infrared thermographic observations of rock failure. *Comprehensive Rock Engineering Principles, Practice and Projects*. Pergamon Press, Elsevier Science: Oxford, 1993; Vol. 4, Chapter 26, pp. 715–730.

[43] Sparks PR, Menzies JB. The effect of rate of loading upon the static and fatigue strengths of plain concrete in compression. *Magazine of Concrete Research* 1973 **25**(83):73–80.

Chapter 7

Civil Infrastructure Load Models for Structural Health Monitoring

Udo Peil

Institute for Steel Structures, University of Technology Carolo-Wilhelmina at Braunschweig, Braunschweig, Germany

1 INTRODUCTION

The present and the future behavior of structures in the civil, mechanical, or aerospace engineering field can often be precisely assessed by means of certain monitoring measures. Structural health monitoring (SHM) is a general method that can be used for the assessment of local or global damage, the evolution of damage, and the prediction of the lifetime of a structure under certain loading conditions. Conventional lifetime assessment methods are not very reliable. Differences between theoretically calculated and observed lifetimes may differ by one order of magnitude [1]. Lifetime assessment by means of parallel SHM can considerably improve the accuracy of the prediction [2, 3]. SHM enables the determination of a wide range of monitoring measures including

- deflections and strains, e.g., of newly built structures due to special test loads to compare measured and theoretical results;
- system behavior for system identification;
- threshold crossings;
- input values for the assessment of the lifetime of structures;
- maintenance planning; and
- traffic analysis and control.

Any SHM measure is based on a precise knowledge of the system. The knowledge may be gained by conscientiously performed theoretical investigations, e.g., by means of the finite-element method (FEM), taking into account the actual system parameters and boundary conditions based on thorough inspections of the structure. The modeling must be done with high accuracy, since an oversimplified model of a structure may hide or fake weak points in the structure. Because it is difficult to determine certain system parameters precisely, e.g., boundary conditions (grade

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

of a fixed connection) and material parameters (stiffness, mass, or damping), the theoretical results must be verified.

Verification is accomplished by means of system identification. Defined loads are applied to the structure, and the response is monitored. The results can be used to update the physical model parameters. This type of identification is called *parametric identification*. Nonparametric identification is used if the system model is unknown. In this case, simple black-box models, e.g., fitted least-square functions, can be used to describe the system behavior. The system identification procedure is based on a comparison of the system response to a given action. The type of action needed depends on the response of interest. For example, a concrete structure only reacts to chemicals if there is a special reaction between the chemical substances and the surface of the structure. In any case, before establishing an SHM measure, the system must be identified by certain monitoring measures to begin a lifetime assessment with the best knowledge regarding the structure. For this purpose, precisely defined loads are needed. In Figure 1, the block titled Load model I reflects this knowledge.

To be able to predict the lifetime of the structure, a suitable load model, which may cause future damage, is required to describe the future actions on the structure because they may cause future damage. The models of codes are not usable, in general, because they describe ultimate load situations, which usually never occur in reality. Lifetime predictions must be formulated under typical actions, which are able to precisely describe reality in terms of time dependency and amplitudes and their statistics. These load models are included in the block titled Load model II in Figure 1.

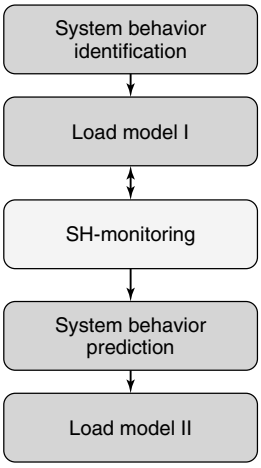


Figure 1. Use of load models.

The prediction of future system behavior and/or the lifetime of a structure only makes sense if the future actions are predictable, as with traffic on a bridge or wind or wave loads on a structure. Unpredictable events such as earthquakes, tsunamis, and blast loads are not suitable for monitoring and lifetime prediction. Wind effects due to vortex or self-excited vibrations on structures lead to the same lack of predictability. Since these types of loads are always dangerous events for a structure, which could cause very early fatigue damage, these loads must be prevented, e.g., by adding damping to the structure [4]. Therefore, these unpredictable events do not play a role in SHM and are not discussed here.

In this article, both types of loads mentioned in Figure 1 are briefly discussed:

- load model I: loads for system identification;
- load model II: realistic loads (traffic, turbulent wind) for the description of realistic future actions.

Load model I is usually needed for a single test when an SHM measure is being selected. It can be used again if, after a couple of years, changes in the structural parameter are being assessed, for instance, to check the predicted lifetime provided by Load model II. If any change in the system behavior is to be assessed, e.g., because of damage evolution, repeated or continuous measurements must be performed. With the aid of the updated measurement results, the system damage prediction models can be adapted to the modified situation (and, if necessary, verified using a procedure that applies load model I). These system damage prediction models are then called *adaptive or smart models* [2]. They can be identical to classical system or damage models, but the models are often much simpler because they only need to predict over a short period of time, e.g., up to the next measurement. Since the adaptive system or damage models can take into account any system change or change in expected actions, they are more reliable than theoretical prediction models, which predict the system behavior over a long time period starting at $t = 0\,\mathrm{s}$.

2 WEAK-POINT ASSESSMENT AND INHERENT DAMAGE

Irrespective of the type of structure to be monitored, the assessment of critical weak points is one of the most important steps in the procedure. Weak points or weak spots are areas of the structure that are prone to damages or where possible damages cause nontolerable consequences. Weak points of older structures, usually designed with very different safety levels, are normally well known and can be determined by existing structural calculations or by experience. They can simply be identified by means of theoretical modeling and parallel system identification measures.

New structures, however, show an equally distributed safety level over a large number of critical details. The critical details can then be detected by means of probabilistic methods, leading to a smaller number of measuring points. The procedure for reliability-oriented determination of weak points classifies critical weak points as points that contribute the largest part to the overall failure probability of the structure [2, 5, 6]. These points must be monitored. To determine the failure probability, the description of

- limit state functions and limit values;
- stochastic model;
- mechanical model; and
- analysis of sequences of events and fault trees

are prerequisites. For details see [6].

When the weak points are determined, the material state near the weak points must then be checked using nondestructive testing (NDT) techniques, such as ultrasonic, X-ray, magnetic tests, etc. NDT can find macrodamage in an order of magnitude of a 10th of a millimeter. Much smaller failures, e.g., microcracks due to fatigue loads on a steel structure, are difficult to measure, although this damage is inherent in any older, i.e., not virgin, structure. The assessment of inherent damage is a difficult part of the identification process. In general, two possibilities are available in general to determine the damage state: a theoretically, model-based and an experimentally based procedure (Figure 2). In this figure, three NDT-methods are shown, which seem to be able to give information about the inherent damage state. Modeling of former actions is based on load or action models of the past. This approach is also discussed in the next section.

3 LOAD MODELS FOR SYSTEM IDENTIFICATION

3.1 General remarks

If the overall system response to certain actions is to be determined, the system must be loaded so that a sufficient system response is generated. Therefore, the loads applied must have the same order of magnitude as the maximum expected loads. In civil engineering, these loads are prescribed by national load codes. The

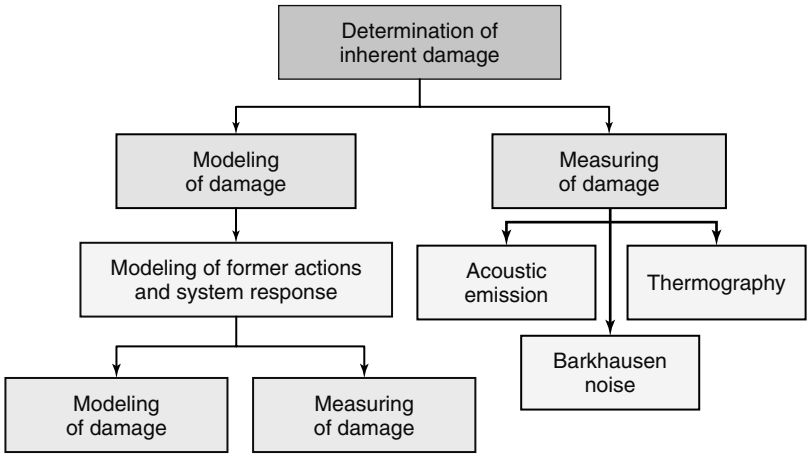


Figure 2. Determination of the inherent damage in a metallic structure.

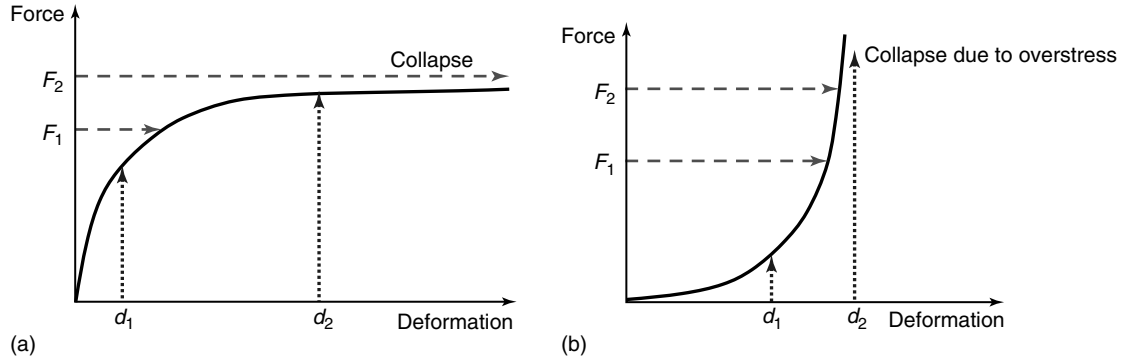


Figure 3. Effect of load and deformation control.

application of large loads may lead to a structural problem. If the structural behavior is completely unknown, e.g., an old bridge with unknown material behavior, the application of loads in the form of large weights (trucks or tanks if a bridge is tested) may pose a risk of global or local system collapse. Because of gravity, loads act continually, independent of any deformation due to dynamic loads. This type of loading of a structure is referred to as *force controlled*. If the gradient of the load deflection curve decreases (Figure 3a), which is usual in most cases, a complete collapse could easily occur if the load F_2 is larger than the ultimate load. For example, supposed a single-span girder is loaded by a rising water level in a tank. When the full plastic moment of the girder (or the lateral buckling load) is reached, the girder completely collapses because of the weight of the water still acting during the collapse process.

In such cases, so-called deformation-controlled loading is required. Deformations can be induced with hydraulic jacks or the like. The jack force causes a deflection of the structure. If the structure becomes weaker because of damage or nonlinearities, the force is automatically reduced and the given deflection is maintained (Figure 3a). This situation offers the possibility to measure the nonlinear behavior of a structure. If the deformations of the structure are increased by means of the hydraulic jacks, the required hydraulic force is automatically adjusted.

If the gradient of the load deflection curve increases (Figure 3b), which is typical for cable structures, which show a hardening effect during deflection, then a deformation-controlled procedure with a given deflection d_2 could possibly lead to damage due to material overstressing.

Despite the (rare) problems with hardening systems, a deformation-controlled test procedure is usually the best way to test a structure, but it is, in general, more expensive than a force-controlled procedure, especially if large loads must be applied. Large bridges, for instance, need expensive additional foundations for the tension bar elements and the hydraulic jacks. The use of trucks or tanks (force controlled) is much less expensive, in general.

3.2 Static loads

For investigation of the overall response of large structures such as the main girders of road bridges, many heavy trucks or tanks are often used. Railway bridges are loaded by engines (Figure 4). When statically undetermined structures are tested, the effect

Figure 4. Test loading of a railway bridge.

of the influence lines or areas could be used to increase the response of the structure by spanwise adapted loads. When substructures of the system, e.g., longitudinal or cross girders of bridges, are investigated, usually one or two trucks or engines may be sufficient.

Of course, the single loads must be known. Thus trucks and tanks must be weighted beforehand. The exact load due to engines is generally known.

In Germany, tests are carried out by building a special bridge loading truck, which is used for the large number of small bridges with a span <18 m, [7, 8]. This truck can strain the bridge with up to 100 t. Figure 5 explains the procedure. Figure 5(a) shows the special truck on the road. When arriving at the test site, the rear axles are braked and the traction engine is pulled out (Figure 5b). In this case, the bridge is loaded only by a small load. Figure 5(c) shows the final test position. The weight of the truck can be increased by 20 t of water to a total load of up to 100 t. If higher loads are needed, the truck must be anchored at both ends by tension bars to the supports.

One should keep in mind that such loading tests give good results for the purpose of system identification. An extrapolation of the system behavior under ultimate loads is difficult or impossible, owing to the effects of nonlinear system and, in particular, nonlinear material behavior.

Static tests of smaller structures are less expensive because the ultimate loads usually decrease. If single loads are used, the procedure is straightforward. The application of (approximately) continuous loads is more difficult. These loads can be realized, for instance, by a few hydraulic jacks, which are hydraulically coupled so that they do have the same

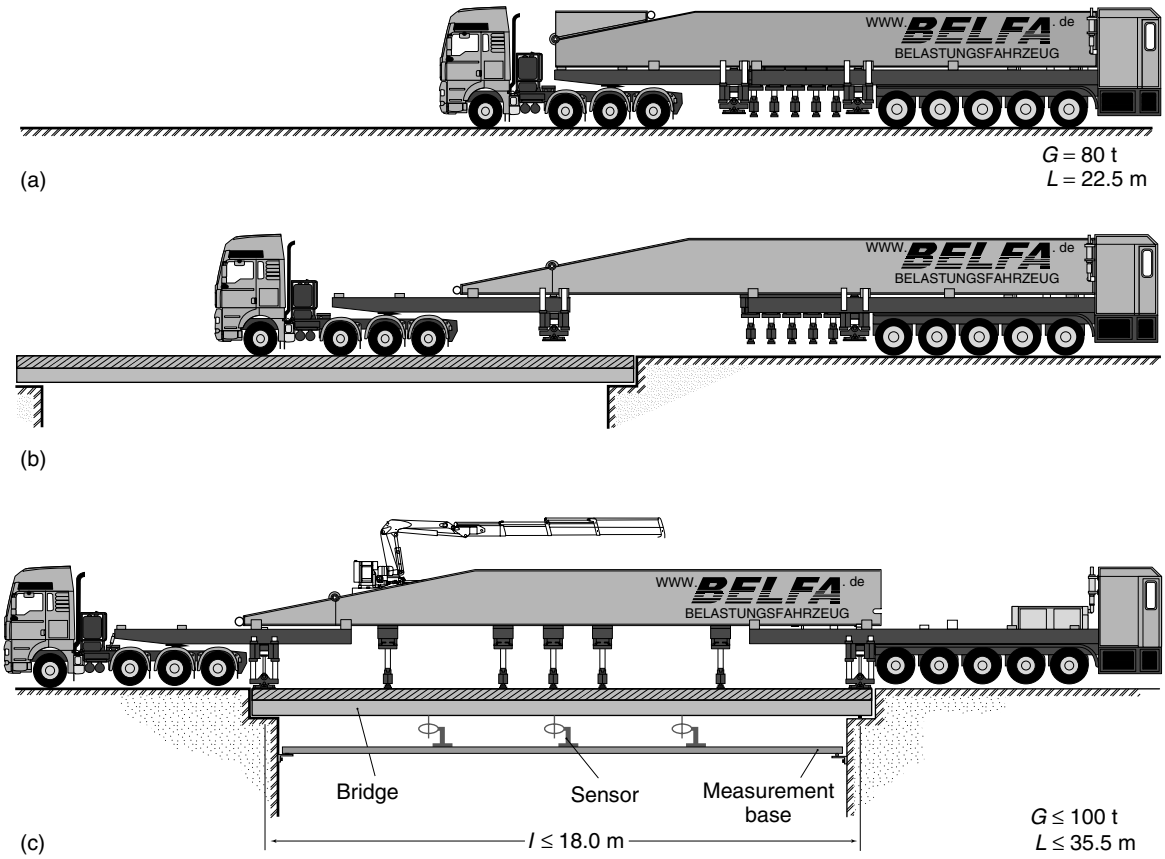


Figure 5. Special loading truck Belfa. [Reproduced with permission from Ref. 7. © Ernst und Sohn, 2001.]

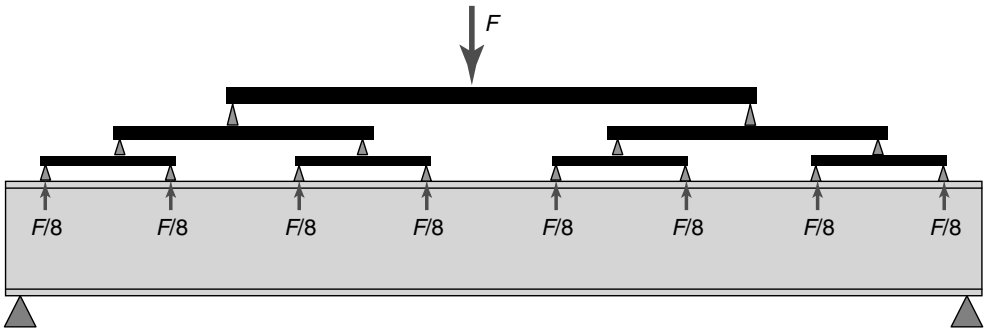


Figure 6. Loading with equal single loads.

oil pressure. However, if the jacks show different friction behavior, the activated loads are different. If a precise loading situation is needed, a load-distributing system can be used that leads to equal single loads as shown in Figure 6.

It is obvious that the loading system becomes more complicated and expensive as more single loads are needed, keeping in mind that a stabilization structure must be added to avoid lateral instability.

In such cases, it is much easier to use the hydrostatic (or the aerostatic) principle. If the load is applied by means of a pressure cushion, the loading system becomes very simple. Figure 7 shows a beam under uniform loading. The girder is turned upside down. It is loaded at its supports by means of a load-distributing girder and a hydraulic jack. The beam is pressed into a water-filled fire hose [9]. Thus, a perfect continuous constant load is produced. The aerostatic principle can be used in a similar way.

3.3 Dynamic loads

3.3.1 Large structures or global monitoring

Stationary excitation

The simplest way to produce a stationary excitation is the installation of an electrodynamic exciter. It is the simplest solution if tests could be performed in the laboratory, as is typical in the field of mechanical or aeronautical engineering, because those structures are small enough to be tested in the lab. The exciter system consists of a signal generator, a power amplifier and the exciter itself. In the lab, electromagnetic exciters with frequencies from 2 to 10 000 Hz, force amplitudes from 10 to 500 N, and dynamic deflections up to 2 cm are used. The exciters can be anchored at the ground or at large masses, which can carry the loads by inertia forces. This configuration must, of course, be conscientiously dynamically investigated before testing.

Figure 7. Constant continuous loading by means of the hydraulic principle.

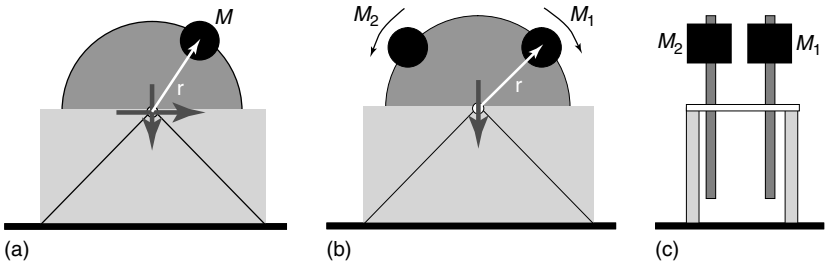


Figure 8. Excitation with unbalanced masses.

If the frequencies are much lower (<5 Hz), hydraulic jacks are used, which could produce much higher forces and deflections of the structure. This configuration is typically used for earthquake shaking tables.

The excitation could be stationary (harmonic excitation) or nonstationary (transient signals, swept sine, noise, etc.), depending on the target of the investigation.

Civil engineering structures can often not be tested in a lab. The use of a hydraulic exciter is often difficult because the exciter could not be anchored against the ground. In such cases, a dynamic exciter based on unbalanced masses is used to generate a stationary dynamic excitation. If just one mass is used, a rotating radial force occurs. The force has a horizontal and a vertical harmonic component, which shows a phase shift of $\pi/4$ (Figure 8a). It is often inconvenient to work with two force components. If two unbalanced forces, rotating in opposite directions are used, only one resulting force component occurs and the second one is internally balanced as shown in Figure 8(b).

The forces are proportional to the square of the circular frequency of rotation Ω (equation 1):

$$F = M \cdot r \cdot \Omega^2 \tag{1}$$

Stationary excitations have advantages because the energy is concentrated on one single frequency, thus the ratio between the signal and the noise is large. Since the transfer function can be measured directly, the nonlinear behavior of the system or frequency-dependent damping can be measured directly.

The effort can be substantial, in particular, if low frequencies must be excited. Since the circular frequency Ω is low, large masses are needed to generate sufficient force components. The rotational speed control must be precise enough to localize the very narrow resonance peaks that are exhibited by lightly damped systems. Sinusoidal excitation tests are time consuming as well. After changing the frequency, a new stationary state must be adapted before the measurements can be performed. Owing to feedback effects, it is often difficult to leave the resonance frequency because the rotational frequency persists at the resonance, if the power of the engine is not large enough.

One can reduce some of the above-mentioned disadvantages if a stationary excitation is used that is not purely sinusoidal but instead is swept sinusoidal with a frequency that changes slowly with time (Figure 9). This procedure also requires a precisely controlled unbalanced-mass engine and large masses if low frequencies are to be applied. The transfer function can then no longer be measured directly.

Transient excitation

Because of the above-mentioned disadvantages of the method of unbalanced masses, transient (from

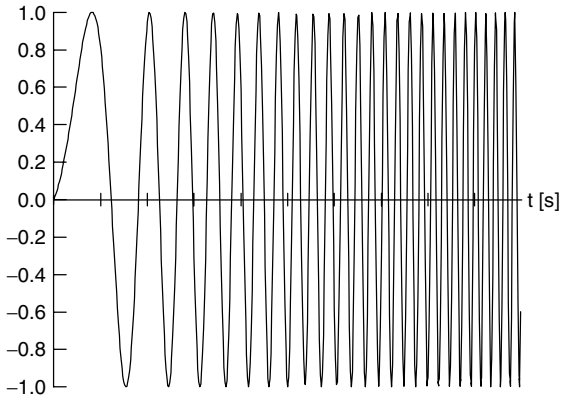


Figure 9. Swept sine.

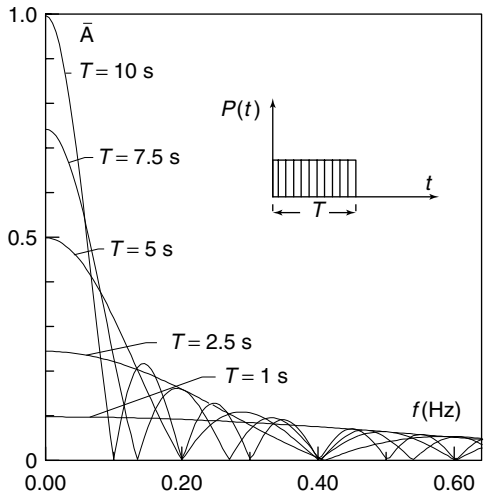


Figure 10. Normalized frequency spectra depending on the duration of a load pulse.

latin: *trans ire* — to pass) excitations are used in most cases. Depending on the size of the structure and of its eigenfrequencies, different methods may be used, bearing in mind, however, that short time signals cause a broadband frequency response and vice versa. A short load-time process, therefore, causes a broad frequency spectrum with higher frequencies, and a long lasting load-time process causes a narrow frequency spectrum with many low frequencies. Figure 10 shows different normalized frequency spectra depending on the duration of a rectangular load pulse.

Electromagnetic or hydraulic exciters can be used in the field as well. These approaches have already been discussed in the section on "Stationary excitation".

Smaller structures are usually excited by means of a hammer impact. Special impact hammers allow the load-time history to be measured. This information is not necessary, if only eigenfrequencies and the damping are to be measured. The eigenfrequencies can easily be determined from the peaks of the computed frequency spectrum and the damping is estimated from the time history by analyzing the decrease in amplitude. If more than one eigenfrequency is excited, the time history may need to be band-pass filtered to determine the frequency-dependent damping values [10, 11]. The information of the impact-force time history, measured by a special impact hammer, is, of course, important, if the system identification method requires the time history as an input.

As already mentioned, a short hammer impact generates a frequency spectrum with high frequencies and low frequencies possess low amounts of energy; thus, the low eigenfrequencies of the structure are only excited a little. Therefore, the hammer impact method is used if small structures with higher eigenfrequencies are investigated.

Large structures are often excited by means of sudden load relief, i.e., a load jump. This approach could be implemented, for instance, by cutting an auxiliary rope under a certain rope force. The frequency content of such a load relief jump shows large low-frequency content in the spectrum. This can also be derived from Figure 10. Therefore, it is ideal for exciting large structures with low frequencies.

If the method with auxiliary ropes is not practicable, other excitation mechanisms must be used. One possibility is to use special blasting cartridges, which generate a reaction like a rocket. If the cartridge is mounted on a special device, the time history of the force could be measured.

Ambient excitation

The use of ambient excitation, such as wind, waves, and traffic, is often a very quick and inexpensive method to assess the main dynamic properties of a structure. The main problem is that the excitation input is unknown. Keeping in mind that at a resonance peak only the accompanying eigenmode responds significantly, the problem can be partly solved.

Assuming that only one eigenmode responds at the resonance frequency and that the damping is small as usual, the excitation has no phase shift, thus the imaginary parts of the solution can be neglected. In this case, the response at two points (i.e., two degrees of freedom) can be expressed as follows (equations 2 and 3):

$$\boldsymbol{u}_{\mathbf{a}}(\omega_j) = \Phi_{aj} \cdot y_j(i\omega) \quad \text{and} \quad \boldsymbol{u}_{\mathbf{b}}(\omega_j) = \Phi_{bj} \cdot y_j(i\omega) \tag{2}$$

with

$$y_j = \frac{\boldsymbol{\Phi}_j^{\mathbf{T}} \times \boldsymbol{P}(i\omega)}{i \cdot 2 \cdot \xi_j \cdot k_j} \tag{3}$$

as the modal response in the jth eigenfrequency. Φ_{aj} is the value of the eigenvector $\boldsymbol{j}$ at point a, $\boldsymbol{P}$ is the unknown excitation vector, ξ_j is the damping value at the jth eigenfrequency, and k_j is the modal stiffness (equation 4):

$$k_j = m_j \cdot \omega_j^2 \tag{4}$$

If the ratio

$$\frac{\boldsymbol{u_a}}{\boldsymbol{u_b}} = \frac{\phi_{aj}}{\phi_{bj}} \tag{5}$$

is formed, one can see from equation (5) that the modal response, which contains the unknown excitation vector $\boldsymbol{P}$, is eliminated by reduction. Thus, the components of the eigenmode can be determined at $a = 1, 2, 3, \ldots$ degrees of freedom related to a reference degree of freedom at b [12].

This relation even holds true if, in the case of stochastic excitation, instead of the Fourier-transformed time histories, the power spectral densities (PSD) are used. For this purpose, the PSD are determined from the response time history. Information about this task can be found in [13]. The peaks of the PSD locate the eigenfrequencies. The modal masses cannot be determined, because the excitation forces are unknown.

The damping of the structure at ambient excitation can be determined from the autocorrelation function (Figure 11). If the structure is excited by a stochastic process in a frequency range that contains the main eigenfrequencies of the structure, the system response depends on its mechanical transfer function. If the excitation follows a white noise spectrum,

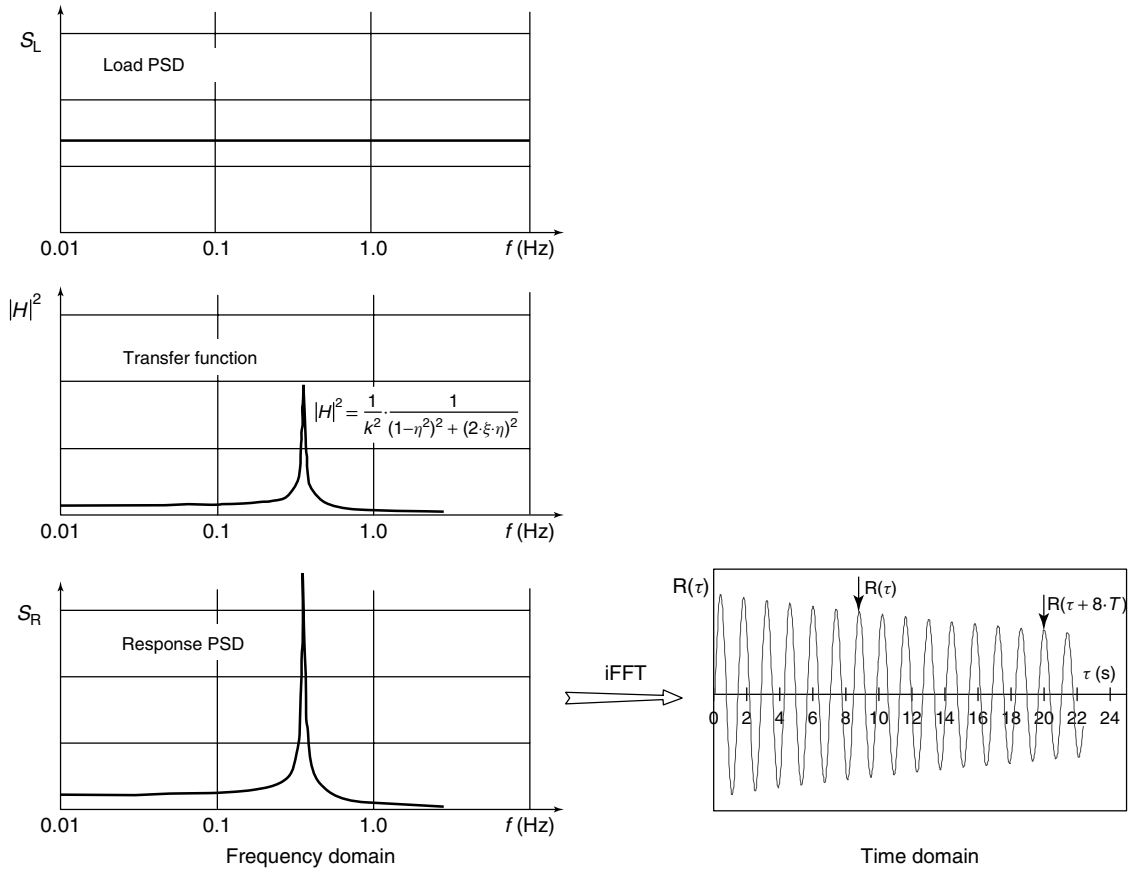


Figure 11. Damping assessment at ambient excitation.

the response function is proportional to the transfer function due to the multiplication of the load PSD and the system transfer function. In practice, it is sufficient that the load PSD is approximately constant around the eigenfrequencies. It is well known that the autocorrelation of the system response is the transformation of the PSD of the system response into the time domain (Wiener–Khintchine relation [13]). The autocorrelation of a system, which vibrates mainly at its main eigenfrequency, looks like a typical damped vibration function. Thus, the damping is determined in the usual manner (equation 6):

$$\delta = \frac{1}{k} \cdot \ln \frac{R(\tau)}{R(\tau + k \cdot T)} \tag{6}$$

where k is the number of cycles with the period T between the measured amplitudes of u ($k = 8$ in Figure 11).

In practice, this method is a very powerful one because the damping of large structures can be measured without expensive measurement devices. It is important to note that the method is a statistical one, which leads to reliable results only if the law of large numbers is observed. This law means that the measurement period must be long. Large structures with eigenfrequencies of about 0.4 Hz need measurement periods of 1 h or more. If the measurement periods are too short, accidental numbers will be the result.

3.3.2 Small structures or local monitoring

Impact loads

Small structures or local parts of larger structures can be dynamically excited by a hammer impact as mentioned above. This is a simple and inexpensive procedure. Owing to the higher eigenfrequencies of small structures, the higher frequency content of a hammer impact is perfect for the excitation of these eigenfrequencies. If the hammer allows the tip pressure time history to be measured, then the input for a parallel calculation of the forced vibration is available (Figure 12).

In some cases this method is not optimal, especially when repeated or continuous measurements are scheduled over the lifetime or repeated monitoring (and loading) is impossible owing to large heights or owing to dangerous environments (the inner parts of reactors). In these cases, it could be advantageous to use piezo elements.

Figure 12. Impact hammer.

Diverse piezoelectric materials of different dimensions, made as double-sided electroded discs, are available. The amount of the impact depends on the size of the piezo element, the direction of polarization, and on the electrical voltage applied to the piezo [14, 15]. For monitoring limited areas of flat structures, it is recommended to use simple circular discs with a diameter of 10–30 mm and a thickness of less than 0.5 mm (Figure 13). A two-component epoxy is used to bond the elements with a certain pressure to the structure. If the structure is metallic, it can be used as an electrical conductor. In this case, the epoxy must be made electrically conductive by adding copper powder with a particle size $<100\,\mu$m. Finite element (FE) simulations show that the thin epoxy layer can be regarded as a stiff bond [16].

Because of the free boundary conditions at its rear, the thickness mode of vibration of the piezo discs does not affect the structure in any significant way. In fact, in this case, the radial component caused by the lateral strain of the piezo is important. If the vertical component is used, additional masses at the rear must be added to permit a reaction supported by the inertia forces.

The elements can be used as actuators and/or as sensors. This is a great advantage because it reduces the costs.

A hybrid amplifier is used for excitation of the piezo elements [16]. This amplifier meets both requirements for improved control of the piezos compared to a regular voltage amplifier: with this equipment, the actuators with a relative small capacitance can be driven with signal frequencies of up to 200 kHz.

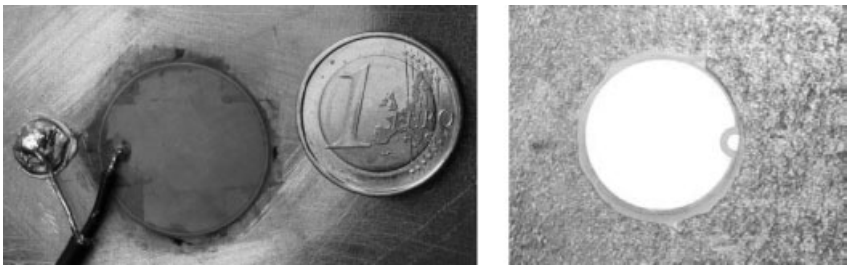


Figure 13. Piezo discs.

Ultrasonic loads

Ultrasonically excited infrared thermography (UT) is a relatively new NDT technique for detecting flaws and damages in different materials. The underlying idea of the technique is the thermal activation of defective locations, such as cracks, delaminations, and kissing bonds by high-intensity ultrasonic mechanical excitation in the frequency range $>20\,000$ Hz [14]. In the vicinity of defects, the elevated dissipation causes a local rise of the surface temperature, which is detected full field and noncontacting by highly sensitive infrared (IR) cameras [15]. A defect can clearly be distinguished from the undamaged structure if the presence of the defect causes a higher dissipation of mechanical energy than the surrounding material.

Numerous potential fields of application for this innovative NDT technique have been explored in the past. This includes flaw detection in laminates and composite materials, crack detection in lightweight metal and ceramic structures, as well as application in the automotive and aircraft industries. A major advantage of ultrasonically excited thermography is its fast applicability, the full-field resolution, and the selectivity in finding defects. Additionally, the equipment is reasonably lightweight for mobile testing. Despite its early successes, the technique is currently still under development. Numerous issues concerning the ultrasound excitation, measurement approach, and data processing of the IR images need to be addressed prior to a broader application of the technique. The application and further development of the technique for the detection of cracks and fatigue damages in metallic civil engineering structures is subject to current research activities. The primary concerns involve the optimization of mechanical excitations in the ultrasonic frequency range and the corresponding data processing techniques.

The excitation sources commonly applied are standard high-intensity ultrasonic converters used for metal and plastic welding [14]. The converters are designed to operate at a single frequency or within narrow frequency bands. Typical working frequencies are 20 000, 40 000, and 60 000 Hz. Some specialized ultrasonic generators allow the converters to be operated over a larger frequency range, i.e., 15 000–25 000 Hz. The extension of the frequency range is linked with some technical obstacles because the converters themselves are optimized to operate under resonance conditions at a single frequency. For good defect detection, variability of the excitation frequency f_U is of special importance. Mechanically, the system, composed of the component to be tested and the ultrasonic excitation, constitutes a two-degree-of-freedom system. Tests showed that similar to lower frequency vibrations, high energy input can be achieved if the vibration system is operated under resonance conditions. Resonance occurs if the excitation frequencies coincide with a natural frequency of the tested component, a practice that explains the selection of high frequencies for defect detection. If only a single frequency is used during excitation, a stationary vibration pattern develops. This causes poor defect detection capabilities of UT in the zero crossings of the vibration mode. The problem can be addressed by shifting or wobbling of the ultrasonic frequency during testing.

The aim of IR-image data processing is the detection and separation of defects from the thermal background radiation of the remainder of the structure. A promising approach is the so-called lock-in technique. For this purpose, the power output of the excitation source is modulated by a modulation frequency, f_m, which is much smaller than f_U. The amplitude modulation of the excitation leads to temporal periodic variations of the dissipated energy and the

corresponding temperature changes in the vicinity of defects. This variation of the surface temperature is captured by an IR-image sequence over a certain time interval and is analyzed on-line or off-line with respect to the periodic variations. A discrete Fourier transformation (DFT) of the temperature evolution of each pixel can be used to separate the temperature changes at the modulation frequency, whereas contributions at other frequencies are filtered out during data processing. The DFT can be used to calculate the amplitude of the temperature evolution and its phase shift with respect to the modulation cycle; these can be easily displayed as amplitude or phase images. The phase shift, especially, promises good results for defect recognition because it is not influenced by variations in the local surface emissivity.

4 LOAD MODELS FOR LIFETIME PREDICTION

4.1 General remarks

If the lifetime of structures is to be assessed, models of the structural behavior are needed. Usually three types of models are necessary:

- load or action model
- system model
- damage model.

In this article, the different types of load models are discussed. The load or action models can be deterministic or stochastic. Usually, nearly all loads are stochastic, but the variance may be very small, so they appear to be deterministic. Water loads acting on gates, for instance, of locks in a canal, do have very small variance as the water level changes between the lowest and the highest point. Such loads may be treated as deterministic.

Wind or wave loads and typical road traffic show a large variance. The scattering between very light vehicles and extremely heavy ones is large, for example. The same holds true for railway traffic. Although the sequence of trains is known by the time schedule, the single-axle load is not [17].

The scattering of the loads is increased by the dynamic system behavior. A load that passes a bridge causes vibration of the structure due to the deflection of the structure under the moving load and due to the additional pavement roughness.

In the following section, typical load or action processes and possible models are discussed.

4.2 Traffic model

4.2.1 Load generation

The generation of a load sequence of road traffic is difficult owing to the stochastic character of the loads, the widely scattered traffic behavior, the dependence on the local site, and the nonlinear effect caused by feedback of traffic to the drivers. Nonlinear effects are usually not taken into account. A sufficient micro-traffic model should, therefore, take into account the

- distribution of vehicle weights
- sequence of vehicle types
- time distance between vehicles.

These input values are stochastic variables. The statistical data of vehicle types may be estimated from vehicle registration data from the authorities. If no data are available, WIM (weigh-in-motion) data can be measured to get the input data, for instance, for a bridge. WIM, first proposed by Moses [18], has been applied worldwide for more than 20 years [19]. On the other hand, data can be achieved by means of strain measurements of, e.g., cross girders of bridges [20–22]. If measurements are not available, estimated load distribution functions must be used.

Typical vehicle sequences (e.g., clustering of trucks) and typical temporal distances between vehicles must be taken into account as well [2]. Experiments have shown that, for instance, the influence of a traffic time history with clustered trucks is large [2]. The fatigue of a specimen loaded with a force time history taking into account typical truck clustering is just 50% compared with a force time history, which shows exactly the same statistics, but includes stationary distributed trucks over time. The vehicle statistics (load distribution, clusters, and temporal vehicle distances) depend on the type of road (highway, country road, and city road) and on the local situation. The statistics of a certain road type can also be used for other similar cases, e.g., for other bridges of the same road. If the statistical input is known, artificial load–time histories can be generated by means of the Monte Carlo method.

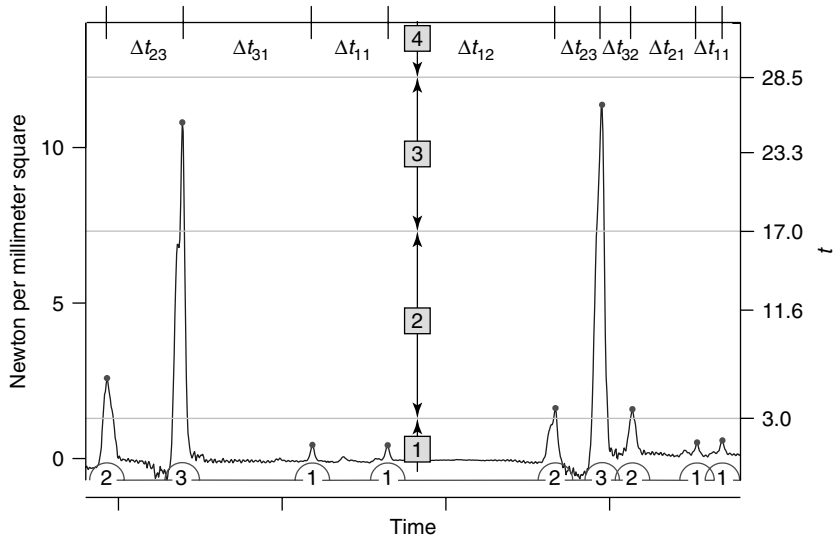


Figure 14. Time history of traffic and vehicle assignment.

In the first step, the measurement records must be analyzed. Figure 14 shows a small part of a long-term record. Heavy trucks and lighter cars are easy to identify. In the whole record, four different types of vehicles are identified: (i) cars, (ii) light or empty trucks, (iii) ordinary, and (iv) heavy trucks. The question is how to separate these different classes. To do this, the density function of the overall traffic is needed. As an example, Figure 15 shows the histogram of vehicle loads measured at a cross girder of a highway bridge. In the next step, four weighted normal distributions are used to fit the measured histogram. It is obvious that the overall traffic can be described well by the sum of four normally distributed vehicle types [2].

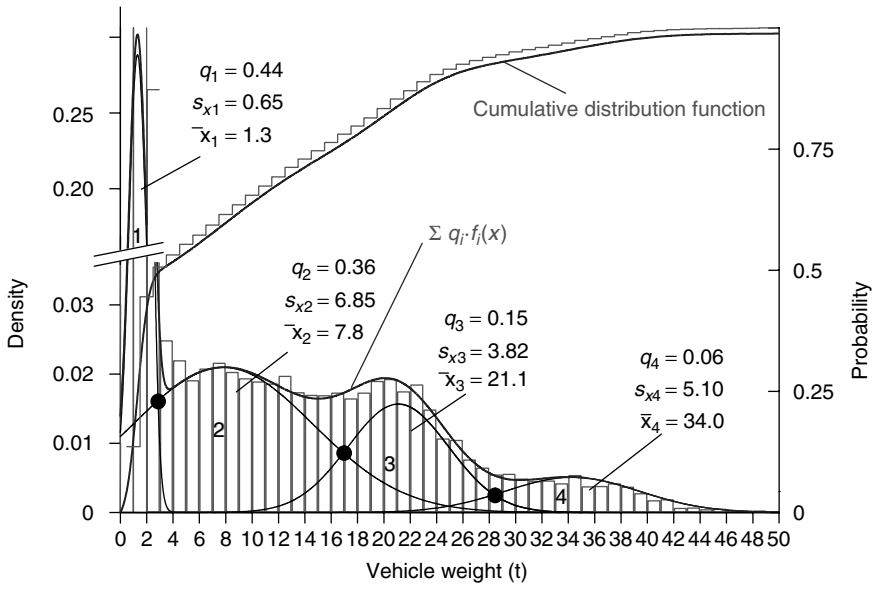


Figure 15. Load distribution and fitted combination of density functions.

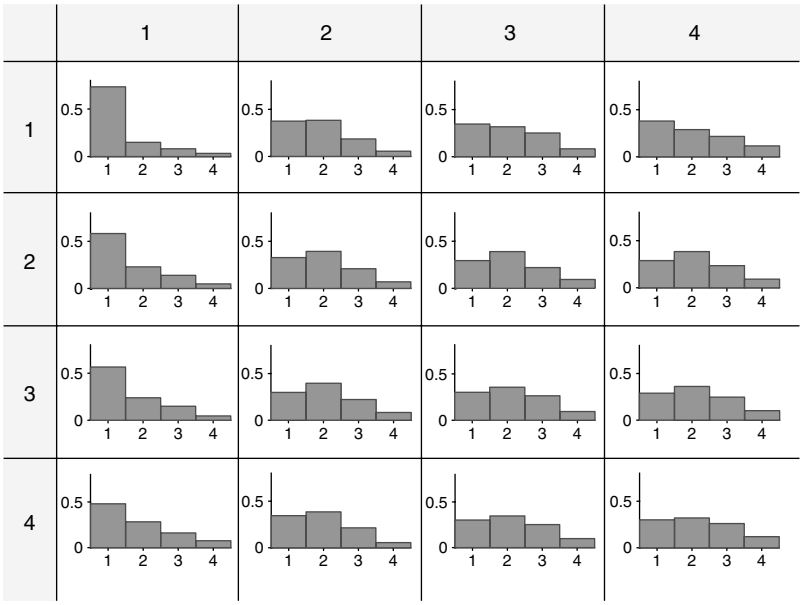


Figure 16. Matrix of sequence densities.

On the basis of the discriminate analysis method, e.g. [23], the intersections of the normal distributions represent the borders of the four accompanying classes. The intersection points in this example are 3, 17, and 28.5 t (Figure 15). With this class definition, the traffic stream plotted in Figure 14 can be classified into four classes. In the next step, the traffic stream is transposed into a stream of numerals, which define the adjacent classes, (lower part of Figure 14).

To classify the sequence and the temporal distances of the traffic (expressed by the numerals), an n-dimensional matrix can be used. First, the description of the vehicle subsequence is discussed (Figure 16). The rows belong to the class of the first vehicle, the columns to the subsequent one. The elements of the matrix then contain a histogram of the four vehicle classes. The histogram of element i, j shows the probability of occurrence of the next vehicle in that sequence. It is obvious that the probability of occurrence of succeeding trucks is high (see the high mode of the distributions in elements with a column number greater than 1). If n sequences of vehicles are to be taken into account, the matrix becomes $(n-1)$ dimensional.

A similar approach is used to describe the temporal distances of the vehicles. Figure 17 shows the density functions of the temporal distances in seconds between two successive vehicles sorted in a similar matrix. The row of a matrix element indicates the actual vehicle type and the column the next one. Elements (3,1) or (4,1) indicate, for example, cars (type 1) waiting impatiently for an opportunity to overtake trucks, as the modal value of the distributions in these elements is remarkably low. If n sequences of vehicles are to be taken into account, the matrix becomes n-dimensional.

With these data, an artificial traffic stream can be generated using the Monte Carlo method. A random generator generates vehicle types that always consider the most recent sequence generated on the basis of the densities shown in Figure 16. Temporal distances are chosen according to the densities in Figure 17. The actual load of a vehicle is then randomly determined on the basis of the corresponding density of the forecast vehicle type.

4.2.2 Strain-generating procedure

Although it is not a typical question of load generation, a few remarks are given about the generation of adjacent strains in a structure, e.g., a bridge [24]. The structure is usually modeled by the FEM

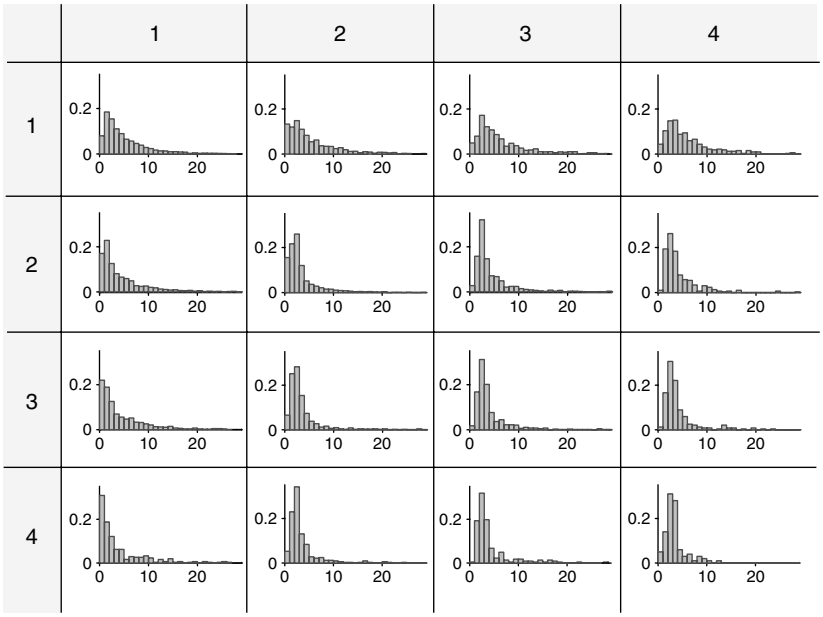


Figure 17. Matrix of temporal distance densities (time in [s]).

(Figure 18). In a preceding modal analysis, eigenvalues and eigenmodes are extracted as inputs to the iterative solution scheme. The different types of vehicles are idealized as damped 2-mass systems for each wheel. They consist of a wheel mass, a mass containing the corresponding part of vehicle body and payload, spring and damper between both, and another spring representing the tire tangent stiffness (Figure 19).

It should be noted that the use of the modal method implies that the passing vehicle masses are small compared with the structure mass because the vehicle masses are not included in the system modal parameters. If the loads are high (e.g., railway) and the bridge is short, i.e., light, then an expensive full time history calculation may be necessary.

A large number of measurements show that it is permissible to define the roughness of roadway surfaces to be a Gaussian, stationary, and ergodic process characterized by a power spectral density function. In [25] the exponential function

$$\phi(\Omega) = \phi(\Omega_0) \cdot \left(\frac{\Omega}{\Omega_0}\right)^{-w} \tag{7}$$

was proposed. Equation (7) describes the PSD by a reference frequency Ω_0 and the dimensionless

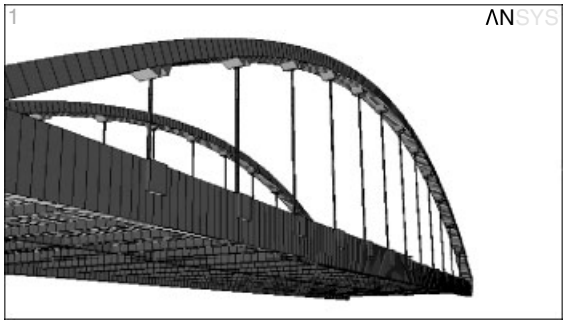


Figure 18. Finite-element model.

waviness w. Figure 20 shows some PSD functions of pavements of different conditions [24].

For use in numerical simulations, a discrete realization of a roughness function is needed. This realization can be generated from a finite Fourier series with random phase, where the amplitude $\hat{u}_i$ of each Fourier term is determined (equation 8) in accordance with the spectrum

$$\hat{u}_i = \sqrt{2 \cdot \Delta\Omega_i \cdot \phi(\Omega_i)} \tag{8}$$

Owing to the higher intensity of roughness at low circular frequencies (larger wavelengths), the smaller frequency interval $\Delta\Omega_0$ is chosen at this

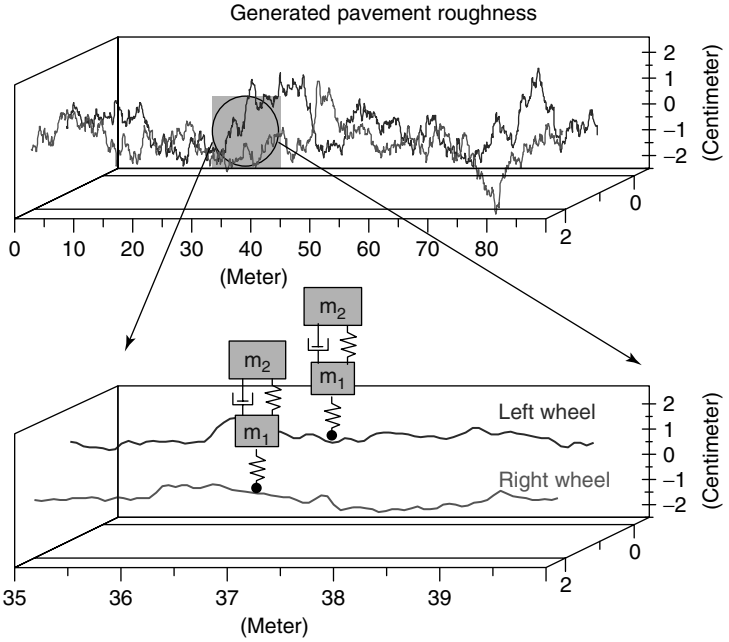


Figure 19. Wheels as 2-mass systems.

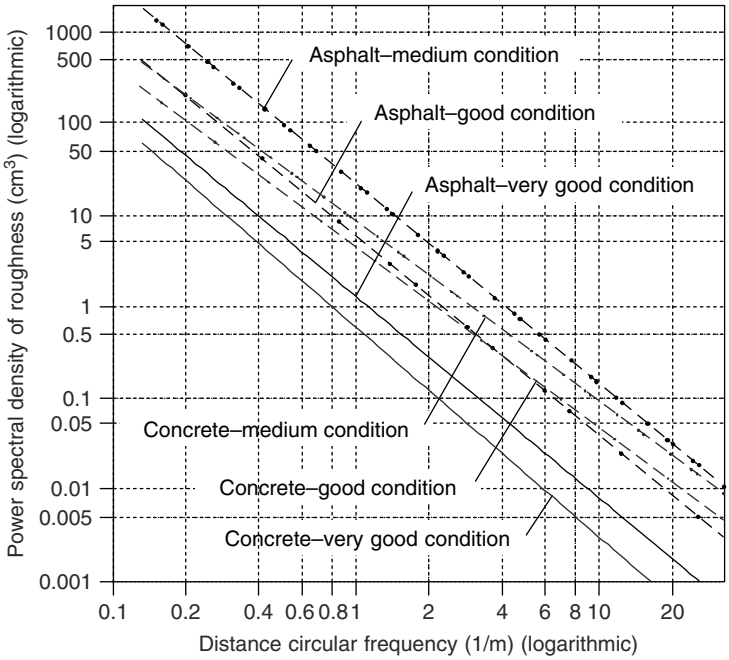


Figure 20. Power spectral densities.

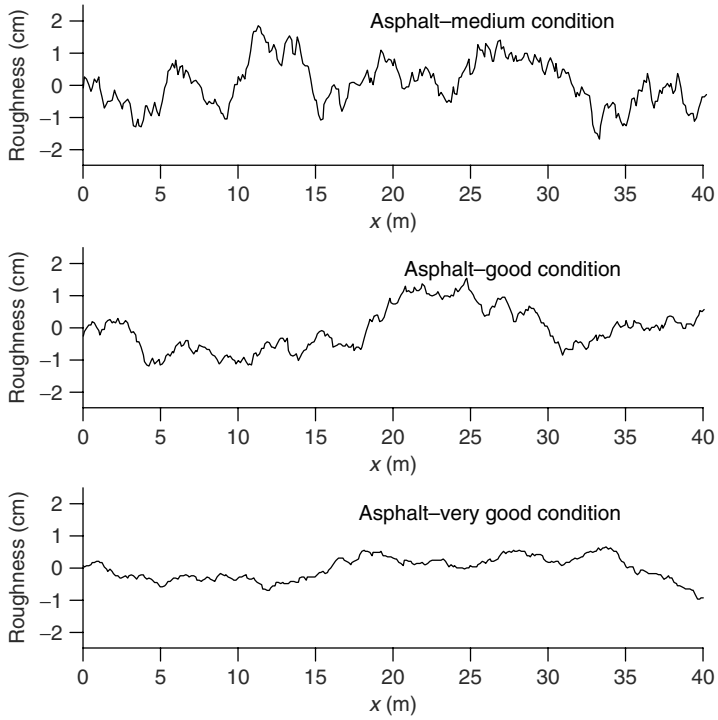


Figure 21. Pavement roughness profiles.

point. Figure 21 shows three generated roughness profiles as an example.

Because the system changes at each time step, the submodels of the bridge and vehicles must be treated separately. The coupling forces of both systems are calculated and an iteration must be performed until the Euclidean norm of the changes of the interactive forces fall below a given threshold. The solution of this iteration is saved for each time step in modal coordinates. Subsequently, displacements, velocities, accelerations, or stresses can be calculated from these data for nodes that are of particular interest.

4.3 Wind models

4.3.1 Introduction

The natural wind is the most important loading case for tall structures like skyscrapers, towers, masts, and chimneys. The wind process is highly dynamical, and the wind forces change in time and in space due to the roughness-induced gusts or turbulence [4]. In many codes, the wind load is statically modeled on the basis of an envelope of the maximum-gust wind speed. This assumption is on the safe side for freestanding structures, as it is the worst-case loading. Thus, it cannot be used for the assessment of the lifetime of a structure.

Because of the time-dependent wind forces, slender structures respond with vibrations. Furthermore, even a gust-free, laminar wind can excite vibrations. Examples are vortex-induced or self-excited vibrations like galloping or flutter. The latter occur if the structure is not rigid, but can respond to wind effects with deformations [4]. The structure itself may then control the wind force and the problem becomes aeroelastic. Since these events may cause long lasting vibrations with large amplitudes, they could cause structural collapse or early fatigue damage. Thus, they must be prevented, e.g., by adding damping to the structure. They do not play a significant role in SHM and thus will not be discussed here.

Vibrating structures, especially those made of steel, are prone to fatigue, even if the stresses are very

much lower than the yield stress. Thus, the adequate description and modeling of the stochastic turbulent wind process is a prerequisite for a lifetime assessment of a structure.

4.3.2 Models for gust-induced vibration

Description of the wind process

The natural wind changes in time and space. One can illustrate the process by imagining that a constant flow between high- and low-pressure zones is superimposed by wind balls (gusts) of any diameter and rotation direction near the rough earth surface. The rougher the surface, the more turbulent or gusty the wind is and vice versa. Figure 22 shows a measured wind profile. It is obvious that the wind speed increases with height.

The time record of a wind speed is normally split up into two parts: the mean wind speed and the superimposed fluctuating part due to turbulence (equation 9):

$$u(z,t) = \overline{u}(z) + u'(z,t) \tag{9}$$

Figure 23 describes the situation: the actual wind speed is the sum of the mean wind speed and the superimposed gusts.

The increasing profile of the mean wind with height can be described by a logarithmic or a power law [2, 26]. Both profiles are rather similar at lower heights (<100 m); over 100 m the power law is closer to reality [27]. The logarithmic law reads (equation 10)

$$\overline{u}(z) = \overline{u}(10) \cdot \frac{\ln \dfrac{z}{z_0}}{\ln \dfrac{10}{z_0}} \tag{10}$$

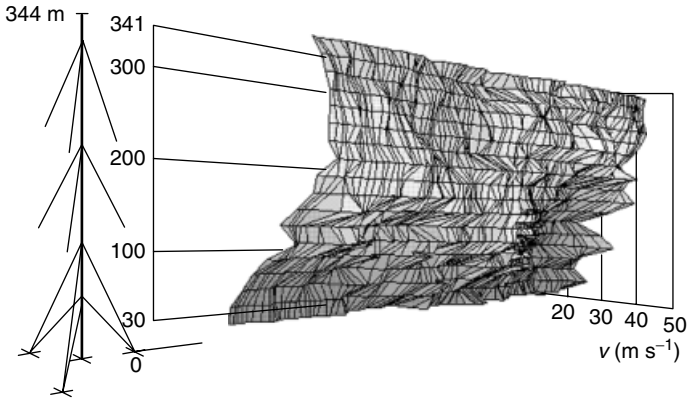


Figure 22. Wind field measured over 20 s.

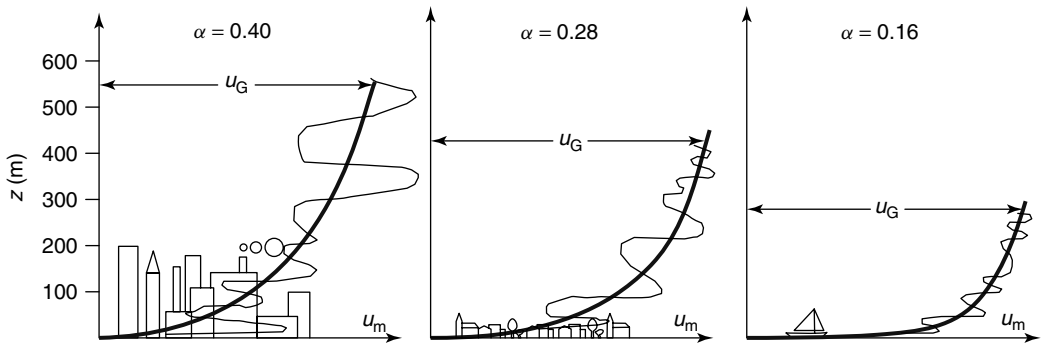


Figure 23. Wind speed depending on roughness and height.

z is the height at which the speed should be determined, and z_0 is a so-called roughness height depending on the roughness of the surface (Table 1). $u_m(10)$ is the mean reference speed at 10-m height, usually given by codes or meteorologists. For greater heights, the power law is closer to reality, and it reads as follows (equation 11):

$$\overline{u}(z) = \overline{u}(10) \cdot \left(\frac{z}{10}\right)^{\alpha} \tag{11}$$

where α is an empirical exponent depending on the roughness on site; its value can be taken from Table 1 [28].

If the influence of the changing roughness is to be investigated in more detail, the mixing of profiles must be taken into account [29, 30]. In the framework of this short article, no more details can be provided.

The turbulent wind is a stochastic process. For the lifetime assessment of a (vibrating) structure, the statistical input must be known. As already mentioned, the wind process is split up into a constant mean wind speed and a fluctuating speed. The structure response due to the mean wind speed is purely static. The remaining fluctuating part of the wind process results in stochastic random vibrations of the structure. It is not the aim of this article to deal with the problems of random vibrations; for a description of these, the reader is referred to, e.g. [31]. This article discusses just the stochastic load characteristics.

Table 1. Description of the terrain category

Terrain category	Description	z_0 (m)	α	ε	z_{min}
I	Open sea, lakes with minimum 5-km free area in wind direction, flat country without obstacles	0.01	0.13	0.13	2
II	Terrain with fences, single farm houses, trees, agricultural terrain	0.05	0.16	0.26	4
III	Suburbs, industrial or trade areas, forests	0.3	0.22	0.37	8
IV	Cities with a minimum 15% of built-up area with buildings with a mean height more than 15 m	1	0.3	0.46	16

Figure 24 shows three records of measured wind speeds. The passing gusts cause a correlation between the wind speed record at a certain height level and between the speeds of neighboring levels. The correlations can be described by auto- and cross-correlation functions. Figure 25 shows auto- and cross-correlation functions as an example. The symmeric function is the autocorrelation $R_{xx}(\tau)$ of the record at level a. The skew functions are the cross-correlations $R_{xy}(\tau)$. The broader the correlation function, the longer the signals are correlated, the larger the mean gusts. The maxima of the cross-correlation functions are shifted a little. The time shift indicates the time difference a gust needs to reach the other level. Figure 24 illustrates this correlation. The increase of wind speed indicates a gust. The height level 66 m is hit by a gust at about 50 s. The height level 30 m is reached by the same gust due to the ball shape about 4 s later, and both functions show the largest correlation if they are shifted by 4 s.

The auto- and the cross-correlations decrease with increasing time or distance. This trend is the consequence of the limited gust diameter. The value of the autocorrelation function at $\tau = 0$ is the variance

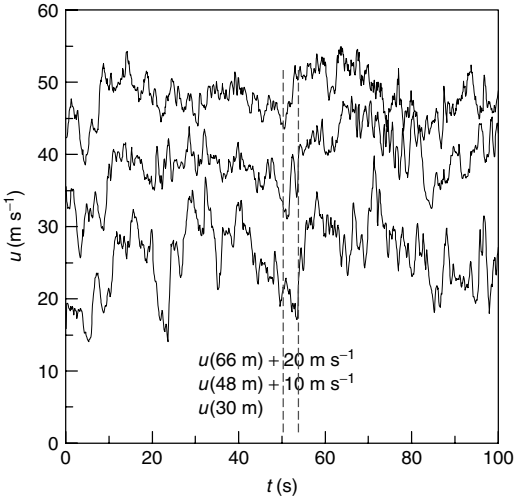


Figure 24. Wind speed records.

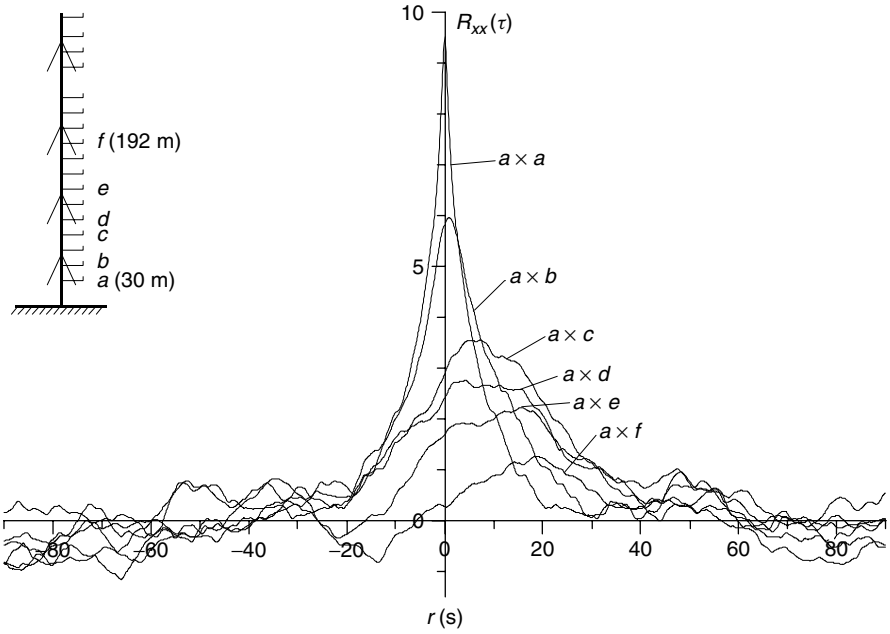


Figure 25. Measured auto- and cross-correlations.

of the stochastic wind process. For more information about the determination and interpretation of correlation functions, see e.g., [32].

If the correlation functions are Fourier-transformed into the frequency domain, the PSD function S_u is obtained (Wiener–Khintchine relation). An estimation of the PSD can be obtained by taking the Fourier transform of the measured records and then squaring the resulting PSD. A PSD describes the frequency-dependent variance of the stochastic process. The integral over the whole PSD gives the variance of the stochastic wind process.

In the past, many PSD functions S_u have been proposed [4, 26, 27]. Some of them are based on physical background, some are empirical functions, and others are limited to small heights (Figure 26). In this article, it is not possible to go into the details of the PSD functions.

The Eurocode [28] prescribes a PSD as follows (equations 12 and 13):

$$R_N = \frac{f_{1,x} \cdot S_u(n_{1,x})}{\sigma_u^2} = \frac{6,8 \cdot N_{1,x}}{(1 + 10,2 \cdot N_{1,x})^{5/3}} \tag{12}$$

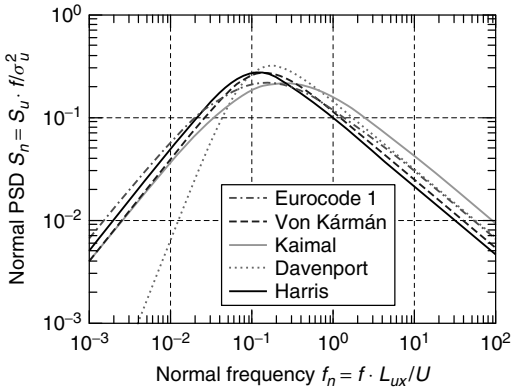


Figure 26. PSD of wind speed.

with

$$N_{1,x} = \frac{f_{1,x} \cdot L_{\mathrm{i}}(z_{\mathrm{eff}})}{\overline{u}(z_{\mathrm{eff}})} \tag{13}$$

where $f_{1,x}$ is the lowest eigenfrequency of the structure, ${\sigma_{\mathrm{v}}}^2$ is the variance of the wind process, $u_{\mathrm{m}}(z_{\mathrm{eff}})$ is the mean wind at the effective height z_{eff}, and L_{i} is the integral length scale. The integral length

scale is the mean gust diameter. It is assessed by integrating the autocorrelation function, i.e., determining the area under the function. This value is divided by the maximum at $\tau = 0$ (equation 14). The resulting rectangle has the same area as the autocorrelation function.

$$L_{\mathrm{i}}(z) = \frac{\int_{\tau=0}^{\infty} R_{xx}(\tau) \cdot \mathrm{d}\tau}{R_{xx}(0)} \tag{14}$$

The integral length scale L_{i} results from (equation 15)

$$\begin{aligned} L_{\mathrm{i}}(z) &= 300 \cdot (z/300)^{\varepsilon} \quad (L_{\mathrm{i}}, z \text{ in meter}) \\ &\qquad \text{for } z_{\min} \le z \le 300 \text{ m} \\ L_{\mathrm{i}}(z) &= 300 \cdot (z_{\min}/300)^{\varepsilon} \quad (L_{\mathrm{i}}, z_{\min} \text{ in meter}) \\ &\qquad \text{for } z \le z_{\min} \\ L_{\mathrm{i}}(z) &= 300 \text{ m} \quad \text{for} \quad z \ge 300 \text{ m} \end{aligned} \tag{15}$$

with

ε ... Exponent, (Table 1)

$z_{\min}$ (Table 1).

The variance ${\sigma_{\mathrm{v}}}^2$ of the wind process can be determined from the turbulence intensity I_u, which is defined to be the variation coefficient of the stochastic process (equation 16):

$$I_u(z) = \frac{\sigma_u(z)}{\overline{u}(z)} \tag{16}$$

This function can be expanded as follows (equation 17):

$$\begin{aligned} I_u(z) &= I_u(10) \left(\frac{z}{10}\right)^{-\alpha} \\ I_u(10) &= \frac{1}{\ln(10/z_0)} \end{aligned} \tag{17}$$

The turbulence intensity is provided in the codes [28, 30]. It should be noted that all codes present values for the maximum wind load situation. If—for lifetime assessment calculations—even lower wind speeds are taken into account, the values should be used with some reflection. In [27] values of nonextreme wind situations are given.

If the structure is hit by a turbulent wind at more than one point (which is usual), then the cross-correlations between the wind processes between these points become important. Coherence is introduced as a dimensionless characteristic number of the correlation of two spectra. It is given by (equation 18):

$$\gamma_{i,j} = \sqrt{\frac{|S_{i,j}(f)|^2}{S_i(f) \cdot S_j(f)}} \tag{18}$$

In practical cases, the auto-PSD is available from codes, but not the cross-PSD. Thus, the coherence must be determined on the basis of a proposal from Davenport [33] (equation 19):

$$\gamma_{ij}(f) = \mathrm{e}^{-\frac{f \cdot C \cdot \Delta z}{\overline{\mathrm{u}}_{10}}} \tag{19}$$

For the so-called decay constant, C, values between 5 and 12 are used. The coherence can be determined in any direction of space. In such cases, the coordinate Δz must be replaced by Δx or Δy. The decay constants depend on the direction [4, 26]. In equation (19), f is the frequency and $\overline{\mathrm{u}}_{10}$ the mean wind speed at 10-m height.

With the known coherence, the cross-PSD can now be determined (equation 20):

$$S_{ij} = \gamma_{ij}(f) \cdot \sqrt{S_i(f) \cdot S_j(f)} \tag{20}$$

For a structure with more than one point under wind loading, the auto- and cross-PSD are written in a spectral matrix. Equation (21) gives an example for three loaded points (Figure 27).

$$S_{ij}(f) = \begin{bmatrix} \tilde{\mathrm{S}}_{11} & \tilde{\mathrm{S}}_{12} & \tilde{\mathrm{S}}_{13} \\ \tilde{\mathrm{S}}_{21}^{*} & \tilde{\mathrm{S}}_{22} & \tilde{\mathrm{S}}_{23} \\ \tilde{\mathrm{S}}_{31}^{*} & \tilde{\mathrm{S}}_{32}^{*} & \tilde{\mathrm{S}}_{33} \end{bmatrix} \tag{21}$$

The tilde shows that the elements are complex values and the star indicates the conjugated complex values.

In the next step, the wind force must be calculated from the wind speed. Slender, linelike structures are usually investigated under the assumption of a

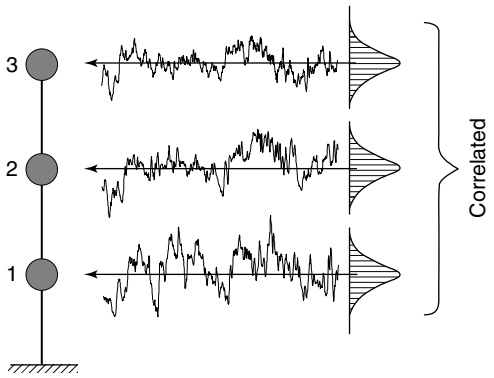


Figure 27. Structure with three correlated wind attacks.

quasi-steady theory (equation 22):

$$F_{\mathrm{w}}(t) = \frac{\rho}{2} \cdot C \cdot A \cdot u(t)^2 \approx \frac{\rho}{2} \cdot C \cdot A \cdot \overline{\mathrm{u}}^2 + \rho \cdot C \cdot A \cdot \overline{\mathrm{u}} \cdot u'(t) \quad (22)$$

Wind forces due to quadratic terms are neglected. This model simplifies in practice because it assumes that all wind gusts are equally transferred into wind forces. Owing to different gust diameters, which could be interpreted as different frequencies, the transfer function is frequency dependent. The smaller the gusts, the lower the resultant wind force, and the faster the gusts pass, the higher the frequency [4, 34]. Thus, the aerodynamic coefficient in the second part of equation (22) decreases with the frequency. To calculate the PSD of the wind forces, the PSD of the wind speed turbulence must be multiplied by the aerodynamic transfer function $\chi_{\mathrm{aero}}(f)$ (equation 23):

$$S_{\mathrm{iw}}(f) = (\rho \cdot C \cdot A \cdot \overline{\mathrm{u}})^2 \cdot \chi_{\mathrm{aero}}(f)^2 \cdot S_i(f) \quad (23)$$

Investigation in the frequency domain

To analyze the effects of wind loading in the frequency domain, the wind process is decomposed into a mean wind speed and the superimposed fluctuating part due to turbulence (equation 24):

$$w(z,t) = \overline{w}(z) + w'(z,t) \quad (24)$$

The determination of the system response under the static mean wind force $\overline{w}$ is performed by a static calculation and results in a static response $\overline{A}$.

The second step is the determination of the system response to the dynamic stochastic process. The dynamic behavior of the vibrating system can be described in the frequency domain by its mechanical transfer function [35]. Although this is not what this article focuses on, the methodology is briefly summarized. The transfer function is determined by calculating the steady dynamic response of a state variable of a structure under a variable harmonic load with constant frequency. If the calculation is performed for every single frequency in the range of interest, the dynamic transfer function for this state variable is obtained. The calculation is performed with standard dynamic analysis methods.

Multiplication of the PSD of the wind load $S_{\mathrm{W}}(f)$ and the dynamic transfer function results in the PSD of the response (equation 25):

$$S_A(f) = |H(f)|^2 \, S_{\mathrm{W}}(f) \quad (25)$$

If the mechanical transfer function is determined for a state variable deflection or for an internal force, the PSD of the response describes a deflection or an internal force. Because of the influence of the square of the mechanical transfer function, resonance peaks have a considerable effect on the response PSD. As the integral of the wind speed, PSD is equal to the variance of the wind speed process, the integral over the response PSD is equal to the variance of the response process (equation 26). The whole procedure is graphically described in Figure 28.

$$\sigma_A^2 = \int_0^\infty S_A(f) \, \mathrm{d}f \quad (26)$$

The overall response can then be determined by addition of the result of the static solution under mean wind and the dynamic response under turbulent wind (equation 27).

$$A = \overline{\mathrm{A}} + g \cdot \sigma_A \quad (27)$$

The so-called peak factor g depends on the level of probability of occurrence of the peak values; for extreme wind situations it is usually chosen to be about 3.5. If lower wind speeds are under investigation, the factor is different.

If structures with more than one degree of freedom are to be investigated, the influence of the cross-spectra must be taken into account. Equation (25) becomes a matrix formulation in these cases. The

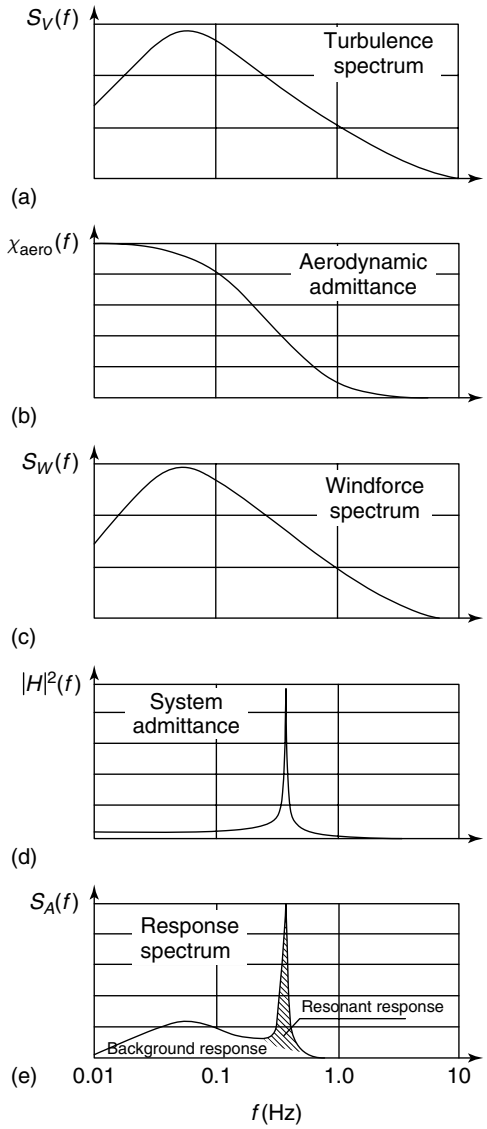


Figure 28. Procedure for the determination of turbulent wind effects.

general procedure is very similar, and is discussed in more detail in [35].

Owing to the assumption that the response contributions at all frequencies can be superimposed, the frequency-domain method is restricted to linear systems. If the nonlinearities are small, they often can be taken into account by means of linearization. For this purpose, the nonlinear response of the mean wind load situation is calculated taking into account the nonlinearities by means of a nonlinear static calculation. The system parameters of the state under mean wind are then used for a linear dynamic calculation, resulting in a tangential vibration around the operation point (Figure 29) with W as a wind force and f as a deflection. If the nonlinearities are small, the deviation from the real solution is small.

Investigation in the time domain

Investigations in the time domain are performed if structural nonlinearities (caused, for instance, by cable sag of guyed structures, certain material nonlinearities, etc.) play a significant role. Any time domain calculation requires that time-dependent forces be defined, i.e., wind forces. The wind forces are determined from the wind speeds, usually neglecting the influence of the frequency-dependent aerodynamic transfer function, because it is defined in the frequency domain and not in the time domain. If this influence is to be taken into account, fitted indicial functions calculated from the frequency-dependent aerodynamic transfer function can be used [36]. For the generation of an artificial wind speed record, the whole wind process is split up into a constant mean wind speed part and a fluctuating one. The following derivation is focused on the fluctuating part. The total wind speed record is then achieved by superimposing the mean wind speed to the fluctuating part.

The process of generating artificial wind speed records begins with a given statistical input that satisfies the total wind speed record. Input parameters include the variance of the process, the mean wind speed, the PSD, and the decay factors of the coherence functions.

For the generation of wind speed records, many numerical procedures are available:

- superposition of harmonic waves with different amplitudes and stochastic phases;
- filtering of a white noise with autoregressive filter (AR, ARMA models);
- data composition by means of wavelets.

Each method has its advantages and disadvantages. ARMA models required large storage space in the computer, but they are very fast. Wave superposition, on the other hand, requires more CPU time, but less storage space. The effectiveness of ARMA models strongly depends on the filter order and the chosen

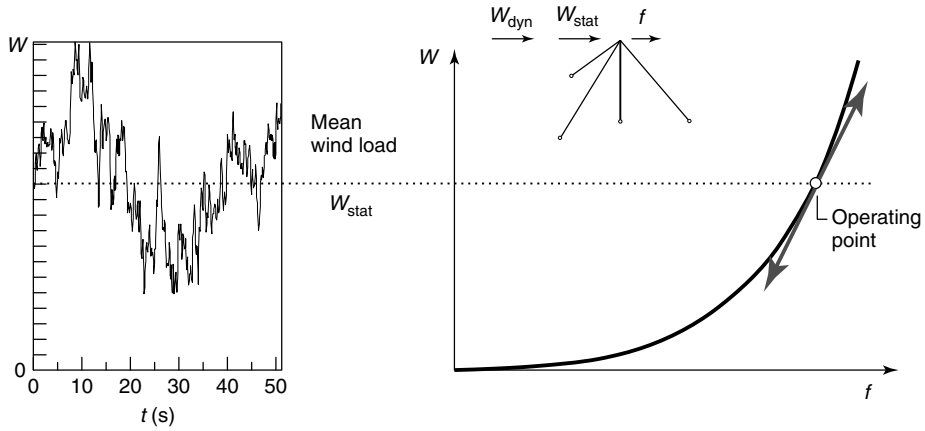


Figure 29. Linearization around the operating point.

time step. The wave superposition method seems to generate time records that correspond better to the statistical input values. Thus, the wave superposition method is explained in more detail; it is based on a fast algorithm [37, 38].

The wind speed time history at node j results from a type of cascade algorithm, starting from the wind process at node i, which depends only on its autospectrum. The wind speed process $u'_j(t)$ at every other node is generated by the addition of another independent process (equation 28).

$$u'_j(t) = \sum_{m=1}^{j} \sum_{n=1}^{N} |H_{jm}(f_n)| \cdot \sqrt{2 \cdot \Delta f} \cdot \cos(2\pi f_n t + \Theta_{jm}(f_n) + \Phi_{m,n}) \tag{28}$$

$|H_{jm}(f_n)|$ are the amplitudes of the harmonic components at node j, which could be determined from the spectral density matrix via a Cholesky decomposition. The elements $H_{jm}(f_n)$ must meet the condition (equation 29)

$$\mathbf{S}(f) = \mathbf{H}(f) \cdot \mathbf{H}^{*\mathrm{T}}(f) \tag{29}$$

In equation (28) Δf is the frequency distance between the discrete frequencies f_n and f_{n+1}. The phase angle Θ results from the complex cross-PSD and can be determined from the imaginary and real parts of $H_{jm}(f_n)$ by means of equation (30):

$$\Theta_{jm} = \arctan \frac{\Im\{H_{jm}(f)\}}{\Re\{H_{jm}(f)\}} \tag{30}$$

The spectral density matrix for this purpose is usually set to be real. Only a phase angle of $\Theta = 2 \cdot \pi \cdot f \cdot \Delta x / \bar{u}$ is taken into account (Taylor's hypothesis) in equation (28). The phase angle $\Phi_{m,n}$ of the mth process with frequency f_n in node j is a uniformly distributed random number between 0 and 2π.

The superposition of the harmonic components is performed with much less effort if it is done in the frequency domain. A summation over all components m results in a Fourier transformation of the time history in node j (equation 31):

$$c_j(f_n) = \sum_{m=1}^{j} |c_{jm}(f_n)| \cdot \mathrm{e}^{-i(\Theta_{jm}(f_n) + \Phi_m(f_n))} \tag{31}$$

The back transformation into the time domain yields (equation 32):

$$u_j(t) = \sum_{k=0}^{N-1} c_j(f_n) \cdot \mathrm{e}^{i2\pi nk/N} \tag{32}$$

If the number of Fourier coefficients is a power of 2, the calculation can be accelerated using the algorithm of Cooley and Tukey [39]. The amplitude of the harmonic wave with frequency f_n of the wind

component of the node m in node j can be determined by equation (33):

$$|c_{jm}(f_n)| = \frac{1}{2} \cdot |H_{jm}(f_n)| \cdot \sqrt{2 \cdot \Delta f} \qquad (33)$$

The factor 1/2 results from the spectral function defined over the whole two-sided frequency band up to the frequency of

$$f = \frac{N-1}{T} \qquad (34)$$

with N equal to the number of discrete values in the time domain and T equal to the length of the time record (equation 34).

The real and imaginary parts of the Fourier coefficients of the component m in node j are (equations 35 and 36):

$$\begin{aligned} \Re\{c_{jm}(f_n)\} &= 2 \cdot |c_{jm}(f_n)| \cdot \cos(\Theta_{jm}(f_n) \\ &\quad + \Phi_m(f_n)) \\ \Im\{c_{jm}(f_n)\} &= 2 \cdot |c_{jm}(f_n)| \cdot \sin(\Theta_{jm}(f_n) \\ &\quad + \Phi_m(f_n)) \end{aligned} \qquad (35)$$

with

$$c_{jm}(f_n) = \frac{1}{2} \cdot (\Re\{c_{jm}(f_n)\}) - \mathrm{i} \cdot \Im\{c_{jm}(f_n)\} \qquad (36)$$

Use of the Euler identity $\mathrm{e}^{-ix} = \cos(x) + i \cdot \sin(x)$ results in the complex Fourier coefficients (equation 37):

$$c_{jm}(f_n) = |c_{jm}(f_n)| \cdot \mathrm{e}^{-i(\Theta_{jm}(f_n)+\Phi_m(f_n))} \qquad (37)$$

2^M points in the time domain result in $2^{M-1} + 1$ points in the frequency domain. Compared with a pure time domain calculation as it is usually performed, a reduction of about 10 s in the CPU time is achieved using this method.

If the node distances are fairly large, a reduction of the wind loads at the nodes of interest due to the decreasing correlation can be taken into account [38].

The computation results in a couple of correlated wind speed records, which are used as inputs for a time domain calculation (Figure 27).

Finally, the mean wind speed at the different sites must be superimposed.

5 CONCLUSIONS

Any monitoring measure needs actions or loads to produce strains in the structure to be monitored. In this article, advice and references are given for the right choice of actions or loads and how to model these loads. Two types of loads are generally necessary for SHM measures. At the beginning of the SHM procedure, precise system identification is usually needed to adjust the theoretical model of the structure or parts of it. If a life time prediction of the structure should be performed, other types of load models must be used, which take into account the probability of occurrence of realistic load situations. Both types are discussed.

In this article, only traffic and wind load models are discussed in detail. Unexpected loads such as earthquakes, blasts, etc., cannot be predicted very accurately; thus, a lifetime assessment is not feasible. A similar situation exists with wind-induced vibrations, e.g., due to vortex shedding, galloping, flutter, etc. These vibrations are usually always dangerous for a structure and may cause heavy fatigue or other types of damage; therefore, wind-induced vibrations must be prevented rather than predicted.

REFERENCES

[1] Schütz W. Fatigue life prediction by calculation: facts and fantasies. In *Structural Safety & Reliability*, Schueller GI, Yao JTP (eds). Baalkema: Rotterdam, 1994, pp. 1125–1131.

[2] Peil U. Assessment of bridges via monitoring. *Structure and Infrastructure Engineering* 2005 **2**:101–117.

[3] Peil U, Mehdianpour M, Frenz M. Fatigue prediction of steel structures by means of monitoring and testing. In: *Life Cycle Cost Analysis and Design of Civil Infrastructure Systems*, Frangopol DM, Furuta H (eds). ASCI-Publ, 2001, pp. 222–238.

[4] Dyrbye C, Hansen SO. *Wind Loads on Structures*. John Wiley & Sons: New York, 1996.

[5] Peil U, Frenz M, Hosser D, Dehne M. Life time estimation of steel structures and assessment of critical details. *Proceedings of International*

Conference of Structural Faults and Repair. London, 2003.

[6] Peil U. Life time prolongation of civil engineering structures via monitoring. *Proceedings of the 4th International Workshop on Structural Health Monitoring*. Stanford, CA, 2003; pp. 64–78.

[7] Opitz H, Quade J, Schwesinger P, Steffens K. *Das Belastungsfahrzeug BELFA für die Tragsicherheitsbewertung von Massivbrücken und Abwasserkanälen*. Bautechnik, 2001.

[8] Belfa—Cooperative research project. http://www.belfa.de.

[9] Peil U, Ruff D. Development of lightweight girders made of sandwich elements. *Proceedings of the 5th Structural Specialty Conference*. CSCE, 2006; pp. 83–89.

[10] Box GEP, Jenkins GM. *Time Series Analysis. Forecasting and Control*. Holden Day: San Francisco, CA, 1976.

[11] Link M, Vollan A. Identification of Structural System Parameters from Dynamic Response Data. *Zeitschrift für Flugwissenschaft und Weltraumforschung, ZFW* 1978 **2**:165–174.

[12] Krätzig WB, Meskouris K, Link M. Baudynamik und systemidentifikation. *Der Ingenieurbau—Baustatik, Baudynamik*. Ernst & Sohn: Berlin, 1996, pp. 365–517.

[13] Ariarathman ST, Schueller GI, Elishakoff I. *Stochastic Structural Dynamics*. Elsevier Applied Science Publishers, 1988.

[14] Keats Wilkie W. Low-cost piezo-composite actuator for structural control application. *SPIE's 7th Annual International Symposium on Smart Structures and Materials*. Newport Beach, CA, 2000.

[15] Heinzmann A, Hennig E, Kolle B, Richter S, Schwotzer H, Wehrsdorfer E. *Properties of PZT Multilayer Actuators: Actuator 2002*. PI Ceramic GmbH: Lederhose, 2002.

[16] Peil U, Loppe S. An approach for monitoring plane structures with low-cost transducers. *Proceedings of the 6th International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2007; 1004–1011.

[17] Andkjær N, Agerskov H, Vejrum T. Fatigue damage accumulation in steel bridges under highway random loading. *Proceedings of the 1st European Conference on Steel Structures Eurosteel 95*. Athens, 1995.

[18] Moses F. Weigh-In motion system using instrumented bridges. *Transportation Engineering. Journal of ASCE* 1979 **105**(TE3):233–249.

[19] Tatsuya O. BWIM systems using truss bridges. In *Bridge Management Four*, Ryall M, Parke G, Harding J (eds). Guildford/Surrey, 2000, pp. 378–385.

[20] Peil U, Mehdianpour M. Life cycle prediction via monitoring. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 1999; pp. 723–730.

[21] Sokolik A. Experimental investigation of traffic load on highway bridges. *Proceedings of the Conference of IABSE*. Copenhagen, 1993; pp. 85–92.

[22] Waubke H, Baumgärtner W. Traffic load estimation by long-term strain measurements. *Proceedings of the Conference of IABSE*. Zurich, 1993; pp. 427–434.

[23] Deichsel G, Trampisch HJ. *Clusteranalyse und Diskriminanzanalyse*. Gustav Fischer Verlag: Stuttgart, 1985.

[24] Peil U, Mehdianpour M, Scharff R. Life time assessment of existing bridges. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. Palo Alto, CA, 2001; pp. 365–383.

[25] Drosner S. *Beitrag zur Berechnung der dynamischen Beanspruchung von Brücken unter Verkehrslasten*, Dissertation. RWTH Aachen, 1989.

[26] Simiu E, Scanlan RH. *Wind Effects on Structures*. John Wiley & Sons: New York, 1986.

[27] Peil U, Telljohann G. A wind turbulence model based on long-term measurements. *Proceedings of the 10th International Conference of Wind Engineering*. Elsevier: Copenhagen, 1999.

[28] Eurocode 1991 Basis of Design and Action on Structures Part 4 Wind Actions.

[29] Troen I, Petersen EL. Department Meteorology and Wind Energy. *European Wind Atlas*. Riso National Laboratory, 1990.

[30] DIN 1055–4, *Action on Structures—Wind Loads*. Beuth Verlag GmbH: Berlin, 2005.

[31] Panofsky HA, Dutton JA. Atmospheric turbulence. *Models and Methods for Engineering Applications*. John Wiley & Sons: New York, 1981.

[32] Bendat JS. *Nonlinear System Analysis and Identification from Random Data*. John Wiley & Sons: New York, 2000.

[33] Davenport AG. Telford Services. The application of statistical concepts to the wind loading of structures. *Proceedings of the Institution of Civil Engineers* 1961 **19**:449–472.

[34] Peil U, Behrens M. Aerodynamic admittance models checked by full scale measurements. *Proceedings of 11th International Conference on Wind Engineering*. Lubbock, TX, 2003; pp. 1769–1776.

[35] Fabian L. *Zufallsschwingungen und ihre Behandlung*. Berlin, NY, Springer-Verlag: 1973.

[36] Peil U, Clobes M. Beschreibung der instationären Übertragung der Windturbulenz mittels transienter Funktionen—numerische Untersuchungen an abgespannten Masten. In: *Praktische Anwendungen in der Windingenieurtechnik*, Band 10, 2007, pp. 73–88.

[37] Peil U, Clobes M. Ersatzlastverfahren zur Böenwirkung auf abgespannte Maste nach E DIN 4131—Validierung und Parameterstudie. In: *Praktische Anwendungen in der Windingenieurtechnik*, Band 10, 2007, pp. 213–228.

[38] Clobes M. *Identifikation und Simulation instationärer Übertragung der Windturbulenz im Zeitbereich*, Dissertation. TU-Braunschweig (Germany), 2008.

[39] Cooley JW, Tukey JW. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computations* 1965 **19**:297–301.

FURTHER READING

Abramov OV. *High Intensity Ultrasonics*. Gordon and Breach Science Publishers: Amsterdam, 1998.

Breitenstein O, Langenkamp M. Lock-in thermography. *Basics and Use for Functional Diagnostics of Electronic Components*. Springer-Verlag: Berlin, NY, 2003.

Chen X, Matsumoto M, Kareem A. A time domain flutter and buffeting analysis of bridges. *Journal of Engineering Mechanics* 2000 **126**:7–16.

Frangopol DM, Peil U. Life-cycle assessment and life extension of structures via innovative methods. *Proceedings of the 3rd European Workshop Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2006; pp. 49–56.

Jones NP, Shi T, Ellis H, Scalan RH. System-identification procedure for system and input parameters in ambient vibration surveys. *Journal of Wind Engineering and Industrial Aerodynamics* 1995 **54/55**:91–99.

Peil U, Nölle H, Wang Zh. Nonlinear dynamic behaviour of guys and guyed masts under turbulent wind load. *Journal of the International Association for Shell and Spatial Structures* 1996 **37**:77–88.

Chapter 8
Static Damage Phenomena and Models

Shankar Sundararaman
Ray W. Herrick Laboratories, Purdue University, West Lafayette, IN, USA

1 INTRODUCTION

Models are used to describe the state of a structural material/component using a mathematical relationship based on observed measurement data and/or physics-based laws. Structural models can be classified as follows: (i) static and dynamic models, (ii) time and frequency-domain models, (iii) linear and nonlinear models, (iv) elastic, viscoelastic, plastic, and viscoplastic models (material-based), (v) micro, meso, and macromechanical models, and (vi) analytical models including three-dimensional elastic, plane strain, and plane stress models, empirical and numerical models such as finite element, finite difference, boundary element, and cellular models. A detailed description of different types of structural models is given in [1].

Damage phenomena can be analyzed using either static or dynamic models. Dynamic damage models trace the growth and retardation of damage. The use of static damage models enables the study of the structural response to damage while freezing variations due to damage growth and/or retardation. By solving the structural inverse problem, measurement data can be used to extract physical and modeling parameters that change due to damage.

Most structural models are based on an assumption of linear elasticity, which indicates that the stress is independent of the loading history [1]. Viscoelastic models are based on a cumulative loading history. Both elastic and viscoelastic models are limited to small deformations. Plasticity theory typically incorporates large deformations, usually past the yield criterion, through the use of a deviatoric stress tensor in addition to elastic volumetric stress, kinetic strain hardening, isotropic strain hardening, and other quantities. Additionally, plasticity theory describes physical mechanisms associated with irreversible behavior

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

observed in real structures, such as loading and unloading conditions leading to different structural states, due to changes in dislocation densities. From a design viewpoint, plasticity is concerned with identifying the maximum load that a structure can sustain without excessive yielding [2]. Material phenomena observed in other types of materials often reside outside the scope of plasticity theories that incorporate plastic anisotropy, martensitic states, time and temperature-dependent plastic strained materials (viscoplastic), elastic hysteresis, and directionality of strain hardening. Some details regarding these concepts are outlined in **Damage Evolution Phenomena and Models** [2, 3].

In the simplest description, the loss of structural integrity and reduction in performance can be identified on the basis of the sensory perception that a detrimental mechanical change has occurred. This perception can be visual/optical (for example, large cracks), sonic (for example, brake squeal), tactile (for example, surface roughness), or olfactive (for example, malodorous gas leaks). However, the concept of structural integrity as diagnosed by these conventional techniques is limited to identifying only when a component has failed or can no longer be reliably used. The need for developing objective indicators to improve safety and reliability of component operation, lower costs, facilitate maintenance scheduling, and monitor inaccessible structural locations resulted in the development of several nondestructive techniques. Physics-based structural models facilitate the continued and iterative development of these techniques by reconciling the input and response parameters from the experimentally observed phenomena. The parameters monitored and the structural model studied is ultimately dependent on the nondestructive technique used for structural diagnostics.

1.1 Static damage models

Broadly, the following macrostructural damage and defect phenomena are of interest:

1. Metals—corrosion, cracks (*see* **Free and Forced Vibration Models**, Figure 21), dents, manufacturing imperfections, residual stresses, inclusions, and fatigue damage (*see* **Free and Forced Vibration Models**, Figure 12);
2. Composites, laminated or otherwise—delaminations, debonds, fiber pullout, fiber breakage, and matrix cracking.

These damage types are typically identified in the literature by monitoring variations in geometrical dimensions, density, elastic modulus, damping, cyclic stress amplitude, wave speed, impedance, microhardness (a quantity based on durability and small-scale shear modulus), and/or electrical resistance changes [3]. Each of these monitored quantities produces different levels of effectiveness in identifying brittle, ductile, creep, and fatigue damage. Surveys of models presented in the literature that describe the cumulative effects of fatigue damage on metals and composites are provided in [4, 5].

It is important to emphasize that "static" damage can be monitored using static analysis methods such as geometrical measurement and microhardness measurements, as well as strain and dynamic analysis techniques such as standing wave/vibration methods and traveling wave methods. The term "static damage" is typically used to refer to electrostatic-induced damage or damage caused by static loads [6]. Here, the term is used in a more general sense to refer to all damage models that do not consider growth and retardation of damage. Static damage models can be classified as geometrical (associated with spatial changes, including, length, width, thickness), material (associated with modulus, density, damping), thermodynamic (humidity, temperature), and electromagnetic (resistivity, permittivity) models. The focus of most structural models is on geometrical and/or material parameters, while the thermodynamic quantities are usually classified as environmental factors. However, there has been increasing emphasis in the incorporation of thermodynamic and electromagnetic factors in structural health monitoring damage models to reduce model uncertainty.

Apart from damage types, damage models are also developed on the basis of the nondestructive testing measure administered to identify damage. Other factors that drive the use of static damage models include the form of the measured data including strain, pressure, and acceleration, and the instrumentation used including strain gauge, accelerometer, laser vibrometer, linear variable differential transformer, rotary variable differential transformer, force transducer, pressure transducer, thermocouple,

Table 1. Schematic of material and geometrical changes associated with different static damage types (adapted from [7])

Damage	Material	Geometrical model	Material model
Corrosion, oxidation	M	Thickness reduction	Density and/or stiffness change
Cracks, notch	All	Moment of inertia, cross-sectional area reduction	Local damping and impedance change
Buckling	All	Moment of inertia, cross-sectional area reduction	Global impedance reduction.
Creep	All	Geometrical distortion	Modulus change
Fastener preload loss; joint loosening	All	Cross-sectional area, length change	Impedance (stiffness)
Plastic deformation	M	Geometrical distortion	Stiffness and mass change
Microstructural degradation	All	Not applicable	Local modulus and damping change
Fiber pullout	LC	Fiber separation (gaps in fiber length)	Damping, stiffness, and density change
Fiber breakage	LC	Fiber distortion (changes in position of fibers)	Damping, stiffness, and density change
Matrix cracking	LC, Ce	Not applicable	Damping, stiffness, and density change
Debonding	Co, LC, Ce	Layer separation (thickness change)	Impedance reduction
Delamination	Co, LC	Local layer separation (thickness change)	Impedance change

M, Metals and alloys; LC, Laminated composite; Ce, Ceramic; Co, Generic orthotropic and composite materials (metallic/ceramic/LC); All = M, LC, Ce, Co.

hygrometer, electromagnetic sensors, X-ray radiographs, infrared cameras, and tomograms. Material models can be further divided into material property models that describe individual changes in physical parameters and impedance models. In all instances, there are changes in the stresses and strains that are manifested as changes in the material property matrices. Some of the methods used for modeling different damage types are described in Table 1. More details of geometrical and material models of some of these damage types are provided in [7]. Although these models provide an idea of the effect of a particular type of damage, the lack of information of modeling parameters because of uncertainties in the material stiffness properties, extent of material anisotropy, spatial spread of defects, and defect shape often results in insufficient modeling accuracy.

Defect types, mechanisms, and static damage models along with suitable references to the literature and some details of the historical background are provided in the remainder of this article.

2 DAMAGE DETECTION USING NONDESTRUCTIVE TECHNIQUES

The development of structural models can be traced to the need for information on the necessary geometrical and material parameters such as stiffness and mass density that ensure safety and structural integrity. A historical background of structural models from the time of Archimedes to the middle of the twentieth century is provided in a review book by

Timoshenko [8]. This book covers development of static and dynamic models associated with elasticity, material fatigue, brittle fracture, bending, elastic stability, impact, and vibrations. More recent overviews of dynamics-based structural diagnostic methods for wave propagation (*see* **Fundamentals of Guided Elastic Waves in Solids**) and vibration (*see* **Free and Forced Vibration Models**) methods are provided in **Damage Measures** and in the literature review articles by Doebling *et al.* [9] and Sohn *et al.* [10]. In general, both static and dynamic-based structural models make use of changes in mechanical properties associated with geometry, stiffness, mass density, and/or damping to model and diagnose damage.

Handbooks on nondestructive testing [11, 12] provide an overview with chronological developments of conventional nondestructive testing measures such as magnetic particle inspection, radiographic testing, ultrasonic inspection, eddy-current testing, thermal infrared, and acoustic emission (AE). In these methods, parametric changes associated with elastic properties (stiffness), tensile properties (yield strength, ultimate tensile strength, fatigue endurance, ductility), hardness, fracture toughness, impact resistance, stress intensity factor, S–N curves, creep range, thermal properties (specific heat, thermal expansion, thermal conductivity, thermal diffusivity), electrical properties (resistivity, permittivity, flux), and optical properties (spectral absorption, refractive index) are monitored [13]. Additionally, loading (*see* **Civil Infrastructure Load Models for Structural Health Monitoring**), environmental (temperature, pressure, humidity), and boundary conditions are also monitored to get a complete picture of structural integrity. A summary of some of the structural and damage models used in nondestructive tests is provided in Table 2. More details with a brief literature review are provided in the subsections that follows the description below.

It is evident that in most of these tools, damage is correlated to a visible change in a geometrical parameter. Most of the other parameters can be directly or indirectly related to the material properties of the structure such as mass, stiffness, or damping. For example, electrical equivalent circuits can be constructed wherein a transmission line model is used to describe electromagnetic wave propagation through layers having different capacitance (equivalently stiffness), inductance (mass), and resistance (damping) [14]. Electromechanical and thermomechanical (*see* **Thermomechanical Models**; **Thermal Imaging Methods**) models rely on the coupling between mechanical quantities with electromagnetic and thermodynamic parameters, respectively.

Table 2. Nondestructive tests, structural and damage models

Nondestructive test	Structural/damage model
Liquid penetrant testing	Chemical, geometry
Electromagnetic testing including eddy current, magnetic particle inspection, magnetic resonance imaging	Geometry, dielectric constant, magnetic susceptibility, charge density, current density, electric conductivity, electric polarization
Radiographic testing and tomography	Geometry, mass density, radioactive decay, wave speed
Modal impact testing	Mass, stiffness, damping, natural frequency, geometry
Ultrasonic testing	Mass, stiffness, damping, structural acoustic wave speed, impedance, coefficient of thermal expansion, geometry.
Infrared and thermal testing	Thermal emissivity, electrical conductivity, geometry, mass, stiffness, damping, impedance.
Leak testing	Geometry

2.1 Radiography

Radiography methods were some of the earliest techniques used for assessment of structural components on an industrial scale. Radiographic testing uses X rays, gamma rays, or neutron rays to inspect components for identifying surface and hidden nonplanar flaws in castings, welds [15], and forgings. Flaws in components are identified on the basis of

differential absorption of radiation and the resultant changes in exposure levels on a radiographic film. The energy absorbed by the component depends on the thickness and density of the material [11, 16]. Typical flaws in a weld [11] that can be evaluated include cracks that appear as dark, irregular, linear segments, lack of fusion that appears as sharp edges, incomplete penetration that appears straight, dark, and linear with sharp edges, inclusions wherein the shade of color is dependent on material properties of the inclusion, and porosity where the defect region is darker than the surrounding material.

2.2 Ultrasonic testing

Ultrasonic testing makes use of concentrated high-energy acoustic waves generated using a pulser–receiver and transducer in frequency ranges typically between 1 and 50 MHz to identify flaw existence and flaw dimensions, determine geometric dimensions, and characterize material properties [17–19]. The ultrasonic transducer can be operated in contact or noncontact mode and in either pulse–echo or through-transmission modes. In addition to the development of transducer and pulser–receiver technology, developments in systems, control, robotics, digitization, and signal processing have facilitated the development of this methodology to active use of ultrasonics in the inspection of multiple, large components. Ultrasonic testing has been used to detect defects in welds [20], inspect railroads [21], identify corrosion and disbonds [22], and estimate material properties [23].

2.3 Acoustic emission

AE testing uses changes in signal strength associated with the sudden release of energy emitted because of a changing stress field ([24]; **Acoustic Emission**). This changing stress field can be attributed to growth in structural damage. Prosser *et al.* [25] used AE to detect damage in cross-ply graphite/epoxy composite test specimens. Wevers [24] noted that AE provided good capabilities to identify fiber breakage, delaminations, matrix cracking, and debonds in a loaded composite as compared to other conventional nondestructive evaluation (NDE) techniques.

2.4 Vibration-based methods

Adams *et al.* [26] identified vibration as a tool for detecting damage in the form of resin-bound shear cracks. The technique was found to be more effective than ultrasonic attenuation or radiographic transmission methods. Cawley and Adams [27] located defects in composite structures using changes in natural frequencies through a detailed sensitivity analysis study. Yao [28] reviewed the techniques used at that time for studying structural damage and reliability assessment. These and other vibration-based techniques primarily relate changes in modal parameters such as natural frequency, mode shape, and damping to the mass, stiffness, and damping parameters associated with a structure.

2.5 Optical methods

Optical methods include approaches such as photoelasticity, holography, and Moirè methods [29]. These methods identify material changes based on variations in transmission intensities and phase changes, diffraction properties, and interference fringe patterns. Structural changes associated with delamination, crack, thermal and residual stress, creep, and fracture can be studied using this approach.

2.6 Thermography

Thermal methods such as thermography [30, 31] are used to detect and measure the infrared range of the electromagnetic spectrum emitted from objects. Structural components and defects have different specific heat constants, thermal diffusivity, and thermal conductivity. This difference is amplified when the components are hotter. Thermal wave propagation methods such as lock-in thermography [32] and pulsed thermography [33] have been used to detect delaminations, corrosion, surface cracks, and voids. Combined approaches such as vibrothermography where thermal wave propagation is used in conjunction with elastic wave propagation, has been used to effectively detect microcracks [34].

2.7 Electromagnetic testing

Electromagnetic induction was developed for railroad inspection in the late 1920s by Drake and Sperry. Electromagnetic techniques [35, 36] make use of induced electric current and/or magnetic field to test the response of a structural component. Eddy-current testing is one of the widely used electromagnetic methods of inspecting structures for surface and near-surface cracks [37], corrosion [38], delaminations [39], and other structural defects. The technique is used to identify defects in electrically conducting materials by correlating the measured impedance with calibrated defect dimensions. In certain cases, a pulsed eddy-current method has been used to speed up the process of structural testing using this technique. Apart from eddy-current methods, other electromagnetic methods [40] have been studied by researchers. Chen *et al.* [41] identified corrosion in pipelines using a magnetoresistance sensor for magnetic flux leakage testing. Lewis *et al.* [35] identified surface-level stress corrosion cracking using an alternating current field measurement technique. Collins *et al.* [42] and Michael *et al.* [43] identified surface cracks using an alternating current potential drop method.

2.8 Magnetic particle inspection

Magnetic particle inspection methods [44] are used to detect magnetic flux leaks from surface and near-surface flaws in ferromagnetic materials. These flux leaks can be traced to permeability variations within test specimens [45]. Magnetoelastic methods such as Barkhausen noise are used for detecting damage through inductive measurements of noise-like signals. Sablik *et al.* [46] modeled a welded specimen to assess creep damage using finite element analysis and identified that the heat-affected zone in welds have reduced permeability resulting in lower electromotive force (emf) than the parent material while creep damage produces additional emf reduction. While most of the magnetic particle methods have focused on ferromagnetic metals, researchers have also identified damage in ceramic and composite composites using nuclear magnetic resonance spectroscopy methods.

3 SCHEMATIC APPROACH TO MODELING DAMAGE

In the context of static damage models, a discussion of damage in the linear elastic regime will suffice because damage evolution (both time and stress state) is frozen and the material state can be considered as piecewise linear in the operating range of loads and induced strains. The dynamics of a structural system with damage, based on the description of Lemaitre [3] and Fritzen [47], among others, can be written as

$$\begin{aligned}\mathbf{M}(\mathbf{x}, t, \mathbf{q_d}, \mathbf{q_e})\ddot{\mathbf{x}} + \boldsymbol{\chi}(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{q_d}, \mathbf{q_e})\dot{\mathbf{x}} \\ + \bar{\boldsymbol{\kappa}}(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{q_d}, \mathbf{q_e})\mathbf{x} + \boldsymbol{\varepsilon}(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{q_d}, \mathbf{q_e}) \\ = \mathbf{F}(\mathbf{x}, t, \mathbf{q_d}, \mathbf{q_e}) \qquad (1)\end{aligned}$$

where $\mathbf{q_d}$ is the material and geometrical damage parameter vector, $\mathbf{q_e}$ is the thermodynamic parameter vector, $\mathbf{F}$ is the external force applied to the system, $\mathbf{M}$ is the mass matrix, $\bar{\boldsymbol{\kappa}}\mathbf{x}$ is the equivalent elastic stiffness term, $\boldsymbol{\chi}\dot{\mathbf{x}}$ is the equivalent linear velocity-dependent damping term, and $\boldsymbol{\varepsilon}$ consists of the residual restoring forces dependent on the time-varying (t) displacement ($\mathbf{x}$) and velocity ($\dot{\mathbf{x}}$) states.

The damage parameter is obtained using a nonlinear differential equation of the form,

$$\dot{\mathbf{q}}_\mathbf{d} \doteq \mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, t, \mathbf{q_d}, \mathbf{q_e}, \mathbf{F}) \qquad (2)$$

where the damage evolution vector is dependent on the states of the system, material and geometrical damage parameters, and the cumulative loading history. Damage usually evolves at different, usually slower, rates than the static/dynamic mechanical processes associated with interrogation using external forces [47]. For this reason, the evolution rates of these parameters are set to zero (for example, $\dot{\mathbf{q}}_\mathbf{d} = 0$). Static damage models facilitate the interpretation of measurement data by providing an understanding of the geometrical and material property variations associated with damage. The variation in the response variables such as displacement $\mathbf{x}$, velocity $\dot{\mathbf{x}}$, and/or acceleration $\ddot{\mathbf{x}}$ for the damaged component with respect to a baseline response variable for the undamaged component is used for diagnosing damage. The response due to the presence of

damage is often identified by computing the difference response, Δy,

$$\Delta y = y_{\text{damage}} - y_{\text{baseline}} \tag{3}$$

where y refers to one or more of the response parameter variables ($\mathbf{x}$, $\dot{\mathbf{x}}$, and/or $\ddot{\mathbf{x}}$).

On the basis of the above description, equation (1) can be simplified to obtain the well-known linear dynamic model,

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{F}(\mathbf{x}, t, \mathbf{q_d}) \tag{4}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ and $\mathbf{K}$ are the linear elastic damping and stiffness matrices respectively, and $\mathbf{F}$ is the external force. Damage is modeled as small changes in the mass ($\Delta\mathbf{M}$), stiffness ($\Delta\mathbf{K}$), and/or damping ($\Delta\mathbf{C}$) matrices.

3.1 Schema for modeling structural damage

The choice of the schema for different modeling approaches is based on the phenomena modeled. A typical schema for model development can be described as

1. identification of the damage mechanism such as cracking and/or corrosion that needs to be diagnosed and the nondestructive methodology such as vibration analysis that is adopted;
2. identification of mechanical (stiffness, density, geometry), electromagnetic and thermal parameters, loading conditions including tensile or compressive loads, and other constraint factors such as boundary conditions associated with the structure of interest (for example, a steel plate specimen of dimensions $1 \times 1 \times 0.02\,\text{m}^3$) in the context of the modeling method (for example, lumped mass and distributed stiffness vibration model);
3. modeling of structure and preliminary model validation through experimental comparisons such as modal testing using analytical or numerical approaches including lumped parameter, Rayleigh Ritz, and finite element approaches;
4. sensitivity study of the model through model variations, including variations in thickness, material properties, and boundary conditions;
5. modeling of the damage mechanism and model validation through experimental comparisons of healthy and damaged specimens;
6. sensitivity study of variations in damage parameters, including material properties such as stiffness and mass and geometrical parameters such as shape, size, and location of defect within the context of the model.

Variations to the above schema abound in the literature. For instance, thermomechanical models such as the techniques described in **Thermomechanical Models** could start out with a description of empirical structural models and progress toward damage identification based on correlation of phenomenological observations with known thermomechanical models; for example, variations in thermal conductivity and thermal dissipativity because of local inhomogeneities in the form of flaws. While it is not necessary to follow each step outlined above for the successful implementation of damage models, it is important to note that these models need to be validated adequately using experimental methods (in the lab and/or field) to improve reliability, iteratively modify model parameters and/or model techniques, identify potential roadblocks, and identify bounds for model applicability.

4 STRUCTURAL DAMAGE PHENOMENA

Models for damage can be determined by either solving the forward problem sequentially in a feedback cycle or by solving the inverse problem and predicting the material, thermal, electromagnetic, and geometrical parameters that affect measured data. Several techniques are available in the literature for estimating the material properties of a structure and that of the damage site. These include vibration-based methods such as embedded sensitivity functions [48], restoring force models [49], and ultrasonic-based [23, 50, 51] material property determination techniques.

4.1 Corrosion

Corrosion is an environment-induced deterioration of a structure that primarily occurs through electrochemical reactions [52, 53]. Corrosion is classified according to geometry as uneven and even

global corrosion (also called *uniform corrosion*) and local corrosion (wide, medium, and narrow pits; crevices in the interior of a structural component). It is also classified according to the forms of corrosion as galvanic corrosion (potential differences between components), high-temperature corrosion, stress corrosion cracking (accompanied by little material loss but with loss of strength), and fretting corrosion. Corrosion can also be classified on the basis of the environmental conditions, which can consist of atmospheric (reinforcing corrosion includes suspended salts and water vapour) conditions, underground (soil) conditions, suspensions in liquid, or high temperatures. The most common form of corrosion is rust, which is caused by the electrochemical oxidation of iron. Corrosion can be investigated using visual, X-ray (geometric details of subsurface flaws), ultrasonic (geometric discontinuities, material impedance variations, which include flaw detection, thickness determination, pulse–echo, transmission, and resonance), electromagnetic (based on electrical and magnetic properties of the fluid), liquid penetrant, and magnetic particle inspection.

Corrosion reactions consist of oxidation (loss of electrons with increase in positive charge) or reduction (gain of electrons with increase in negative charge) reactions involving an exchange of electrons that occur in an electrolyte medium. For instance, metals and alloys corrode when exposed to humidity and oxygen. Corrosion involves the loss of material associated with the metal reacting with hydroxyl ions (or, equivalently water and oxygen) and typically results in the formation of localized regions with loss of material in the form of a pit. Apart from oxygen and water, other favorable environmental conditions that tend to accelerate corrosion rates include the presence of metal cations, urea (and nitrogen oxide groups), sulfide (and sulfur oxide groups), bicarbonate, and halide groups [54, 55], increased temperature, increased pressure (indirect influence through increased solubility of gases in liquids), and the presence of microorganisms. In favorable environments, external factors such as high temperatures and severe tensile stresses tend to accelerate the rate of corrosion. The alkalinity or acidity of the medium surrounding the structural material of interest also influences corrosion. Localized corrosion spots can also lead to changes in stress concentrations resulting in the initiation of cracks.

Researchers have investigated the use of different static damage models for studying corrosion. The electrochemical nature of corrosion suggests that it can be monitored using electrochemical energy methods such as variations in electrical resistivity. Brown and Barnard [56] used a three-dimensional finite difference model based on the electrochemical relations to model localized corrosion in a zinc–aluminum galvanized steel alloy. Terrien *et al.* [57], among others, modeled corrosion as a small notch and studied the interaction of Lamb waves with different types of notch-like defects using a two-dimensional finite element analysis. In fact, corrosion is most often modeled as a mass loss or a geometrical change (for example, thickness reduction). Other aspects, including the energy sink behavior of corrosion spots, can be modeled as localized impedance increases around the corrosion spot (local impedance is calculated as the ratio of input force to the velocity $\mathbf{Z}(\omega) = \mathbf{C} + j(\omega\mathbf{M} - \mathbf{K}/\omega)$) due to increases in damping in the corroded region. Apart from density and geometrical changes, the wave speed across a corrosion spot also changes because of local changes in mass.

There are certain considerations that must be kept in mind when modeling corrosion damage. Some researchers model corrosion as a mass or a density change. It is important to note that a reduction in mass does not translate to an equivalent reduction in density and changes in the two parameters have differing trends [58]. In fact, corrosion models based on density changes need to incorporate changes in both geometry and density. The type of model needs to reflect the response variable being analyzed and the nature of the analysis techniques used. For instance, microhardness measurements require appropriate microstructural models, which consider localized variations in shear modulus, to adequately reconcile the experimental observations. The differences in micromechanic and macromechanic behavior will then need to be suitably incorporated into the model.

4.2 Cracks

Cracks (*see* **Free and Forced Vibration Models**, Figure 21) can be caused by multiple factors ranging from large cycles of loads during normal operation

to spikes in loading cycles due to sudden events in defective materials (for example, built-up residual stresses during manufacturing, the adoption of incorrect manufacturing methods, corroded spots, and/or dislocations). Cracks tend to initiate and propagate as microcracks or dislocation densities (initial flaws) from regions of high stress concentration. Significant focus of the work on crack modeling has been carried out in relation to fracture mechanics with a focus on crack initiation, crack growth, and acceleration of crack growth [59–61]. Most of these methods monitor crack growth associated with plastic deformation at the crack tip and the effects of fatigue loading applied to a cracked structure in the form of cycles of tensile and compressive loading. From the point of view of a static damage model (including propagating and standing wave models), a crack is associated with a localized geometrical discontinuity, mass change, stiffness loss, and/or damping change. Changes in the material and geometrical properties also affect the impedance (including attenuation and amplitude changes) and speed of the propagating waveform. On the basis of the observed changes in one or more of these parameters, the actual changes in physical parameters such as stiffness and damping can be obtained. Several crack models are available in the literature. The reader is referred to **Damage Evolution Phenomena and Models** for additional details.

4.3 Delamination and debonding

Delamination and/or debonding can be caused by the loss of adhesion of adjacent layers of curved sections such as laminated sections of tubular sections and shells, abrupt changes of section properties such as those associated with bonded and bolted joints, residual stress buildup due to differential thermal and hygroscopic (moisture) expansion/contraction rates, imperfect curing procedures during manufacturing, or the application of severe loads during operation [62–66]. The distinction between delamination and debonding is contextual. Delamination or debonding refers to the separation of layers of a layered composite material due to the inability of the specimen to resist the bending stiffness and loading conditions. Debonding can viewed as the surface separation of multilayered components, while delamination can refer to the separation of an internal layer of a multilayered component [53]. Studies of delamination mechanics have typically focused on damage growth models based on the descriptions provided by fracture mechanics models [67–69] such as J-integral [70] and virtual crack closure [71] techniques and damage mechanics models [72] such as cohesive zone models [73]. An example of delamination is illustrated in Figure 1. From the point of view of a static damage model, the delaminated (and debonded) site is modeled as a separation accompanied with an air gap between two material layers. This aspect suggests that there is a loss of mass, stiffness, and damping in the layer of interest and geometrical deformity in the overall structure. A delamination model should, in addition, include bending stiffnesses in a direction transverse to the plane of the layer to ensure that the model is different from that of a damage model for void space in material layers and to take into account the effect of stretched layers.

4.4 Fiber pullout, fiber breakage and matrix cracking

Matrix cracks are frequently observed flaws in composite specimens associated with pit-like indentations along the surface of the specimens [66]. In most

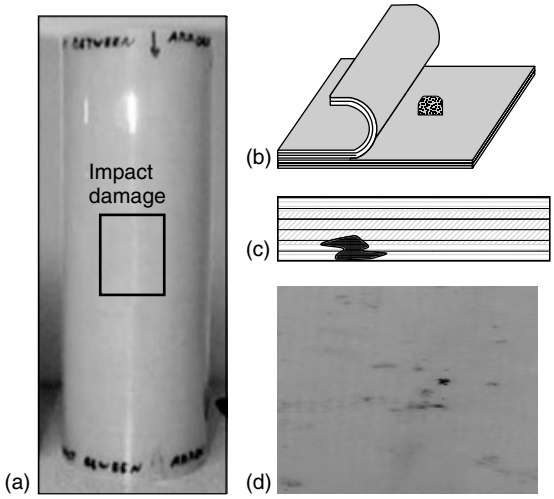


Figure 1. (a) Filament wound composite cylinder with impact damage, (b) laminated composite with delamination seed [Adapted from Purekar, 2006; reproduced by permission of A. Purekar], (c) fiber breakage, and (d) matrix cracking.

instances, matrix cracks are not defects that compromise the structural integrity of the specimen. In laminated composites, impact loading or manufacturing flaws can result in fiber pullout, fiber breakage, and matrix cracking. High tensile loads can lead to fiber separation and result in loss of structural integrity at the fiber pullout location. Fiber breakage (Figure 1c) can also result in similar phenomena except that there is no real separation, but there is, instead, a loss of stiffness along some of the in-plane dimensions (assuming fibers are primarily in one plane). The damage due to matrix cracking, fiber breakage and fiber pullout leads to a modification of the Voight–Reuss mixing criterion and result in a lower matrix or fiber volume and altered stiffness [53].

4.5 Fretting in crevices: slips, loose joints, and fasteners

In nonunitized structures, the more prevalent forms of damage include loosening and cracking of bolted joints and fasteners. The presence of fasteners tends to increase rubbing or fretting during component operation, leading to the formation of slips due to material degradation in the form of a loss of material in threaded joints, erosion, microcracking, loss of preload, etc. The built-up stresses during loading cycles create tension loads on the bolt and compression loads on the joints. Threaded connections can give rise to a number of defects in the joints including thread failures, corrosion, and loose joints because of insufficient torque at the time of threading in the components. Loose joints result in lowered stiffness (translational and torsional) and greater damping at the connection due to the loss in preload and increase in the nonlinear nature of the restoring forces. Damage in the bolt results in a stiffness reduction and can be associated with decreased cross-sectional areas, increased region of contact and an amplification of sliding tendencies of the bolt in the connection. Abrasion or cracking due to fretting/rubbing could also occur in the joints themselves. This aspect can potentially increase tendencies for bolt impacts within the connection joint leading to further bolt and connection damage. The gap between the fastener and the connection can also be modeled as a change in damping. Bolt loosening, loose joints, and damage at fasteners leads to nonlinear changes under loading associated with Coulomb friction [75]. More details about this damage type are discussed in **Free and Forced Vibration Models** [75].

4.6 Creep

Creep is a deformation mechanism associated with the long-term stress accumulation that results in permanent structural deformation. Creep is dependent on time, temperature, and the stress applied and is often associated with structural systems that are subjected to high-temperature loads. Because of the nature of creep and the timescales involved, the effects of creep deformation can best be modeled through dynamic damage models or through the incorporation of residual stresses in static damage models. Additional details are discussed in **Damage Evolution Phenomena and Models**.

4.7 Buckling

Buckling is an elastic instability mode associated with the failure of a material when it is subjected to compressive loads. Buckling is typically accelerated in the presence of material and/or load eccentricity [76]. The prominent forms of buckling include columnar buckling, thermoelastic buckling due to differential heating, and dynamic buckling (including nonlinear snap-through and snapback buckling (Figure 2, [77])). Modeling of buckling requires large displacement analysis and often piecewise linear models are employed to model buckling one load step at a time to avoid numerical instabilities associated with modeling large displacements. For example, finite element models are used to perform a load-controlled step-by-step analysis of models wherein the specimen curvature is changed at the end of each step. Buckling can also be modeled as a reduction in axial direction stiffness of cryogenic vessels that are subjected to permanent damage due to imposed off-axis shock loading [7]. Buckling is also modeled as a small change in curvature (with the change in curvature affecting the natural frequency [78]) and the wave speed [79]. The reader is referred to **Damage Evolution Phenomena and Models** for additional details.

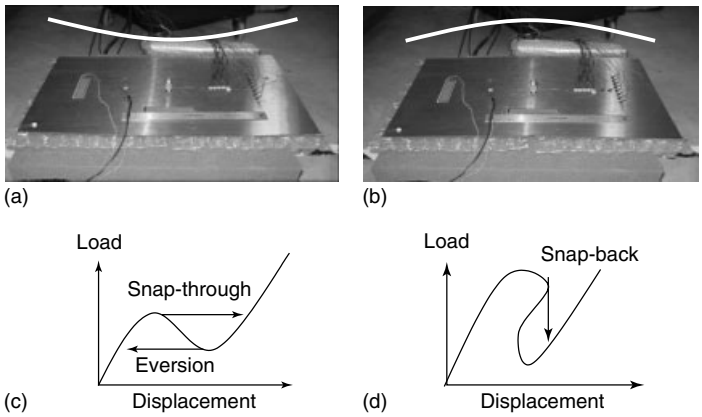


Figure 2. Snapshot of the gamma Ti–Al sheet with the reflected light at the center of the specimen showing the two different states of the specimen: (a) concave up and (b) convex up [77] (c,d) snap-through and snap-back buckling schematic. (adapted from [77]).

4.8 Penetration and plastic deformation

Plastic deformation, as the name suggests, is a large displacement that leads to irreversible material and geometric changes. In the context of a static model, plastically deformed specimens can be modeled as materials with different geometrical and material properties in the localized plastic zone region. Plastic deformation is induced by large acoustic, vibration, or thermal loads such as hail and extreme heat that lead to irreversible changes in the specimen structure. Figure 3 shows an example of this type of damage [80].

4.9 Welds, weld defects

Two metal pieces can be joined together by either riveting or welding. In general, it is preferable to join parts through welding because much of the damage at crevices can be minimized. Welding involves the permanent joining or fusing of two metallic specimens [81]. Welding results in thickness changes at the interface associated with the weld as well as a buildup of stresses at the weld interface and phase-transformations with changes in the grain structure in the welded region (thermomechanically affected and heat-affected zones are created). The changes in phase result in small changes in the overall impedance and more significant changes in the residual stresses associated with the weld formation. Welding also results in geometrical distortion of the welded specimen and increased/decreased volume of material in the weld zone. Additionally, improper weld formations in the form of metallic inclusions, voids, porosities, slag inclusions, spatter, and other defects such as cracks, rough weld formations, under-cut and excessive reinforcements can be introduced during welding because of exposure to the environment and the foreign materials used during welding. The influence of welds and weld defects can be modeled using changes in geometry, stiffness, density, and damping in suitable analytical and numerical analysis methods.

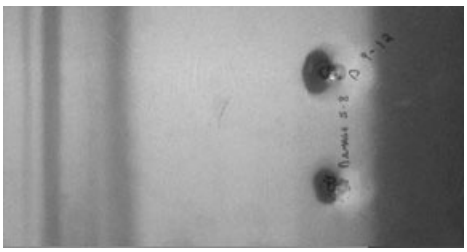


Figure 3. Plastic deformation and penetration resulting from impact damage [White, 2006; Reproduced by permission of J.R. White].

4.10 Residual stresses

The measurement and modeling of residual stresses is difficult. To select the proper residual stress model, measurement data are needed. Dilthey *et al.*

[82] provided a summary of work in the use of finite element simulations for the measurement of residual stresses. In the NDE context, residual stress measurements are evaluated either on the basis of the second-order mean stress effects or through thermal techniques such as thermal evaluation for residual stress analysis (TERSA). TERSA is carried out by heating with a laser, which provides a pseudodynamic signal [83]. The results from this analysis can then be used to incorporate residual stresses in models. Sundararaman and Adams [84] modeled a plate with a center weld and incorporated the changes associated with residual stresses as a small change in the material stiffness and density, on the basis of the assumption that the thickness of the welded region is identical to that of the parent material.

5 STATIC DAMAGE MODELS

The methods used for nondestructive testing can be associated with a change in geometric, material, electrical, and/or thermodynamic parameters. Damage models can make use of multiple studies of healthy and damage cases based on analytical, experimental or numerical (finite difference [85], finite element, boundary element, spectral element [86], and meshless methods [87]) approaches. Detailed literature surveys of nondestructive models that use finite element approaches are provided in [88–90]. Damage can also be studied using discrimination models that focus on differences between a baseline and expected behavior due to the presence of a defect. Examples of such discrimination methods include the expression in equation (3) to the more detailed descriptions presented in **Thermomechanical Models** and [91]. Typically, damage studies make use of a combination of discrimination models in conjunction with structural models of healthy and damage cases.

A diverse set of models discussed in the literature are included in the section below as an illustration of different types of damage models and their applicability to identify different damage phenomena.

5.1 Example 1: modeling cracks using stiffness models

Damage growth models such as linear elastic fracture mechanics and cohesive zone models consider the effects of damage at different points during the cycle of a component and correlate it with models. Cohesive zone models, used for simulating crack initiation and growth in fracture mechanics applications, describe the cohesive forces/tractions that occur when a material is stressed using an approach that deals with the nonlinear zone ahead of the structural discontinuity. For example, the formulation by Yang and Cox [92] is used to describe tractions directly instead of by the use of springs, thereby avoiding potential traction concentrations at mesh discontinuities. In addition to modeling crack growth, these models can also be used to model delaminations and other structural deformations such as microbuckling. While useful information can be gleaned from these growth models for identifying damage, a more realistic approach for modeling static damage is to directly incorporate local changes in stiffness and mass and to employ suitable structural models. Christides and Barr [93] modeled a cracked beam using a reduced stiffness *EI*. The stiffness reduction in the cracked region was expressed using

$$EI(x) = \frac{EI_0}{1 + (I_0 - I_C)I_0/I_C \exp(-2\alpha|x - x_C|/d)} \tag{5}$$

where I_0 and I_C are the second moments of areas associated with the undamaged beam and at the crack, respectively, d is the crack depth, α is an experimentally derived constant and set to 0.667 here, x is the position along the beam, and x_C is the crack position. Alternatively, the region associated with the crack can be modeled using a localized reduction in stiffness extending only to the cracked region or over a triangular region in the localized neighborhood of the crack [94]. Friswell and Penny [95] considered some of the above crack models and compared the model results with experimental vibration data and observed that despite the inherent simplicity of the models, these models adequately modeled the effects of a crack. A typical approach used to identify structural damage at a specific element, j, can be described by studying the kth frequency and can be obtained using [96],

$$\frac{\partial f_k}{\partial \alpha_j} = \frac{1}{8\pi^2 f_k} \cdot \frac{\{\phi\}_k^T [k_j] \{\phi\}_k}{\{\phi\}_k^T [\mathbf{M}] \{\phi\}_k} \tag{6}$$

where, f_k is the kth frequency, $\Delta\alpha_j$ describes the change in a suitable parameter, $[k_j]$ is the stiffness matrix, **M** is the mass matrix, and $\{\phi\}_k$ is the kth eigen vector of the structural system. Data for this approach can be obtained from structural models in standardized finite element packages. Apart from the expression in equation (6), other approaches can also be used to discriminate between baseline and damage states of the structure.

5.2 Example 2: influence of moisture uptake models to crack densities

Patel [97] provided a model to study the moisture uptake of a laminated composite with cracks in multiple directions by combining ideas from continuum damage mechanics and thermodynamics. The model represents an extension of the method presented in [98]. The governing equation for diffusion is described as

$$\frac{\partial m}{\partial t} = \frac{\partial}{\partial X_i}\left(D_i \frac{\partial \mu}{\partial X_i}\right) \tag{7}$$

where $\mu = \rho_s \dfrac{\partial \phi}{\partial m}$ is the chemical potential of the moisture in a polymer, ρ_s is the mass density of the polymeric solid, ϕ is the Gibbs potential, D_i is the diffusivity in a given direction i, m is the mass, t is time and f_i is the moisture flux. The author, in addition, considered the presence of only one nonzero normalized stress. The damage component is based on a damage model for cracks presented by Talreja [99] as

$$d_i = \frac{\kappa_i(m,T) h_c^2 \delta_{ii}}{h}, \quad i = 1, 2 \tag{8}$$

where $i = 1$ corresponds to transverse cracking and $i = 2$ corresponds to longitudinal cracking, $\kappa_i(m,T)$ is an experimental constraining parameter, h_c is the crack thickness, h is the laminate thickness, and δ_{ii} is the crack density in the direction i. The form of Gibbs potential is obtained from Roy [98] and incorporated into the model to obtain the governing equation for an unstressed orthotropic laminate with uniform temperature and damage distributions. The form of the equation is dependent on the parameters described in equations (7) and (8). Additional constants are then evaluated by estimating multiple moisture diffusivities at different states of laminate cracking. Static damage estimates can then be obtained at different points in the usage cycle.

5.3 Example 3: electrical resistivity as a damage discriminator

Irving and Thiagarajan [100] and Chung [101] used an electrical potential technique to characterize fatigue damage in carbon fiber composite materials. In the electromechanical analogies presented in the literature [14], resistance, capacitance, and inductance are the corresponding analogs to damping, compliance, and mass, respectively. Any structural changes associated with damage in the fiber or the matrix results in a loss of conduction and an increased resistance measure. Damage is indicated by an increase in resistivity. This method is used for identifying small damage of the dimensions of the fibers in a laminated composite. The resistivity can be described using

$$\rho = RA/L \tag{9}$$

where R is the sample resistance, A is the cross-sectional area, and L is the sample length. Irving and Thiagarajan [100], in addition, identified that the longitudinal resistivity was an order of magnitude higher than the transverse resistivity.

5.4 Example 4: thermoelastic models

Schmidt and Hattel [102] used a three-dimensional thermoelastic finite element model to study friction stir welds. The elastic portion of the model comprised of mass density, ρ, damping, c, and stiffness, k, components as described in the equation,

$$\rho\ddot{x} + c\dot{x} + kx = F \tag{10}$$

where F is the body force and x is the displacement. The thermal response is expressed using the thermomechanical diffusion equation,

$$\rho c_P \dot{T} = (kT_{,i})_{,i} + \eta s_{ij} \dot{\varepsilon}_{ij}^{pl} \tag{11}$$

where T is the temperature, c_P is the specific heat capacity, k is the thermal conductivity, η is the plastic

energy dissipation fraction, s_{ij} is the deviatoric stress tensor, and $\dot{\varepsilon}_{ij}^{pl}$ is the plastic strain rate tensor. The surface flux, $q_{\text{surf}} = \mu p \dot{\gamma}$, where μ is the friction coefficient, p is the pressure, and $\dot{\gamma}$ is the slip rate, is used to describe the contact boundary condition at the surface nodes. Suitable expressions are then used to appropriately discretize equations (10) and (11). This model can then be used to monitor the deformation field at different points in time during the welding process. For instance, Schmidt and Hattel [102] monitored the formation of defects such as voids for different parametric states. Note that this model is a damage evolution model and to glean static damage results, time-frozen snapshots need to be captured to evaluate the damage state of the structural component.

While the above examples provide a flavor for different damage models that are used to diagnose different sets of damage, they are by no means complete. A detailed description requires a full rendition of the sequential steps outlined in the schema for damage models described previously in this article. A case study in the form of a sensitivity analysis is presented in the next section involving the use of a numerical model based on a wave propagation model.

6 CASE STUDY: SENSITIVITY ANALYSIS OF DAMAGE PHENOMENA

6.1 Numerical modeling using a wave propagation model

The focus of the remainder of this article will be to study the sensitivity of model parameters to varying material (mass, spring, damping) and geometrical parameters. These parameters, either separately or in combination, are associated with each of the damage types listed in this article. In this article, component responses are simulated using static damage models based on changes in density, stiffness, damping and/or geometrical variations at macrostructural dimensions (>1 mm; >5% variations in density, stiffness, damping). Specifically, propagating elastic waves are simulated using a local interaction simulation approach (LISA)/sharp interface method (SIM) [84, 103–109]. The LISA/SIM has been developed for modeling propagating waves using a cellular model approach to obtain iterative difference relations and is used to illustrate the interaction of interrogating signal waveforms with different levels of changes in geometry, stiffness, and density in thin plate specimens. In these specimens, elastic bulk waves interact through multiple reflections between the top and bottom surfaces of plate specimens resulting in guided Lamb waves (*see* **Fundamentals of Guided Elastic Waves in Solids** for additional details).

6.2 Implementation

The implementation of numerical models requires an assessment of the model development (description of suitable material and geometrical models, environmental and operating conditions, damage model, and sources of input), study of model limitations (stability, accuracy, and convergence), and model validation (analytical and experimental validation). The three-dimensional linear elastodynamic equation for the propagation of bulk waves in orthotropic, inhomogeneous media is the basis for the simulation model used in this section. The equation of motion can be described as

$$\partial_l (S_{klmn} w_{m,n}) + F_k = \rho \ddot{w}_k + \chi_k \dot{w}_k \qquad k, l, m, n = 1, 2, 3 \tag{12}$$

where $S_{klmn}(x_1, x_2, x_3)$ is the stiffness tensor, $w_k(x_1, x_2, x_3, t)$ is the kth displacement, $F_k(x_1, x_2, x_3)$ is the body force applied at a local point, $\rho(x_1, x_2, x_3)$ is the material density, $\chi_k(x_1, x_2, x_3)$ is the material attenuation factor based on a proportional, viscoelastic damper, and the comma in $w_{m,n}$ indicates differentiation. In the absence of a stress field (for example, a stress field resulting from residual stresses), $S_{klmn}(x_1, x_2, x_3)$ represents the second-order elastic constants tensor. For an orthotropic solid, the stiffness tensor can be written as

$$S_{klmn} = \begin{pmatrix} \sigma_1 & \lambda_{12} & \lambda_{13} & 0 & 0 & 0 \\ \lambda_{12} & \sigma_2 & \lambda_{23} & 0 & 0 & 0 \\ \lambda_{13} & \lambda_{23} & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_{13} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_{12} \end{pmatrix} \tag{13}$$

where λ_{pq} and μ_{rs} are the Lame constants, σ_p is the bulk stiffness, and the subscripts p, q, r, and s vary between 1 and 3 and are indicative of the stiffness dimension. Each pair of indices (k, l) and (m, n) in equations (5) and (6) can be replaced by a single index based on the Voight–Kelvin convention,

$$V(i, j) = i\delta_{ij} + (1 - \delta_{ij})(9 - i - j); \quad i \to k, m; \;\; j \to l, n \tag{14}$$

Propagating waves are modeled by using a discrete form of equation (3) and solving for the difference relationships using a LISA in conjunction with the SIM. The iterative difference relations are obtained using the method outlined in [105]. The final difference relation form is obtained for a rectangular cuboid grid. It was assumed that the displacements and the material properties of the cell are continuous although the adjacent cells at an interface can have significantly different material properties. Note that both displacement and stress continuity are maintained at these interfaces in the LISA/SIM technique. The three-dimensional relations at each point $P(i, j, k)$ are obtained for a rectangular cuboid grid. Each displacement degree of freedom is individually determined based on the stiffness, density, and geometrical properties of the neighboring grids using an iterative difference relation. A typical iterative difference relation is given by

$$u^{i,j,k,t+1} = -\bar{q}_1^2 u^{i,j,k,t-1} + 2\bar{q}_1 u^{i,j,k,t} + \frac{\bar{q}_1 \Delta t^2}{8\bar{\rho}} \times \sum_{a,b,c=\pm} \begin{bmatrix} -2u^{i,j,k,t}\left(\tilde{\eta}_1^2\tilde{\sigma}_1 + \tilde{\eta}_2^2\tilde{\mu}_{12} + \tilde{\eta}_3^2\tilde{\mu}_{13}\right) \\ + 2\tilde{\eta}_1^2\tilde{\sigma}_1 u\,(1^a) + 2\tilde{\eta}_2^2\tilde{\mu}_{12} u\,(2^b) \\ + 2\tilde{\eta}_3^2\tilde{\mu}_{13} u\,(3^c) \\ + ab\tilde{\eta}_1\tilde{\eta}_2\big[\left(\tilde{\lambda}_{12} + \tilde{\mu}_{12}\right)\left(v(6_b^a) - v\right) \\ + \left(\tilde{\lambda}_{12} - \tilde{\mu}_{12}\right)\left(v(2^b) - v\,(1^a)\right)\big] \\ + ac\tilde{\eta}_1\tilde{\eta}_3\big[\left(\tilde{\lambda}_{13} + \tilde{\mu}_{13}\right)\left(w(5_c^a) - w\right) \\ + \left(\tilde{\lambda}_{13} - \tilde{\mu}_{13}\right)\left(w(3^c) - w(1^a)\right)\big] \end{bmatrix} \tag{15}$$

where $a, b, c = \pm 1$, and $\bar{a}, \bar{b}, \bar{c} = -1$ when $a, b, c = -1$ and $\bar{a}, \bar{b}, \bar{c} = 0$ when $a, b, c = 1$, $\bar{q}_k = (1 + \Delta t/2 \sum_{a,b,c} \tilde{\chi}_k/\tilde{\rho})^{-1}$ is the transmission factor, $\bar{\rho} = \frac{1}{8}\sum_{a,b,c} \tilde{\rho}(a, b, c)$, (u, v, w) are the displacements in the three directions (x, y, z), $\tilde{\sigma}_k = \sigma_k(a, b, c)$, $\tilde{\eta}_1 = \eta_1(a, b, c) = 1/\Delta x$, $\tilde{\eta}_2 = \eta_2(a, b, c) = 1/\Delta y$ and $\tilde{\eta}_3 = \eta_3(a, b, c) = 1/\Delta z$. Similar expressions hold for other quantities with a tilde.

The discretization parameters for the numerical model are obtained by ensuring that the minimum criterion for stable solutions is met. This stability criterion, referred to as the Courant–Friedrich–Lewy (CFL) criterion, constrains the grid spacing for a given time step by

$$\text{CFL criterion}: c_{\max}\Delta t\sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2}} \leq 1 \tag{16}$$

where $c_{\max}$ is the maximum wave speed (usually the longitudinal wave speed), Δt is the time step, and Δx, Δy, and Δz are the spatial steps in the three Cartesian coordinates x, y, and z.

6.3 Model validation

A broadband rectangular pulse with a 2-MHz signal bandwidth was imparted at the center of a $609.6 \times 609.6 \times 2\,\text{mm}^3$ plate with grid dimensions of approximately $1 \times 1 \times 1\,\text{mm}^3$ using the three-dimensional LISA model for a homogeneous, isotropic aluminum plate. An aluminum plate structure with elastic modulus 70 GPa, density $2700\,\text{kg m}^{-3}$, and Poisson's ratio 0.334 was modeled. Data was collected at 608 nodal locations approximately 1 mm apart along the center of the plate structure. The numerical results obtained were then compared with the results obtained analytically from the Rayleigh–Lamb dispersion curves. The results in Figure 4(a) indicate a match between the modes obtained numerically and analytically. The lines that emanate backwards and forwards from the right edge in Figure 4(a) are spatially aliased segments of the different dispersion curves.

Next, a plate structure of dimensions $530 \times 306 \times 6.5\,\text{mm}^3$ was discretized into grids of dimension $1 \times 1 \times 1.625\,\text{mm}^3$. Dispersion curves were first obtained numerically to verify the material properties obtained through experimental analysis. A sinc pulse was used as the input in both the experimental and numerical

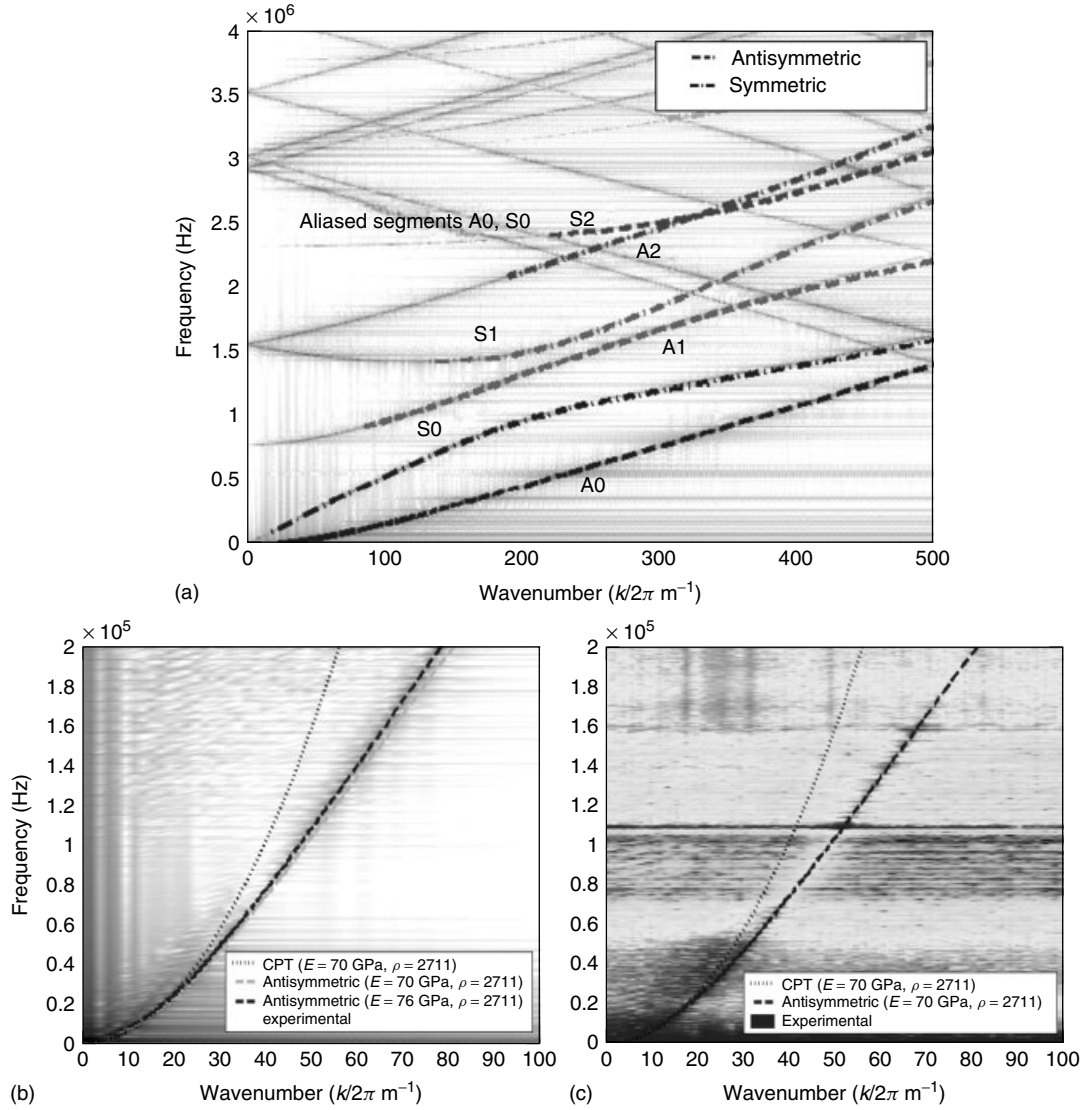


Figure 4. Comparisons of analytical Rayleigh–Lamb dispersion curves with (a) numerical dispersion curve results for a $609.6 \times 2\,\text{mm}^2$ aluminum section with approximate grid spacing of $1/8 \times 1/8\,\text{mm}^2$, (b) numerical dispersion curve results for 1-mm sensor spacing in a $530 \times 306 \times 6.5\,\text{mm}^3$ aluminum section with grid spacing of $1 \times 1 \times 1.625\,\text{mm}^3$ and (c) experimental results for 5-mm sensor spacing.

studies. The sampling frequency was set on the basis of the CFL criterion as 10 MHz and the signal was input at the left edge and center of the plate specimen (5, 153, 0) mm. The results obtained were then compared with the results obtained analytically from the Rayleigh–Lamb dispersion curves and the classical plate theory. The results indicate that a match is obtained between the antisymmetric (A_0) mode and the experimentally measured dispersion curves. However, it is seen in Figure 4(b) that the estimated stiffness is about 10% higher than the modeled stiffness. This result suggests that the coarse nature of the grid structure results in grids with higher stiffness. The experimental results for the edge of the aluminum

Table 3. List of specimens used

Plate	Al-1	Al-2	Or-1	LC-1
Material	Aluminum	Aluminum	Orthotropic	Laminated orthotropic
Dimensions (mm)	$300 \times 300 \times 2$	$400 \times 200 \times 4$	$450 \times 250 \times 3$	$400 \times 200 \times 4$
Grid (mm)	$1 \times 1 \times 0.5$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$
Density ($kg\,m^{-3}$)	2710	2710	1575	1600
Stiffness (GPa) (along 0°)				
E	70	70		
σ_1	104.5	104.5	85.9	155.43
σ_2	104.5	104.5	85.9	16.34
Σ_3	104.5	104.5	16.34	16.34
Λ_{12}	52.1	52.1	3.72	3.72
Λ_{13}	52.1	52.1	6.5	3.72
Λ_{23}	52.1	52.1	6.5	4.96
μ_{23}	26.2	26.2	6.59	3.37
μ_{13}	26.2	26.2	6.59	7.48
μ_{12}	26.2	26.2	7.48	7.48
Laminate	n/a	n/a	n/a	$[0_{10}/90_{10}]_S$
Sampling frequency (MHz)	20	10	10	20
Data length	8000	4000	4000	8000
Signal (kHz)	80	40	160	40, 150, 300
	(150, 5, 0); (5, 295, 0)	(200, 5, 0)	(225, 245, 0); (5, 125, 0)	(200, 5, 0)
Actuator (mm)	8-mm diameter, modeled as a Gaussian pulse in space.			
	Hanning window modulated 5 wave sine wave (model in time).			
Input signal	Directed in the in-plane directions.			

plate yielded similar results and, in this instance, matched the analytical results (Figure 4c). These comparisons indicate that the numerical methodology can be used to model real specimens.

6.4 Damage simulation studies conducted

The damage simulation studies were conducted on four different specimens. Elemental cells were defined on the basis of the material properties listed in Table 3. Additional grid points equivalent to one extra cell surrounding the entire medium were defined with the material properties of air (density $1.3\,kg\,m^{-3}$ and stiffnesses of 1/10 000th of that of the surrounding medium) to denote a traction-free boundary. The actuator was modeled such that a Gaussian pulse was sent into the plate structure over a diameter of 8 mm. The sensor was modeled to extract a response measurement as an output average of all points over a diameter of 8 mm. The sampling frequency was chosen based on the CFL criterion (equation 16). A listing of the classes of damage studied is provided in Table 4.

The damage sites in the studies described in this work were identified by using the difference data as calculated in equation (3) because the damage sites involve localized changes in stiffness, density, and damping with dimensions smaller than a wavelength of the interrogating waveform.

6.5 Discussion

An excitation was introduced at "Actuator-1" in a plate with dimensions $300 \times 300 \times 2\,mm^3$ (Al-1).

Table 4. Parametric sensitivity analysis study of damage in specimens

Figure	Description
	Al-1: $300 \times 300 \times 2\,mm^3$ aluminum plate with $1 \times 1 \times 0.5\,mm^3$ grids
6	Baseline studies.
7	[Cases 1–2]—multiple-site damage at two actuator locations D1 (150, 79, 0); D2 (80, 157, 0); D3 (248, 248, 0) mm. Changes in stiffness (D1: +10%, D2, D3: 0.01%) and change in density (D1: +5%, D2: 0.05%).
8	[Cases 3–8]—superposition of the above multiple-site damage at two actuator locations
9(a, b)	[Cases 9–12]—surface damage: damage size study (length: 3–9 mm in steps of 2 mm × width: 3 mm × thickness: 1 mm); location (60, 125, 0.5) mm; surface notch modeled with material properties of air.
9(c, d)	[Cases 13–16]—interior damage: change in stiffness and/or density; location (60, 125, 1) mm; 1.2 × stiffness, 0.8 × stiffness, 1.2 × stiffness 0.9 × density, 1.3 × density.
	Al-2: $400 \times 200 \times 4\,mm^3$ aluminum plate with $1 \times 1 \times 1\,mm^3$ grids
7	[Case 25]—multiple site damage at (170, 79, 0); (80, 157, 0); and (318, 108, 0) mm grids.
10	[Cases 17–19]—through the thickness damage: damping change at (240, 125, 2) mm (cases 1–3: 0.95, 0.95, 0.9 transmission factors as described in equation (15)); Case-2 damping study also includes 0.9 × stiffness and 0.9 × density.
	Or-1: $450 \times 250 \times 3\,mm^3$ orthotropic plate with $1 \times 1 \times 1\,mm^3$ grids.
11	[Cases 20–21]—multiple site surface damage at two actuator locations (140, 104, 0); (250, 25, 0) and (340, 175, 0) mm.
	LC-1: $400 \times 200 \times 4\,mm^3$ laminated composite plate with $1 \times 1 \times 1\,mm^3$ grids
12	[Cases 22–24]—impedance change at three frequencies (40, 150, and 300 kHz; single site subsurface damage with material properties of air) (240, 125, 1) mm.

The out-of-plane displacement responses at different sensor locations are shown in Figure 5(a). Additionally, the x-, y-, and z-displacement responses at one sensor location (200, 5, 2) mm are shown in Figure 5(b). The incident and reflected waveforms are clearly seen in the out-of-displacement response pattern, while it is not easy to discern the different waveforms from the in-plane displacement profiles.

A second aluminum plate with dimensions $400 \times 200 \times 4\,mm^3$ (Al-2) was also studied. The response at 120 μs captured for Al-1 (Figure 6a) is also shown for Al-2 (Figure 6d). The differences in the two profiles were observed here. The effects of multiple-site damage centered at three locations (Cases 1 (Al-1), 25 (Al-2); Table 4), interrogated using the actuator at (150, 5, 0) mm (Al-1) and (200, 5, 0) mm (Al-2) and after a difference response has been computed, are shown in Figure 6(b), (c), (e), and (f). Local changes in stiffness and density were expected to result in corresponding changes in the impedance and wave speed, resulting in waveform distortion. Figures 7(b), (c), (e), and (f) have directional conical spread signal patterns rather than the expected circularly symmetric patterns. This aspect can be traced to the dependence of a signal response at a particular point in the spatial domain on the original signal's response profile at that instant. In other words, at 120 μs, the original signal (Figure 6a, d) was directed toward the top boundary with some reflected profiles radiating in from the left- and right-hand boundaries. The difference response shares this pattern and was primarily illuminated along the excitation leading to directional difference profiles. Another reason for the lack of circular symmetry can be traced to the effects of discretization. Only five grid points were used along the thickness dimension in the simulation because of computational restrictions (runtime, memory, hard-disk space) even though the preferred number of grid points along the thickness dimension is between 8 and 20. Any reduction in accuracy is a

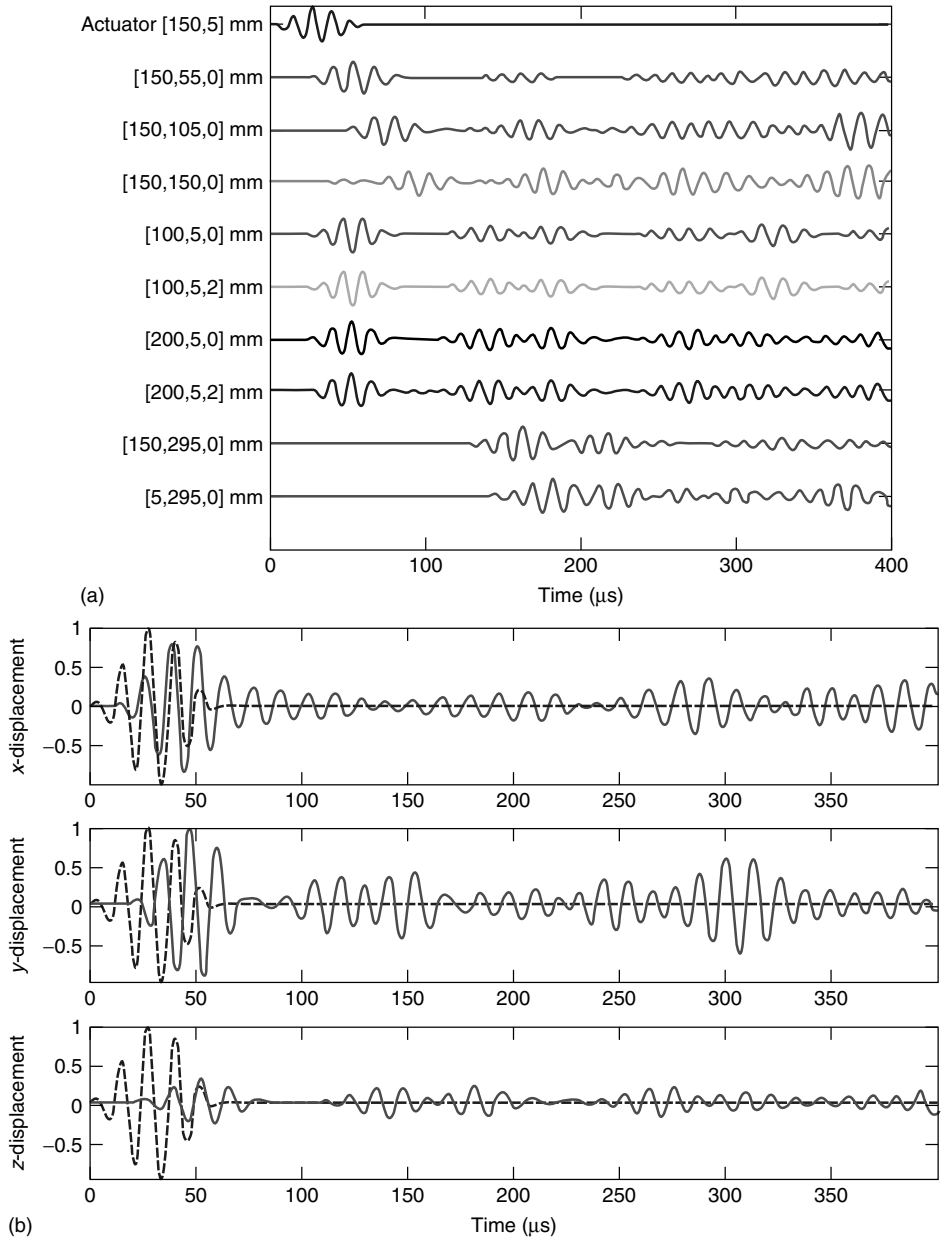


Figure 5. Displacement profiles at (a) different sensor locations and (b) at (200, 5, 2) mm in a $300 \times 300 \times 2\,\text{mm}^3$ aluminum plate subjected to a Hanning window excitation.

trade-off for modeling large specimens using three-dimensional analysis in serial computers.

The three damage sites considered in Figure 6 were then modeled separately and the difference patterns were recomputed for each case. The sensed response shows visible differences for the different types of damage in the plate. The patterns observed in Figure 7(a) indicate that the difference responses for each damage site were distinct and that the damage site closest to the actuator responds first followed

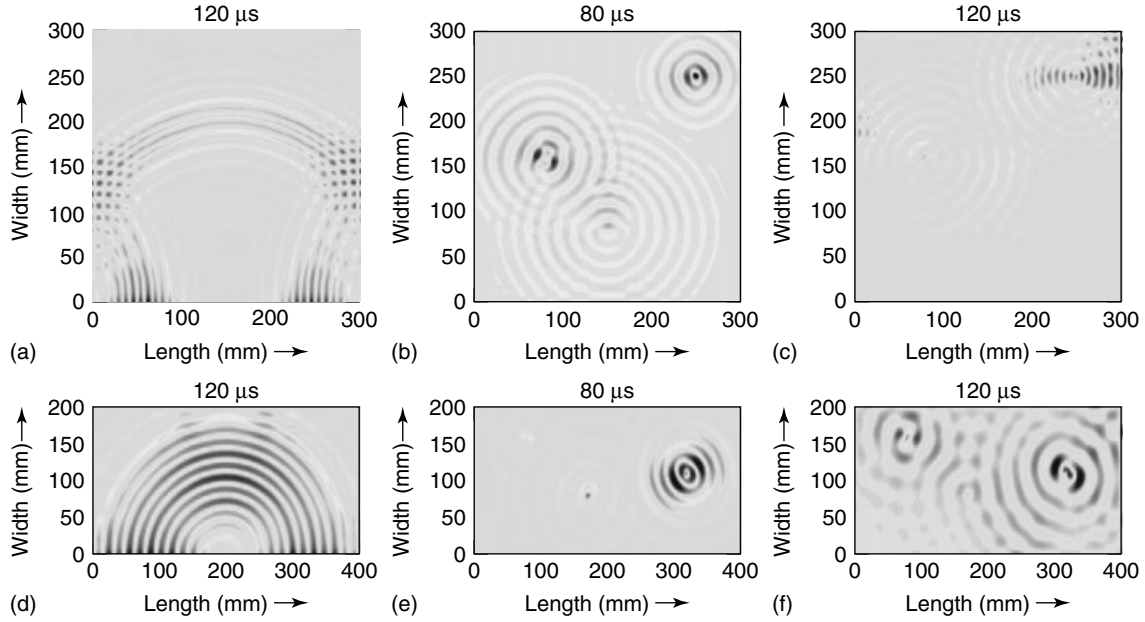


Figure 6. Damage identification through difference data. Time slices in $300 \times 300 \times 2\,\text{mm}^3$ (80 kHz) and $400 \times 200 \times 4\,\text{mm}^3$ (40 kHz) aluminum plates. (a, d) baseline data at 120 μs, (b–f) difference data at 80 and 120 μs.

by other locations in increasing distances from the actuator location. The relative amplitude levels of individual damage sites were dependent on the extent of damage and the closeness of the damage site to the actuator and sensor locations and the boundaries of the specimen. The differences from the three damage sites were summed and the results were compared with the multiple-site damage study in Figure 6. When the difference results were scaled by a factor of 10, it was found that the results in Figure 7(b) were significantly higher. This shows that the presence of any imperfections in a specimen makes it difficult to identify additional damage sites.

The peak difference response (both out of plane and in plane) at two sensor locations were then studied for all four cases (three damage sites individually and multiple-site damage instance) and the results are shown in Figure 7(c) and (d). The out-of-plane damage site showed the anticipated linear trends at both sensor locations. However, the characteristics of the in-plane difference responses showed slightly different trends. For instance, the x-displacement response showed a trend reversal for the multiple-site damage instance from the single site damage at D3 (Figure 7d). This aspect is seen because the difference response of the cumulative damage from the multiple damage instances is less than that of the single damage site. However, the constructive and destructive influence of signals from the damage sites (treating damage as a secondary source) were not considered, nor was the choice of the damage metric (here, peak amplitude), the location of the sensor, or the cumulative sum of the displacements in all three dimensions (both y- and z-difference responses are significantly higher). These aspects need to be kept in mind when using sensor responses to correlate with damage.

To study the influence of the size of damage, a rectangular damage region was introduced in plate Al-1 and the local stiffness and density were modified. The results were then computed for both in-plane and out-of-plane displacements. All three displacement components showed linear increases in damage with increasing size of damage (Figure 8a, b, Cases 9–12). The slope of this trend was higher for the x-displacement than either the y- or the z-displacements. This result was obtained because the damage site varied linearly along the x dimension. In

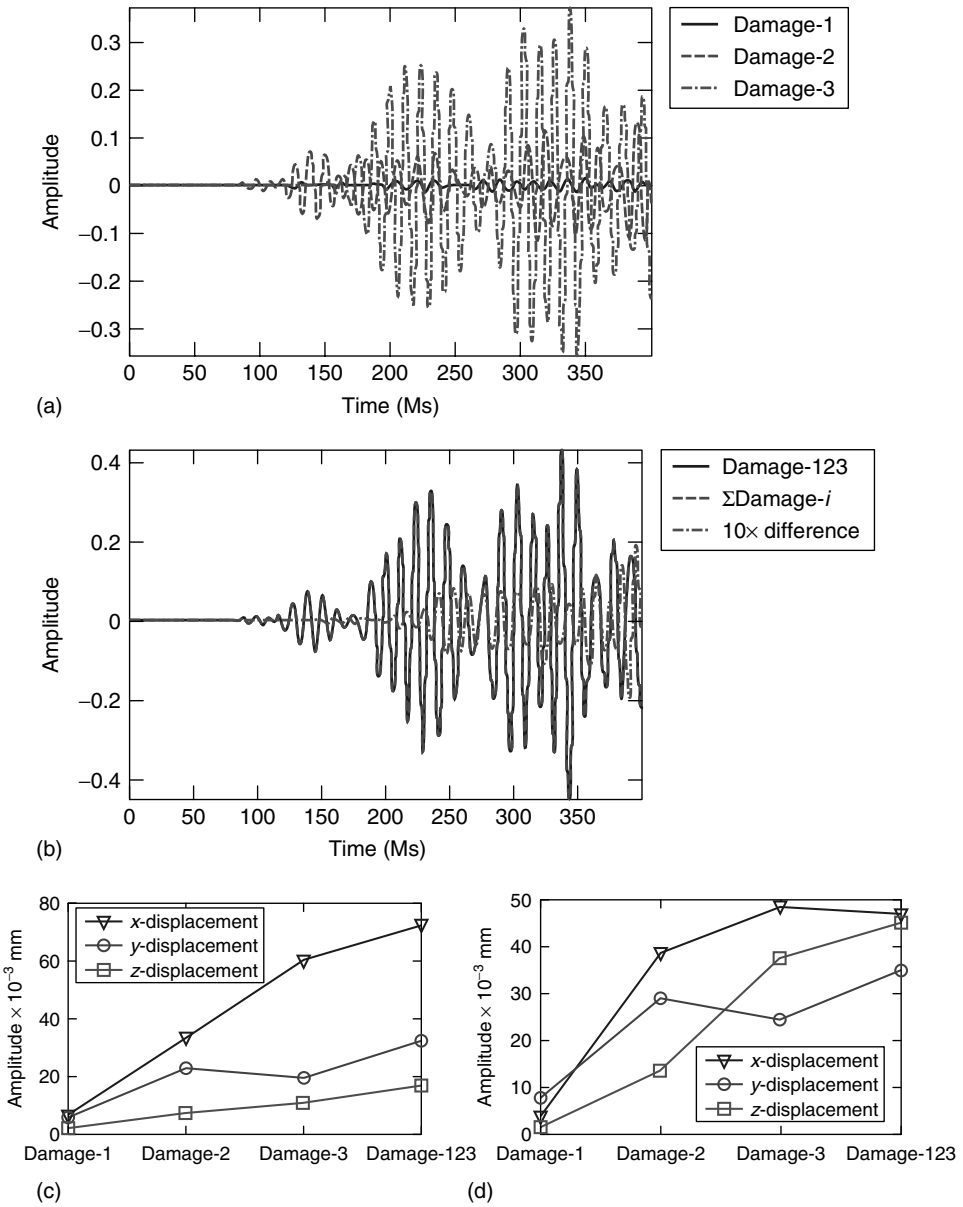
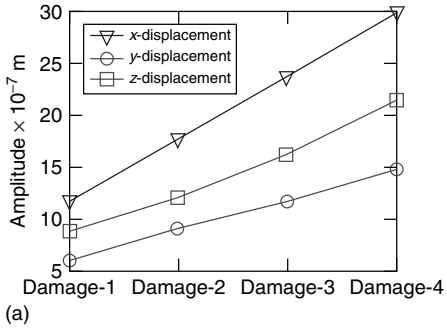


Figure 7. Single site damage for $300 \times 300 \times 2\,\text{mm}^3$ isotropic plate studying the superposition effects of damage identification using (a, b) an out-of-plane displacement comparison, (c, d) peak in-plane and out-of-plane displacements; (a, b, c) sensor at (295, 5, 0) mm and (d) sensor at (5, 245, 0) mm; actuator at (150, 5, 0) mm.

this instance, the study was carried out at a sensor location, which is positioned along different x and y dimensions than the actuator and damage locations. In the study by Sundararaman and Adams [107], the sensor was placed along the same y dimension as the actuator, resulting in negligible variation in the peak y-displacement response with increases in damage along the x dimension.

Next, the influence of changes in local material properties was studied for a given size of

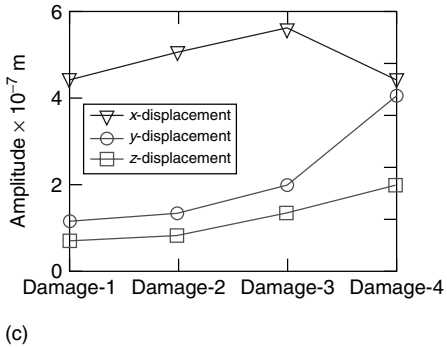


(a)

Location (60,125,0.5) mm
Stiffness 0.01% medium
Density 0.1% medium

Damage	Dimensions (mm)
1	(3,3,1)
2	(5,3,1)
3	(7,3,1)
4	(9,3,1)

(b)



(c)

Location (60,125,0.25) mm
Dimensions: (9,3,0.5) mm

Damage	Stiffness	Density
1	120%	Same
2	80%	Same
3	120%	90%
4	Same	130%

Change in material properties in relation to the rest of plate.

(d)

Figure 8. Peak in-plane and out-of-plane displacement comparisons in a $300 \times 300 \times 2\,\text{mm}^3$ aluminum plate for (a) impedance study, (b) damage sizing study. (c,d) Tables of parameters used for generating the results displayed in Figure 9 (a,b) respectivrly. Sensor at (5, 245, 0) mm; actuator at (150, 5, 0) mm.

damage (Figure 8c,d, Cases 13–16). The peak in the difference in displacement response indicated increasing trends for the damage sequence described in Figure 8(d) for the y- and the z-displacements. On the other hand, the response in Figure 8(c) for the x-displacement indicated a decrease in the amplitude of the peak difference displacement for the last damage case. This result is similar to the result in [107] and highlights the influence of sensor location on the ability to discern actual trends associated with a given damage site. Overall, the net change in displacement patterns suggested that an increase in density leads to a greater change than a similar percent increase in stiffness. The results in Figure 8(a) and (c) also indicated that the in-plane displacements were dependent on changes in the wave speed although the out-of-plane displacement was strongly dependent on the mechanical impedance (product of the density and the wave speed).

Next, to study the effect of damping, the local transmission factors were modified in a $400 \times 200 \times 4\,\text{mm}^3$ aluminum plate. It was evident that a change in the transmission factor resulted in a significantly higher amplitude change than changes associated with stiffness and/or density. Two transmission factors were studied (Figure 9, Cases 17–19 in Table 4) with one of the transmission factors also accompanied by a decrease in both stiffness and density. It was observed that the net decrease in both stiffness and density accompanied by a decrease in the transmission factor resulted in an overall decrease in the peak difference amplitude for both sensor locations. In general, decreasing the transmission factor drastically altered the peak difference amplitude response.

The effects of damage identification in an orthotropic plate ("Or-1") of dimensions $450 \times 250 \times 3\,\text{mm}^3$ involving changes in unit cell properties as described in Tables 3 and 4 were also studied. The results from actuators located at (5, 125, 0) mm and (225, 245, 0) mm are shown in Figure 10(a)–(c) and Figure 10(d)–(f) respectively. The in-plane difference waveforms had highly directional responses with elliptical patterns, while the out-of-plane difference waveforms exhibited

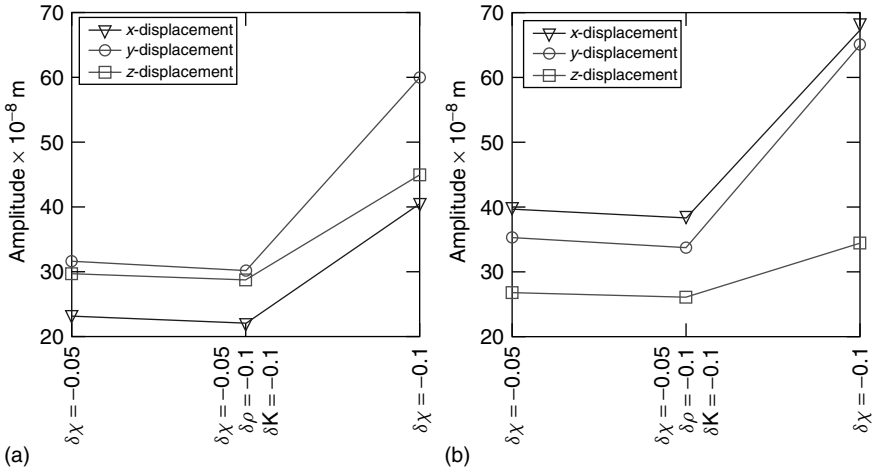


Figure 9. Peak in-plane and out-of-plane displacement comparisons in a $400 \times 200 \times 4\,\text{mm}^3$ aluminum plate for damping transmission factor variations with actuator located at (200, 5, 0) mm and sensors located at (a) (395, 5, 0) mm and (b) (200, 55, 4) mm.

circular patterns. The results described in Figure 10 were further exaggerated because of the grid profile chosen with no grid description at the center of the plate specimen. In general, it is not advisable to model guided waves without a center grid point; here, the results were used more for illustration of an exaggerated case of anisotropy. While the directionality observed was not surprising, the out-of-plane displacement response indicating circular profiles was somewhat counterintuitive. Within individual circular and elliptical profiles, the amplitude levels along different directions were different, depending on the location of the source signal at a certain point in time.

Finally, a laminated composite plate ("LC-1") of dimensions $400 \times 200 \times 4\,\text{mm}^3$ involving changes in unit cell properties as described in Tables 3 and 4 was studied. Excitations at three different signal frequencies (40, 150, and 300 kHz) were studied to investigate the influence of different modes. Single site damage was introduced into the plate specimen with no additional change introduced due to damping and involved subsurface damage in the form of an air gap between layers to mimic a separation mechanism such as delamination. It was observed that changes in signal frequency resulted in significant increases in the peak difference amplitude response (Figure 11a–c) along all three displacements. The rate of increase in the peak difference amplitude was higher for the out-of-plane displacements than the in-plane displacements. Additionally, out-of-plane time slices were also displayed as shown in Figure 11(d)–(f). The initial wavefront with slight directionality at 40 kHz gave way to circular wavefronts at 150 kHz and circular wavefronts with more complex patterns at 300 kHz. This aspect indicated that for subsurface defects, the higher frequencies (with more complex mode behavior) were more effective for damage identification.

7 SUMMARY

This article provides a brief introduction into nondestructive testing and variables of interest, damage mechanisms of interest, and examples that illustrate the use of different damage models. Additionally, it incorporates a detailed sensitivity analysis study that models damage as changes associated with localized material and geometrical properties.

The sensitivity analysis study is carried out using a time-domain numerical model based on the LISA/SIM that models propagating waves based on material property and geometrical changes. This exercise is undertaken without consideration of thermodynamic fluctuations (temperature, humidity) and large displacement considerations, models of which requires additional input parameters and the use of

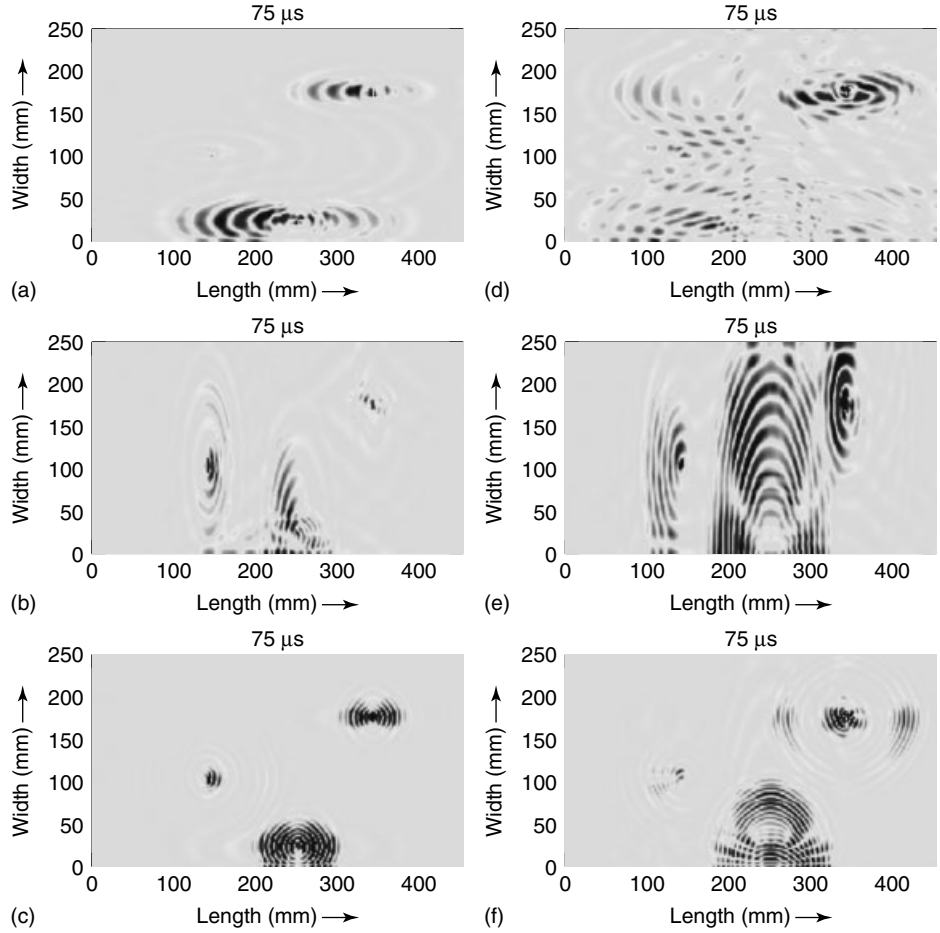


Figure 10. $450 \times 250 \times 3\,\text{mm}^3$ orthotropic plate specimen showing the displacement profiles of the difference between a damaged specimen and a pristine specimen subject to an excitation at 160 kHz corresponding to an actuator at (a, b, c) (5, 125, 0) mm and (d, e, f) (225, 245, 0) mm. Responses are obtained for all three displacements i.e., (a, d) are the x-displacements, (b, e) are the y-displacements and (c, f) are the z-displacements.

other modeling techniques. These changes in material and geometrical properties can be used to potentially improve the capabilities of a structural health monitoring methodology (for example, sensor and actuator location, displacement direction monitored). Localized changes are expressed in terms of changes in stiffness, density, and damping with geometrical dimensions not exceeding the wavelength of the waveform excited were considered in four different plate specimens. The four plate specimens consisted of plates of different sizes and material properties (isotropic, orthotropic, and heterogeneous orthotropic). It is shown that the models can be used to identify the extent of damage using time slices over the entire plate or through sensor locations at different points on the plate specimen. Different actuation, sensor and damage locations, and plate dimensions are considered. The effect of damage size is dependent on the sensor, actuator, and damage position in addition to the relative size of the damage to the wavelength of the excited waveform. Multiple-site damage is studied and the interaction of responses reflected and/or diffracted from the damage site is shown to be sometimes greater than small changes in stiffness and density. The benefits of using a three-dimensional model are shown by the differing trends

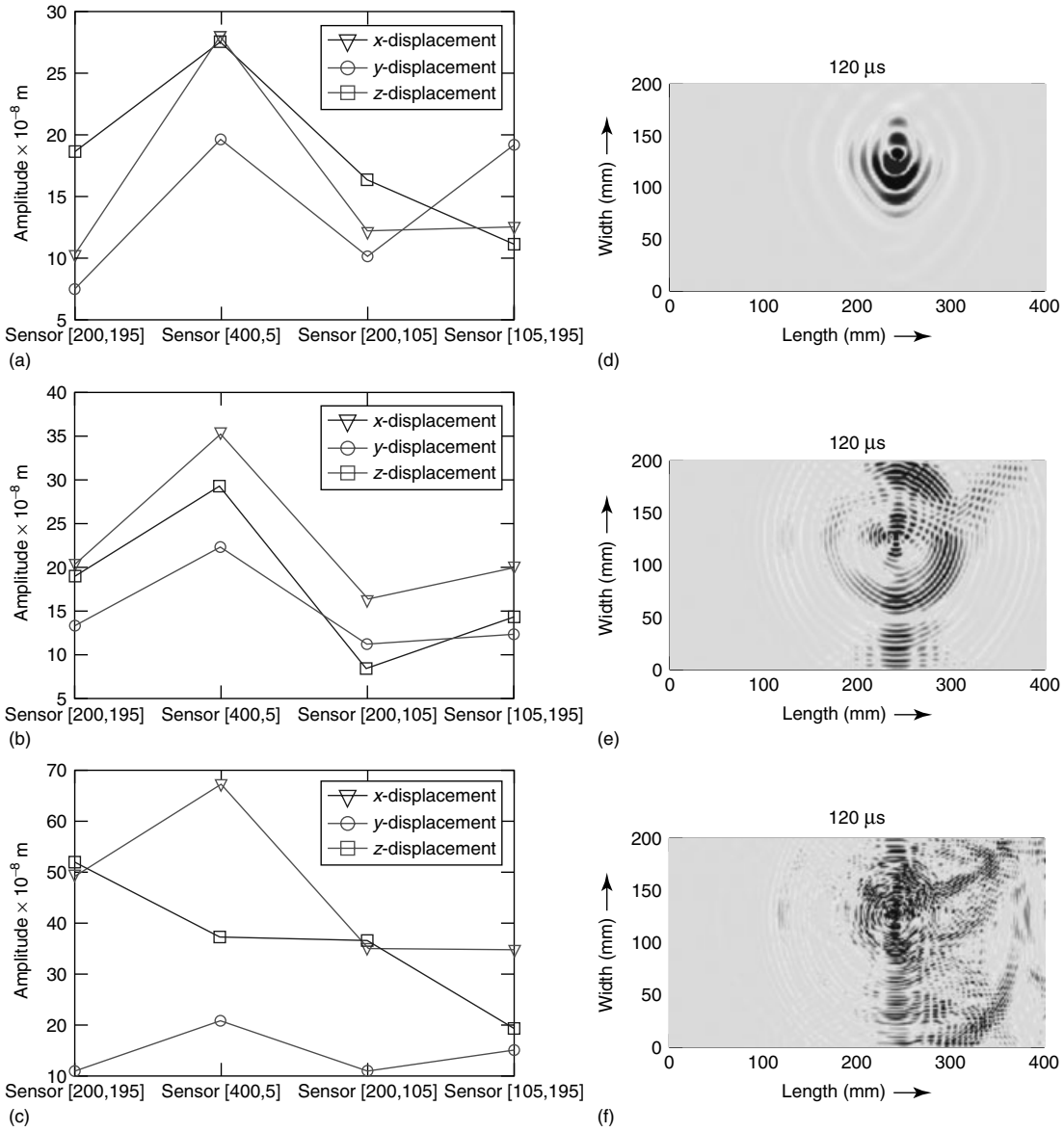


Figure 11. Laminated composite multimodal study for separation identification based on the difference out-of-plane displacements. (a, d) 40 kHz, (b, e) 150 kHz, (c, f) 300 kHz in all three directions, (a–c) peak difference in amplitude at four sensor locations, (d–f) time slice results.

due to changes in stiffness, density, and damping. Damage studies are also carried out in heterogeneous specimens. Structural heterogeneities are introduced into the plate specimen in the form of laminates along the out-of-plane dimensions. Here, the material property of the laminated heterogeneity is modeled to be the same as that of an orthotropic material (graphite epoxy).

The case study is organized along the lines of the general schema for modeling static damage. Specifically, this case study illustrates the need for validating models using experimental testing. One benefit of

studying static damage models is that it provides additional information in correlating observed changes in data with changes in the material property matrices. Static damage models are driven by the nondestructive tool used and the damage mechanism of interest.

ACKNOWLEDGMENTS

The work related to the numerical model was performed using computing facilities at the Ray W. Herrick Labs, Purdue University. The author would like to thank Purdue University Engineering Computing Network for providing access to the workstation from which the simulation results were obtained.

RELATED ARTICLES

A Simplified Damage Model for SHM Metallic and Composite Structures

Modeling of Lamb Waves in Composite Structures

Multiple-model Structural Identification

REFERENCES

[1] Atluri SN, Kobayashi AS. In *Mechanical Responses of Materials, Book Chapter in Handbook on Experimental Mechanics*, Kobayashi AS (ed). Society for Experimental Mechanics: Bethel, CT, 1993.

[2] Dieter GE. In *Mechanical Metallurgy–SI Metric Edition*. Adapted by Bacon D (ed). McGraw-Hill Book Company: London, 1988.

[3] Lemaitre J. *A Course on Damage Mechanics*. Berlin, Springer-Verlag, 1996.

[4] Fatemi A, Yang L. Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. *International Journal of Fatigue* 1998 **20**(1):9–34.

[5] Kaminski M. On probabilistic fatigue models for composite materials. *International Journal of Fatigue* 2002 **24**(2):477–495.

[6] Payan J, Hochard C. Damage modelling of laminated carbon/epoxy composites under static and fatigue loadings. *International Journal of Fatigue* 2002 **24**(2–4):299–306.

[7] Adams DE. *Health Monitoring of Structural Materials and Components*. John Wiley & Sons: New York, 2007.

[8] Timoshenko SP. *History of Strength of Materials*. Dover Publications Inc., New York; Originally Published: McGraw-Hill, New York, 1983, 1953 Reprint.

[9] Doebling SW, Farrar C, Prime MB, Daniel WS. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics: A Literature Review*, Los Alamos National Laboratory, LA-13070-MS, 1996.

[10] Sohn H, Farrar CR, Francois MH, Devin DS, Daniel WS, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Los Alamos National Laboratory Report, LA-13976-MS, 2003.

[11] Hellier C. *Handbook of Nondestructive Evaluation*. McGraw-Hill, 2001.

[12] Shull PJ. *Nondestructive Evaluation: Theory*. Techniques, and Applications: Marcel Dekker, 2002.

[13] Smith EH, with specialist contributors. *Mechanical Engineer's Reference Book, Twelfth Edition*. Butterworth-Heinemann: Woburn, MA, 1994.

[14] Kinsler LE, Frey AR, Coppens AB, Sanders JV. *Fundamentals of Acoustics, Fourth Edition*. John Wiley & Sons, 2000.

[15] Halmshaw R. *Introduction to the Non-destructive Testing of Welded Joints*. Woodhead Publishing: Cambridge, 1996.

[16] Bossi RH, Iddings FA, Wheeler GC. *Nondestructive Testing Handbook, Vol. 4: Radiographic Testing, Third Edition*. American Society for Nondestructive Testing: Columbus, OH, 2002.

[17] Krautkramer J, Krautkramer H. *Ultrasonic Testing of Materials*, Springer-Verlag: New York, 1990.

[18] Rose JL. *Ultrasonic Waves in Solid Media*, Cambridge University Press: Cambridge, 1999.

[19] Wu T-T. Elastic wave propagation and nondestructive evaluation of materials. *Proceedings of the National Science Council ROC(A)* 1999 **23**(6):703–715.

[20] Hardt DE, Katz JM. Ultrasonic measurement of weld penetration. *Welding Journal* 1984 **63**(9): 273s–281s.

[21] Bray DE, Najm M. The effect of material deformation on the velocity of critically refracted shear waves in railroad rail. *Proceedings of the 1983 Ultrasonics International Conference*. Halifax, 12–14 July 1983; pp. 13–19.

[22] Chahbaz A, Mustafa V, and Hay DR. Corrosion detection in aircraft structures using guided Lamb waves. *NDTnet* 1996 **1**(11): http://www.ndt.net/article/tektrend/tektrend.htm, Online Journal.

[23] Rogers WP. *Elastic Property Measurement Using Rayleigh-Lamb Waves*. Research in Nondestructive Evaluation, 1995, pp. 185–208.

[24] Wevers M, Listening to the sound of materials: Acoustic emission for the analysis of material behavior, *NDT & E International* 1997 **30**(2):99–106.

[25] Prosser WH, Jackson KE, Kellas S, Smith BT, McKeon J, and Friedman A. Advanced, waveform based acoustic emission detection of matrix cracking in composites, *Materials Evaluation* 1995 **53**(9):1052–1058.

[26] Adams RD, Walton D, Flitcroft JE, Short D. *Vibration Testing as a Nondestructive Test Tool for Composite Materials, Composite Reliability*, ASTM STP 580. American Society for Testing and Materials, 1975, pp. 159–175.

[27] Cawley P, Adams RD. The location of defects in structures from measurements of natural frequencies. *Journal of Strain Analysis* 1979 **14**(2):49–57.

[28] Yao JTP. Damage assessment and reliability evaluation of existing structures. *Engineering Structures* 1979 **1**:245–251.

[29] Mackerle J. Finite-element modelling of nondestructive material evaluation, an addendum: a bibliography (1997–2003), *Materials Science and Engineering* 2004 **12**:799–834.

[30] Maldague XPV. *Nondestructive Evaluation of Materials by Infrared Thermography*. London, Springer-Verlag, 1993.

[31] Ibara-Castanedo C, Genest M, Piau J-M, Guibert S, Bendada A, Maldague XPV, Active infrared thermography techniques for the nondestructive testing of materials. In *Ultrasonic and Advanced Methods for Nondestructive Testing and Material Characterization*, Chen CH (ed). World Scientific; Chapter 14.

[32] Sakagami T, Kubo S. Applications of pulse heating and lock-in thermography to quantitative nondestructive evaluation. *Infrared Physics and Technology* 2002 **43**(3–5):211–218.

[33] Maldague X, Ziadi A, Klein M. Double pulse infrared thermography. *NDT and E International* 2004 **37**(7):559–564.

[34] Meola C, Carlomagno GM. Recent advances in infrared thermography. *Measurement Science and Technology* 2004 **15**:R27–R58.

[35] Lewis M, Michael DH, Lugg MC, Collins R. Thin-skin electromagnetic fields around surface-breaking crack in metals. *Journal of Applied Physics* 1988 **64**(8):3777–3784.

[36] Rao BPC, Jayakumar T, Raj B. Electromagnetic NDE techniques for materials characterization. In *Ultrasonic and Advanced Methods for Nondestructive Testing and Material Characterization*, Chen CH (ed). World Scientific, 2007; Chapter 11.

[37] Bowler J. Eddy-current interaction with an ideal crack. I. The forward problem, *Journal of Applied Physics* 1994 **75**(12):8128–8137.

[38] Fitzpatrick GL, Thome DK, Skaugset RL, Shih EYC, Shih WCL. Magneto-optic/eddy current imaging of aging aircraft: a new NDI technique. *Materials Evaluation* 1993 **51**(12):1402–1407.

[39] Grimberg R, Premel D, Savin A, Le Bihan Y, Placko D. Eddy current holography evaluation of delamination in carbon-epoxy composites. *Insight (UK)* 2001 **43**(4):260–264.

[40] Zoughi R. *Microwave Non-Destructive Testing and Evaluation*. Kluwer Academic Publishers, 2000.

[41] Chen L, Que P-W, Jin T. A giant-magnetoresistance sensor for magnetic flux leakage nondestructive testing of a pipeline. *Russian Journal of Nondestructive Testing* 2005 **41**(7):462–465.

[42] Collins R, Dover WD, Michael DH. In *Potential Drop Techniques, Nondestructive Testing*. Sharpe RS (ed). Academic Press: New York, 1985.

[43] Michael DH, Waechter RT, Collins R. The measurement of surface cracks in metals by using a.c. electric fields. *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences* 1982 **381**(1780):139–157.

[44] Lovejoy DJ. *Magnetic Particle Inspection: A Practical Guide*. Chapman & Hall, 1993.

[45] Mandayam S, Udpa L, Udpa SS, Lord W. Invariance transformations for magnetic flux leakage signals. *IEEE Transactions on Magnetics* 1996 **32**(3):1577–1580.

[46] Sablik MJ, Jiles DC, Govindaraju MR. Finite element modeling of creep damage effects on a

magnetic detector signal for a seam weld/HAZ-region in a steel pipe. *IEEE Transactions on Magnetics* 1998 **34**(4):2156–2158.

[47] Fritzen C-P. Recent developments in vibration-based structural health monitoring. In *The Proceedings of the 5th International Workshop on Structural Health Monitoring 2005: Advancements and Challenges for Implementation*, Structural Health Monitoring 2005, Chang Fu-Kuo (ed). DEStech Publications: Lancaster, PA, 2005; pp. 42–60.

[48] Yang C, *Experimental Embedded Sensitivity Functions for Use in Mechanical System Identification*, Doctoral Thesis. School of Mechanical Engineering, Purdue University: West Lafayette, 2004.

[49] Haroon M, and Adams DE. Identification of damage in a suspension component using narrow-band nonlinear signal processing techniques. *Proceedings of the SPIE* 2007; Vol. 6532, p. 1–12.

[50] Heller K, Jacobs LJ, Qu J. Characterization of adhesive bond properties using Lamb waves. *NDT and E International* 2000 **33**:555–563.

[51] Balasubramanian K, Rao NS. Inversion of composite material elastic constants from ultrasonic bulk wave phase velocity data using genetic algorithms. *Composites Part B* 1998 **29B**:171–180.

[52] Grandt A. *Fundamentals of Structural Integrity*. John Wiley & Sons, 2004.

[53] Jata K, Parthasarathy T. Physics of failure. *Proceedings of the First International Forum on Integrated System Health Engineering and Management in Aerospace, NASA Workshop on Health Management*. Napa, CA, 7–10 November 2005.

[54] Hajeeh M. Estimating corrosion: a statistical approach. *Materials and Design* 2003 **24**(7):509–518.

[55] McCafferty E. Sequence of steps in the pitting of aluminum by chloride ions. *Corrosion Science* 2003 **45**:1421–1438.

[56] Brown SGR, Barnard NC. 3D computer simulation of the influence of microstructure on the cut edge corrosion behaviour of a zinc aluminium alloy galvanized steel. *Corrosion Science* 2006 **48**:2291–2303.

[57] Terrien N, Osmont D, Royer D, Lepoutre F, Deom A. A combined finite element and modal decomposition method to study the interaction of lamb modes with micro-defects. *Ultrasonics* 2007 **46**:74–88.

[58] Jiang H, Adams DE, Jata K. Material damage modeling and detection in a homogeneous thin metallic sheet and sandwich panel using passive acoustic transmission. *Structural Health Monitoring* 2006 **5**(4):373–387.

[59] Anderson TL. *Fracture Mechanics: Fundamentals and Applications*, *Third Edition*. CRC Press, 2005.

[60] Nataraju M, Adams DE, Rigas EJ. Nonlinear dynamical effects and observations in modeling and simulating damage evolution in a cantilevered beam. *Structural Health Monitoring* 2005 **4**(3): 259–282.

[61] Sanford RJ. *Principles of Fracture Mechanics*. Prentice Hall: Upper Saddle River, NJ, 2003.

[62] Jih CJ, Sun CT. Prediction of delamination in composite laminates sublected to low velocity impact. *Journal of Composite Materials* 1993 **27**(7):686–701.

[63] Pagano NJ, Schoeppner GA. Delamination of polymer matrix composites: problems and assessment. In *Comprehensive Composite Materials*, Kelly AZC (ed). 2000; Vol. 2, pp. 433–528.

[64] Bolotin VV. Delaminations in composite structures: its origin. *Buckling Growth and Stability, Composites Part B-Engineering* 1996 **27**(2): 29–145.

[65] Bolotin VV. Mechanics of delaminations in laminated composite structures. *Mechanics of Composite Materials* 2001 **37**(5):367–380.

[66] Soutis C, Beaumont PWR. *Multi-Scale Modelling of Composite Material Systems: The Art of Predictive Damage Modeling*. Woodhead Publishing: Cambridge, 2005.

[67] Griffith AA. The phenomena of rupture and flow in solids, philosophical transactions of the royal society. *London, Series A* 1921 **221**:163–198.

[68] Irwin GR. Analysis of stresses and strains near the end of a crack transversing a plate. *Journal of Applied Mechanics* 1957 **24**:361–366.

[69] Rybicki EF, Kanninen MF. A finite element calculation of stress intensity factors by a modified crack closure integral. *Engineering Fracture Mechanics* 1977 **9**:931–938.

[70] Rice JR. A path independent integral and the approximate analysis of strain concentration by notches and cracks. *Journal of Applied Mechanics* 1968 **35**:379–386.

[71] Krueger R. *The Virtual Crack Closure Technique: History, Approach and Applications*, NASA/CR-2002-211628, 2002.

[72] Dugdale DS. Yielding of steel sheets containing slits. *Journal of the Mechanics and Physics of Solids* 1960 **8**:100–104.

[73] Hillerborg A, Modeer M, Petersson PE. Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements. *Cement and Concrete Research* 1976 **6**:773–782.

[74] Purekar AS. *Piezoelectric Phased Array Acousto-Ultrasonic Interrogation of Damage in Thin Plates*, Ph.D. Dissertation. Department of Aerospace Engineering, University of Maryland: College Park, MD, 2006.

[75] Adams DE, Farrar CR. Identifying Linear and Nonlinear damage using frequency domain ARX models, *Structural Health Monitoring, An International Journal* 2002 **1**(2) 185–201.

[76] Timoshenko SP, Gere JM. *Theory of Elastic Stability, International Edition*. McGraw-Hill Book: Singapore, 1963.

[77] Sundararaman S., *Numerical and Experimental Investigations of Practical Issues in the Use of Wave Propagation for Damage Identification*, Doctoral Dissertation. School of Mechanical Engineering, Purdue University: West Lafayette, IN, 2007.

[78] Soedel W. *Vibrations of Plates and Shells*. Marcel Decker, 1994.

[79] Kess HR, Sundararaman S, Shah CD, Adams DE, Walsh SM, Pergantis C, Triplett M. Identification of impact damage in S-2 glass composite missile casings using complementary vibration and wave propagation approaches. *Experimental Mechanics* 2007 **47**(4):497–509.

[80] White JR. *Impact and Thermal Damage Identification in Metallic Honeycomb Thermal Protection System Panels using Active Distributed Sensing with the Method of Virtual Forces*, Masters Thesis. School of Mechanical Engineering, Purdue University: West Lafayette, IN, 2006.

[81] Jata KV, Semiatin SL. Continuous dynamic recrystallization during friction stir welding of high strength aluminum alloys. *Scripta Materialia* 2000 **43**(8):743–749.

[82] Dilthey U, Reisgen U, Kretschmer M. Comparison of FEM simulations to measurements of residual stresses for the example of a welded plate: a state-of-the-art report. *Modelling Simulation in Materials Science and Engineering* 2000 **8**:911–926.

[83] Rauch BJ, Rowlands RE. *Thermoelastic Stress Analysis, Book Chapter in Handbook on Experimental Mechanics*, Kobayashi AS (ed). Society for Experimental Mechanics: Bethel, CT, 1993.

[84] Sundararaman S, Adams DE, Simulation of lamb wave propagation in a C458 Al-Li friction stir welded plate. *Proceedings of the SAMPE Conference*. Baltimore, MD, 2007; pp. 1–15.

[85] Balasubramanyam R, Quinney D, Challis RE, Todd CPD. A finite-difference simulation of ultrasonic lamb waves in metal sheets with experimental verification. *Journal of Physics D-Applied Physics* 1996 **29**(1):147–155.

[86] Nag A, Roy Mahapatra D, Gopalakrishnan S. Identification of delamination in a composite beam using a damaged spectral element. *International Journal of Structural Health Monitoring* 2002 **1**(1):105–126.

[87] Srinivasan V and Subbarayan G. Hierarchical Partition of Unity Constructions for Meshless Optimal Design in the Presence of Cracks. *In Proceedings (CD-ROM) of the 17th International Conference on Computer Methods in Mechanics (CMM)*, Lodz-Spala, Poland, 19–22 June 2007.

[88] Mackerle J. An information retrieval system for finite element and boundary element literature and software. *Engineering Analysis with Boundary Elements* 1993 **11**:177–187.

[89] Mackerle J. Finite element modelling of non-destructive material evaluation: a bibliography (1976–1997). *Modelling and Simulation in Materials Science and Engineering* 1999 **7**:107–145.

[90] Mackerle J. Finite element modelling and simulation of indentation testing, a bibliography (1990–2002). *Engineering Computations* 2004 **21**:23–52.

[91] Zimmerman D, Kaouk M. Structural damage detection using a minimum rank update theory, *Journal of Vibrations and Acoustics, Transactions of the ASME* 1994 **116**(2):222–231.

[92] Yang Q, Cox B. Cohesive models for damage evolution in laminated composites. *International Journal of Fracture* 2005 **133**(2):107–137.

[93] Christides S, Barr ADS. One dimensional theory of cracked Bernoulli-Euler beams. *International Journal of Mechanical Science* 1984 **26**(11/12): 639–648.

[94] Sinha JK, Friswell MI, Edwards S. Simplified models for the location of cracks in beam structures using measured vibration data. *Journal of Sound and Vibration* 2002 **251**(1):13–38.

[95] Friswell MI, Penny JET. Crack modeling for structural health monitoring. *International Journal of Structural Health Monitoring* 2002 **1**(2):139–148.

[96] Fox RL, Kapoor MP. Rates of changes of eigenvalues and eigenvectors. *AIAA Journal* 1968 **6**:2426–2429.

[97] Patel SR. *Durability of Advanced Woven Polymer Matrix Composites for Aerospace Applications*, MS Thesis. Virginia Polytechnic Institute and State University: Blacksburg, VA, 1999.

[98] Roy S, Xu W. Modeling of diffusion in the presence of damage in polymer matrix composites. *International Journal of Solids and Structures* 2001 **38**(1):115–126.

[99] Talreja R. Life prediction of composite structures by damage mechanics, collection of technical papers AIAA ASME ASCE AHS ASC structures. *Structural Dynamics, and Materials Conference* 1998 **1**:345–351.

[100] Irving PE, Thiagarajan C. Fatigue damage characterization in carbon fibre composite materials using an electrical potential technique. *Smart Materials and Structures* 1998 **7**:456–466.

[101] Chung DDL. Structural health monitoring by electrical resistance measurement. *Smart Materials and Structures* 2001 **10**:624–636.

[102] Schmidt H, Hattel J. A local model for the thermomechanical conditions in friction stir welding. *Modelling and Simulation in Materials Science and Engineering* 2005 **13**:77–93.

[103] Delsanto PP, Whitcombe T, Chaskelis HH, Mignogna RB. Connection machine simulation of ultrasonic wave propagation in materials I: the one-dimensional case. *Wave Motion* 1992 **16**:65–80.

[104] Delsanto PP, Schechter RS, Chaskelis HH, Mignogna RB, Kline RB. Connection machine simulation of ultrasonic wave propagation in materials II: the two-dimensional case. *Wave Motion* 1994 **20**:295–314.

[105] Delsanto PP, Schechter RS, Mignogna RB. Connection machine simulation of ultrasonic wave propagation in materials III: the three-dimensional case. *Wave Motion* 1997 **26**:329–339.

[106] Agostini V, Delsanto PP, Genesio I, Oliviero D. Simulation of lamb wave propagation for the characterization of complex structures. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2003 **50**(4):441–448.

[107] Sundararaman S, Adams DE. Modeling guided waves for damage identification in isotropic and orthotropic plates using a local interaction simulation approach. *ASME Journal of Vibration and Acoustics* 2008 **130**(4):041009 (16).

[108] Sundararaman S, Adams DE., Accuracy and convergence using a local interaction simulation approach in one, two and three dimensions, accepted for publication in the *ASME Journal of Applied Mechanics* 2008.

[109] Lee BC, Staszewski WJ. Modelling of lamb waves for damage detection in metallic structures: part II. Wave interactions with damage. *Smart Materials and Structures* 2003 **12**:815–824.

FURTHER READING

Adams DE, Nataraju M. A nonlinear dynamical systems framework for structural diagnosis and prognosis. *International Journal of Engineering Science* 2002 **40**(17): 1919–1941.

Chahbaz A, Gauthier J, Brassard M, Hay DR. Ultrasonic technique for hidden corrosion detection in aircraft wing skin. *Proceedings of Third Joint FAA/DoD/NASA Conference on Aging Aircraft*. Albuquerque, NM, 20–23 September 1999.

Dimarogonas AD. Vibration of cracked structures: a state of the art review. *Engineering Fracture Mechanics* 1996 **55**:831–857.

Graff K. *Wave Motion in Elastic Solids*. Dover Publications: New York, 1991.

Hellen TK. On the method of the virtual crack extension. *International Journal for Numerical Methods in Engineering* 1975 **9**:187–207.

Ida N. *Numerical Modeling for Electromagnetic Non-Destructive Evaluation*, Springer, 1995.

Kessler SS. *Piezoelectric-based In-Situ Damage Detection of Composite Materials for Structural Health Monitoring Systems*, Doctoral Thesis. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology: Cambridge, 2002.

Leonard S, Atherton DL. Calculations of the effects of anisotropy on magnetic flux leakage detector signals. *IEEE Transactions on Magnetics* 1996 **32**(3):1905–1909.

Miller JB. NMR imaging of materials. *Progress in Nuclear Magnetic Resonance Spectroscopy* 1998 **33**(3–4): 273–308.

Nataraju M. *A Transitionary Nonlinear Dynamics Approach for Modeling and Simulating Damage Evolution in a Cantilevered Structure*, MS Thesis. School of Mechanical Engineering, Purdue University: West Lafayette, IN, 2003.

Ostachowicz W, Krawczuk M. On modeling of structural stiffness loss due to damage. *DAMAS 2001: 4th International Conference on Damage Assessment of Structures*. Cardiff, 2001; pp. 185–199.

Rao BPC, Rao CB, Jayakumar T, Raj B. Simulation of eddy current signals from multiple defects. *NDT and E International* 1996 **29**(5):269–273.

Sundararaman S, Haroon M, Adams DE, Jata K. Incipient damage identification using elastic wave propagation through a friction stir welded Al-li interface for cryogenic tank applications. *Proceedings of the European Workshop on Structural Health Monitoring*. Munich, 2004; pp. 525–532.

Talreja R. *Damage Mechanics of Composite Materials*. Elsevier Science, 1994.

Tay TE, Shen F. Analysis of delamination growth in laminated composites with consideration for residual thermal stress effects. *Journal of Composite Materials* 2002 **36**(11):1299–1320.

White J, Adams DE, Jata KV. Modeling and material damage identification of a sandwich plate using MDOF modal parameter estimation and the method of virtual forces. *The Proceedings of the International Mechanical Engineers Congress and Exposition*. Orlando, FL, 5–11 November 2005.

Yang C, Adams DE, Yoo S, Kim H-J. An embedded sensitivity approach for diagnosing system-level vibration problems. *Journal of Sound and Vibration* 2004 **269**(22): 1063–1081.

Chapter 9
Damage Evolution Phenomena and Models

Alten F. Grandt Jr.
School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN, USA

1 INTRODUCTION

Although structural failures are rare, those that do occur are often related to preexistent manufacturing or service-induced damage and/or more severe service usage than anticipated. Such failures can be prevented though a combination of damage tolerant design, nondestructive inspections, and structural health monitoring (SHM). Damage tolerance is the ability of a structure to resist fracture from preexistent cracks for a given period of time and is the first line of defense for insuring that unanticipated damage does not lead to premature failure. Periodic inspections provide a second defense by determining the maximum size of life-limiting defects that could be present at a given time. Since structural damage can grow during service, inspection intervals are based on the time it would take for undetected damage to grow to failure under the anticipated service loads. Thus, failure prevention must consider not only the current damage state but also the expected rate of damage growth due to future loading conditions (including environmental influences). The latter task would include SHM of accumulated loads and environmental exposure experienced in service.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 OVERVIEW OF STRUCTURAL FAILURE MODES

Typical failure modes are briefly outlined below with the goal of setting the context for SHM requirements. It is important here to distinguish between pristine structures manufactured to "ideal" standards and components that have been subjected to service wear and tear or those that could contain initial manufacturing defects. Indeed, the initial structural and material condition has a tremendous influence on which failure modes limit structural performance. Since one must expect damage to develop and grow in service, an initially "perfect" structure will likely develop fatigue cracks, corrosion, or some other form of trauma that could ultimately lead to failure. The

fact that the limiting failure modes often change with use must be considered by SHM.

2.1 Deformation and creep

All structures are made from deformable materials that lead to changes in component dimensions caused by applied loads. Many materials behave in an elastic manner when forces are small and this results in deflections that are readily predicted by concepts governing the strength of materials [1–4]. These elastic deformations are controlled by the stiffness of the material and are recoverable when the component is unloaded (i.e., it returns to its original dimensions). Although elastic deformations are often not fatal by themselves, there are many instances when they can lead to structural "failure". Excessive deformations can, for example, eliminate tight clearances, resulting in interference between moving components that prevent a machine from performing properly. Another example is the aerodynamic "flutter" that can result in resonant vibrations that deteriorate to unstable conditions and/or lead to fatigue cracking.

If the applied forces exceed a given value for a particular material, the loaded member may continue to deform without fracturing and experience additional "inelastic" deformations that lead to permanent changes in shape. The maximum load that can be applied without causing inelastic deformation is related to the material yield strength. Although most structures are designed to prevent yielding, it should be noted that the component may continue to carry additional load and that inelastic deformations often provide a "safety factor" that is important for damage tolerant structures.

In many structural metals, yielding is specified by the maximum shear stress (Tresca) or von Mises yield criteria. The *maximum shear stress (Tresca) criterion* assumes that yield occurs in a ductile material when the maximum shear stress at the point of interest exceeds a critical material value. Stated mathematically, yield occurs when

$$\tau_{\max} \geq \text{material constant} = \frac{\sigma_{ys}}{2} \tag{1}$$

Here the material constant is half the tensile yield strength σ_{ys}, and $\tau_{\max}$ is the maximum shear stress determined at the point of interest. It is readily shown, for example, that $\tau_{\max} = \frac{1}{2}|\sigma_1 - \sigma_3|$, where σ_1 and σ_3 are the largest and smallest of the three principal stresses at the given point.

The *von Mises yield criterion* states that yield occurs when the following condition is met.

$$\begin{aligned}\sqrt{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_x - \sigma_z)^2 + 6(\tau_{xy}^2 + \tau_{xz}^2 + \tau_{yz}^2)} &\geq \sqrt{2}\sigma_{ys} \\ = \sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_1 - \sigma_3)^2 + (\sigma_2 - \sigma_3)^2} &\geq \sqrt{2}\sigma_{ys}\end{aligned} \tag{2}$$

Here again σ_{ys} is the tensile yield strength, (σ_x, σ_y, σ_z, τ_{xy}, τ_{xz}, τ_{yz}) are the six components of stress at the point of interest, and σ_1, σ_2, and σ_3, are the three principal stresses at that point.

Creep is a time-dependent distortion that occurs when a member continues to deform under extended application of a sustained static load. When the load is removed, some elongations may be recovered (elastic), whereas some deformations may remain at zero load (inelastic). These latter changes in dimension may be permanent or the member may gradually return to its original state after an additional period of time as these deformations continue to "relax". Creep is an important failure mode for metal components that operate for long periods at elevated temperatures and high loads (e.g., turbine engine blades). In addition, many polymers creep at room temperature.

2.2 Buckling

Buckling is a failure mode unique to slender members loaded in compression or shear. In this case, an instability develops that results in lateral deflections that, in turn, cause an additional bending moment that leads to further deflections and increased bending. Buckling can occur at very small elastic loads and is controlled by the material stiffness, the

unsupported length, the component cross-sectional moment of inertia, and the type of end support. Buckling is an important failure mode for thin, compression-dominated structures (e.g., upper wing skins in aircraft, building support columns). Although there are several forms of elastic and inelastic buckling, insight into this failure mode is given by examination of the elastic Euler buckling equation given by

$$P_{cr} = \frac{\pi^2 EI}{(kL)^2} \tag{3}$$

Here P_{cr} is the critical buckling force, E is the elastic modulus, I is the moment of inertia for the column cross section, L is the unsupported column length, and k is a dimensionless coefficient that depends on the constraint applied to the ends of the specimen. In general, k decreases as more constraint is provided at the column ends. Note that the column buckling load increases linearly with the elastic modulus E and the cross-sectional moment of inertia I but is inversely related to the square of kL (known as the *effective column length*). Thus, buckling resistance is greatly increased by shortening the column length and/or increasing the end constraint. Detecting loss of end constraint would be an important SHM task.

2.3 Corrosion

Corrosion is material degradation due to chemical attack and can occur in several forms: galvanic (dissimilar metal), pitting, exfoliation, intergranular attack, filiform, or, when combined with the presence of a tensile stress, stress corrosion cracking. This time-dependent failure mechanism is highly dependent on the particular combination of structural material and environment combination in question and is often accelerated by increasing temperature. Corrosion is a complex chemical phenomenon that can cause general thickness loss (and a corresponding increase in stress) as well as stress concentrations (i.e., pits and other localized areas of attack) that lead to fatigue cracking or fracture. Corrosion is difficult to predict, and its prevention requires careful materials selection, protective coatings, and periodic maintenance. Although corrosion can occur independent of applied loading, it frequently acts in conjunction with static or cyclic loading (e.g., stress corrosion or corrosion fatigue) to represent a particularly dangerous failure mode. Detecting the accumulation of corrosive fluids and/or break down of environmental protection systems would be key SHM tasks.

2.4 Fatigue

Cyclic fatigue is the failure mode associated with *repeated* loading and is one of the main life-limiting factors for mechanical devices. Consider, for example, a structural member that is subjected to repeated application and removal of remotely applied load cycles, as shown in Figure 1. Although the peak values of the maximum and minimum load per cycle may be tension or compression and may change during the component life (i.e., variable-amplitude loading), no single load application is large enough to fracture new components.

After repeated load cycles, however, small cracks will form, often at multiple locations in the structure. At the outset these cracks may be benign, but they do extend slowly after repeated cycling, and eventually some coalesce into a dominant crack(s) that continues to grow in a stable manner. Finally, the dominant crack reaches a size that causes fracture, and the member fails in a sudden, catastrophic manner. The

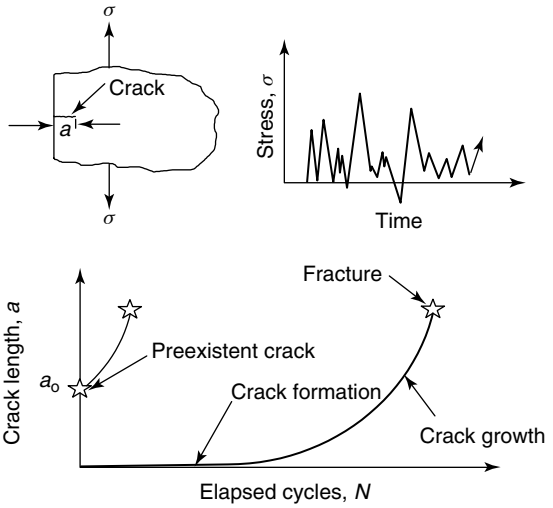


Figure 1. Schematic representation of fatigue failure mechanism, showing crack formation and propagation following a period of cyclic loading.

final fracture surface often exhibits a "brittle" appearance and may demonstrate characteristic markings that are remnants of the crack tip position at various stages in life. These "beach marks" are visible to the naked eye and are one of the distinguishing characteristics of a fatigue fracture. More detailed examination of the fracture surface under a microscope may reveal additional patterns of closely spaced lines, or "striations," that record the crack tip position after each cycle of loading. The striation spacing is the local fatigue crack growth rate, and if present, striations offer conclusive proof of a fatigue failure.

The fatigue failure mode is one of the main concerns for SHM, and is emphasized in this article. Note several points about this important failure mechanism. First, fatigue requires *cyclic* loading to occur. The number of cycles that can be applied before fracture takes place (i.e., the fatigue life) depends on the amplitude and mean value of the cyclic load, the sequence of applied loads (when variable-amplitude load histories are involved), and the component shape and material. The condition of the specimen surface is extremely important, as rough surfaces, notches, or material impurities can rapidly lead to cracks, eliminating the "crack formation" portion of the fatigue life. Thus, SHM should report evidence of structural or material trauma that could lead to fatigue cracking.

Residual stresses can also have a significant influence on the fatigue life. Intentional introduction of compressive residual stresses by shot peening or cold working can be quite beneficial, for example, and is a common technique for extending the fatigue life of a component. On the other hand, residual tensile stresses, such as those introduced by manufacturing processes, can be extremely detrimental to fatigue.

2.5 Fracture

Final catastrophic failure results when the member cleaves into two or more parts. Although fracture may occur in pristine structures subjected to overloads, it is frequently initiated at smaller loads by preexistent flaws or by service-induced damage (e.g., fatigue cracking and/or corrosion). Fracture can occur in a ductile manner that requires considerable expenditure of energy, but can also happen suddenly with little warning. This latter result may exhibit "brittle" behavior without evidence of plasticity and is a particularly dangerous failure mode. The fracture resistance of a material is characterized by its "toughness", a quantity that can be a function of temperature and loading rate, as well as geometric constraint to yielding (e.g., specimen thickness).

3 FATIGUE DAMAGE MODELS

This section briefly outlines life prediction models for the fatigue damage mechanism. The stress–life and strain–life methods are "crack nucleation" approaches that assume crack-free structures, and focus on the time that it would take for a fatigue crack to form. The damage tolerance or fatigue crack growth approach assumes that the component already contains a subcritical crack, and one is interested in determining the number of cycles that it takes for it to grow to final fracture. The initial crack size is a key for these analyses and is often based on the size that could be missed by a nondestructive inspection. As will be seen, fatigue life depends strongly on the applied stresses, and, in addition to component condition, determination of the actual load history is an important parameter to be obtained from SHM.

3.1 Crack nucleation

The *stress–life method* is the original approach to fatigue developed during the latter portion of the nineteenth century. The stress–life or S–N curve shown in Figure 2 is obtained from subjecting "smooth" specimens to a constant-amplitude stress. Here the applied stress amplitude is plotted against the measured fatigue life N_f, which could be the number of cycles to form a "detectable" crack, but is often simply the total number of cycles to fracture the small test coupon.

Some materials (steels in particular) demonstrate an endurance limit S_e, defined as the largest stress amplitude that can be applied without having a fatigue failure. It must be emphasized, however, that while endurance limits are an important measure of fatigue resistance, this property is measured in small, polished specimens subjected to constant-amplitude loading, and may not accurately reflect the fatigue behavior of large components that contain initial

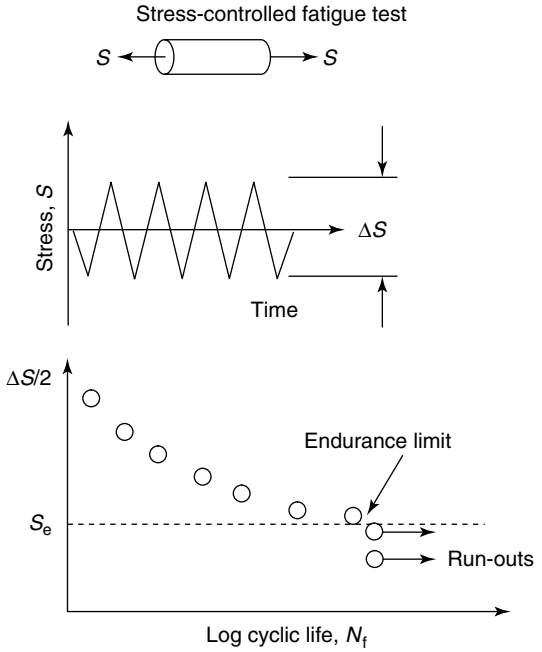


Figure 2. Schematic stress–life (S–N) curve showing relationship between applied stress amplitude $\Delta S/2$ and fatigue life N in small specimen subjected to constant-amplitude loading.

material and/or structural imperfections, residual stresses, variable-amplitude loads, and so on.

Test data above the endurance limit (i.e., the finite life regime) may be modeled by equation (4).

$$\frac{\Delta\sigma}{2} = \sigma'_f(2N_f)^b \tag{4}$$

Here $\frac{\Delta\sigma}{2}$ is the applied stress amplitude[a], $2N_f$ is the fatigue life measured in reversals (1 cycle = 2 reversals), and σ'_f (fatigue strength coefficient) and b (fatigue strength exponent) are empirically determined material constants. Since the basic S–N curve is obtained from small, polished specimens subjected to completely reversed constant-amplitude loading, this expression is quite limited in scope, and in practice one must account for specimen size, stress gradients, surface condition, mean stress, notches, and variable-amplitude loading for most practical applications [3–6].

The *strain–life approach* was developed in the 1950s and 1960s [7, 8] to handle shortcomings of the stress–life method associated with plastic deformation encountered during low cycle fatigue (LCF). This method involves testing smooth specimens under constant-amplitude "strain control", and measuring the number of *reversals* to failure $2N_f$. Although the applied strain amplitude $\Delta\varepsilon/2$ is kept constant, the corresponding stress amplitude $\Delta\sigma/2$ needed to maintain the strain limits may change as the material undergoes inelastic deformation. Materials may initially "cyclically soften" or "cyclically harden", although stable stress–strain behavior is usually observed by mid-fatigue life.

The *cyclic stress–strain* curve relates the steady-state stress amplitude $\Delta\sigma/2$ with the completely reversed applied strain amplitude $\Delta\varepsilon/2$ for a series of tests. It may be modeled for subsequent fatigue analyses by dividing the total applied strain amplitude $\Delta\varepsilon/2$ into elastic and plastic components.

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\varepsilon_{\text{elastic}}}{2} + \frac{\Delta\varepsilon_{\text{plastic}}}{2} \tag{5}$$

Equation (4) and Hooke's law give the *elastic* component of strain in equation (6), where E is the elastic modulus:

$$\frac{\Delta\varepsilon_{\text{elastic}}}{2} = \frac{\Delta\sigma}{2E} = \frac{\sigma'_f}{E}(2N_f)^b \tag{6}$$

Now, solving for the plastic strain amplitude and relating it to the stable stress amplitude $\Delta/\sigma 2$ with another power law defines the empirical constants K' (cyclic strength coefficient) and n' (cyclic strength exponent) in equation (7).

$$\frac{\Delta\sigma}{2} = K'\left(\frac{\Delta\varepsilon_{\text{plastic}}}{2}\right)^{n'} \tag{7}$$

Equations (5–7) are now combined to give the cyclic stress–strain curve:

$$\frac{\Delta\varepsilon}{2} = \frac{\Delta\sigma}{2E} + \left(\frac{\Delta\sigma}{2K'}\right)^{1/n'} \tag{8}$$

Fatigue life behavior is obtained by relating the total applied strain amplitude or the plastic strain amplitude to the lives of individual test specimens. The plastic strain amplitude $\Delta\varepsilon_{\text{plastic}}/2$ may be empirically related to fatigue life (given in reversals

$2N_f$) as follows:

$$\frac{\Delta\varepsilon_{plastic}}{2} = \varepsilon'_f(2N_f)^c \tag{9}$$

This expression defines the fatigue ductility coefficient ε'_f and the fatigue ductility exponent c. Now, as given by equation (10) and shown schematically in Figure 3, combining expressions gives the total strain amplitude in terms of fatigue life.

$$\frac{\Delta\varepsilon}{2} = \frac{\sigma'_f}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c \tag{10}$$

This summation is the key to the powerful strain–life approach to fatigue, and effectively combines the original S–N curve with the plastic strain–life approach. In the LCF regime, elastic strains are small and fatigue life is dominated by the plastic strain. In the high cycle fatigue (HCF) regime, the applied loads are much smaller and the plastic strains are negligible, so that fatigue behavior is controlled by the elastic strain term.

So far, discussion has been limited to completely reversed cycling of smooth specimens without the presence of mean stresses. Mean stress does, however, influence both stress–life and strain–life behavior, with compressive mean stresses being beneficial and tensile means detrimental. Notches also significantly influence fatigue behavior, and are problematic for fatigue analysis. Further discussion of mean stress on fatigue behavior, along with complications of notches is, however, beyond the scope of this paper (see [3–6]).

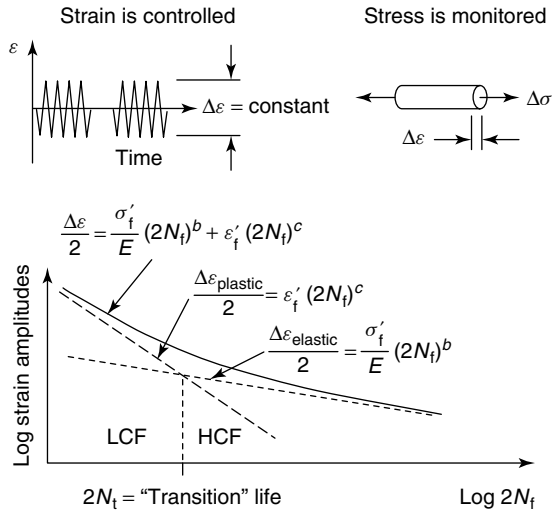


Figure 3. Schematic strain–life curve showing relationship between elastic, plastic, and total strain amplitude with fatigue life $2N_f$ in small specimen subjected to constant-amplitude loading.

3.2 Fatigue crack propagation

This section overviews linear elastic fracture mechanics (LEFM) concepts for crack growth analysis and describes the use of the stress intensity factor to characterize critical and subcritical crack growth. The stress intensity factor K is the LEFM parameter that relates remote load, crack size, and structural geometry, and may be expressed in the form

$$K = \sigma\sqrt{\pi a}\beta \tag{11}$$

Here σ is the applied stress, a is the crack length, and β is a dimensionless factor that depends on crack length and component geometry.

It should be emphasized that the LEFM stress intensity factor K used in the study of the strength of materials. (Note that the stress *intensity* factor has units of stress $\times$ length$^{1/2}$ while the stress *concentration* factor K_t is the dimensionless ratio of local stress at a notch divided by the remote stress.) The stress intensity factor has a rigorous mathematical definition in the context of elastic crack tip stress fields and is based on the fact that all cracks have a characteristic square root singularity at the crack tip [4, 9]. For practical engineering analysis purposes, stress intensity factor solutions have been obtained for many crack geometries, and several handbook compilations are available [10–12].

As a fracture criterion, LEFM employs the experimental observation that "brittle" materials fracture when the stress intensity factor reaches a "critical" value:

$$K = \sigma\sqrt{\pi a}\beta = K_c = \text{constant at fracture} \tag{12}$$

Here K_c is a thickness-dependent material property called the *fracture toughness* of the material and is the limiting value of the stress intensity factor that causes catastrophic fracture in all components made from the same material. Fracture toughness values

for many structural materials are reported in material handbooks [13]. Note that since K relates load, crack length, and structural geometry this simple fracture criterion allows one to correlate fracture measurements from laboratory specimens with fracture of actual structural components.

The fracture mechanics approach to fatigue is based on work by Paris [14], who demonstrated that the cyclic range in stress intensity factor ΔK controls the fatigue crack growth rate per cycle $\mathrm{d}a/\mathrm{d}N$. Here ΔK is the difference between the maximum and minimum stress intensity factors applied during a particular cycle of loading and is given by

$$\begin{aligned}\Delta K &= K_{\max} - K_{\min} \\ &= \left(\sigma_{\max} - \sigma_{\min}\right)\sqrt{\pi a}\beta \\ &= \Delta\sigma\sqrt{\pi a}\beta \end{aligned} \tag{13}$$

Now $\Delta\sigma$ is the cyclic stress range, a is the current crack length, and β is a dimensionless function of crack size and geometry as before.

Standard procedures [15] are available to establish the relationship between the applied ΔK and the measured fatigue crack growth rate per cycle $\mathrm{d}a/\mathrm{d}N$ in constant-amplitude fatigue tests with small laboratory specimens. This experimentally determined $\mathrm{d}a/\mathrm{d}N - \Delta K$ curve may be effectively treated as a material property and is also recorded in material handbooks [13]. Note that the ΔK parameter accounts for the stress range, crack size, and crack geometry (i.e., the β term) for the problem of interest.

When collected over a wide range of crack growth rates, $\mathrm{d}a/\mathrm{d}N - \Delta K$ curves for many materials have the characteristic sigmoidal shape shown schematically in Figure 4. A vertical asymptote is observed when $K_{\max} = K_{\mathrm{c}}$ since fracture occurs at that point. There may also be an asymptote at low ΔK levels, designated as the fatigue threshold stress intensity factor ΔK^{th}. Below ΔK^{th} cracks do not extend by cyclic loading ($\mathrm{d}a/\mathrm{d}N = 0$), and the specimen would have "infinite" life. The ΔK^{th} value is the fatigue crack growth analog to the endurance limit S_{e} measured in *uncracked* fatigue specimens.

A linear relation between log $\mathrm{d}a/\mathrm{d}N$ and log ΔK is often observed between the upper and lower asymptotes. Paris and Erdogan [14] expressed the crack growth behavior in that region by equation (14).

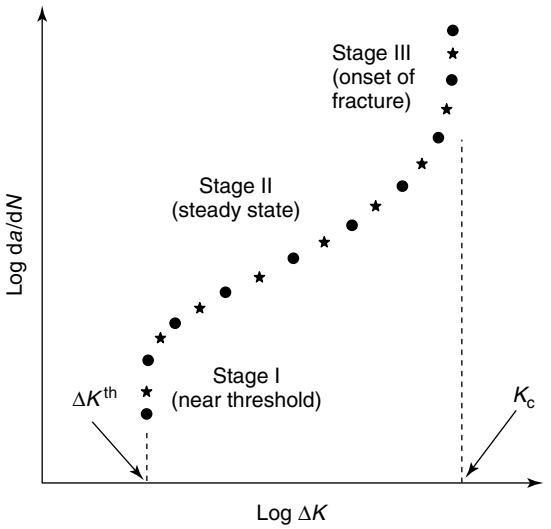


Figure 4. Schematic representation of sigmoidal relationship between fatigue crack growth rate and the cyclic stress intensity factor.

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C\Delta K^{m} \tag{14}$$

Here C and m are empirical material constants obtained for a particular set of material data. Many other more general crack growth equations have been employed to relate $\mathrm{d}a/\mathrm{d}N$ with ΔK [4], yielding equations of the following form:

$$\frac{\mathrm{d}a}{\mathrm{d}N} = F(K) \tag{15}$$

Here $F(K)$ is a mathematical expression that fits $\mathrm{d}a/\mathrm{d}N$ over an appropriate range of ΔK values, including the upper and lower asymptotes. These crack growth models may also account for other loading variables, such as mean stress and temperature.

Returning now to the original objective of predicting the fatigue crack growth *life*, it is a simple task to integrate equation (15) for the total cycles N_{f} required to grow an initial crack of length a_0 to some final size a_{f}. Solving for the cyclic life gives

$$N_{\mathrm{f}} = \int_{a_{\mathrm{o}}}^{a_{\mathrm{f}}} \frac{\mathrm{d}a}{F(K)} \tag{16}$$

It is important to recognize limitations to the stress intensity factor approach described here. It is assumed that K is a valid crack parameter and that crack tip plasticity effects are negligible [4]. Large peak loads applied during the fatigue cycling can introduce large plastic zones, for example, which significantly influence subsequent fatigue crack growth (cause fatigue crack retardation as described in the following section). Sophisticated life analysis procedures have, however, been developed to analyze peak overloads, mean stress, temperature, and environmental influences (i.e., corrosion fatigue) that may occur in service. For variable-amplitude loading, for example, equation (16) cannot be integrated directly, but life is obtained by a cycle-by-cycle summation of crack growth rates for individual loading cycles. Additional details of those procedures are presented in [4].

The LEFM approach may also be applied to the stress corrosion problem where cracks can form and extend under the combined influence of a *sustained* tensile stress and an aggressive chemical environment. Again, laboratory tests are used to correlate the stress corrosion crack growth rate $\mathrm{d}a/\mathrm{d}t$ as a function of the applied stress intensity factor K. In this case, K results from a sustained tensile stress rather than a cyclic load as in fatigue, and crack growth is measured as a function of elapsed time (i.e., hours, days) rather than cycles. Again the curve of log $\mathrm{d}a/\mathrm{d}t$ versus log K assumes a sigmoidal shape between a lower (K_{ISCC}) and upper (K_{c}) asymptote. As before, these data can be represented by the empirical equation

$$\frac{\mathrm{d}a}{\mathrm{d}t} = f(K) \tag{17}$$

where $f(K)$ is some convenient mathematical function chosen to represent the test data. Now, the total time t_{f} required to grow a crack from length a_0 to a_{f} is given by

$$t_{\mathrm{f}} = \int_{a_0}^{a_{\mathrm{f}}} \frac{\mathrm{d}a}{f(K)} \tag{18}$$

Note that different crack geometries and component materials are treated in a manner analogous to computing fatigue crack growth lives.

It is important here to also note the significant effect environment has when combined with cyclic loading. In general, corrosion fatigue crack growth rates can be considerably faster than those observed for cyclic loading in an inert environment. The influence environment plays on fatigue life depends on the cyclic frequency, the shape of the curve of applied load versus time (i.e., the waveform), the temperature, the environment, the crack orientation (with respect to material axes), and, of course, the particular material of interest [4].

4 LOAD HISTORY EFFECTS

Local crack tip plasticity often limits application of the stress–life, strain–life, and LEFM methods, and leads to some peculiar crack growth behavior that can be important for SHM. Consider, for example, the "fatigue crack retardation" phenomenon shown schematically in Figure 5, where application of large tensile overloads can actually *increase* cyclic life. Note that this increase in life (or reduction in $\mathrm{d}a/\mathrm{d}N$) is *not* predicted by the fatigue crack growth rate models discussed in the previous section. Although

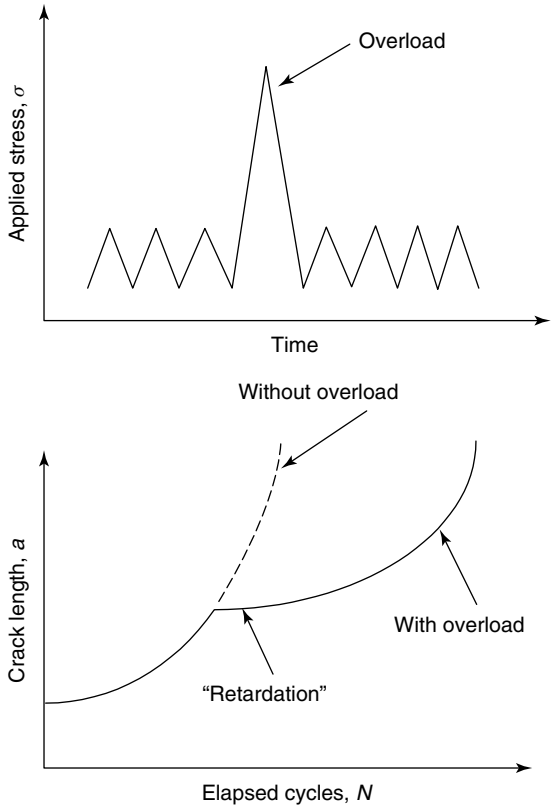


Figure 5. Schematic representation of fatigue crack retardation phenomenon.

the amount of crack delay depends on the particular test material in general, more ductility leads to larger crack tip plastic zones and more retardation. Increasing the magnitude or number of the overloads can also increase the fatigue life and, in some cases, it may be possible to permanently stop subsequent crack growth. (It is assumed here, of course, that $K_{max} < K_c$ during the overload cycle, so that it does not cause fracture.)

Test data showing the influence of overloads on fatigue crack growth lives in 7075-T6 aluminum specimens are given in Figure 6 [16]. Note that in the absence of peak overloads the fatigue life for a center-cracked specimen was less than 10 blocks of 4001 constant-amplitude cycles. Here the constant-amplitude loading during each block varied between 2.7 and 8.0 ksi (18.3 and 55.2 MPa) remote stress. When *one cycle* per block was changed to vary between 2.7 and 14.4 ksi (18.3 and 99.3 MPa), however, the life *increased* to approximately 480 blocks (a 50-fold increase in life).

The sign and sequence of the overload can have a tremendous influence on life, as shown in Figure 7 [17]. Here fatigue crack growth curves are given for

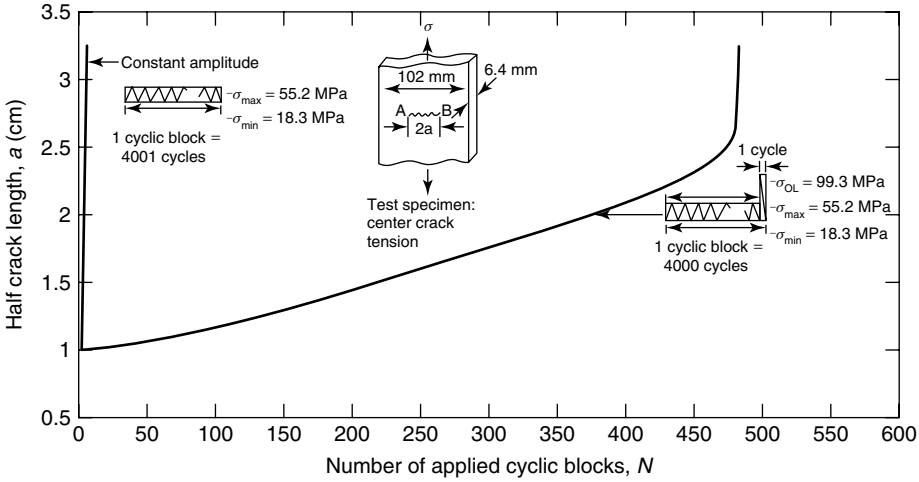


Figure 6. Fatigue crack retardation produced by a single peak overload in aluminum alloy 7075-T6 [16].

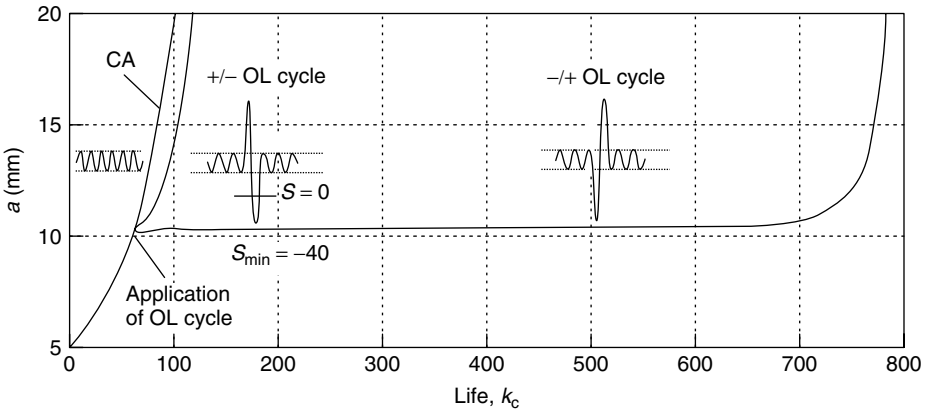
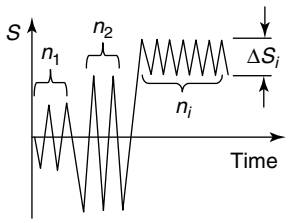


Figure 7. Test data showing the effect of two different overload cycles on fatigue crack growth in 2024-T3 aluminum. Mean stress is 80 MPa in all cases. Stress amplitude is 25 MPa for constant-amplitude cycles and 120 MPa for overload cycles [17].

three center-cracked 2024-T3 aluminum specimens. Compare the crack growth behavior after a 0.4-in. (10-mm) crack length for these three cases. In the first instance, constant-amplitude cycling continued between 8.0 and 15.2 ksi (55 and 105 MPa) remote stress, and failure occurred within 50 000 additional cycles (curve CA in Figure 7). Another specimen, however, was subjected to a single peak under load of −5.9 ksi (−40 MPa) followed by a minimum peak overload of +29 ksi (200 MPa) at that time and then returned to the original constant-amplitude cycling. Note that this specimen lasted nearly 700 000 more cycles, an order-of-magnitude life increase compared to the original specimen without the overload cycle. When a third specimen was subjected to an overload cycle that was identical to that applied to the second specimen, *except that the order* of the −5.8 ksi and the +29 ksi (−40 MPa and +200 MPa) underload/overload peaks were *reversed*, the retardation effect was nearly eliminated, and the specimen had approximately the same life as for the original constant-amplitude loading. Thus, subtle changes in load sequencing can be very significant when the fatigue crack retardation phenomenon is present and pose a particular challenge for SHM (i.e., both magnitude *and* sequence of applied loads need to be considered).

Now consider the important issue of variable-amplitude fatigue life prediction by the stress–life or strain–life methods. Although this topic might best be approached by fatigue tests with the actual load history of interest, such experiments are cost prohibitive for all possible loadings, so there is great interest in predictive methods that employ the basic, completely reversed constant-amplitude data. Again, however, variable-amplitude fatigue analysis is a complex subject that can only be overviewed here, and the reader is referred to [3–6] for additional details.

Recall that the effects of constant-amplitude loading are described by equations (4) and (5) for the stress–life and strain–life approaches. Now consider the more general case where the applied load consists of various *blocks* with different stress amplitudes and mean stresses (or strains) or, in the limit, completely random loading. The simplest approach here is to assume that fatigue damage accumulates in a "linear" manner that is independent of load history. This assumption may be stated by equation (19) and is known as *linear cumulative damage* or *Miner's rule*.



Miner's rule: $\Sigma(n_i/N_{if}) = 1$

n_i = number of applied cycles of stress range ΔS_i

N_{if} = fatigue life for ΔS_i cycling only

Figure 8. Definition of Miner's rule for variable-amplitude loading.

$$\sum \frac{n_i}{N_{if}} = 1 \tag{19}$$

As shown in Figure 8, n_i is the number of *applied* cycles with the ith loading block that has a given stress range ΔS_i and mean stress, and N_{if} is the fatigue life that would result if only the ith block of loading were applied to the specimen. Basically, equation (16) states that n_i/N_{if} is the percentage of fatigue life exhausted for the ith loading block, and that when the damage summation for all blocks equals 1 (i.e., 100%), the total fatigue life is expended. Note that N_{if} for the ith block incorporates the effect of both mean stress and stress range but does not include potential load history effects (i.e., residual stresses) that could be developed by prior loading blocks. Miner's rule can also be employed with the strain–life approach by simply expressing the fatigue lives in terms of reversals (i.e., $2n_i$ and $2N_{if}$) and using the strain–life relationship (equation 10) to compute the percentage of damage accumulated during each loading block.

Although Miner's rule has seen wide use in industry, it has significant limitations that can lead to inaccurate results. The chief shortcoming is the failure to recognize that the *sequence* in which loads are applied can have a significant influence on life through development of residual stresses. Recalling the load history effects associated with fatigue crack growth, it is not surprising that similar issues also exist for the fatigue "nucleation" phenomena.

Crews [18], for example, reports the results of fatigue tests with 12 × 35-in. (30 × 87.5-cm)

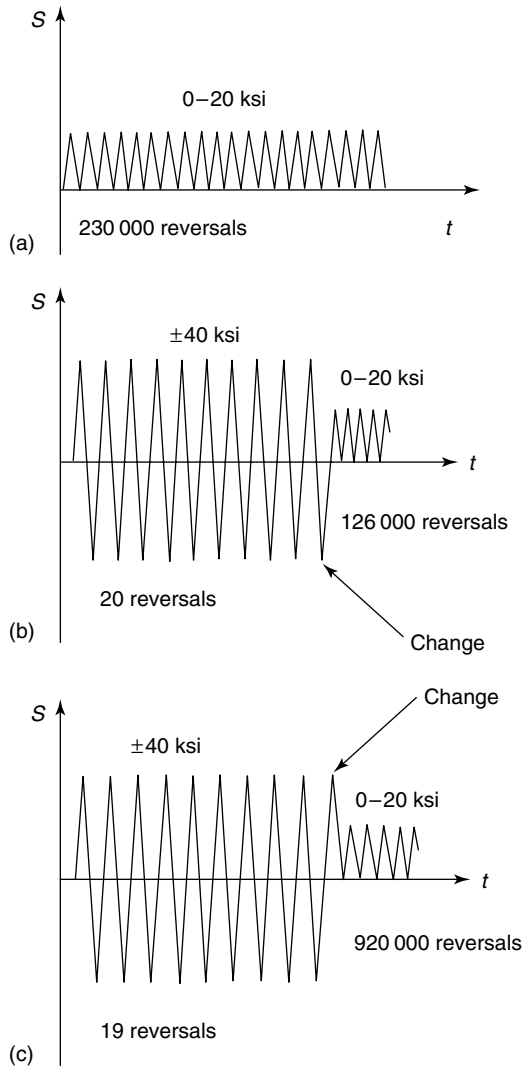


Figure 9. Stress–time histories showing effect of load sequence on fatigue life for notched aluminum specimens tested by Crews.

2024-T3 aluminum plates that contained 2-in. (5-cm) diameter open holes. As shown in Figure 9, one reference specimen subjected to a constant-amplitude remote stress that varied from 0 to 20 ksi (0–138 MPa) had a fatigue life of 230 000 reversals. Two identical specimens were subsequently tested to the same stress history, except for a few initial reversals of ±40 ksi (275 MPa) stress. Although similar, these two "high–low block" loads differed in one critical detail: the sign of the last large peak before the load amplitude was decreased. This subtle change in loading sequence had a large influence, however, on fatigue life. When 20 reversals of ±40 ksi (±138 MPa) were applied before reducing to the 0–20 ksi (0–138 MPa) load *after* the *last large compressive* (40 ksi = 276 MPa) peak, the specimen had a fatigue life of 126 000 reversals, less than the reference case without the large preloads. In a third specimen, the high–low load transition was made *after* the last 40-ksi (138-MPa) *tensile* peak and only involved 19 reversals of the larger preload. That specimen resisted 920 000 additional 0–20 ksi (0–138-MPa) reversals, over seven times as many as the prior sequence, and four times as long as the reference 0–20 ksi (0–138 MPa) loading. Thus, the large preloads actually *increased* the fatigue life as compared to the original case without the large loads. Clearly, load sequence effects play a significant influence on fatigue life that is not predicted by Miner's rule, which would give essentially the same total life for both high–low load sequences.

As shown schematically in Figure 10, the high–low load block sequencing phenomenon results from mean stresses created by overloads. Here two completely reversed strain blocks are applied to a smooth test specimen, and the resulting stress history required to maintain the strain peaks is shown. Note that the first large strain block develops a stable hysteresis loop, and continued cycling about this loop results in the large, *completely reversed* stress history shown.

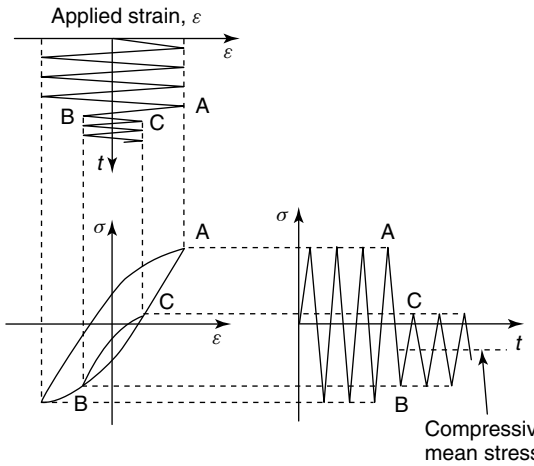


Figure 10. Strain–life sequence showing development of compressive mean stress under high–low strain loading sequence where the last large strain peak is tensile.

However, when the strain amplitude is reduced after a tension peak (i.e., the ABC sequence shown in Figure 10), subsequent strain-controlled cycling follows the smaller hysteresis loop BC contained in the original large loop. Note that this sequence develops a *compressive mean stress* for subsequent completely reversed cycling at the smaller strain amplitude. This compressive mean stress will increase the fatigue life for the small-amplitude strain block. If, instead, the sequence for changing from high to low strain amplitudes is reversed, so that the last large peak before changing amplitude is *compressive*, a deleterious tensile mean stress is created [4].

Thus, it is possible to develop tensile or compressive mean stresses during sequences of completely reversed strain cycling. These mean stresses may be estimated by the cyclic stress–strain curve (equation 8) for *simple* block loadings and incorporated into Miner's rule by modifying the N_{if} term in the denominator of equation (19) to account for the presence of a tensile or compressive mean stress. Details for these types of calculations are given in [3, 5]. Although not discussed here, cycle counting methods are another important aspect of this procedure for determining variable-amplitude fatigue lives (see [3, 19]).

5 PROPOSED WITNESS SAMPLE APPROACH FOR SHM

One goal of SHM programs is to ensure damage tolerance and durability by refining inspection and maintenance intervals based on actual usage. Although those intervals often employ design load assumptions, experience has shown that actual service differs significantly from that projected during design. The fact that individual members within a fleet may encounter large variations in load severity, combined with sequencing effects, indicates that individual structures will accumulate fatigue damage at significantly different rates. This issue is further complicated by the wide variation in the thermal–chemical environments encountered during service, so that the potential for corrosion and/or corrosion fatigue can also vary significantly. As discussed in the previous section, fatigue damage accumulation is particularly sensitive to the magnitude and *sequence* of applied loading, making fatigue monitoring problematic for SHM.

SHM uses data acquisition and reduction tools to measure loads applied to individual structures to predict potential damage in key structural components. Current tracking techniques employ mechanical and electronic strain gauges, acoustic emission monitors, counting accelerometers, flight load recorders, pilot logs and other devices to measure and record applied loads. This load information must then be analyzed by complicated computer algorithms to determine the extent of fatigue damage seen by the structure. These techniques require large capital, extensive effort, and many assumptions to reduce the data to fatigue life, and could introduce errors that lead to inspection intervals that are either too long or too short. Thus, a simplified SHM approach that is sensitive to parameters that control fatigue life is needed.

One method proposed by the author for SHM is the "witness sample" or "fatigue crack gauge" concept [20–25] for real time assessment of the potential for imminent failure due to service-induced damage accumulation. As shown schematically in Figure 11, the witness sample technique is a potentially simple and inexpensive prognosis sensor. The witness sample is simply a precracked coupon that is attached to a load bearing component. Since the "crack gauge" receives a predictable fraction of the component loads, it is also subjected to damage accumulation scaled to that seen by the parent structure. Moreover, since the gauge and structure are subjected to the same thermochemical environment, corrosive aspects of crack growth are automatically accounted for.

Now the crack gauge effectively acts as an analog computer that measures loads applied to the structure, determines their effect on fatigue crack growth, and responds with an output (gauge crack extension) that is directly related to damage formation in the component of interest. Thus, monitoring the gauge crack length gives a real time measure of the "potential" for crack growth in the structural member. It should be emphasized that although the gauge crack is physically present, the *structural* crack is a fictitious metric whose initial size and location are chosen to be consistent with the fracture mechanics worst-case initial crack assumption specified by the appropriate damage tolerant design criteria.

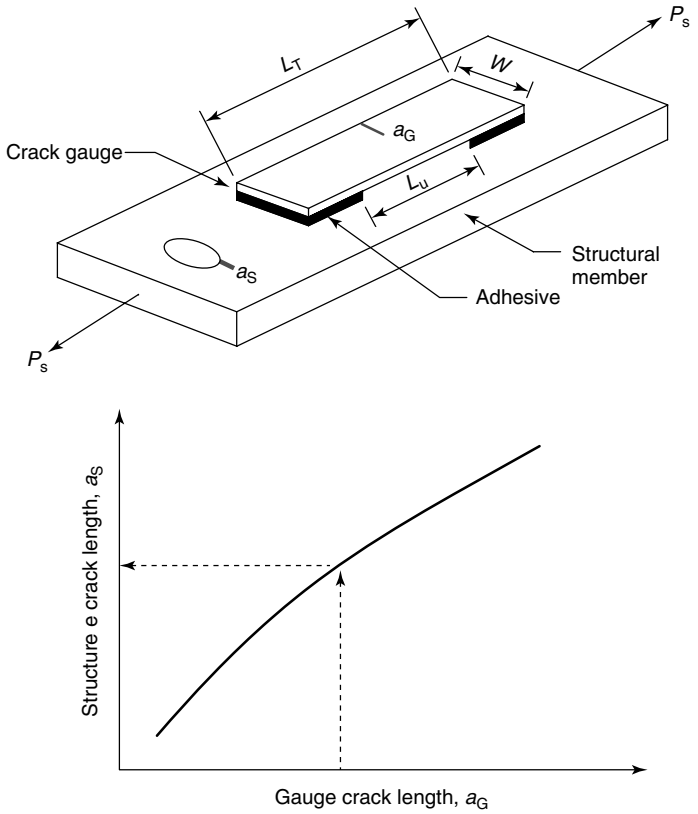


Figure 11. Schematic view of crack gauge mounted to structural component and transfer function relating measured gauge crack growth with extension of assumed structural flaw.

The theoretical basis for the fracture mechanics based "transfer functions" to relate gauge and structural cracks is discussed in [20–25] along with techniques for designing a crack gauge to give a desired response for the application of interest. Extensive laboratory testing has verified the transfer function calculations and has shown the crack gauge approach to have a firm foundation in fracture mechanics theory. Briefly, the theoretical model assumes the coupon is attached to the structure (by adhesive bonding, for example) in such a way that when a remote stress, σ_S, is applied to the structure, an effective stress, σ_G, is transferred to the gauge. This relationship can be represented by

$$\sigma_G = f\sigma_S \tag{20}$$

It has been shown that the load transfer function, f, depends on parameters such as gauge geometry and material properties, but not on the stress level. The load transfer function can be found by stress analysis, based on either theoretical or experimental techniques. Noting that at any point in time the gauge and structure have witnessed the same number of cycles, N, equation (16) can be solved for cyclic life, N:

$$N = \int_{a_{iS}}^{a_{fS}} \frac{da}{F_S(\Delta K)} = \int_{a_{iG}}^{a_{fG}} \frac{da}{F_G(\Delta K)} \tag{21}$$

Here a is crack length, the subscripts i and f refer to initial and final crack lengths, and the subscripts S and G refer to structure and gauge, respectively. In particular, a_{fG} would be the current crack length measured in the witness sample, and a_{fS} is the corresponding crack length in the structure (i.e., the SHM metric of interest).

Assuming the Paris law of equation (14) for the crack growth rate function $F(\Delta K)$, equations (20) and (21) give

$$N = \int_{a_{iS}}^{a_{fS}} \frac{da}{C_S(\Delta\sigma_S\sqrt{\pi a}\beta_S)^{m_S}} = \int_{a_{iG}}^{a_{fG}} \frac{da}{C_G(f\Delta\sigma_S\sqrt{\pi a}\beta_G)^{m_G}} \quad (22)$$

Note that crack length a is the variable of integration and neither f nor β depends on the cyclic load $\Delta\sigma_S$. Now if the gauge and structure have the same crack growth exponent, $m_S = m_G = m$ (a reasonable assumption if the gauge and structure are made of similar materials), $\Delta\sigma_S$ cancels and equation (22) reduces to

$$\int_{a_{iS}}^{a_{fS}} \frac{da}{C_S(\sqrt{\pi a}\beta_S)^m} = \int_{a_{iG}}^{a_{fG}} \frac{da}{C_G(f\sqrt{\pi a}\beta_G)^m} \quad (23)$$

While this expression no longer determines the cyclic life, N, it can be numerically integrated to provide information on the current structure crack length (a_{fS}) provided an initial structural crack length is assumed (i.e., the usual initial damage tolerant analysis assumption). Thus, equation (23) provides a "transfer function" that relates measured crack length in the coupon ($a_{fG} \equiv a_G$) with the current length ($a_{fS} \equiv a_S$) of an *assumed* structural crack as shown schematically in Figure 11. Note that all stress level terms effectively cancel, so that the relationship between the gauge and structure crack lengths is *independent* of loading. Indeed, a distinct advantage of this model over current fatigue tracking methods is that it requires no knowledge of the load history (i.e., the current gauge crack length a_{fG} is a measure of the actual applied load history). Experimental data verifying the accuracy of equation (23) as a transfer function are given in [20–25].

An alternate version of the witness sample transfer function model that employs two crack gauges is described in [24]. This "double gauge" procedure entails matching the output from two independent witness samples. In this case, however, one knows the current crack length in both coupons, and can solve equation (22) for an effective stress $\Delta\sigma_S$ that has been experienced by the structure. Once this effective stress has been determined, one would again use the output from one (or both) of the witness samples to determine the current length of the assumed structural crack. The advantage of this approach is that more general fatigue crack growth models may be employed for subsequent analysis than the limited Paris law.

6 CONCLUDING REMARKS

The objective of this article has been to introduce some of the challenges dealing with damage formation and growth that must be considered when designing an SHM system. Guaranteeing structural integrity requires anticipation of all possible failure mechanisms and then providing adequate resistance to these various threats. Whereas every effort is made to ensure high-quality construction, experience has shown that it is impossible to build complex mechanical devices that are free of initial manufacturing or material flaws. Moreover, accidental damage may occur during use, along with fatigue, corrosion, or other forms of material degradation. Although damage tolerant designs protect against such trauma by keeping stress levels small, employing damage resistant materials, and providing crack-arresting structural features, continued SHM during service is an important step for scheduling maintenance actions in the most cost-effective manner.

The foundations for structural integrity have been compared with the stability provided by a "three-legged stool". One leg is a rigorous inspection program that locates manufacturing or service-induced damage before it can lead to failure. Another leg is superior residual strength that ensures the catastrophic fracture load always exceeds the largest applied service load. The third leg of the damage tolerance stool is a long subcritical crack growth life that gives many opportunities to locate and repair any damage that does develop before it grows to a size that causes final fracture. These three legs work together to ensure structural integrity throughout the planned service life.

Since individual usage can fluctuate widely from the fleet average, however, different damage states can develop in particular structures at a given time, leading to a variation in required inspection intervals. Fatigue damage is especially sensitive to actual usage, and is difficult to predict without detailed knowledge of the unique service loads experienced by a given

component. SHM provides an extra dimension to the "three-legged stool" by providing the means to tailor inspection intervals and maintenance actions to individual members. Thus, SHM provides the means to maintain integrity in the most cost-effective manner.

END NOTES

[a.] Here the "engineering" stress S is replaced with the corresponding "true" stress σ (see [3–5]).

REFERENCES

[1] Riley WF, Sturges LD, Morris SH. *Statics and Mechanics of Materials: An Integrated Approach, Second Edition*. John Wiley & Sons, 2002.

[2] Gere JM. *Mechanics of Materials, Fifth Edition*. Brooks/Cole, 2001.

[3] Dowling NE. *Mechanical Behavior of Materials, Engineering Methods for Deformation, Fracture, and Fatigue, Third Edition*. Pearson Prentice Hall, 2007.

[4] Grandt Jr AF. *Fundamentals of Structural Integrity: Damage Tolerant Design and Nondestructive Evaluation*. John Wiley & Sons, 2004.

[5] Bannantine JA, Comer JJ, Handrock JL. *Fundamentals of Metal Fatigue Analysis*. Prentice-Hall: Englewood Cliffs, NJ, 1990.

[6] Stephens RI, Fatemi A, Stephens RR, Fuchs HO. *Metal Fatigue in Engineering, Second Edition*. John Wiley & Sons: New York, 2000.

[7] Tavernelli JF, Coffin Jr LF. Experimental support for generalized equation predicting low cycle fatigue. *Transactions of the ASME, Journal of Basic Engineering* 1962 **84**(4):533.

[8] Manson SS. Discussion of Reference 7 above. *Transactions of the ASME, Journal of Basic Engineering* 1962 **84**(4):537.

[9] Williams ML. On the stress distribution at the base of a stationary crack. *Journal of Applied Mechanics* 1957 **24**:109–114.

[10] Tada H, Paris P, Irwin G. *The Stress Analysis of Cracks Handbook*. Paris Productions: St. Louis, MO, 1985.

[11] Rooke DP, Cartwright DJ. *Compendium of Stress Intensity Factors*. Her Majesty's Stationary Office: London, 1976.

[12] Murakami Y (ed). *Stress Intensity Factors Handbook*. Pergamon: New York, 1987.

[13] Skinn DA, Gallagher JP, Berens AP, Huber PD, Smith J. *Damage Tolerant Design Handbook, A Compilation of Fracture and Crack Growth Data for High Strength Alloys*. CINDAS/JSAF CRDA Handbooks Operation, Purdue University: West Lafayette, IN, 1994.

[14] Paris P, Erdogan F. A critical analysis of crack propagation laws. *Journal of Basic Engineering* 1963 **85**:528–534.

[15] ASTM Standard E 647, Standard test method for measurement of fatigue crack growth rates. *Annual Book of ASTM Standards*. American Society for Testing and Materials: West Conshohocken, PA, 2002; Vol. 03.01.

[16] Bucci RJ. Selecting aluminum alloys to resist failure by fracture mechanisms. *Engineering Fracture Mechanics* 1979 **12**(3):407–441.

[17] Schijve J. Fatigue crack growth under variable-amplitude loading. *ASM Handbook, Fatigue and Fracture*. ASM International: Materials Park, OH, 1996; Vol. 19, pp. 110–133.

[18] Crews Jr JH. Crack initiation at stress concentrations as influenced by prior local plasticity. *Achievement of High Fatigue Resistance in Metals and Alloys*, ASTM STP 467. American Society for Testing and Materials: West Conshohocken, PA, 1970; p. 37.

[19] ASTM Designation E-1049-85, Standard practice for cycle counting in fatigue analysis. *Annual Book of ASTM Standards, Metals–Mechanical Testing: Elevated and Low-Temperature Tests; Metallography*. American Society for Testing and Materials: West Conshohocken, PA, 2000; Vol. 03.01.

[20] Gallagher JP, Grandt Jr AF, Crane RL. Tracking crack growth damage in US Air Force Aircraft. *Journal of Aircraft* 1978 **15**(7):435–442.

[21] Ashbaugh NE, Grandt Jr AF. Single-edge-cracked crack growth gauge for monitoring possible fatigue crack growth. *Service Fatigue Loads Monitoring, Simulation and Analysis* 1979 **ASTM STP 671**: 94–117.

[22] Ori JA, Grandt Jr AF. Single-edge-cracked crack growth gauge. *Fracture Mechanics* 1979 **ASTM STP 677**:533–549.

[23] Dumanis-Modan A, Grandt Jr AF. Development of a side-grooved crack gauge for fleet tracking of fatigue damage. *Engineering Fracture Mechanics* 1987 **26**(1):95–104.

[24] Dumanis, A, Grandt Jr AF. Development of a double crack growth gauge algorithm for application to fleet tracking of fatigue damage. *Proceedings International Committee on Aeronautical Fatigue 21st Conference, 15th Symposium*. Jerusalem, Israel, June 1989.

[25] Gates MD, Grandt Jr AF. Crack gauge approach to monitoring fatigue damage potential in aircraft. *1997 Society for Experimental Mechanics Spring Conference on Experimental and Applied Mechanics*. Bellevue, Washington, DC, 2–4 June 1997.

Chapter 10
Failure Modes of Aerospace Materials

Kumar V. Jata[1], Ajit Roy[1] and Triplicane A. Parthasarathy[1,2]

[1] *Air Force Research Laboratory, Wright Patterson Air Force Base, Dayton, OH, USA*
[2] *UES Incorporated, Air Force Research Laboratory, Wright Patterson Air Force Base, Dayton, OH, USA*

1 GENERAL REMARKS ON FAILURE PROCESS

Substantial research has been performed in the last few decades on understanding of nucleation and growth of cracks in materials, and as a result, failure modes on laboratory coupons are well understood for metallic alloys, ceramics, ceramic matrix composites (CMCs), organic matrix composites (OMCs), and carbon–carbon (C–C) composites. However, failures in the actual components in near-operational environments are still not well understood. The topic is too broad to cover in this article and, therefore, the discussion is restricted to laboratory coupon research in metals and OMCs.

Many of the structural components in use today are manufactured from metallic alloys. Aluminum alloys are most commonly used as airframe materials although the recent trend has been to use composites. Ti alloys are also used but in hotter sections of the airframe, and also as fan and compressor blades in turbine engines. Ni-based alloys (single crystals and equiaxed microstructures) are used in the high-temperature section of the turbine engine. In case of metals, it is important to have a good knowledge of chemical composition, primary processing and secondary processing methods used to produce the material and components, grain size and inclusion (impurity) content, and the morphology, shape, and size of strengthening precipitates. Grain size has a major role in determining the strength, ductility, fracture toughness, threshold for fatigue crack propagation rate, and creep. Strengthening precipitates determine strength, low-cycle fatigue life, fatigue crack growth resistance, and environmental fracture resistance. Inclusion content depends on the purity of an alloy. Large inclusion size is detrimental to fatigue crack initiation and result in low fracture toughness.

 ISBN: 978-0-470-05822-0.

The 7000-series and 2000-series aluminum alloys are commonly used as structural materials for airframe fuselage construction and wing structure. The 7000 series is high in strength and the 2000 series belongs to the damage tolerant class. Zn, Mg, and Cu are the strengthening elements for the 7000 series, whereas Mg and Cu are for the 2000 series; for both alloys Ti and Cr/Zr/Mn control grain size. Ti controls grain size during casting the ingot and Cr/Zr/Mn controls grain size during primary processing and solution heat treatment.

In aluminum alloys, primary processing varies depending on whether the product form is a forging, sheet, or a thick plate. Depending on this primary processing, fracture varies. For example, in the case of a thick plate, commonly employed for bulkheads and wing structure, grain size and chemical composition from the mid thickness of the plate to the top or bottom surface can result in a strength variation and crack propagation direction.

Secondary processing refers to the solution heat treatment temperature (400–550 °C depending on the alloy) and the aging treatment applied (120–170 °C depending on the alloy and desired temper) to attain desirable strength and damage tolerant characteristics in precipitation-hardened alloys. The solution treatment temperature, along with the amount of grain size refining elements listed above, can affect recrystallization and grain growth. Also, different aging temperatures and times can affect the precipitate distribution, size, and coherency of the precipitate. All these factors, i.e., grain size and precipitates, affect the strength, fracture toughness, and crack growth properties of the material.

The impurity content in alloys, primarily Si, and Fe in Al alloys, greatly affects ductility, fracture toughness, and nucleation of fatigue cracks. Obviously, higher impurity content results in inferior properties.

Besides the chemistry and heat treatments, defects arising from improper machining during component manufacturing can play havoc in failures. If sharp radii are present in the structure, stress concentration results causing fatigue crack nucleation if the structure is under fatigue loading or can become sites for promotion of brittle fracture during static loading in an otherwise ductile material.

In almost all metallic structures, large inclusions such as those formed from impurity content can give rise to premature crack nucleation under fatigue loading conditions. In very high pure alloys, slip bands can take over the nucleation of fatigue cracks. The load range then influences further crack growth. The later sections elaborate these points. The thickness of the structure can also play a role by influencing whether the structure would be under plane-strain or plane-stress conditions. Plane-strain conditions reduce the strain required for fracture. Industrial applications, therefore, minimize the use of thick structures to avoid plane-strain conditions unless required for load-carrying purposes. However, in some cases such as airframe bulkheads of fighter aircraft, it is not possible to avoid thick structures. Fatigue cracks initiate from the bolt-holes and the critical crack length to fracture (the crack length that corresponds to fracture toughness of the alloy) must be evaluated on thick sections of the alloy rather than thin sections. In the case of thin structures where plane-stress conditions exist, failure mode will shift from a catastrophic plane strain (with very little ductility) to a more slant fracture or a ductile fracture.

Gaseous species in the environment accentuate fracture. A very common static fracture that occurs is in the case of landing gear steels, where an improper electroplating process can lead to hydrogen embrittlement. (Hydrogen is adsorbed/absorbed into an alloy, such as steels, during electroplating and thus must be removed by a post-heat-treatment process.) Some alloys, high-strength aluminum alloys and high-strength steels, are susceptible to stress corrosion cracking (SCC) in the presence of aqueous environments such as high humidity and sodium chloride solution. A premature and unpredictable fracture may also occur when the structure is under fatigue load and environment-assisted crack growth (corrosion fatigue) can rapidly occur.

Airframe structures also employ Ti-6Al-4V and Ti-6Al-2Sn-4Zr-6Mo alloys but on a much smaller scale owing to prohibitive cost. Currently, a number of Ti alloy airframe applications require welding and the welding process of large structures for service conditions is difficult. During welding, tensile residual stresses are introduced. Their removal is critical as they promote unpredictable failures. Nickel alloys, used as engine materials, are much more complex than the aluminum class of alloys used in the airframe. Extensive database covering the failure modes for in-service environments are discussed

at length in handbooks such as the ASM Metals Handbook on fractography.

In the following section, physics of failures for metals is introduced and summarized.

2 PHYSICS OF FAILURE OF METALS

2.1 High-level classification

At a high level, failure modes in metals or alloys can be broadly classified under three categories: (i) deformation, (ii) fracture, and (iii) material loss (Table 1).

2.1.1 Deformation

A metallic alloy can fail to perform an assigned function owing to excessive deformation either arising from low-yield strength or through creep at high temperatures. Most of the plasticity associated with yield is athermal, whereas in the case of creep it is thermal. In both cases, plasticity plays a key role as a failure mode even though fracture has not occurred because the component is prevented from performing its function. Creep deformation resulting from applied stress and high temperature is a function of time. Many of the components that undergo creep deformation are designed not to exceed a certain amount of deformation under a given applied load over a specified time interval. The amount of creep deformation allowed in a design can lie in either the primary creep regime or secondary regime.

2.1.2 Fracture

Fracture is a very broad term; however, at a higher level it can be classified as static, dynamic, and creep rupture. Static fracture occurs when the load is applied monotonically, thus overloading or overstressing the material. Cyclic loads are absent in static fracture situations. Creep rupture (not deformation as discussed above) is the second major mode of failure that occurs in high-temperature components. Rotating engine parts such as turbine blades are influenced by centrifugal forces at high temperatures for prolonged periods. Excessive creep deformation may initially occur but continued exposure may lead to creep rupture, which may be owing to either a poor selection of materials or a poor definition of the operational environment. Creep rupture is a parameter commonly employed in the design of pipes carrying liquid through a boiler or a nuclear reactor. Creep in the pipe not only reduces the cross section of the pipe but also promotes the formation of small cavities over time at the operating temperature. The cavities form either intragranularly or at the grain boundaries, which eventually link up resulting in failure through creep rupture. The formation of cavities is a thermally activated process.

2.1.3 Dynamic fatigue

Major attention has been focused over the last few decades on the field of fatigue as the majority of failures occur under fatigue loading. When a component is subjected to cyclic stress, fatigue failure occurs with plasticity accumulation. Fatigue properties can be determined by using smooth bars that

Table 1. Classification of Failure Modes in Metallic Alloys

Deformation	Yield	Athermal plasticity	f (stress)
	Creep	Thermally activated plasticity	f (temperature, time, stress)
Fracture	Static	Athermal/acyclic fracture	f (stress)
	Creep rupture	Thermally activated cavitation leading to fracture	f (temperature, time, stress)
	Dynamic fatigue	Cyclic plasticity leading to nucleation/growth of cracks	f (cyclic stress, cycle#)
Material Loss	Corrosion	Athermal loss of material	f (time, environment)
	Oxidation	Thermally activated	f (time, temperature, O_2 pres)

can be subjected either to stress-controlled or strain-controlled cyclic loads. Smooth bars represent failure in the absence of cracks in the structure. In stress-controlled fatigue testing (high-cycle fatigue), applied stress is below the yield stress and very little plasticity is accumulated. In the case of strain-controlled fatigue (also known as *low-cycle fatigue*), applied stresses are beyond the yield stress of the material and the accumulated plasticity is considerably higher than that in high-cycle fatigue, resulting in fewer cycles to failure. Fracture mechanics tests are conducted for the evaluation of fatigue crack resistance in the presence of a sharp crack. Current fracture mechanics approaches use atomically sharp cracks or microstructural inhomogeneities as crack nucleation sites. In all fatigue situations, the applied stress range (maximum stress minus minimum stress) plays a key role. However, in the fracture mechanics approach, the stress intensity range is the parameter that extends the crack and crack growth rates, which dictate the remaining life. The number of cycles to failure are computed for a given crack size to reach a critical size. The presence of small cracks (equivalent to the size of a microstructure unit such as grain size) is detrimental to engine components; therefore, much attention has been focused in the literature on studying small crack behavior in engine component materials.

2.1.4 Material loss

The last failure mode listed in Table 1 is the loss of material that can occur owing to either corrosion or high-temperature oxidation. Loss of material in a corrosive environment is referred to as *general* or *uniform corrosion* and is strongly dependent on the environment and time of exposure. Many times, components undergoing such corrosion are either repaired by grinding out the corrosion or replaced after a certain amount of material loss has occurred beyond the amount allowed in the design of the component. Note that the time of exposure in corrosive environments is a key parameter. Many other corrosion mechanisms exist and these are discussed in the next level of failure processes. Material loss at high temperatures usually involves material oxidation. For a first-order approximation, material loss by such a mechanism depends on the partial pressure of O_2 and time and temperature of exposure. In this case, loss of material could be rapid as compared to the material loss experienced in the corrosion process.

2.2 Second-level classification

In this section, the factors associated with the first-level failure mechanisms, namely, yield, creep, static fracture, creep rupture, dynamic fatigue, corrosion, and oxidation, are discussed. Here, the goal is to provide a discussion of the microstructural parameters such as grain size, strengthening precipitates, and second-phase particle content that control these material failure modes.

2.2.1 Deformation

Depending on the service temperature, the two material properties that can lead to failure are the yield strength and creep resistance. If one needs to monitor failures due to large plastic deformation, then the athermal plasticity of the material in question must be considered. From the microstructure point of view, the logical question is "what material factors control the athermal plasticity?" For traditional structural engineering materials, the yield strength is inversely proportional to the square root of the grain size (a fundamental microstructure parameter) through the Hall–Petch relation:

$$\sigma_y = \sigma_0 + kd^{-n} \tag{1}$$

where σ_y is the yield strength, σ_0 is the yield strength of a single crystal, d is the grain size, k is the Hall–Petch coefficient, and n is the Hall–Petch slope, typically equal to 0.5.

In conventional airframe materials such as the aerospace aluminum alloys, grain size is controlled by special alloying elements [1], which have a low solid solubility limit at high temperatures and, hence, precipitate out during casting of the alloys. For example, titanium is normally added to castings to control the grain size. In addition, small amounts of manganese, chromium, scandium, and zirconium are added to control grain size and grain shape. Zirconium and scandium are the most effective elements because they form coherent dispersoids; for example, Al_3Zr dispersoid is a second phase that is dispersed in the matrix. They are also very effective in preventing

recrystallization. Apart from the grain size, a second factor that controls plastic deformation is the presence of second-phase particles. These second-phase particles are termed *strengthening phases* or *precipitates*. Additions of zinc and copper in combination with magnesium result in the formation of precipitates that provide the necessary strengthening in aluminum alloys. Many high-strength alloys (7xxx and 2xxx series) used in military and commercial aircrafts contain these elements.

Dislocations are line defects, which propagate resulting in plastic deformation. In an alloy with precipitates, the dislocations either cut through these obstacles or bypass them by leaving loops of dislocations around them. When dislocations shear these precipitates, the strength of a material increases with the increase in volume fraction and radius of the precipitate. When the precipitate size exceeds a "cutoff radius", it is too large to be cut by dislocations. This causes the dislocations to bypass the precipitates during deformation resulting in lower strength. The strength in the cutting regime (i.e., the regime where dislocation cuts through the precipitate) is determined by a variety of factors. The most prominent effect arises from the antiphase boundary (APB) energy in the case of ordered precipitates. The coherency strain (mismatch in lattice parameter between matrix and precipitate) plays a significant role in some systems. Other factors include the volume fraction, size of the precipitates, and friction stress in the matrix and precipitate. Equations describing these multiple effects can be quite complex. In the case of the bypass regime, the strength is dependent on fewer factors as given below:

$$\sigma_{\mathrm{y}} = \frac{\alpha G \mathbf{b}}{l} \tag{2}$$

$$l = \frac{2r}{f^{0.5}} \tag{3}$$

where l is the distance between precipitates, r is the radius of the precipitates, f is the volume fraction, a is a geometrical factor typically equal to 0.5, G is the shear modulus, and $\mathbf{b}$ is the Burgers vector of the dislocation, typically half the lattice parameter.

Thus, the second-phase/precipitate particles lose their ability to contribute to yield strength of the material when they exceed a certain size. In many aluminum alloys, for example, this loss in strength can occur at prolonged exposures to high temperature (approximately 150 $^{\circ}$C) owing to severe coarsening of the precipitates (severe over aging), with the level of coarsening dependent on the individual alloy and the precipitate-forming elements. Many of the designated tempers such as T3, T6, T8, and T7 refer to the extent to which the alloys have been aged at temperatures prior to use. Each temper thus corresponds to a prescribed strength, and each temper is suitable for different applications. However, at present, an on-line monitoring approach for structures that accounts for differences in these factors does not exist.

2.2.2 Creep deformation and rupture

Failures at high temperatures (<0.5 of melting point of material in kelvin) due to excessive deformation caused by creep are not uncommon. Components are not usually permitted to deform beyond a certain amount of strain or thermal plasticity over a specific period of time under load. However, in some instances the amount of creep is not critical, but failure due to creep rupture has to be avoided (such as in the case of a pipe carrying a hot fluid where a certain amount of deformation can be tolerated, but not rupture of the pipe). Creep rupture data, i.e., stress versus time to creep rupture, at desired temperatures are typically used. As shown in Table 1, to avoid such failures, stress, time, and temperature are key variables, and their effects need to be understood.

The dominance of a specific creep mechanism in any alloy is dependent on the applied stress and temperature [2]. Constitutive equations for each creep mechanism are readily available, and by solving these equations for a desired temperature and stress range, the strain rates can be plotted as isostrain rate contours in an Ashby map (Figure 1).

At high stress and high temperatures, creep rates are determined by a "power law", or "dislocation creep", regime where the creep rates vary with stress rapidly, as given by

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = A \exp\left(-\frac{Q}{RT}\right)(\sigma - \sigma_{\mathrm{th}})^n \tag{4}$$

where ε is the creep strain, t is the time, A is a material constant, Q is the activation energy for

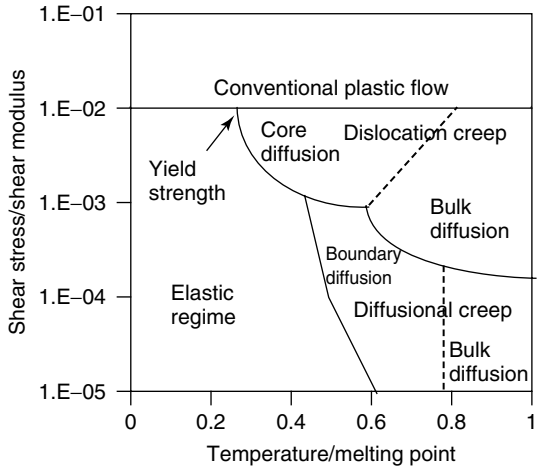


Figure 1. A notional creep deformation mechanism map showing isocreep strain rate contours, and creep mechanisms that can occur in metallic alloys under applied stress and temperature.

creep, R is the universal gas constant, σ is the applied stress, and σ_{th} is a threshold stress that depends on the microstructural state of the material.

At lower stresses, creep rates vary linearly with stress and are said to be undergoing "diffusion creep". Within this regime, at higher temperatures, the creep rate is determined by

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = 10\frac{\sigma\Omega}{kT}D_{\mathrm{L,o}}\exp\left(-\frac{Q_{\mathrm{L}}}{RT}\right)\frac{1}{d^2} \qquad (5)$$

where Ω is the atomic volume, k is the Boltzmann's constant, $D_{\mathrm{L,o}}$ is the preexponent for bulk diffusion of atoms, Q_{L} is the activation energy for bulk diffusion, and d is the grain size. At lower temperatures, the creep is limited by "Coble" creep and the rate is given by

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = 47\frac{\sigma\Omega}{kT}(\delta_{\mathrm{b}}D_{\mathrm{b,o}})\exp\left(-\frac{Q_{\mathrm{b}}}{RT}\right)\frac{1}{d^3} \qquad (6)$$

where $D_{\mathrm{b,o}}$ is the preexponent for grain-boundary diffusion, δ_{b} is the grain-boundary width, and Q_{b} is the activation energy for grain-boundary diffusion.

The isostrain rate contours in a plot of applied shear stress (normalized with respect to shear modulus) versus operating temperature (normalized with respect to melting temperature in kelvin) represent the boundaries that delineate different creep mechanisms. Figure 1 shows that at low stresses and high temperatures, Nabarro–Herring creep deformation dominates, where atomic diffusion takes place from boundaries normal to the stress axis to boundaries parallel to the stress axis. Vacancies diffuse in the opposite direction, i.e., from boundaries parallel to the stress axis to boundaries perpendicular to the stress axis. It is important to note here that the microstructural factor that governs the creep strain rate in this regime is proportional to $1/d^2$. Even at lower stresses, creep deformation occurs through vacancy migration but along grain boundaries, and the strain rate is very sensitive to the number of grain boundaries through a $1/d^3$ relationship. There is no dislocation motion in the high-temperature and low-stress regimes, and no grain elongation occurs. Grains, however, do rotate resulting in loss of the original texture and sliding of grain boundaries occur. At higher stress levels, creep is controlled by dislocation movement and is independent of grain size; this aspect is indicated as the dislocation creep regime in Figure 1. Grains here elongate and new textures form. Currently, in order to decipher the specific mechanism that is occurring in a component, specimens from the component are sectioned and prepared for scanning or transmission electron microscopy. Voids, grain size/orientation modification, and formation of new textures can then be studied and documented for any material component after the failure has occurred.

2.2.3 Static fracture

In aircrafts, there are very few metallic components that fail purely owing to a static overload mechanism. However, overload fracture can occur after a crack has first grown by a different fracture mechanism, for example, by stress corrosion. In such a case, a crack grows to a critical size by such a mechanism, and then the final fracture occurs by overload. Rather than using percent elongation or percent area reduction, fracture toughness ($K_{1\mathrm{c}}$) is used as an engineering indicator for a material's ability to withstand overloads under brittle conditions or plane-strain loading conditions.

In order to obtain $K_{1\mathrm{c}}$, a sharp crack is imbedded in a fracture mechanics specimen and fracture toughness

is evaluated using an approach outlined in the American Society for Testing of Materials (ASTM) Standard E399. J_{1c} (elastic–plastic fracture mechanics parameter) is used when the material exhibits considerable plasticity during loading and crack extension; therefore, linear elastic fracture mechanics is not applicable. The tearing modulus defined as dJ/da represents crack growth toughness and has been evaluated recently as an additional parameter to evaluate material fracture resistance. For thin sheet materials, such as the gauge used for aircraft fuselage, R-curves (crack resistance behavior) are generated on very large specimens according to the ASTM standards.

Figure 2 summarizes various microstructural parameters that govern the fracture toughness and tearing modulus for precipitation-hardened aluminum alloys routinely used for aircraft structures [3]. It may be observed that fracture initiation toughness (K_{1c}, J_{1c}) and crack growth toughness (also known as *tearing modulus*, T_r) decrease with an increase in yield strength. The figure also attempts to illustrate that for constant yield strength, as microstructural parameters such as grain size, strengthening precipitates, and impurity content are refined (reduced in size or extent), K_{1c}, J_{1c}, and T_r increase. Fine second-phase particles improve homogeneity of slip, thereby improving fracture toughness parameters. Homogeneity of slip also promotes ductile failure, void nucleation, and growth. Larger grain-boundary precipitates and impurity particles (also known as *inclusions*) decrease toughness through promotion of brittle fracture. Thus, grain size, strengthening phases, and second-phase particle distribution play a key role in fracture initiation, and crack growth toughness of many engineering alloys.

2.2.4 Fatigue

Fatigue is one of the dominant modes of failure, and has been investigated more thoroughly than any other fracture mechanism; numerous papers have been published in the technical literature. When a stress-controlled or high-cycle fatigue test is performed on a smooth or notched specimen, the number of fatigue cycles to failure represents the fatigue life. The cycles to failure normally depend on the R-ratio (the ratio of minimum to maximum applied stress in a cycle $= \sigma_{min}/\sigma_{max}$). In the case of high-cycle fatigue, the maximum applied stress is kept below the yield stress of the material. Cracks usually initiate on the surface either at the site of a large inclusion(s) or at a grinding or a tool mark due to improper machining of the component. In high-cycle fatigue applications, compressive residual stresses are usually induced on the surface either through shot peening or most recently by laser shock peening to increase the component life; however, the compressive residual stresses usually decay with fatigue cycling, so shot peening is not a permanent solution to the fatigue failure problem. Research has shown that the key

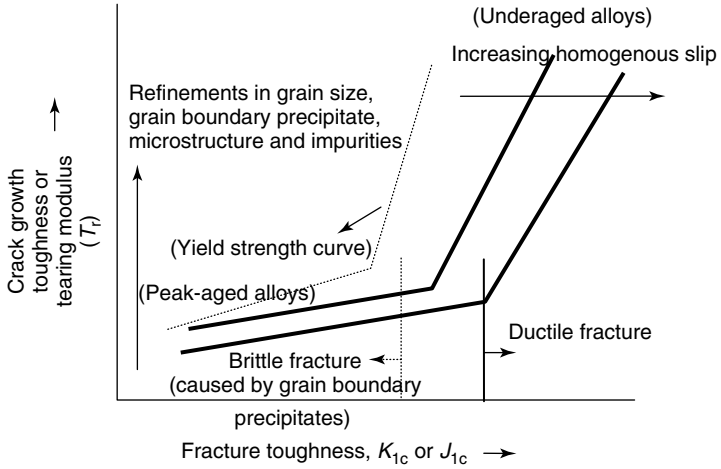


Figure 2. Relationship of crack initiation toughness and tearing modulus (or crack growth toughness) to various microstructural parameters in precipitation-hardened Al alloys.

microstructural parameter in high-cycle fatigue is the size of the largest inclusion or the range of size of inclusions that participate in the initiation of fatigue fracture.

In strain-controlled fatigue, a specimen is cycled within a prescribed strain range that could be below or above the yield of the material. Cyclic plasticity is accumulated as the specimen is cycled and the total accumulated plastic deformation dictates the (low-cycle) cycles to failure. Here, the plasticity is accumulated through slip or dislocation movement from the bulk of the material to the surface. Cracks initiate in slip bands and eventually link up causing failure. The Coffin–Manson law relates the applied plastic strain amplitude, $(\Delta\varepsilon_p/2)$, to the number of strain reversals to failure, $2N_f$, as given by the equation

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon_f'(2N_f)^c \tag{7}$$

where ε_f' is the fatigue ductility coefficient (the value of the plastic strain amplitude when failure occurs in one strain reversal) and c is the fatigue ductility exponent (obtained by the slope of the line that relates the plastic strain amplitude to the cycles to failure).

However, a more important problem in structural fatigue is failure through fatigue crack propagation. Here, the assumption is that when a component is inserted into service after nondestructive evaluation (NDE) inspection, a crack population below the resolution limit of the NDE equipment or due to human error will go undetected. Cracks below the resolution limit of the NDE equipment are assumed to be present in the structure and then the fatigue crack growth law for that particular material in question is applied to estimate the number of cycles that will be needed to grow the crack or cracks to a critical crack size and cause failure. For airframe structures, it is common to assume that cracks below the size of 30 mil (0.762 mm) will go undetected. This number usually gets larger when the structure becomes complex from the point of view of number of material layers and the presence of sealants. As shown in Figure 3, many complex aerospace structures can be made up of multiple layers, sometimes up to five, separated by corrosion prevention compounds known as *sealants*. In addition, the multilayers can also be separated by shims with the structure held together by fasteners, which may or may not consist of the same material as the multilayers. All these factors make the detection, location, and sizing of the cracks extremely difficult, particularly when the NDE inspection is performed without disassembling the structure.

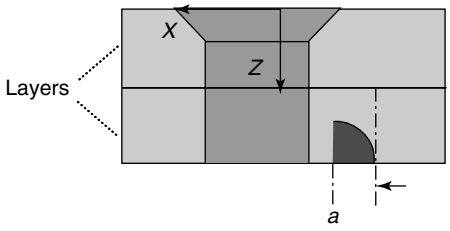


Figure 3. Aircraft lap joint, splice plates with a corner crack near the fastener.

Unlike airframe structures, a small flaw present in a rotating turbine engine component can create a major loss of the entire engine or even loss of the aircraft. Thus, the study of small cracks (cracks whose dimensions are of the same order as that of the grain size of the material or less) is of extreme importance in materials that are used for gas turbine engines. A number of investigations have been performed to understand the growth behavior of small cracks with respect to microstructure, residual stress, and grain size. The reader is referred to the article, [4], as a starting point on this subject. Fatigue crack growth rates (da/dN, where da is the increment in crack size and dN is the increment in number of applied load cycles) for larger cracks for a few common materials are shown in Figure 4 as a function of the stress intensity range ($\Delta K = \Delta\sigma\pi\sqrt{a}$), also termed the *crack driving force* [2]. When the crack growth rate decreases to very small values below 10^{-10} m/cycle, the corresponding stress intensity range is designated as ΔK_{th} and known as the *fatigue threshold stress intensity range*. In Figure 4, the curve designated as "striation model" refers to the *Paris law* and is determined by equation (7). Many aircraft structural components are designed for fatigue crack growth rates corresponding to those in Region II of the fatigue crack growth curve where Paris' law is obeyed and the fatigue crack extends by a striation mechanism given by equation (9) below,

$$\frac{da}{dN} = (\Delta K)^m \tag{8}$$

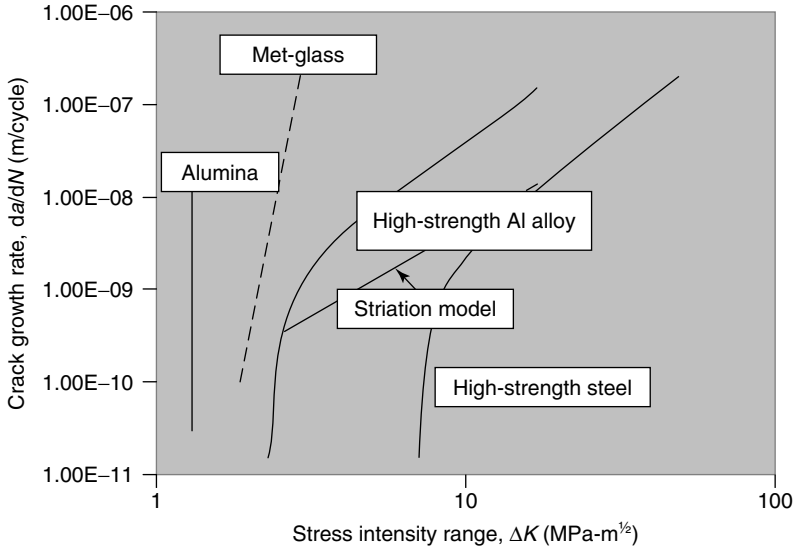


Figure 4. Fatigue crack growth rate versus stress intensity range of a number of materials. Striation models use the Paris law equation.

$$\delta = \beta \left(\frac{\Delta K^2}{E\sigma_{ys}} \right) \tag{9}$$

where δ is the crack-tip opening displacement related to crack growth rate per cycle in the material and β is a constant. In Figure 4, brittle materials such as amorphous glass and alumina exhibit very small crack opening displacement and, hence, need lower driving force or stress intensity range to extend the fatigue crack. In the case of materials such as high-strength steels and aluminum alloys, the stress intensity range for the cracks to grow is larger. Note that the crack opening displacement for a given stress intensity range is related to elastic modulus (E) and yield strength (σ_{ys}) of the material.

As discussed above, for the lower end of the stress intensity range, crack growth rates asymptotically decrease to very small values (10^{-11} m/cycle) and the stress intensity range reaches a fatigue threshold stress intensity ΔK_{th}. In this regime, fatigue crack growth rates are strongly influenced by microstructural parameters such as grain size and second-phase particles. In engineering materials, it is often difficult to isolate the effects of grain size from other microstructural parameters such as second-phase particles. Nevertheless, in precipitation-hardened aluminum alloys, it has been shown that for a uniform grain size, fatigue crack growth resistance is superior when strengthening phases are coherent (underaged alloys) and is inferior when they are incoherent (overaged Al alloys). These effects of strengthening phases are related to slip characteristics and nonlinearity of the crack path in the material. When particles are coherent, crack-tip plasticity is accumulated through planar slip and the fatigue cracks follow the slip planes resulting in a path that has many tilts. In the case of incoherent particles, nonplanar slip dominates, which results in a relatively straight crack path. The out-of-plane crack path (made up of tilted cracks) provides a higher fatigue crack growth resistance. Similarly, the fatigue threshold stress intensity range has also been shown to have a direct correlation with particle spacing and particle volume fraction. Particles with lower volume fractions and larger spacing decrease out-of-plane cracking resulting in larger fatigue thresholds [5–7].

Recently, a comparative study of grain size effects on crack propagation in nickel was performed [8]. The study showed that nickel containing nanosize grains had the least amount of nonlinearity in the crack path and exhibited the fastest crack propagation rates. Fatigue cracks in microcrystalline and conventional nickel with larger grains propagated with much more out-of-plane tilts providing slower fatigue crack

growth rates. In addition, fatigue threshold intensity factors for a number of important engineering alloys were compiled and analyzed and it was shown that the threshold stress intensity range increases with grain size [9]. Thus, the subject of fatigue is vast and the microstructural relationship to fatigue crack growth rate is clearly material dependent. Therefore, for prognosis, it is essential to know the material and its microstructural state so that the fatigue crack growth law corresponding to that particular material and material state in question can be applied.

2.2.5 Corrosion

Uniform corrosion is the term that is used in the corrosion literature and does not necessarily imply the loss of material through chemical reactions. The term is used even though the material loss is different at different locations. The corrosion does not impact structural integrity initially, but can do so if a large amount of material loss occurs. There are at least seven forms of corrosion [10] and of these the ones that affect the engineering structures include pitting, environmentally assisted cracking (EAC), and intergranular corrosion. SCC, corrosion fatigue, and hydrogen embrittlement fall under the category of EAC and, when they occur, can lead to very costly failures.

In the case of SCC, an existing crack in a material propagates under the combined influence of stress and a corrosive environment. The fracture toughness of the material is not altered owing to the environment but an existing crack can reach the critical crack size ($K_{1C} = \sigma\sqrt{\pi a_{cr}}$) through the combined influence of stress and environment. Fracture mechanics test methods exist to evaluate the threshold, K_{1scc} or K_{1EAC} (where EAC refers to environmentally assisted cracking, and the subscript "1" refers to crack opening mode 1, below which stress corrosion cracks will not propagate). The two regimes, "Region I" and "Region II", of SCC in fracture mechanics specimens or structures containing atomically sharp cracks are shown in Figure 5. Numerous reports exist in the literature for Al and Ti alloys where microstructure has been shown to play a dominant role in controlling the fracture mode. High-strength aerospace aluminum alloys

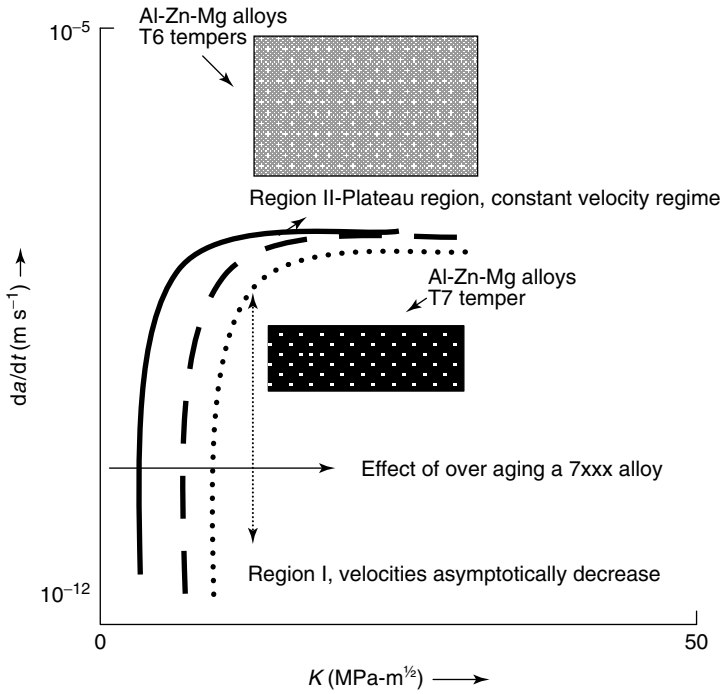


Figure 5. A schematic showing stress corrosion crack growth velocities as a function of stress intensity in high-strength Al alloys. Also stress corrosion crack growth rates are compared in T6 and T7 tempers for 7xxx alloys (notional).

have been investigated thoroughly for SCC in 3.5% NaCl (sodium chloride) solution. This electrolyte is of much interest since chloride ions contained in NaCl solution degrade these alloys significantly. A compilation of stress corrosion crack growth rates for 7079, 7039, 7049, 7075, and 7050 shows that the crack growth velocities (da/dt in $m\,s^{-1}$) are almost similar in the T7 temper, approximately $8 \times 10^{-10}\,m\,s^{-1}$. T7 is the overaged temper. However, many of the older aircraft (aging aircraft) used 7049, 7079, and 7075 in T6 temper that corresponds to a high-strength condition (peak aged) and were not optimized for SCC. In alloy 7079-T6, the SCC cracks grow at $10^{-5}\,m\,s^{-1}$ and at $10^{8}\,m\,s^{-1}$ in 7075-T7. Note that in this temper the stress corrosion crack velocity is much faster than $8 \times 10^{-10}\,m\,s^{-1}$.

However, the literature shows that the effect of aging (of aerospace aluminum alloys) on the SCC resistance is not that straightforward. It depends on the chemical composition of the alloy. For example, in some high-strength aerospace aluminum alloys, it has been shown that overaging (i.e., microstructure containing incoherent precipitates) can improve SCC thresholds, whereas in some alloys the thresholds were not affected; instead the plateau velocity in Region II (Figure 5) of the stress corrosion crack growth rate versus stress intensity was drastically improved [11]. For aluminum–lithium (Al–Li) alloys, the addition of trace elements (zinc or indium) has been shown to improve corrosion resistance or K_{1scc}. The overaged alloys also exhibit improved fatigue endurance limits under high-cycle fatigue (stress-controlled fatigue) conditions. Under corrosion-fatigue conditions in fracture mechanics samples, crack growth rates are found to be a function of cycling frequency. At low frequencies, environmental contributions increase and the crack growth rate increases. Recently, for Ti-8Al-1Mo-1V alloys in 3.5% NaCl solution, it has been shown through analysis of corrosion-fatigue data that the environmental contribution to fatigue increases as the frequency decreases [12]. A parameter that represents environmental contribution was plotted as a function of frequency and it was noticed that this parameter increased linearly as the frequency decreased from 15 to 3 Hz. At 3 Hz, the environmental contribution parameter leveled off. This value was found to be equivalent to the SCC threshold, K_{1scc}, of the alloy. Stress corrosion cracks can be identified as they branch out; however, careful analysis and experience are required to distinguish these cracks from others. Many of the fracture mechanisms and example fractographs can be found in [2, 10].

Hydrogen can also affect crack growth rates in many alloys including Ti-based alloys, which are of interest to the aerospace industry. In these alloys, hydrogen can accelerate crack growth rates by lowering the cleavage fracture stress and/or by forming titanium hydrides. The amount of hydrogen required to embrittle the Ti-based alloys depends on the volume fractions of various phases, stress state at the crack tip, temperature, and crack-tip strain rate. A large body of information on this subject has been published between 1973 and 1990 and the reader is referred to the textbooks referenced in this article.

2.2.6 Oxidation

In metallic alloys, material loss at low temperatures is dominated by the chemical action of the environment as discussed under corrosion. At higher temperatures, oxidation is the primary source of material loss. The rate at which a metal oxidizes varies from one metal to another depending upon the chemistry of elements that make up the alloy. In addition, the microstructural features also affect the oxidation rate. In a well-engineered alloy, the composition is tailored in order to form a very dense, tenacious, and adherent oxide film on the surface that prevents further oxidation by forming a physical barrier between the oxygen and the underlying material. For example, aluminum oxide has a very low permeability for oxygen. However, aluminum itself is a low-melting metal. Nickel has a high melting point and retains strength at higher temperatures, but nickel oxide is not a good barrier to oxygen owing to high permeability. Thus, an alloy of nickel and aluminum was conceived of in order to take advantage of the high-temperature capability of nickel and the oxidation barrier formation property of aluminum.

Real engineering alloys are much more complex than just containing two elements. Thus, the oxidation behavior is quite sensitive to actual composition as well as the microstructure of an alloy. During service, the alloy may change in composition, especially at the surface and in the microstructure throughout the material. For example, an engineering alloy of molybdenum and silicon with boron additions makes use of

the high-temperature capability of molybdenum and the low permeability of oxide of silica for the barrier [13, 14]. The silica barrier is formed using the silicon in the alloy, thus depleting the surface of the alloy in silicon. If the silica surface layer is damaged, the silicon content in the substrate may be insufficient to form a barrier once again to protect the base material. If a method by which the composition of the substrate underneath an oxide layer can be measured, then it is possible to predict the remaining life of the alloy. Properties that are sensitive to chemistry need to be identified and technologies that are suitable for measurement need to be invented for such applications.

In the current generation of nickel-based superalloys, the alloys are protected against oxidation by a layer of zirconia. This layer is, however, permeable to oxygen. The oxygen reacts and forms alumina beneath the thermal barrier. As this alumina layer grows in thickness, the thermal barrier becomes unstable and eventually spalls. Obviously, this spalling results in overheating of the Ni-based alloy underneath. It would be of great use if a method to measure the thickness of the alumina that forms under the thermal barrier layer becomes available. The dependence of the alumina layer thickness on the remaining life of the barrier layer is already known; thus, what remains is to come up with a method to measure *in situ* the alumina thickness beneath the thermal barrier layer.

3 FAILURE CHARACTERISTICS IN POLYMER MATRIX COMPOSITES

Polymer matrix composites (PMCs) are primarily made of two or more material phases of which the reinforcing phase in the form of fibers is held together by the binding material, commonly called as *matrix material*. Besides the fiber and matrix phases, other phases (such as nanoconstituents, interleave, fiber–matrix interface, etc.) are sometimes incorporated in composites for tailoring their specific properties or strength. The fiber and matrix materials possess distinctly different material characteristics, both in modulus and strength. Their distinctly different respective material properties cause stress concentration in the vicinity of the fiber and matrix interfaces resulting in complex failure modes, such as matrix failure, fiber breakage, fiber–matrix interface failure, etc. The relative degree of the difference in the fiber and matrix moduli and strength, along with their geometric configuration, dictates the initiation of these different failure modes. It is known that the interface response (either at the fiber–matrix or interply level of layered or three-dimensionally reinforced composites) critically influences the failure process. Engineered fiber–matrix interfaces are also implemented with controlled chemistry, and more recently with nanoengineering, to control the failure process more effectively. There have been extensive studies to analyze and to understand the physics of the failure process and failure modes of PMC since 1960s [15–18]. Although a great deal of understanding has been generated in the literature in developing PMC failure theories, the past development of PMC failure theories, due to the complex nature of the competing failure modes in composites, rely heavily on the extensive material testing, in which many a case becomes cost prohibitive for material insertion in structural components. Thus, recent research thrust on PMC failure understanding is toward developing phenomenological failure theories, minimizing material testing. A brief summary of the physics of the past and recent PMC failure theory development is discussed below.

Unlike metals, PMCs hardly exhibit any plasticity. The primary mechanism of energy absorption in PMC, beyond its usual elastic deformation, is through damage accumulation in the form matrix cracking, delamination between the lamina (plies) interfaces, and fiber breakage. A representative failure scenario in PMC of matrix cracks, delamination initiation, and fiber breakage, due to a tensile load, is illustrated in Figure 6. Typically in monotonous loading, the matrix crack, visible at the midsection at the bottom of the micrograph in Figure 6, grows with loading increase and hits the lamina interface and gets diverted at the lamina interface to initiate fracture surface at the lamina interface (known as *delamination*), as indicated in Figure 6. The presence of any large area of delaminated fracture surface at the lamina interface impairs composite performance (degrades structural rigidity and buckling resistance), especially under compressive and fatigue loadings. The accumulation of fiber breakage accelerates the failure process in PMCs and thus needs to be avoided

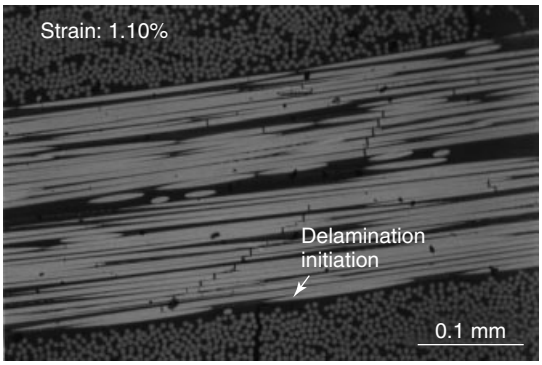


Figure 6. Damage in a $[90_2/0_2]_s$, laminate under a tensile load at a strain level of 1.1% indicating matrix cracks, initiation of delamination, and progression of fiber cracking prior to final failure.

while designing structural components with PMC. For designing composite parts, appropriate failure theories are required to quantify the damage accumulation process and assess these competing failure modes' effect on the final failure.

Owing to the presence of multiple phases in composites, the prediction of failure in PMC is rather much more involved than that of monolithic materials. The maximum stress and maximum strain criteria are very popular in predicting matrix cracking phenomena. The shear-lag theory is also commonly used in predicting the accumulation of the matrix cracking damage. Quantifying the delamination damage in PMC is rather a complex stress analysis problem. Fracture mechanics approach as well as composites failure under combined stress or strain components have been developed extensively to predict delamination in composites [15, 18].

It is observed that the failure in composites under combined loading, especially under combination of shear and normal stress components, the maximum stress or strain criteria are often not able to predict failure accurately. Tsai and Wu [19] in 1965 proposed interactive failure criteria, especially to predict failure under combined loading. The Tsai–Wu interactive failure criteria is expressed as

$$F_{ij}\sigma_i\sigma_j + F_i\sigma_i - 1 = 0 \qquad (10)$$

where, σ_i or $\sigma_j (i, j = 1, \ldots, 6)$ are the six stress components. The stress coefficients, F_{ij}, are dependent on the material properties and are to be determined by a few tests utilizing the directional material strength data as outlined in [19]. The Tsai–Wu failure criterion is considered to be a macroscopic failure criterion, mostly used for predicting laminate failure strength. Also, the criterion utilizes an interactive factor that is usually set empirically, requiring significant material testing.

The stress field in laminated composites (or PMC) in the vicinity of cutouts, open hole, or free edges, is not uniform. The material property disparity between the lamina near the lamina interface causes gradient stress field (in other words causes stress concentration) and at the free edge (the surface containing the thickness) of the laminate the stress components acting perpendicular to the lamina surface, $\sigma_z(\sigma_3)$, and $\tau_{xz}(\sigma_5)$, of the stress field become singular. An accurate prediction of such gradient and singular stress field is essential for reliable prediction of composite failure with open hole or bolted hole. Pagano and Pipes [20] provided the analytical solution to perform stress analysis to accurately capture the free-edge stress field. The nature of stress gradient such as the free-edge stress field for a $[30/-30/90]_s$ laminate, known to cause a highly gradient free-edge stress field, is shown in Figure 7. The true value of the free-edge stress is, however, influenced by the local material properties at the vicinity of the lamina interface, which is known to be somewhat different from that of the bulk lamina. Thus, even with sophisticated predictive tool developed by Pagano and Pipes [20], the use of pointwise failure criteria, such as Tsai–Wu, maximum stress, or maximum strain, becomes unreliable in open hole or bolted hole failure prediction, i.e., failure prediction in the presence of gradient stress field. To overcome this difficulty, Whitney and Nuismer [21] in 1974 introduced a criterion to average the gradient stress field over a characteristic length for predicting failure. This criterion is popularly known as *Whitney–Nuismer criterion*. The characteristic length in Whitney–Nuismer criterion is determined empirically through testing series of specimens of similar characteristic materials configuration. Thus, the utility of Whitney–Nuismer criterion is now in question owing to extensive materials and testing cost involved in developing strength allowable of materials.

Recently, Iarve *et al.* [22], to quantify the Whitney–Nuismer characteristic length more rigorously, introduced the critical failure volume (CFV)

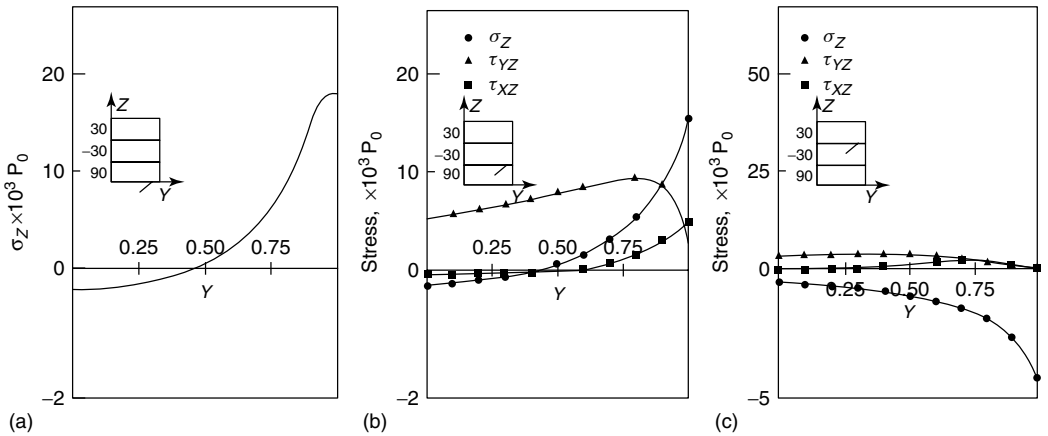


Figure 7. Interlaminar normal and shear stress distribution at ply interfaces of $[30/-30/90]_s$ laminates containing free edge under uniaxial tension loading. (a) [90/90], (b) [−30/90], and (c) [30/−30] interfaces. [Reproduced from Ref. 20. © Sage Publications, 1970.]

element concept. In order to minimize the empiricism in predicting failure in the presence of gradient stress field, they employed the Weibull statistics to tensile strength prediction in laminated composites with open holes. The statistical variability that is expected to exist in gradient stress field arising from the expected variability in the material properties is captured through the Weibull distribution concept. The CFV concept is based on the fact that a material sample under the applied load of a given volume has the following probability of failure under stress σ.

$$f(\sigma, V) = 1 - e^{-\frac{V}{V_0}B(\sigma)} \tag{11}$$

The CFV concept provides a systematic approach toward a methodology of predicting PMC failure in the presence of stress concentration and minimizing the volume of material testing in developing strength allowables.

Fiber–matrix interface failure is one of the prominent failure mechanisms influencing failures in PMCs. Even if the physics of the failure process is known through proper material characterization, unless an accurate value of the interface failure strength is known, optimizing the interface performance for tailored material design of PMC is not achievable. Traditional fiber–matrix interface testing is performed through fiber pullout, where interface failure happens under a combined shear and normal stress field and in the presence of stress concentration. Thus, determination of the interface strength from fiber pullout test is often inaccurate. Recently, Tandon *et al.* [23–25] developed a unique test using cruciform specimen geometry to determine the fiber–matrix interface strength. The cruciform specimen configuration eliminates the stress concentration at the failure surface and failure takes place under a single stress component (radial stress) when the winged arm of the cruciform specimen is perpendicular to the applied load [23]. The winged arm of the cruciform specimen can be oriented to any angle, other than 90°, to impose combined stress field on known ratio, as shown in Figure 8. The angle of the winged arm of the specimen is changed to apply combined normal and shear loading to determine the failure envelope, as shown in Figure 9. The experimental determination of such a failure envelope helps validation of failure criteria applicable to combined stress field, such as proposed by Gosse [26].

3.1 Failure in three-dimensional composites (3D PMC)

In spite of the superior strength-to-weight ratio advantage of PMC over metals, one of the weaknesses in laminated PMC is its poor interlaminar strength due to lack of fiber reinforcement through its thickness. The advent of textile composites utilizing the weaving and stitching (Z-pinning, etc.) technology overcomes this deficiency. PMC of three-dimensional

Figure 8. Geometry of cruciform specimen for fiber–matrix interface strength test. The white line at the center of the specimen reveals interface failure away from free-edge stress field.

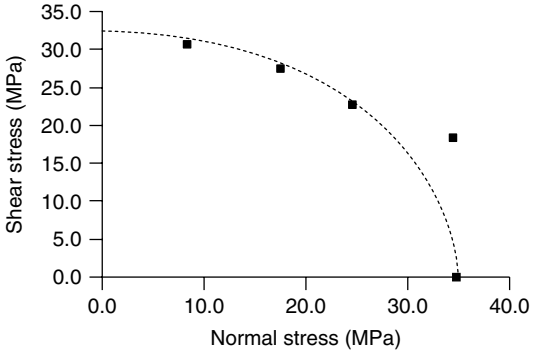


Figure 9. Failure envelope generated from cruciform tests. [Reproduced with permission from Ref. 24. © Begell House Inc., 2004.]

reinforcement (known as *3D PMC*) offers composite structures of complex shapes and contoured parts. Though the three-dimensional reinforcement offers improved through-the-thickness strength attributes, it imposes a complex failure scenario in the composite. During process curing of PMC, the matrix shrinks. In the case of laminated composites, such matrix shrinkage does not cause a major problem because the lamina is mostly able to accommodate the shrinkage, thus avoiding the lamina interface failure. However, in the case of 3D PMC, the interlocking effect imposed by the 3D reinforcement causes the lamina interface failure, as shown in Figure 10. Appropriate stress analysis is required to understand the interface cracking phenomena. Sihn and Roy [27, 28] and others have been developing three-dimensional stress analysis techniques to understand the matrix cracking process in 3D PMC. Accurate stress analysis tools are essential before one can apply failure criteria to understand the physics of the failure process in 3D PMC.

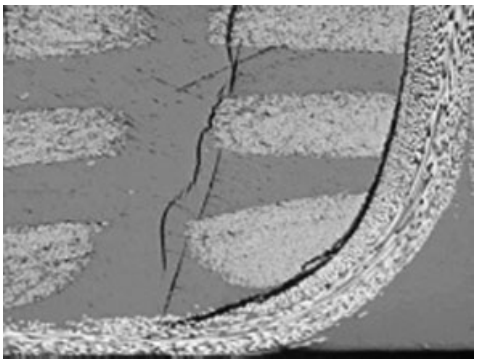


Figure 10. Lamina interface failure due to matrix shrinkage in three-dimensionally reinforced PMC.

3.2 Effect of gradient interface morphology in failure

The interface morphology tailoring between the constituent phases in composites (e.g., fiber–matrix, interlamina, interyarn, nanoconstituents-matrix, etc.) is essential in optimizing composite properties. As discussed earlier, in the case of composite strength, the mismatch of properties between the phases causes stress concentration at the interfaces, which, in turn, causes the initiation of damage and failure. A way of minimizing the mismatch of properties at the interface is demonstrated to reduce the interface stress concentration [29], and hence delay the damage initiation process together with improving composite strength. A gradient interface material morphology is thus desirable to enhance strength as well as other properties (e.g., thermal) of composites. Incorporation of nanoconstituents in composites, at present, potentially enable us to implement the gradient interface morphology at multiple scale level, from nanometer (nanoconstituent interface) to laminate ply interlayer (micrometer scale) [30, 31].

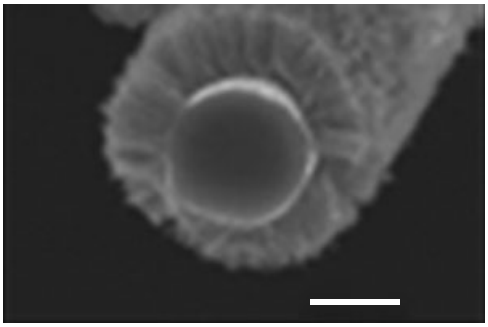


Figure 11. Carbon nanotubes grown on carbon fiber to introduce interface gradient morphology.

The current success of growing nanotubes on fiber is a possible mechanism of introducing gradient morphology at the fiber–matrix interface (Figure 11). The *in situ* polymerization of matrix (thermoset resin) with proper surface functionalization of fibers may provide gradient interface properties, as well. The *in situ* polymerization of functionalized nanofiber interface in epoxy matrix is shown to produce gradient epoxy modulus (qualitatively characterized by electron emission loss spectroscopy (EELS), Figure 13) resulting in suppressing failure at the fiber–matrix interface, as shown in Figure 12. One of the challenges is in quantifying the gradient material modulus at the fiber–matrix interface (of thickness in the order of 20–30 nm), without which the benefit of interface gradient morphology cannot be truly determined.

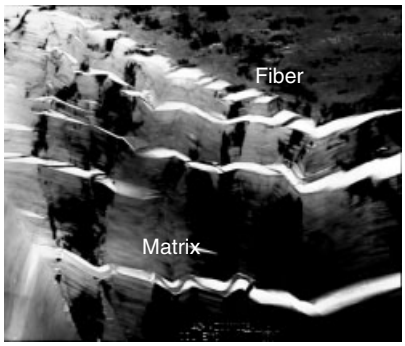


Figure 13. Qualitative characterization of fiber–matrix interface material property with electron emission loss spectroscopy (EELS) by correlating plasmon energy with material modulus. [Reproduced with permission from Ref. 31. © AIAA, 2007.].

4 SUMMARY

Preventing failure of components is an important engineering function in several applications. It is well known that such failures are caused by progressive degradation of materials the components are made of (*see* **Static Damage Phenomena and Models**; **Damage Evolution Phenomena and Models**). For safety and cost considerations, early detection of material damage and monitoring its progression are of primary importance in vehicle or component health monitoring (*see* **Nonlinear Acoustic Methods**; **Time–frequency Analysis**; **Nonlinear Features for SHM Applications**; **Nondestructive**

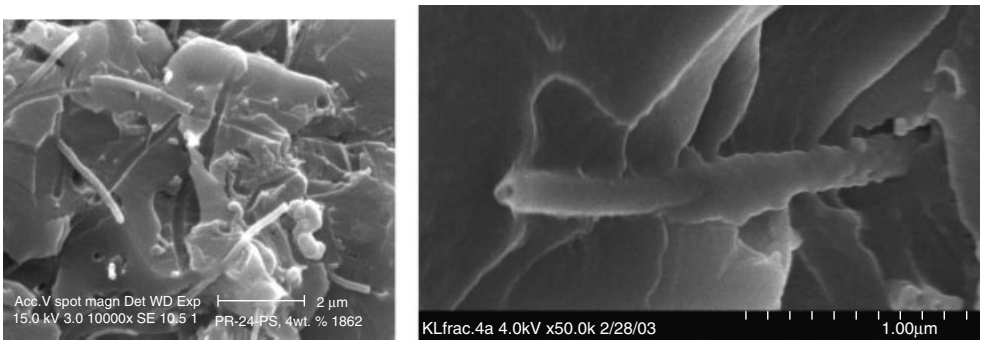


Figure 12. Micrographs of the failure surface of unfunctionalized and functionalized nanofibers in Epon 828. The strength for the case of functionalized fiber apparently increased by 140%. [Reproduced with permission from Ref. 31. © AIAA, 2007.]

Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors). The detection methodologies have been the subject of extensive studies by a community of NDE experts that are different from those that study materials and their damage leading to failure (materials engineering). In this article, the authors assume that if the key principles of each discipline are understood by the others, a synergistic contribution to health monitoring is possible. This article presents the fundamental principles behind failure of engineering materials as introductory material for those who have expertise in NDE and wish to apply their expertise to the advancement of detecting material failures. Briefly, engineering materials can be classified into metals, ceramics, and polymers.

Advanced engineering metallic alloys are often limited by plastic deformation by yield at low temperature or by creep at high temperatures. In some applications, the metallic materials develop sufficient damage under stress and temperature through cavitation that results in premature failure. Thus, methods to detect plastic distortion and internal cavity damage would be of primary interest. In addition, environmental attack often results in the reduction of section resulting in enhanced local stresses leading to failure. Monitoring the extent of environmental attack could possibly be conducted through clever measurements of products of the environmental attack. Similarly, developing structural health monitoring methods to detect early fatigue (precursor fatigue) damage would be of great interest. All these functions should be performed at material, structure, and vehicle levels.

Advanced ceramics are made of a mixture of strong fibers and relatively weak matrices. The weaker matrix often fails early, but the fibers are protected from these matrix cracks providing damage tolerance. Ultimate failure is dominated by failure of the strong fibers. However, the component life is often limited by failure of the weaker matrix. Thermal gradients through material thickness results in shear stresses that result in delamination cracks where layers of fibrous material separate from each other through the failure of the weaker matrix that binds them. Detection of these subsurface failures is vital to predicting the remaining life of CMC components. Other failure mechanisms involve fiber–matrix interface degradation through environmental attack and fiber degradation through grain growth.

In an effort to provide an understanding of the improvement of failure strength of composites, a brief description of the physics of damage initiation and failure process in PMC was discussed. In a nutshell, the existence of material property (especially the modulus) mismatch of the different phases (i.e., fiber, matrix, or other constituents), under any loading conditions, causes stress rise (stress concentration) at the constituent interfaces. This interface stress concentration is the primary driver of the damage initiation and complex failure modes in PMC, in the form of matrix cracking, fiber–matrix failure, fiber breakage, etc. Development of failure theories to reliably predict the complex failure modes in composites has evolved over the last three decades. In a quest to understand the physics of the complex failure modes more accurately, probabilistic failure theories are recently being developed to capture the effect of statistical property distribution of the constituent in composite failure. It is quite obvious that the gradient morphology at the material interface, to minimize the interface stress concentration, provides an opportunity of enhancing composite failure strength. The inclusion of nanoscale constituents (such as, nanotubes, nanofibers, etc.) through nanotechnology is expected to offer such an opportunity.

REFERENCES

[1] Walsh JA, Jata KV, Starke Jr EA. The influence of Mn dispersoid content and stress state on ductile fracture of 2134 type Al alloys. *Acta Metallurgica* 1989 **37**(11):2861–2871.

[2] Hertzberg R. *Deformation and Fracture Mechanics of Engineering Materials*. John Wiley & Sons: New York, 1989.

[3] Jata KV (WL/MLLM), Vasudevan AK. Effect of fabrication and microstructure on the fracture initiation and growth toughness of Al-Li-Cu alloys. *Materials Science and Engineering A: Structural Materials: Properties, Microstructure and Processing* 1998 **A241**(1–2):104–113.

[4] Lankford J, Hudak S. Relevance of the small crack problem to lifetime prediction in gas turbines. *International Journal of Fatigue* 1987 **9**:87–93.

[5] Ritchie RO, Gilbert CJ, McNaney JM. Mechanics and mechanisms of fatigue damage and crack growth in advanced materials. *International Journal of Solids and Structures* 2000 **37**(1–2):311–329.

[6] Jata K, Starke E. Fatigue and fracture behavior in an Al-Li-Cu alloy. *Metallurgical Transactions* 1986 **17A**:1011–1026.

[7] Vasudevan A, Sadananda K, Rajan K. Role of microstructures on the growth of long cracks. *International Journal of Fatigue* 1997 **19**:151–159.

[8] Hanlon T, Tabachnikoya ED, Suresh S. Fatigue behavior of nanocrystalline metals and alloys. *International Journal of Fatigue* 2005 **27**:1147–1158.

[9] Sadananda K, Vasudevan A. Fatigue crack growth mechanisms in steels. *International Journal of Fatigue* 2003 **25**:899–914.

[10] Jones D. *Principles and Prevention of Corrosion*. Prentice Hall: New York, 1992.

[11] Spiedel MO. Stress corrosion cracking of aluminum alloys. *Metallurgical Transactions A—Physical Metallurgy and Materials Science* 1975 **6A**:631–651.

[12] Sadananda K, Vasudevan A. Fatigue crack growth behavior of titanium alloys. *International Journal of Fatigue* 2005 **27**:1255–1266.

[13] Mendiratta M, Dimiduk D, Parthasarathy TA. Oxidation behavior of Mo-Mo3Si-Mo5SiB2 (T2) three phase system. *Intermetallics* 2002 **10**:225–232.

[14] Parthasarathy T, Mendiratta M, Dimiduk D. Oxidation mechanisms in Mo-reinforced Mo5SiB2(T2)-Mo3Si alloys. *Acta Materialia* 2002 **50**:1857–1868.

[15] Jones RM. *Mechanics of Composite Materials*, ISBN 0-07-032790-4. McGraw-Hill, 1975.

[16] Tsai SW, Hahn HT. *Introduction to Composite Materials*, ISBN 0-87762-288-4. Technomic Publishing Company, 1980.

[17] Hyer MW. *Stress Analysis of Fiber-Reinforced Composite Materials*, ISBN 0-07-016700-1. McGraw-Hill, 1998.

[18] Tsai SW. *Composites Design*, ISBN 0-9618090-2-7. Think Composites, 1988.

[19] Tsai SW. *Strength Characteristics of Composite Materials*, NASA CR-224, 1965.

[20] Pipes RB, Pagano NJ. Interlaminar stresses in composite laminates under uniform axial extension. *Journal of Composite Materials* 1970 **4**:538–548.

[21] Whitney JM, Nuismer RJ. Stress fracture criteria for laminated composites containing stress concentrations. *Journal of Composite Materials* 1974 **8**:253–265.

[22] Iarve EV, Mollenhauer D, Whitney TJ, Kim R. *Strength Prediction in Composites with Stress Concentrations: Classical Weibull and Critical Failure Volume Methods with Micromechanical Considerations, Program Review*. Air Force Office of Scientific Research: Monterey, CA, 2007.

[23] Tandon GP, Kim RY, Bechel VT. Fiber-matrix interfacial failure characterization using a cruciform-shaped specimen. *Journal of Composite Materials* 2002 **36**:2667–2691.

[24] Tandon GP, Kim RY, Bechel VT. Construction of the fiber-matrix interfacial failure envelope in a polymer matrix composites. *International Journal for Multiscale Computational Engineering* 2004 **2**(1):65–77.

[25] Foster DC, Tandon GP, Zoghi M. Evaluation of failure behavior of transversely loaded unidirectional model composites. *Experimental Mechanics* 2006 **46**:217–243.

[26] Gosse JH, Christensen S. *Strain Invariant Failure Criteria for Polymers in Composite Materials*, AIAA-2001-1184, 2001.

[27] Sihn S, Roy AK. Three-dimensional stress analysis of textile composites: part I. Numerical analysis. *International Journal of Solids and Structures* 2003 **41**(5–6):1377–1393.

[28] Sihn S, Roy AK. Three-dimensional stress analysis of textile composites: part II. Asymptotic analysis. *International Journal of Solids and Structures* 2003 **41**(5–6):1395–1410.

[29] Pan E, Roy AK. A simple plane-strain solution for functionally graded multilayered isotropic cylinder. *Structural Engineering and Mechanics* 2006 **6**:727–740.

[30] Sihn S, Park JW, Kim RY, Roy AK. Improvement of delamination resistance in composite laminates with nano-interlayers. *SAMPE 2006*. Long Beach, CA, April 30–May 4 2006.

[31] Roy AK. *Role of Gradient and Multiscale Interface Morphology in Three-Dimensional Reinforcements in Composites*, AIAA-2007-2162, 2007.

Chapter 13

Applications of Acoustic Emission for SHM: A Review

Martine Wevers and Kasper Lambrighs

Department of Metallurgy and Materials Engineering, Katholieke Universiteit Leuven, Leuven, Belgium

1 INTRODUCTION

It is a well-known phenomenon that materials can produce audible sound at high loads. The cracking of timber in mineshafts, the crying of tin, the cracking of rocks, and the breaking of bones are some examples. The vibrations produced inside a material are a naturally occurring phenomenon and are called *acoustic emission* (AE). These vibrations are produced when strain energy is suddenly released owing to the occurrence of microstructural changes, resulting in transient elastic waves.

The sources of AE are, generally speaking, local instabilities where transient elastic waves are released because of changes in the local stress field. The AE continues until a local equilibrium is reached. The energy to produce the stress waves comes from the stored elastic energy that is redistributed. Only a small part of this released energy is transformed into AE.

When the elastic stress waves reach the surface of the specimen, small displacements are produced that can be detected by a sensitive piezoelectric transducer. The signals from one or more sensors are amplified and measured to produce data for display and interpretation. The process of generation and detection is illustrated in Figure 1.

By applying AE it should be possible to say when something is happening (detection), where it is happening (localization), and what is happening (discrimination). In general, detection and localization are possible at this moment. In some specific applications, discrimination methods are available.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

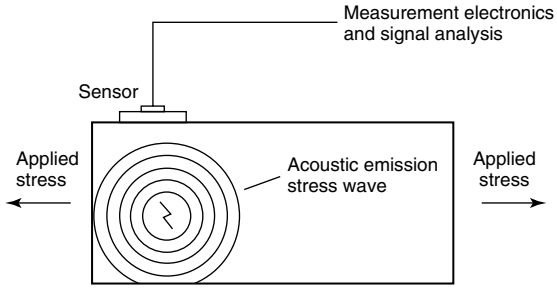


Figure 1. Working principle of the AE method.

The AE technique is somewhat different from other nondestructive test (NDT) methods. The most important differences are as follows [1]:

- The AE technique is a passive method; the technique only detects defects when they are activated by applying a stress that is adequately large. This means the defects have to produce the signals themselves and it is necessary to put an amount of energy into the material to activate the defects.
- The AE technique does not give direct information on the severity of the defects.
- Defects that are not activated by the applied stress will not be detected. Only the most critical defects for the applied loading condition will be activated.
- It is not necessary to place the sensor close to the defect; detection up to tens of meters is possible sometimes. This makes AE suitable for structural health monitoring (SHM) applications.

For SHM, AE is mostly combined with other NDT methods. The AE technique detects and localizes the critical defects, following which other techniques are used for the characterization of the defects.

To date, AE analysis is being used successfully in a wide range of applications including SHM; detection of leaks in storage tanks or piping systems; monitoring welding and other processes; monitoring chemical reactions including corrosion monitoring; and analysis of partial discharges from components subjected to a high voltage. As such, one of the main advantages of AE is its relative sensitivity to a range of phenomena leading to degradation.

2 HISTORICAL BACKGROUND

It is generally accepted that the research of the German scientist Kaiser [2] in the early 1950s represents the beginning of AE as it is known today. He was the first to carry out a systematic investigation of the noises that are generated when materials are deformed. He also discovered the phenomenon of nonrecovery of noises during repeated loading, known as the *Kaiser effect*. Further, he established the relations between the stresses during deformation and the intensity and amplitude of AE waves, which were found to be characteristic for each material. In Kaiser's work, it was already shown that not only stresses to which the material was subjected earlier but also residual stresses, changes in structure, phase transitions, stress relaxation, and decrease in elongation in creep play an important role.

In 1954, Schofield became aware of Kaiser's early work and initiated the first research program in the United States on material engineering applications of AE [3]. He detected, for example, the two different types of AE: continuous and burst emission.

Approximately at the same time, fundamental research on the AE phenomenon began in France as well. In 1958, a first work devoted to the origin of AE waves during the deformation of mild steel was published [4]. It was established that the generation of AE waves begins in the neighborhood of the yield zone and also prior to failure [5, 6].

In the former Union of Soviet Socialist Republics, the first applications of AE were connected with the prediction of coal and gas outbursts in collieries. Here, the seismoacoustic methods for investigating the state of a coal massif in the course of its underground mining were developed. Later, scientific works were published both in the Union of Soviet Socialist Republics and abroad [7–11].

On the basis of the results of the theoretical and experimental investigations mentioned above, the AE technique found its application in industry from the early 1960s on [12–14].

3 TYPES OF ACOUSTIC EMISSION

The AE produced by a source is phenomena and material dependent. The emitted transient elastic waves are basically pulselike. The width and height

of the primitive pulse depends on the dynamics of the source process. Source processes such as microscopic crack jumps and precipitate fractures are often completed in a few microseconds or fractions of a microsecond, so the primitive pulse has a correspondingly short duration. The amplitude and energy of the primitive pulse vary over an enormous range from submicroscopic dislocation movements to gross crack jumps. A distinction can be made between the two types of AE: continuous emission and burst emission (Figure 2).

In *continuous emission*, which is of low energy, the stress wave bursts are irresolvable. Moreover, the amplitude of the emission increases with increasing load. This type of emission can, for instance, be correlated with dislocation movements in metals. In some cases, it becomes necessary to monitor the continuous emission (leak testing).

Burst emission refers to a form of emission of much higher amplitude and energy in which the individual stress wave bursts are seen. It occurs when sources of higher energy are operating. Crack growth is an important example of such a source.

One can identify different factors that influence the detectability (amplitude) of the AE. These factors are summarized in Table 1 and can give an indication whether to use the AE technique for a particular material [1].

AE, in the proper sense, covers the audible frequencies up into the high ultrasonic range. Only a part of this range is used for measurements—normally the ultrasonic frequencies between 50 kHz and 2 MHz are the easiest to detect. At higher frequencies, the intensity of the AE signal is low because of attenuation and electronic field disturbance signals. At lower frequencies, the measurement is generally disturbed by background noise, e.g., vibrations from vehicles and noise from pumps or flowing media.

Table 1. Factors influencing the AE detectability [1]

Factors resulting in higher amplitude signals	Factors resulting in lower amplitude signals
High strength	Low strength
High strain rate	Low strain rate
Cleavage fracture	Shear fracture
Crack propagation	Plastic deformation
Anisotropy	Isotropy
Nonhomogeneity	Homogeneity
Thick section	Thin section
Twinning	—
Lower temperatures	High temperatures
Flawed material	Unflawed material
Martensitic phase transformations	Diffusion controlled transformations
Cast structures	Wrought structure
Coarse grains	Fine grains

4 WAVE PROPAGATION

The elastic stress waves generated by each AE source propagate through the material and any other path toward the AE sensor. This propagation greatly influences the resulting electrical signal from the sensor. Aspects of stress wave propagation that significantly influence the electrical signal are geometric spreading, losses due to material absorption, direct and indirect reflected paths from the AE source to the sensor, and different speeds of propagation along with the dispersion of the stress waves.

Geometric spreading is the loss in signal amplitude owing to the fact that, as the wave travels away from the AE point source in a two- or three-dimensional medium, the total area of material through which the wave front is passing increases. Conservation of energy can be used to calculate the resulting change in amplitude.

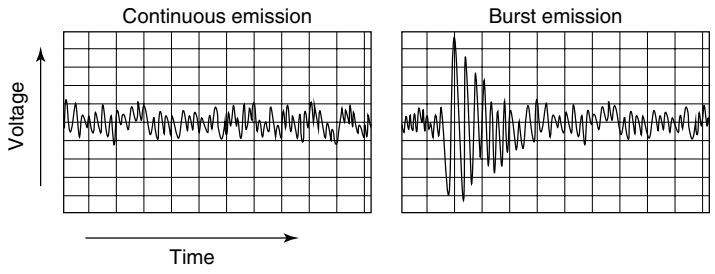


Figure 2. Continuous emission and burst emission.

Loss due to material absorption of the stress wave energy results in an attenuation of the amplitude of the wave as it propagates. This attenuation is more severe for the stress waves at higher frequencies and in viscoelastic materials.

The *different paths, direct and reflected*, of the stress wave in the material can lead to long signal durations measured by the sensor. This means the signal duration primarily results from the continued bouncing around of the stress wave in the material.

AE stress waves have *several components*: longitudinal, transverse, surface (Rayleigh waves), and plate waves (Lamb waves). These components propagate at different speeds and their existence depends on the geometry of the specimen (depending on the thickness of the specimen and the wavelength of the AE stress waves). Typically, two wave packets can be distinguished on the basis of wave speed: a generally lower amplitude first arrival and a higher amplitude second arrival.

For Lamb waves specifically, the plate wave theory states that mechanical waves propagate through plates in two modes: the symmetric or extensional mode and the antisymmetric or flexural mode. The main particle displacement is in the plane of the plate for the extensional mode and perpendicular to the plane of the plate for the flexural mode. The extensional mode generally travels at the highest velocity and is nondispersive in nature, meaning that all frequency components of this mode travel at the same velocity. The flexural mode travels at a lower velocity and is dispersive with the square root of frequency, meaning that the higher frequency components propagate faster than the lower frequency components. In practice, this will lead to a gradual decrease in the amplitude of the flexural mode as it propagates, owing to the spatial separation of the different frequency components [15].

The *dispersion of stress waves* refers to the propagation of different frequency components at different speeds. The net result of dispersion is a spreading in the time domain of the stress wave as a function of the traveled distance.

5 DETECTION

When a physical phenomenon suddenly releases a certain amount of elastic energy, elastic waves are formed; these can be picked up by a sensitive piezoelectric transducer (*see* **Piezoelectricity Principles and Materials**), which is placed on a surface. The key element in an AE resonant sensor is a piezoelectric crystal that converts movement into an electrical voltage. When temperature, strong electrical fields, high mechanical stresses, or radiation influence the piezoelectric properties of the sensor by depolarization, wave guides (rods or wires fabricated from steel, aluminum, or platinum) can be used.

The choice of a sensor depends on its resonance frequency band, the temperature, and its noise level. For sensitivity purposes, one often chooses a resonant sensor in a certain frequency domain coinciding with the frequency content of the AE signals. Broadband sensors with a flat frequency response over a wide frequency domain are often used for frequency analysis of the AE signals.

The AE sensor has to be calibrated to allow signal analysis techniques. The sensor calibration method most widely used is excitation with a standard broadband sensor and noting the resulting frequency response. A great deal of effort has gone into using mechanical sources such as fracture of lead pencils (Hsu–Nielsen source), calculating the transfer function from the source to the sensor locations, and comparing these to the transducer output. The Hsu–Nielsen source is widely accepted as a device to simulate an AE signal.

6 KAISER EFFECT AND FELICITY RATIO

When a sample is loaded and on producing AE, unloaded, and then reloaded in the same direction as the original load, AE often does not begin upon reloading until the previously achieved stress is reached. This phenomenon was first reported by Kaiser [2] and is known as the *Kaiser effect*. The physical basis for this effect is straightforward. AE is usually created by the formation or propagation of a defect under an applied stress, such as a crack, an avalanche of dislocations, decohesion of an interface, a twin, or a martensitic transformation. If the stress to form or propagate such a defect is not reached in the original loading, it will not be reached in the reloading until the applied stress exceeds the initial stress.

Dunegan and Tetelman [16] showed that for materials that obey the Kaiser effect, emission on reloading is still possible, which indicates that structural damage occurred between the first loading and the repeat. Recent test strategies pay much attention to emission that occurs at loads below the previous maximum and to emission that continues when the load is held at a constant level. The evidence is that structurally significant defects will produce AE at reloading before the previously achieved stress is reached. AE that is related to the stabilization of the structure, such as the relief of residual stresses, will tend not to recur upon reloading.

The Kaiser effect is observed in a general sense for many materials. The absence of the Kaiser effect is often expressed in the Felicity ratio. This is the ratio between the applied load at which the AE activity reappears during the subsequent application of loading and the previous maximum applied load. It is important to note that the Felicity ratio can decrease when the material approaches failure.

Careful attention must be paid to the loading schedule if AE testing is to be successful. Procedures for an AE test typically specify the loads that must be applied (relative to the working load or design load) and the upper and lower limits on the loading rate [17–20].

7 SIGNAL ANALYSIS

Once the AE signal has been detected, it is a general procedure to amplify, possibly filter, and then employ some type of signal processing to format and present the AE data for analysis (*see* **Signal Processing for Damage Detection**).

The oldest and easiest way to conduct data analysis is by using a comparator circuit, which generates a digital output pulse whenever the AE signal exceeds a fixed threshold voltage. The relationship between signal, threshold, and threshold-crossing pulses is shown in Figure 3. The threshold level is usually set by the operator; this is a key variable that determines test sensitivity. Depending on instrument design, sensitivity may also be controlled by adjusting the amplifier gain.

A more complex and extended way of data analyses involves the measurement of key parameters

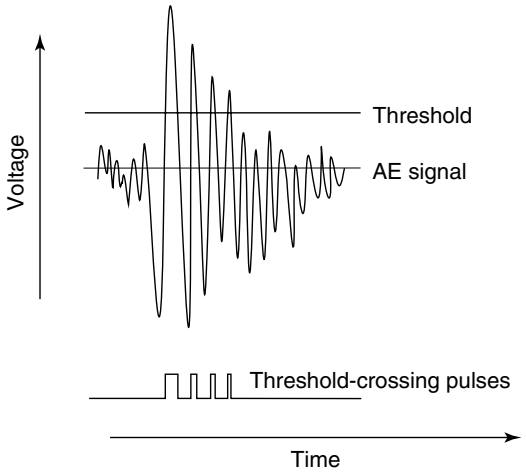


Figure 3. Principle of AE signal detection.

of each hit, that is, each AE signal that crosses the threshold. A digital description of each hit is generated by front-end hardware and is passed in sequence with other hit descriptions through a computer system.

The five most widely used signal measurement parameters are counts (Figure 3), amplitude, duration, rise time, and the measured area under the rectified signal envelope (MARSE) (Figure 4). Some tests are performed with fewer parameters, and some tests use others, such as true energy, counts to peak, average frequency, or spectral moment.

Along with these signal parameters, the hit description passed to the computer typically includes important external variables, such as the time of detection, the current value of the applied load, the cycle count (if it is a cyclic fatigue test), and the current level of continuous background noise.

The *amplitude* is the highest peak value attained by the AE waveform. This is a very important parameter because it directly determines the detectability of an AE signal. Acoustic emission amplitudes are directly related to the magnitude of the source, and they vary over an extremely wide range from microvolts to volts.

The *counts* are the threshold-crossing pulses discussed above (also called *ringdown counts*).

The *MARSE*, also called *energy counts*, is preferred over counts because it is sensitive to amplitude as well as duration, and it is less dependent on threshold setting and operating frequency.

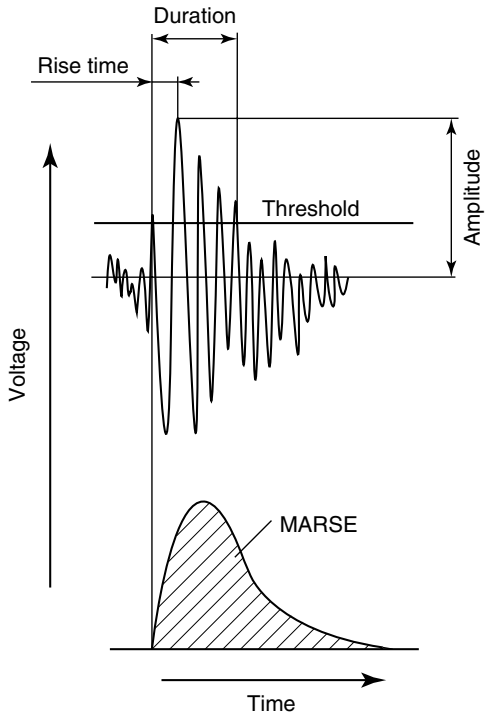


Figure 4. Key parameters of an AE burst signal.

The *duration* is the elapsed time from the first threshold crossing to the last and is directly measured in microseconds. This parameter depends on source magnitude, structural acoustics, and reverberation in approximately the same way as counts.

The *rise time* is the elapsed time from the first threshold crossing to the signal peak.

The counts are normally used for detection purposes, the time of detection for localization and the other parameters for discrimination purposes.

When the continuous emission is of interest (e.g., leak testing), it is sufficient to use AE instrumentation that measures the root mean square (RMS) voltage of the AE in a certain time window.

When a multichannel system is used, caution is needed for the interpretation of the received signals. An individual AE signal may hit just one channel or it may hit many channels, depending on the strength of the signal, the wave attenuation in the structure, and the sensor spacing. Therefore, an early task for the multichannel system is to determine whether a group of closely spaced hits on different channels is from the same source. The second, third, and later hits from one source can either be retained for the purposes of source location or discarded to keep the data clean and simple.

For the presentation and analysis of the AE data, many types of graphs are possible. Broadly, they can be defined as follows:

- history plots that show the course of the test from start to finish;
- distribution functions that show statistical properties of the emission;
- channel plots showing the distribution of detected emissions by channel;
- location displays that show the position of the AE source;
- point plots showing the correlation between different AE parameters;
- diagnostic plots showing the severity of AE indications from different parts of the structure.

More information on the presentation of AE data can be found in the ASM Handbooks online [21].

8 LOCALIZATION

If an AE source present in a material produces burst emission when loaded, localization of this source is possible. Source localization comes in handy when large structures are tested in which AE inspection is used to identify active regions and other NDTs are used to perform a more precise investigation. Large cost savings have been realized through this combination of global AE inspection and focused inspection by other methods.

There are two different types of source localization: zone location and point location.

Zone location only accepts the counts of a sensor that come in first. Hereby, all the activities of a zone around each sensor will be assigned to that sensor.

The reliability of this method is very good but an additional investigation with other NDTs is necessary to determine the exact location of the source.

Point location places the source precisely, by calculating from the relative arrival times of the AE wave at several sensors. Wave velocity is involved in these calculations. The attainable accuracy depends on the wave propagation processes by factors such as geometry, plate thickness, and contained fluids.

These factors alter the wave velocity, leading to errors in source location. In favorable cases, the attainable accuracy is better than 1% of the sensor spacing; in unfavorable cases, worse than 10% [21].

For the reader's convenience, the basic principles of a localization procedure are described using a similar example by Vink [1]. In this example, three sensors are attached to a structure in the form of a right-angled triangle as shown in Figure 5. It is important to note that for some applications other configurations can be more useful.

A calibration method has to be performed first to measure the wave velocity. An artificial AE wave is produced at a known location (the distances between the source and the different sensors are known). From the measured time difference between the excitation of the different sensors, the wave velocity, c, can be calculated using equation (1).

$$\left.\begin{aligned} s_2 - s_1 = \Delta s_{21} = c{\cdot}\Delta t_{21} \\ s_2 - s_3 = \Delta s_{23} = c{\cdot}\Delta t_{23} \end{aligned}\right\} \Rightarrow c \qquad (1)$$

When this sensor configuration is used for source localization, the difference in arrival time between sensors 1 and 2 (Δt_{21}) and between sensors 2 and 3 (Δt_{23}) have to be measured. This data can be used to construct lines with a constant Δt (hyperbola, not given in Figure 5). When the order in which the AE wave reaches the different sensors is registered, only one intersection of the hyperbola remains, indicating the location of the AE source. The AE source location can be calculated using equations (2) and (3).

$$\left.\begin{aligned} s_2 - s_1 = \Delta s_{21} = c{\cdot}\Delta t_{21} \\ s_2 - s_3 = \Delta s_{23} = c{\cdot}\Delta t_{23} \end{aligned}\right\} \Rightarrow \Delta s_{21},\ \Delta s_{23} \qquad (2)$$

$$\left.\begin{aligned} s_1 = \sqrt{(l_{12} - y)^2 + x^2} = s_2 - \Delta s_{21} \\ s_3 = \sqrt{(l_{23} - x)^2 + y^2} = s_2 - \Delta s_{23} \\ s_2 = \sqrt{x^2 + y^2} \end{aligned}\right\} \Rightarrow x, y \qquad (3)$$

For SHM applications, the amount of data that has to be analyzed can be reduced drastically by using a structural neural system (SNS), without loss of information (*see* **Artificial Neural Networks**). The SNS provides an architecture to connect n sensor nodes in series to form neurons, and the neurons in turn interact and pass signals similar to the way in which the biological neural system functions [22–24].

When the source location is not accurate enough, AE tomography [25, 26] can be used, leading to a better source location. AE tomography uses a wavespeed map of the structure rather than the average wavespeed. To apply this method, accurate sensor location, a large number of sensors, and the wavepath duration (time for a signal to travel from the source to the sensor) are required as an input for a specialized iterative AE localization algorithm.

When the requirements for AE tomography are unsuitable for a specific structure or environment, Delta T source location can be used [27]. The Delta T source location provides a novel approach for overcoming particular problems associated with source location in complex structures. This method uses an artificial source to record differences in arrival times from a number of locations to reduce the location error.

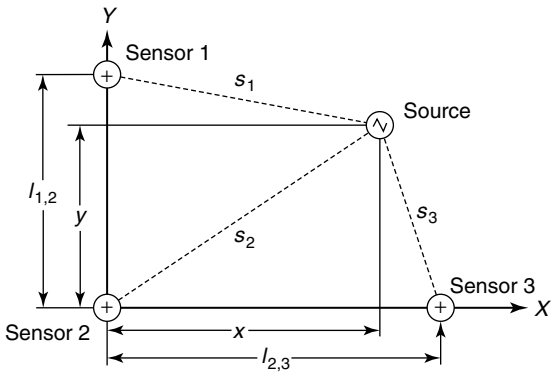


Figure 5. Basic principle of point localization.

9 DISCRIMINATION METHODS

By applying discrimination methods, it is possible to distinguish different kinds of AE sources. Discrimination methods can be divided into statistical and pattern-recognition techniques on one side and source characterization techniques, where the change of signal between the source and the sensor is calculated in, on the other side [28]. For pattern recognition, the AE waveforms can be used to get more information about the source and the transmission path [29].

A good example of source discrimination can be found in composite structures, where fiber fracture, delamination, matrix cracking, and debonding can be discriminated successfully [15, 30].

10 APPLICATIONS OF AE FOR STRUCTURAL HEALTH MONITORING

There are a number of successful applications of SHM via AE, such as the structural testing of aircraft, spacecraft, bridges (*see* **Long-term Monitoring of Dynamic Loads on the Brandenburg Gate**; **Development of a Monitoring System for a Long-span Cantilever Truss Bridge**; **Modular Architecture of SHM System for Cable-supported Bridges**; **Monitoring of Bridges in Korea**; **Bridge Monitoring in Japan**; **Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge**; **Continuous Monitoring of the Øresund Bridge: Data Acquisition and Operational Modal Analysis**; **Condition Compensation in Frequency Analyses—a Basis for Damage Detection**; **Modal Testing of the Vasco da Gama Bridge, Portugal**), bucket trucks, buildings (*see* **Construction Process Monitoring at the New Berlin Main Station**; **SHM Actions on the Holy Shroud Chapel in Torino**; **SHM of a Tall Building**; **Dynamic Response of Buildings of the Cultural Heritage**; **Suspended Roof of Braga Sports Stadium, Portugal**), dams (*see* **Dams**), military vehicles, mines piping systems, pipelines, pressure vessels, railroad tank cars, rotating machinery (*see* **Wind Turbines**; **Large Rotating Machines**; **Gas Turbine Engines**), storage tanks, and other structures. The goal of structural AE testing is to find defects and to assess or ensure structural integrity.

SHM via AE is applied to highlight the regions that threaten the integrity of the structure. When a full-scale test with fixed sensors is performed, AE is normally complemented by other NDT methods, which are used to assist in determining the type and severity of structural significant defects.

A major advantage compared to other NDT methods is that it does not require access to the whole structure. Therefore, it is not necessary to completely remove insulation or internal process fluids, which is typically a major expense in other NDT methods.

When AE tests are performed, it is important to load the structure in such a way that all the structurally significant defects are activated. The continuous monitoring of a structure in service is possible as well, although it is difficult because a small amount of AE from defect growth must be separated from a large amount of noise over a long time period. On the other hand, this approach guarantees the proper loading of the structure.

The stimulation of the structure by a controlled load is more commonly used and can be performed in a short time, typically during maintenance operations. In most cases, this is satisfactorily accomplished by going somewhat above the normally apparent service loads. Attention is needed to the type, magnitude, and rate of the applied load because previously applied stresses will have a very strong influence on the observed AE. Further, it is important to note that this approach does not always work, e.g., when the defects are activated thermally instead of mechanically.

When AE is used for SHM, one should always follow standard test procedures [17, 20].

10.1 Corrosion monitoring

It is shown that AE monitoring of stainless steel is capable of detecting early stages of stress corrosion cracking prior to the growth of a single dominant crack. By measuring AE parameters such as event rate, amplitude, counts, and energy counts, the transition from crack initiation and short crack growth to rapid growth of dominant cracks can be seen clearly [31, 32].

Pitting corrosion can be detected by AE in an austenitic stainless steel. It is hard to detect the initiation process, but once the pits propagate, the AE activity increases [33]. Rise time and counts are seen as discriminating AE parameters for monitoring pitting corrosion of austenitic stainless steel. Time delay and event rate are found to be in very good agreement with the sensitivity of the material toward pitting and with the polarization procedure [34].

Storage tanks from the chemical and petrochemical industry are very often subjected to high stresses, fatigue, aging, and corrosion, reducing their lifespan. Corrosion phenomena related to the aggressiveness of the stored chemical products is often the origin of leaks in these storage tanks. Although successful

online corrosion monitoring of storage tanks is yet to be reported, there are some indications that it is feasible. It is already experimentally validated that the AE produced by the cracking and delamination of the formed corrosion products can be successfully distinguished from noise [35, 36].

Nowadays, the testing of storage tanks with AE is limited to maintenance purposes whereby the structure is overloaded and the AE produced gives an idea of the damage present [37, 38].

The deterioration of high-strength steel tendons of prestressed concrete bridges (*see* **Long-term Monitoring of Dynamic Loads on the Brandenburg Gate**; **Development of a Monitoring System for a Long-span Cantilever Truss Bridge**; **Modular Architecture of SHM System for Cable-supported Bridges**; **Monitoring of Bridges in Korea**; **Bridge Monitoring in Japan**; **Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge**; **Continuous Monitoring of the Øresund Bridge: Data Acquisition and Operational Modal Analysis**; **Condition Compensation in Frequency Analyses—a Basis for Damage Detection**; **Modal Testing of the Vasco da Gama Bridge, Portugal**) progresses because of aging. The failures have been mostly attributed to corrosion because of a severe environment, such as salt attack. In theory, the corrosion process can be monitored by AE owing to the fact that elastic waves are produced due to the cracking and delamination of the corrosion products. In practice, however, these AE signals are hard to distinguish from the environmental noise (traffic). Therefore, the monitoring of high-strength steel tendons of prestressed concrete structures is possible only by detecting individual wire breaks, for which detection and localization is already successful [39, 40].

10.2 Rotating machinery

Rolling element bearing condition monitoring has received considerable attention for many years because the majority of problems in rotating machines are caused by faulty bearings. The classical failure mode of rolling element bearings is localized defects, in which a sizable piece of the contact surface is dislodged during operation, mostly by fatigue cracking in the bearing metal under cyclic contact stressing.

Counts and peak amplitude are conventionally known as good parameters for the detection of defects in rotating machinery [41–43]. It is also reported that the RMS voltage, used to monitor continuous emission, is a good parameter to monitor bearings [44].

10.3 Pressure vessels and pipes

The testing of pressure vessels and pipes is normally done by overloading the structure so that the defects present will be activated and produce AE. This allows defect detection and localization, after which the defects can be validated by other NDT methods and, depending on the criticality of the defect, be repaired.

The failures of metal pressure vessels and pipes are mostly attributed to corrosion [45, 46] or crack growth [47–49]. The measurement of corrosion by AE has already been described. For the measurement of crack growth in pressure vessel steels, it has been shown that the counts, the amplitude, the rise time, the average frequency, and the duration are important parameters for AE detection [49].

10.4 Concrete/buildings

The AE technique has been employed for many years to study damage and microcrack nucleation in concrete materials [50–59]. Here it is shown that AE analysis is, in principle, an effective method of damage assessment in concrete structures. Nevertheless, due to the complexity of most civil engineering structures, the applicability is limited to some isolated cases such as concrete bridges [39, 60, 61] (*see* **Long-term Monitoring of Dynamic Loads on the Brandenburg Gate**; **Modal Testing of the Vasco da Gama Bridge, Portugal**).

The most important AE parameters that correlate with the cracking process in concrete are counts, energy, and duration, together with the frequency characteristics [62]. For reinforced concrete, the most important AE parameters are peak amplitude and energy [56].

Recently, some reports were published on the use of AE methods for successful damage assessment of historical buildings (*see* **Construction Process**

Monitoring at the New Berlin Main Station; **SHM Actions on the Holy Shroud Chapel in Torino**; **SHM of a Tall Building**; **Dynamic Response of Buildings of the Cultural Heritage**; **Suspended Roof of Braga Sports Stadium, Portugal**). A new methodology is proposed, based on the counting of events, to determine the released energy and therefore the conditions of stability or the risk due to the spreading of discrete cracks [61–63].

10.5 Composites

A lot of studies have already been done on the use of AE for composite materials. Not only localization but also discrimination of the AE sources is found possible. It is reported that fiber fracture, delamination, matrix cracking, and debonding can be discriminated successfully [15, 30]. AE energy has shown good general correlation with parameters deduced from stress–strain curves, damage identified with ultrasonic C-scanning, and microscopic analysis of specimens subjected to static and fatigue testing [64].

The most important application of composite materials is in aviation. The condition monitoring and life prediction of the main structures of aircraft have received considerable attention because they are closely related to flight safety. The early detection of fatigue crack initiation and growth is, in general, beyond the capability of various conventional NDT means; AE, however, is more suitable to undertake the task. Although the AE technique is a generally accepted method for testing articles made of composite materials, a number of problems arise with its practical application. So far, there are no test methods that ensure the real-time localization of AE signals in large composite complex-shaped structures. A real-time AE instrument for health monitoring of aircraft structures, if not impossible, is difficult to realize to a great extent. With the help of the combination of AE waveform and parameter analysis, AE can still play an important role in new aviation material studies and in health monitoring of aircraft structures [65, 66].

11 FUTURE PROSPECTS OF AE

It has recently been shown by fundamental research that the formal expectations about the possibilities of AE for SHM were too high. It seems that the relations between the detected AE signal and the AE source are far more complicated than had been assumed before. It will be necessary to fully understand the effects that influence the AE signal, such as characteristics of the AE source, the wave propagation in the material, and the wave propagation at the interface of the sensor.

A part of the research on AE currently focuses on source discrimination methods with artificial intelligence techniques, where the system is "trained" to recognize certain parts of AE signals and connect them to an AE source. An example of such a system is a neural network (*see* **Artificial Neural Networks**), an area in which a lot of research is being carried out at present.

The development of new sensors for AE with a higher sensitivity and robustness could lead to more applications for real-time health monitoring of structures in their total life cycle. An example of such a sensor is the fiber-optic sensor (*see* **Fiber-optic Sensor Principles**; **Fiber-optic Sensors**) with the following advantages over conventional piezoelectric sensors: small diameter and light weight, flexibility, high strength, heat resistance, immunity to electromagnetic interference, durability, and resistance to corrosion. Further these sensors are believed to have a very high sensitivity in an extremely broad frequency range [67].

RELATED ARTICLES

Acoustic Emission
Lamb Wave-based SHM for Laminated Composite Structures

REFERENCES

[1] Vink WJP. *Niet-Destructief Onderzoek*, ISBN 90-407-1147-X (in Dutch). Delftse Universitaire Pers, 1995.

[2] Kaiser J. Erkenntnisse und Folgerungen aus der Messung von Geräuschen bei Zugbeanspruchung von metallischen Werkstoffen. *Archiv für das Eisenhüttenwesen* 1953 **24**(1–2):43–45.

[3] Schofield BH, Barreiss B, Kyrala A. *Acoustic Emission under Applied Stress WADS*, Technical Report 58-194. Lessells and Associates: Boston, MA, 1958.

[4] Lean JB, Plateau J, Bachet C, Crussard C. Sur la formation d'ondes sonores, au cours d'essais de traction, dans des éprouvettes métalliques. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 1958 **246**(20):2845–2848.

[5] Lean JB, Plateau J. Observation des ondes sonores prenant naissance au cours de la déformation plastique des métaux, en relation avec le problème de la rupture fragile de l'acier doux. *Mémoires Scientifiques de la Revue de Métallurgie* 1959 **56**(1):91–99.

[6] Lean JB, Plateau J, Crussard C. Contribution à l'étude du mécanisme de la rupture fragile de l'acier doux. *Mémoires Scientifiques de la Revue de Métallurgie* 1959 **56**(4):427–452.

[7] Shamina OG. Elastic pulses in the fracture of specimens of rocks. *Izvestiia Akademii Nauk SSSR, Seriia Geofizicheskaya* 1956 **5**(5):513–518.

[8] *Control of Sudden Outbursts in Coal Mines [in Russian]*. Gosgortekhizdat: Moscow, 1962.

[9] *Application of Seismoacoustic Methods in Mining Art [in Russian]*. Nauka: Moscow, 1964.

[10] Vinogradov SD. *Acoustic Observations of the Processes of Fracture of Rocks [in Russian]*. Nauka: Moscow, 1964.

[11] Knill IL, Franklin IA, Malone AW. A study of acoustic emission from stressed rock. *International Journal of Rock Mechanics and Mining Sciences* 1968 **5**:87–121.

[12] Green AT, Lockman CS, Haines HK. *Acoustical Analysis of Filament-Wound Polaris Chambers*, Report 0672-01F. Aerojet-General Corporation: Sacramento, CA, 1963.

[13] Green AT, Lockman CS, Brown SJ, Steel RK. *Feasibility Study of Acoustic Depressurization System*, Final Report, NASA CR-55472. Aerojet-General Corporation: Sacramento, CA, 1964.

[14] Dunegan HL. Acoustic emission: a promising technique. *UCID 4643*. Lawrence Radiation Laboratory: Livermore, CA, 1963; pp. 203–238.

[15] Surgeon M, Wevers M. Modal analysis of acoustic emission signals from CFRP laminates. *NDT and E International* 1999 **32**:311–322.

[16] Dunegan HL, Tetelman AS. Acoustic emission. *Research and Development* 1971 **22**(5):20–24.

[17] NEN-EN 13477-1:2001 en, *Non-destructive Testing—Acoustic Emission—Equipment Characterisation—Part 1: Equipment Description*.

[18] NEN-EN 13477-2:2001 en, *Non-Destructive Testing—Acoustic Emission—Equipment Characterisation—Part 2: Verification of Operating Characteristics*.

[19] NEN-EN 13554:2002 en, *Non-Destructive Testing—Acoustic Emission—General Principles*.

[20] NEN-ISO 22096:2007 en, *Condition Monitoring and Diagnostics of Machines—Acoustic Emission*.

[21] ASM, *ASM Handbooks Volume 17: Nondestructive Evaluation and Quality Control, Methods of Nondestructive Evaluation, Acoustic Emission Inspection*, 1989, http://products.asminternational.org/hbk.

[22] Sundaresan MJ, Schulz MJ, Ghoshal A, Martin WN, Pratap PR. Neural system for structural health monitoring. *Smart Structures and Materials 2001: Sensory Phenomena and Measurement Instrumentation for Smart Structures and Materials; Proceedings of the Conference*. Newport Beach, CA, 5–6 March 2001; pp. 130–141.

[23] Kirikera GR, Shindea V, Schulza MJ, Ghoshalb A, Sundaresanc M, Allemang R. Damage localization in composite and metallic structures using a structural neural system and simulated acoustic emissions. *Mechanical Systems and Signal Processing* 2007 **21**:280–297.

[24] Lee JW, Kirikera GR, Kang I, Schulz MJ, Shanov VN. Structural health monitoring using continuous sensors and neural network analysis. *Smart Materials and Structures* 2006 **15**:1266–1274.

[25] Schubert F. Basic principles of acoustic emission tomography. *DGZfP-Proceedings BB 90-CD*, Lecture 58. European Working Group on Acoustic Emission (EWGAE), 2004.

[26] Schubert F. Tomography techniques for acoustic emission monitoring. *European Conference on Nondestructive Testing (ECNDT), Berlin, 2006—We.3.6.2*.

[27] Baxter MG, Pullin R, Holford KM, Evans SL. Delta T source location for acoustic emission. *Mechanical Systems and Signal Processing* 2007 **21**:1512–1520.

[28] American Society for Nondestructive Testing (ASNT), Acoustic emission testing. *Nondestructive Testing Handbook*, ISBN: 1-57117-106-1, Third Edition: Columbus, OH, 2005; Vol. 6, p. 456.

[29] Cole P, Carlos M. Use of advanced A.E. analysis for source discrimination using captured waveforms. *Proceedings of the 27th European Conference on Acoustic Emission Testing*, Cardiff, 2006; pp. 401–406.

[30] Wevers M. Listening to the sound of materials: acoustic emission for the analysis of material behaviour. *NDT and E International* 1997 **30**(2):99–106.

[31] Sung KY, Kim S, Yoon YK. Characteristics of acoustic emission during stress corrosion cracking of inconel 600 alloy. *Scripta Metallurgica et Materialia* 1999 **37**(8):1255–1262.

[32] Shaikh H, Amirthalingam R, Anita T, Sivaibharasi N, Jaykumar T, Manohar P, Khatak HS. Evaluation of stress corrosion cracking phenomenon in an AISI type 316LN stainless steel using acoustic emission technique. *Corrosion Science* 2007 **49**: 740–765.

[33] Fregonese M, Idrissi H, Mazille H, Renaud L, Cetre Y. Initiation and propagation steps in pitting corrosion of austenitic stainless steels: monitoring by acoustic emission. *Corrosion Science* 2001 **43**:627–641.

[34] Fregonese M, Idrissi H, Mazille H, Renaud L, Cetre Y. Monitoring pitting corrosion of AISI 316L austenitic stainless steel by acoustic emission technique: choice of representative acoustic parameters. *Journal of Materials Science* 2001 **36**: 557–563.

[35] Cole P, Watson R. Acoustic emission for corrosion detection. *Advanced Materials Research* 2006 **13–14**:231–236.

[36] Mechraoui S, Amami S, Laksimi A, Benmedakhene S. Acoustic emission analysis of crack growth in a corrosion product. *Advanced Material Research* 2006 **13–14**:237–242.

[37] Riahi M, Shamekh H. Health monitoring of aboveground storage tanks' floors: a new methodology based on practical experience. *Russian Journal of Nondestructive Testing* 2006 **42**(8): 537–543.

[38] Kwon JR, Lyu GJ, Lee TH, Kim JY. Acoustic emission testing of repaired storage tank. *International Journal of Pressure Vessels and Piping* 2001 **78**:373–378.

[39] Yuyama S, Yokoyama K, Niitani K, Ohtsu M, Uomoto T. Detection and evaluation of failures in high-strength tendon of prestressed concrete bridges by acoustic emission. *Construction and Building Materials* 2007 **21**:491–500.

[40] Fricker S, Vogel T. Site installation and testing of a continuous acoustic monitoring. *Construction and Building Materials* 2007 **21**:501–510.

[41] Tandon N, Nakra BC. Defect detection in rolling element bearings by acoustic emission method. *Journal of Acoustic Emission* 1990 **9**(1):25–28.

[42] Bansal V, Gupta BC, Prakash A, Eshwar VA. Quality inspection of rolling element bearing using acoustic emission technique. *Journal of Acoustic Emission* 1990 **9**(2):142–146.

[43] Choudhury A, Tandon N. Application of acoustic emission technique for the detection of defects in rolling element bearings. *Tribology International* 2000 **33**:39–45.

[44] Morhain A, Mba D. Bearing defect diagnosis and acoustic emission. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 2003 **217**:257–272.

[45] Lackner G, Tscheliesnig P. Detection of corrosion damage on pressure equipment with acoustic emission. *Advanced Materials Research* 2006 **13–14**: 127–132.

[46] Kabanov BS, Gomera VP, Sokolov VL, Fedorov VP, Okhotnikov AA. AE testing of refinery structures. *Advanced Materials Research* 2006 **13–14**:133–138.

[47] Svoboda V, Zemlicka F, Brumovsky M. Investigation of fatigue crack growth on material for reactor pressure vessel by acoustic emission. *Advanced Materials Research* 2006 **13–14**:139–146.

[48] Raucher F. Acoustic emission during fatigue testing of pressure vessels. *Advanced Materials Research* 2006 **13–14**:147–152.

[49] Ennaceur C, Laksimi A, Herve C, Cherfaoui M. Monitoring crack growth in pressure vessel steels by the acoustic emission technique and the method of potential difference. *International Journal of Pressure Vessels and Piping* 2006 **83**:197–204.

[50] Maji A, Shah SP. Process zone and acoustic-emission measurements in concrete. *Experimental Mechanics* 1988 **28**(1):27–33.

[51] Berthelot JM, Robert JL. Modeling concrete damage by acoustic emission. *Journal of Acoustic Emission* 1987 **6**:43–46.

[52] Li Z, Shah SP. Localization of microcracking in concrete under uniaxial tension. *Materials Journal* 1994 **91**(4):372–381.

[53] Li Z. Microcrack characterization in concrete under uniaxial tension. *Magazine of Concrete Research* 1996 **48**(176):219–228.

[54] Landis EN, Shah SP. The influence of microcracking on the mechanical behavior of cement-based materials. *Advanced Cement Based Materials* 1995 **2**(3):105–118.

[55] Landis EN. Micro-macro fracture relationships and acoustic emissions in concrete. *Construction and Building Materials* 1999 **13**:65–72.

[56] Mirmiran A, Philip S. Comparison of acoustic emission activity in steel-reinforced and FRP-reinforced concrete beams. *Construction and Building Materials* 2000 **14**:299–310.

[57] Grosse C, Reinhardt HW, Finck F. Signal-based acoustic emission techniques in civil engineering. *Journal of Materials in Civil Engineering* 2003 **15**(3):274–279.

[58] Chen B, Liu JY. Experimental study on AE characteristics of three-point-bending concrete beams. *Cement and Concrete Research* 2004 **34**: 391–397.

[59] Grosse CU, Finck F. Quantitative evaluation of fracture processes in concrete using signal-based acoustic emission techniques. *Cement and Concrete Composites* 2006 **28**:330–336.

[60] Shigeishia M, Colombob S, Broughtonb KJ, Rutledgeb H, Batchelor AJ, Forde MC. Acoustic emission to assess and monitor the integrity of bridges. *Construction and Building Materials* 2001 **15**:35–49.

[61] Carpinteri A, Lacidogna G, Pugno N. Structural damage diagnosis and life-time assessment by acoustic emission monitoring. *Engineering Fracture Mechanics* 2007 **74**:273–289.

[62] Sagaidak AI, Elizarov SV. Acoustic emission parameters correlated with fracture and deformation processes of concrete members. *Construction and Building Materials* 2007 **21**:477–482.

[63] Carpinteri A, Lacidogna G. Damage evaluation of three masonry towers by acoustic emission. *Engineering Structures* 2007, **29**:1569–1579.

[64] Bourchak M, Farrow IR, Bond IP, Rowland CW, Menan F. Acoustic emission energy as a fatigue damage parameter for CFRP composites. *International Journal of Fatigue* 2007 **29**: 457–470.

[65] Sereznov AN, Stepanova LN, Lebedev EYu, Chaplygin VN, Katarushkin SA, Kozhemyakin VL. Studies of the fracture process in composite structural elements based on strain measurements and the acoustic-emission technique. *Russian Journal of Nondestructive Testing* 2004 **40**(9):580–586.

[66] Geng Rongsheng. Modern acoustic emission technique and its application in aviation industry. *Ultrasonics* 2006 **44**:e1025–e1029.

[67] Kageyama K, Murayama H, Ohsawa I, Kanai M, Nagata K, Machijima Y, Matsumura F. Acoustic emission monitoring of a reinforced concrete structure by applying new fiber-optic sensors. *Smart Materials and Structures* 2005 **14**:S52–S59.

Chapter 12

Data Interrogation Approaches with Strain and Load Gauge Sensor Arrays

Michael D. Todd

Department of Structural Engineering, University of California, San Diego, CA, USA

1 INTRODUCTION

Any structural health monitoring (SHM) strategy for any application invariably must employ three fundamental components: (i) a sensor array that captures some form of local or global kinematic or kinetic response of the system at hand (such as a vibration response); (ii) a series of algorithms that process the raw sensor array data in such a way to extract "features" that convey meaningful information about the condition or performance of the system; and (iii) statistical modeling tools that appropriately classify these extracted features into assessment categories, ordered by increasing complexity: detection of damage, location of damage, level of damage, and type of damage. This is known as *the Rytter taxonomy* [1]. One may couple such a strategy for SHM to a *damage prognosis strategy* by employing operational and environmental monitoring, a probabilistic expected loading model, and an appropriate failure or degradation model for any such mechanism being targeted for assessment and prediction.

This overall approach clearly relies on both hardware (sensors and actuators, as appropriate) and software (data processing for feature extraction, statistical classification, model implementation/comparison, etc.) for execution. Both play critical roles: one certainly needs a means to observe system behavior (the role of sensors and measurement systems in hardware). However, there is no such thing as a "damage sensor" or a "condition assessment sensor"; all types of sensors do nothing but measure a system's field response, such as acceleration, pressure, strain, temperature, etc., to stimuli. To perform SHM with the goal of current state assessment—at any level of the Rytter taxonomy—one must mine information from the sensor's field measurements (the role of algorithms in software). Inevitably, this sensor data mining relies on pattern-recognition methods, where one must draw a comparison in some way between baseline data sets and test data sets. Such comparisons may be broadly classified into two classes: (i) direct waveform comparison or (ii) comparison between baseline and fitted model

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

parameters or between baseline and test residual error when using a fixed baseline model. The former class involves extracting direct features from raw waveform data obtained in both a baseline or reference condition and a test condition. Such features could be something as simple as a single peak or rms value over a given time window or a more complex signal-processing scheme such as a bispectral construction. The latter class involves fitting some sort of model to the data sets and comparing how a model fitted appropriately to a baseline data set compares to the same kind of model fitted to a test data set. Alternatively, one may fit a model to a reference or baseline data set only and then use it as a predictor for future data sets, using the resulting residual error as the primary feature set on which to make comparisons.

This section begins with a technology review of strain gauges and load cells primarily used in SHM today, and it follows with a presentation of data interrogation approaches used with these sensor modalities.

2 HARDWARE TECHNOLOGY OVERVIEW

2.1 Strain gauges

Strain measurements are likely the second-most commonly used data type (after acceleration, which is considered in another section) for SHM. Strain is a nondimensional measure of an object's deformation resulting from an applied stress. More formally, strain is defined as the displacement per unit length of the object; strain gauges have a finite gauge length that serves as a normalizing factor. As such, the strain gauge actually gives an average strain reading over the length of the gauge. Strain gauges, at least in their mature forms used most commonly, generally fall into two categories: (i) resistive foil gauges and their variants and (ii) fiber-optic gauges. The most common strain gauge technology is the resistive foil kind; in 1856, Lord Kelvin first noted that the resistive properties of some metallic conductors change as a function of applied strain; this effect was put into practical use by the 1930s. In practice, other properties such as capacitance or inductance also change with applied strain, and the hardware community has commercialized variations on the resistive theme as well, but their sensitivity to other measurands, mounting requirements, and complex circuitry have limited their application. Bulk optical methods, taking advantage of interference patterns produced by optical flats, are very accurate and highly sensitive, but these techniques are delicate and cannot withstand industrial applications in many cases. Finally, it is important to mention piezoelectric patches, which are increasingly being used as strain sensors. However, their initial uses were primarily in actuation or impedance measurement, and this architecture is not considered in this section.

2.1.1 Resistive strain gauges

The typical foil resistive-type gauge consists of a wire grid network (the resistor) embedded on an elastic layer, which in turn is bonded with an epoxy layer on to the object of interest. When the object deforms under load, the deformation is transferred to the wire network, causing resistive changes in the material. Typical materials include copper–nickel alloys, platinum alloys (usually tungsten), nickel–iron alloys, or nickel–chrome alloys, foils, or semiconductor materials. The most popular alloys used for strain gauges are copper–nickel alloys and nickel–chrome alloys. Two further improvements in strain gauge technology include the thin-film strain gauge and diffused semiconductor gauges, but despite these alternative piezoresistive designs, bonded metallic resistance strain gauges have the best reputation and are the most widely used. They are relatively inexpensive, can achieve overall accuracy of better than $\pm 0.10\%$, are available in a short gauge length, have high primary sensitivity, and have only moderate thermal sensitivity. One may operate these gauges in temperatures ranging from cryogenic to those found in jet engine turbines. One may also use them to measure both static and dynamic strain, as the inherent time constants in the material are negligibly small.

The changes in the resistive characteristics of metallic foil gauges to applied strain are very small, requiring signal processing to convert these changes into voltages that normal data acquisition systems may detect and process. The most popular detection networks are the Wheatstone bridge, Chevron bridge, and four-wire ohm circuit, all shown in Figure 1.

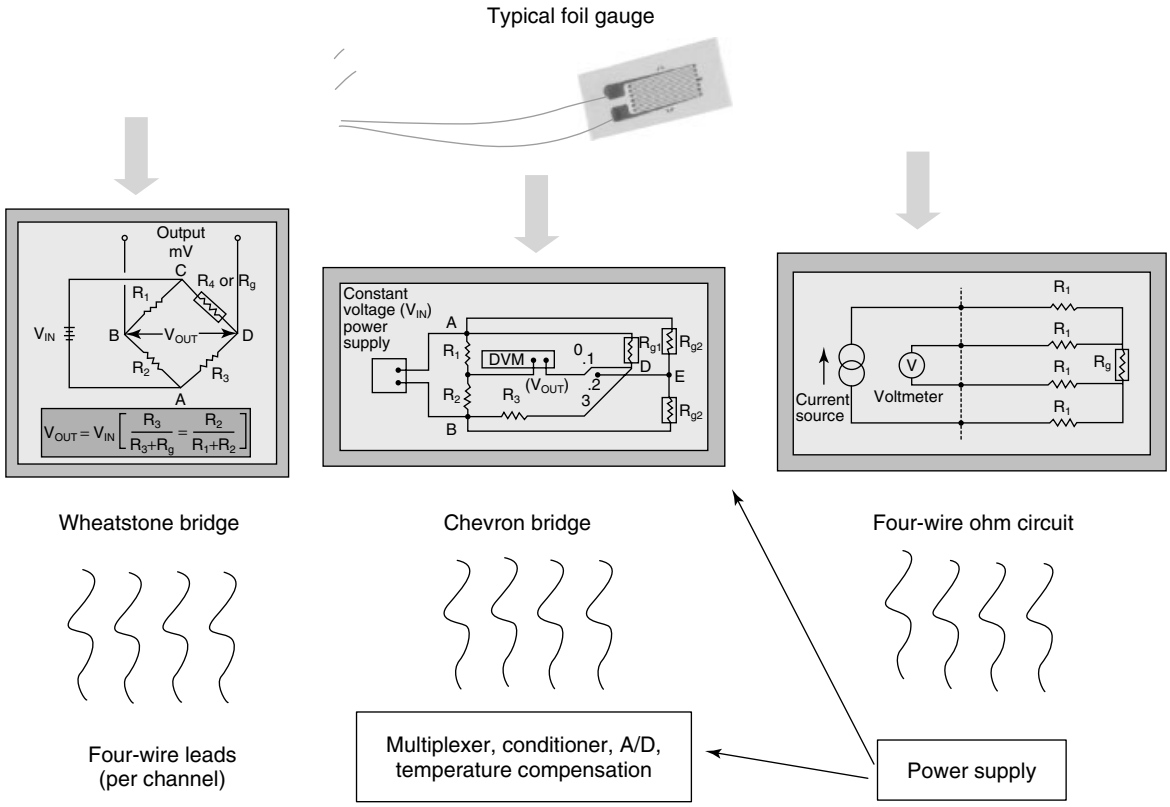


Figure 1. Common signal-processing networks for resistive strain gauges.

2.1.2 Fiber-optic strain gauges

Although foil resistive gauges dominate current market usage, an explosion of commercially available fiber-optic solutions for strain measurement has occurred over the last few years. The fiber-optic communications revolution of the late 1990s led to great improvement in component technologies at lower costs, and the sensor development community piggy-backed on these advances. The two dominant fiber-optic technologies are direct fiber interferometry and fiber Bragg gratings (FBGs) [2]. The former method is older, but its relative complexity (despite several orders of magnitude sensitivity improvement over foil gauges) and low multiplexing capability has limited its use to specialized military and industrial applications. Most commercial systems today take advantage of FBG technology.

FBGs are intrinsic structures that may be photo-written into the fiber. Silica glass, with germanium doping, is absorptive in the ultraviolet optical range, and such irradiation of the fiber core will cause an essentially permanent change in the refractive properties of the core. If the irradiation is performed in a spatially periodic fashion, then a series of gratings (a periodic change in refractive index) is introduced in the fiber core, and these act like a local bandstop optical filter. The physical periodicity is chosen so that when broadband light in the infrared range ($\sim$1300–1600 nm) is propagated down the fiber core, a narrowband component (width $\sim$0.2 nm) is rejected and reflected. The central wavelength of this narrowband component is directly proportional to the spacing in the FBG at that location; so, as that spacing physically changes (because of strain), the reflected wavelength shifts in proportion. In this way, tremendous multiplexing may be achieved: several FBGs may be individually photowritten into a single optical fiber with each one written to reflect at a unique wavelength (Figure 2).

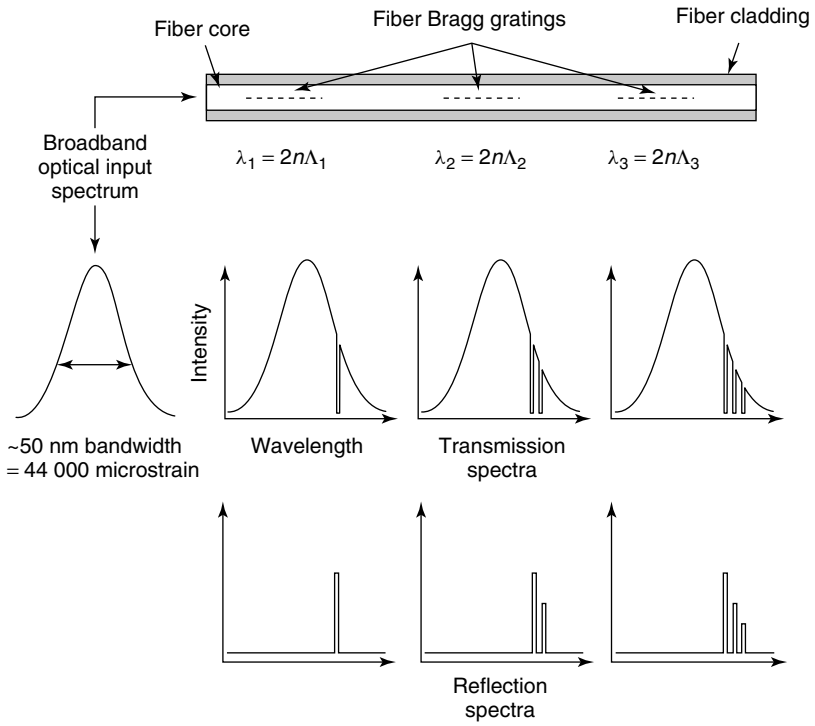


Figure 2. FBG wavelength-encoded multiplexing.

Thus, tracking the individual reflected wavelength shifts allows a direct measurement of strain at the various locations. There are several architectures possible for doing this interrogation, but a dominant approach is shown in Figure 3. Here, broadband light is inserted in an FBG array via a superluminescent diode, and the reflections from the array are coupled back through a Fabry–Perot filter. This device passes only a narrowband wavelength of light (designed to be commensurate with the FBG reflection width of $\sim$0.2 nm) dependent upon the spacing between mirrors in the device. This spacing, and thus the passband, is controlled by applying a rapidly stepped voltage to a piezoelectric driver controlling the mirrors. The passed light signal is sent to a photodetector and differentiated; the zero crossings of the differentiated signal correspond to the peak wavelengths of the reflected light, and correlation between the ramp voltage level and shifts in the zero-crossings results in obtaining the strain for each FBG sensor. The resolution of the voltage ramp and the spectral range of the filter primarily determine the optimal strain resolution, which has been demonstrated on the order of less than 1 microstrain; in some variations on this architecture, strain resolutions on the orders of tens of nanostrain have been reported [3]. This is several orders of magnitude more sensitive than the best foil resistive gauge, but the costs of FBG systems (and specifically, the FBGs themselves, which are $\sim$\$150–200 per sensor) have limited deep market penetration. However, since FBG systems are insensitive to electromagnetic interference, do not create a spark source, are extremely lightweight and nonintrusive, and are highly multiplexible, many application areas—particularly in aerospace structural monitoring—are emerging.

2.1.3 Strain sensing challenges

Despite significant commercialization and use, strain gauges of both types pose some challenges, particularly if being used to supply raw strain data to an SHM or damage prognosis algorithm suite. First, one must apply both strain gauge types to the host material with some form of epoxy. Such epoxies are typically not stiffness-matched with either the

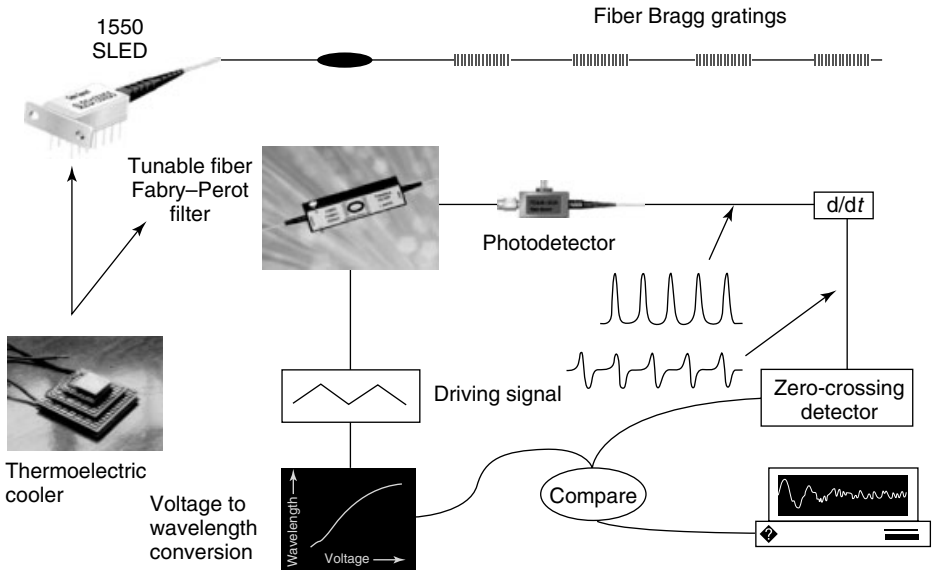


Figure 3. Tunable Fabry–Perot filter FBG interrogation.

gauge material or the host material, which means that strain does not completely transfer between host material and gauge. Typically, the epoxy bond layer will partially absorb some of the material-induced strain in a shear lag effect, leading to a reduced strain reading at the gauge. For simple gauge/host geometries, the community has developed shear lag correction models [4], but these models effectively have their use limited to simple load states and stationary environmental conditions. So, while one may find shear lag correction factors for a sensor array during a laboratory calibration step, it is very common for these factors to change with load level, load application rate, temperature, and humidity, rendering inaccurate strain readings that may introduce errors in subsequent data processing algorithms. Second, all strain gauges and their associated interrogation hardware are inherently sensitive themselves to extraneous measurands, such as temperature. Decoupling of mechanical and thermal strain is a critical issue, and one usually makes recourse to a "dummy" gauge approach whereby a number of gauges are left mechanically debonded from the host structure (thus free to measure temperature alone). Of course, such an approach effectively doubles the sensor requirement, so only a few dummy gauges are typically deployed. For structures such as space-deployable ones where large thermal gradients are observed over the length scales of the structure, this is a big problem, as the thermal correction map cannot accurately couple with the mechanical map, but in many small-scale applications or any application with negligible spatial temperature gradients, this is an acceptable solution. Finally, in the case of fiber-optic gauges, the issue of potential embeddability poses some unique challenges. Because of their small size and self-contained design, fiber gauges may be placed inside composite materials during their fabrication. However, such a benefit leads to problems of ingress/egress reliability, fiber microbending that leads to signal degradation, and the possibility of compromising the host structure's integrity. A number of studies are addressing these issues in fiber sensing [5–8].

2.2 Load/force gauges

The measurement of force or load is a very important component, particularly while doing damage prognosis since loading information is required for developing and updating predictive capability. Load sensors ("cells") fall into two broad classes, both of which enjoy significant market penetration: (i) strain gauge and (ii) piezoelectric. Vibrating wire load cells

Figure 4. Some typical strain gauge load cell transducer designs.

are a third design, but these are primarily used in geotechnical applications rather than SHM.

2.2.1 Strain gauge load cells

The first load cell class is exactly as one may surmise from its label: it consists of a strain gauge, operated as reviewed above, bonded to a specialized transducer that deforms predictably under load. The deformation in the transducer induces strain, which is calibrated through the transducer geometry to applied force. Popular transducer types include bending beam, multiple beam, shear beam, double shear beam, column, membrane, and ring. Some of these designs are shown in Figure 4. One may optimize designs for force detection below a few pound-force up to several thousand pound-force with resolutions dependent upon strain gauge resolution determined by signal processing. Measurement bandwidth ranges from static timescales (with drift compensation) up to whatever sampling limits the data acquisition hardware can sustain.

2.2.2 Piezoelectric load cells

Piezoelectric load cells operate using piezoelectric crystals. Piezoelectric crystals are made from ferroelectric materials (such as lead zirconate titanate, PZT) that induce electric charge (electric dipoles at the molecular level) when mechanically strained. The inertially induced strain on the crystal is contained within a damping block and an inertial seismic mass for frequency response tuning, and the resulting charge is picked up by wiring (Figure 5).

Signal processing is required to amplify and detect this charge, and there are two general architectures: high impedance and low impedance [9].

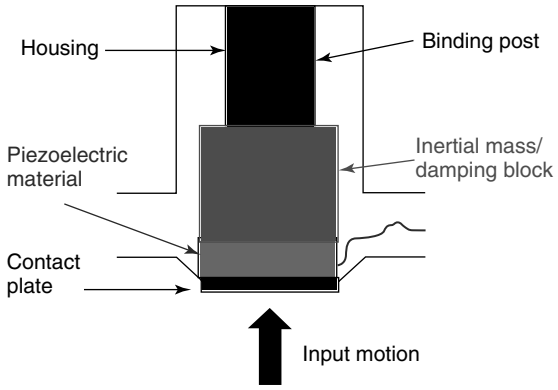


Figure 5. Generic piezoelectric sensor design.

High-impedance designs require a charge amplifier or external impedance converter for charge-to-voltage conversion. The charge amplifier consists of a high-gain inverting voltage amplifier with an Field-Effect Transistor (FET) at its input for high insulation resistance. Low-impedance designs use the same piezoelectric sensing element as high-impedance units, but they also incorporate a miniaturized built-in charge-to-voltage converter. They also require an external power supply coupler, which provides the constant current excitation required for linear operation over a wide voltage range and also decouples the bias voltage from the output. Both the power into and the signal out of the sensor are transmitted over this cable. These designs are shown in Figure 6.

Piezoelectric crystals produce an electrical output only when they experience a change in load; this means they cannot detect true static measurements. Essentially the sensor behaves as a single-degree-of-freedom underdamped oscillator. This means that the sensor is subject to resonance phenomena (frequency-dependent sensitivity in the resonance band, phase distortion, etc.) and operates well as a load cell in a

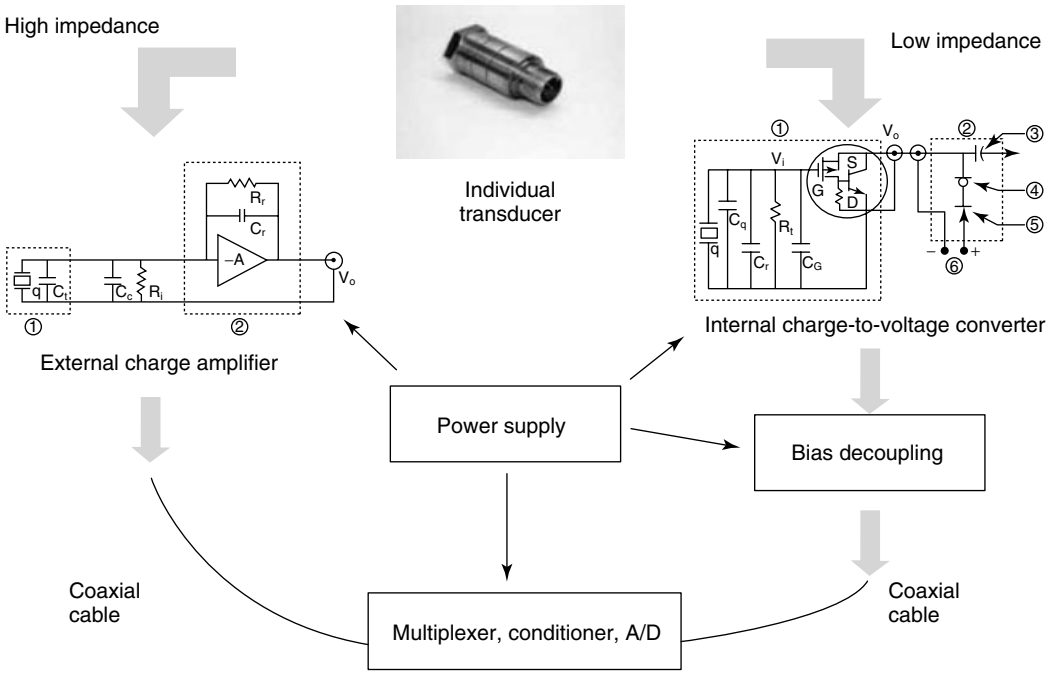


Figure 6. Signal processing for piezoelectric sensors.

finite frequency band below the resonant region, but not all the way to static time scales. Furthermore, these systems have characteristic time constants, which are defined to be the discharge time of the corresponding ac circuit. Time constants are usually partially selectable by resistors in the control circuit, although this depends on the design (low-impedance vs high-impedance). Systems may be designed to measure responses up to many kilohertz with micro-g sensitivity, so these sensors are used very frequently in SHM for measuring small dynamic forces. These load cells work on exactly the same principle as piezoelectric accelerometers, which are the most prolific sensor in current use for SHM applications.

2.2.3 Load sensing challenges

Load cells present their own challenges for effective use. Strain gauge load cells, of course, are subject to the same challenges as "pure" strain gauges described above. There is an added complexity due to the transducer, since the strain gauge itself is affixed to the transducer, but the transducer in turn must be affixed to the system of interest. This is often done with either wax (for temporary measuring), an epoxy (for semipermanent measuring), or through a mounting block (using fasteners). The latter is most desirable, assuming the mounting block's own dynamic response does not interfere with the desired measurement band. The primary issues with any mounting type are (i) maintaining good force transfer from host to sensor and (ii) ensuring that cross-axis sensitivity is minimized. In the latter issue, a sensor that is mounted slightly off of its primary sensing axis (or out of alignment with the desired measurement direction) will not measure what is actually being experienced and will also detect out-of-direction inputs, both of which can significantly corrupt the desired signal.

One other significant challenge with strain gauge load cells is fatigue, which is the accumulation of plastic (unrecoverable) deformation in the strain gauge as load cycles accumulate. All strain gauges suffer from this, but strain gauges used in the direct mode (as above) are not typically used in a highly dynamic, cyclical sensing mode, so fatigue is not as prevalent (other than possible low-cycle, high-amplitude fatigue). However, the dynamic usage

of load cells implies that such sensors have a finite useful life before readings become increasingly inaccurate, distorted, hysteretic, and/or nonlinear.

Piezoelectric load cells do not suffer from strain gauge-related issues, but one must still mount them to the host system in the same ways, so those issues remain. The biggest issue with piezoelectric load cells is their use outside of their calibrated measurement range. As mentioned, these sensors are ac coupled on the low-frequency end (there is significant low-frequency attenuation) and nonlinear beyond a critical high frequency (because of resonance). One has to ensure that force inputs are occurring on time scales commensurate with this frequency range.

3 DATA INTERROGATION STRATEGIES

As mentioned in the introductory remarks, SHM involves more than merely collecting sensor data such as that obtained from a strain gauge or load cell network. One must mine the sensor data for actual information ("features") that directly relates to the damage being assessed. This implies that SHM may be considered as a pattern-recognition problem whereby these features are compared by statistical testing between baseline data sets and test data sets. In the context of this section, strain and/or load gauges provide dynamic response baseline and test condition data, the same feature extraction algorithm(s) are applied to these data, and the features are compared via statistical modeling tools [10] (Figure 7). Feature extraction methods have received a tremendous amount of attention in the literature, and some good reviews may be found in two Los Alamos technical reports [11, 12] and in recent proceedings of the International and European Workshops on Structural Health Monitoring [13, 14]. Statistical modeling for feature discrimination and classification has received significantly less attention in the literature, but the references just cited contain a review of these approaches as well.

Also, as mentioned, feature extraction methods fall into either direct waveform comparisons of some kind or model usage of some kind. The following subsections will consider some examples of these two general types of feature extraction applied to strain/load data.

3.1 Direct waveform comparison

The majority of data interrogation methods with strain/load gauges fall into this group. This is primarily because of two reasons: (i) in general, direct waveform comparison techniques are less computationally burdensome and are more readily interpretable and adaptable and (ii) strain and load measurements are more directly related to damage realizations (strain or loath path changes, for example, when cracks develop and propagate) and existing prognosis models, which often use strain or load as input variables. The civil structural domain—and most notably bridges—has been a prominent application area in this regard [15, 16]. Many bridge

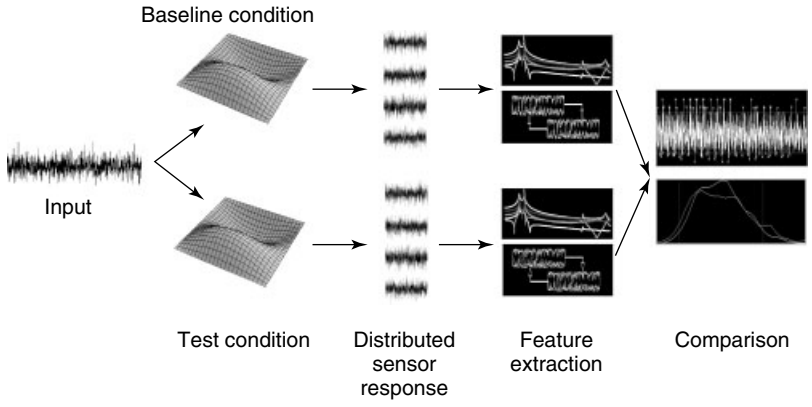


Figure 7. Statistical pattern-recognition approach for SHM.

applications have involved directly embedding strain gauges (typically fiber optic, because of their properties), usually inside specialized tubes [8] that are themselves subsequently placed inside concrete load-bearing members during fabrication. The raw static strain readings are then tracked over time to discover trends, such as seasonal thermal cycling that can lead to wear [17], localized plastic damage due to overload [18], or evidence of a fracture [19].

Researchers have also used direct strain array readings to detect cracks or bond strength verification in composite joints and repair patches (doublers) for aerospace structures [20, 21] and for localizing impacts on composite panels [22]. A number of aerospace applications may be found in [23].

3.2 Model usage for assessment and prognosis

This category involves mining raw strain/load information for damage-sensitive features (rather than directly comparing the waveforms themselves), which often employs using a model of some sort. Model parameters are then compared between baseline and test conditions, or, alternatively, a single model is fit to the baseline and used as a predictor for future test cases, with the resulting error being the feature of interest (i.e., increasing error indicates increasing change to the system). While such techniques typically require more complex computational analysis, they are often tailored for specific damage identification in an application, so they are often more sensitive. Such models may also be coupled to prognosis models, which are models used to predict future system states given the current state and a probabilistic future loading state.

Health assessment models take many forms, e.g., autoregressive models of directly measured time series, modal models of structural response, state space models, mechanics models, or neural networks. These models are broadly classified into physics-based and data-based classes. In the former class, physical first principles are applied to create deterministic relationships among variables that are fitted or tested with data from the sensor arrays. Conversely, data-based models create heuristic relationships among measured data that are tested against each other or a baseline model over time. Researchers have used both model classes with strain/load measurements in a variety of applications, such as detecting loosened bolts in jointed connections [24–26], detecting leaks in composite tanks [27], identifying degradation in bridge girders [28–30], and detecting impact-induced delamination in a composite Unmanned Aerial Vehicle (UAV) wing [31].

Prognosis models are physics-based or empirical (historically based data) models that relate measurable variables to failure criteria for a given failure mode [32], e.g., yielding, fatigue, creep, etc. Most yielding theories (Rankine, Tresca–Guest, St Venant, strain energy, Von Mises–Huber), for example, directly encode strain or stress measurements, so strain and load gauges are needed. High-cycle fatigue models also require stress/strain amplitude and cycle knowledge, so again strain/load gauges are most appropriate. Because of the nonstationary nature of fatigue-inducing stresses, however, several cumulative damage laws (Miner, Henry, Gatts) and subsequent fatigue crack propagation theories (e.g., Paris) have been postulated, all of which require stress as an input variable.

Using damage laws such as these and combining them with current structural health assessments to create damage prognosis technology is relatively new as of this writing. Some recent examples of such integrated technology include using linear cumulative damage laws for assessment of solid rocket motor casing [33], stress cycle curves for detection and prognosis of aircraft landing gear damage [34], and an energy method for fatigue crack detection and growth prediction [35].

RELATED ARTICLES

Signal Processing for Damage Detection

Data Preprocessing for Damage Detection

Piezoelectricity Principles and Materials

Piezoelectric Wafer Active Sensors

Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection

Fiber-optic Sensor Principles

Fiber Bragg Grating Sensors

Novel Fiber-optic Sensors
Wireless Sensor Network Platforms
Sensor Placement Optimization
Sensor Network Paradigms
Hybrid PZT/FBG Sensor System
Loads Monitoring in Aerospace Structures
Loads and Temperature Effects on a Bridge

REFERENCES

[1] Rytter A. *Vibration Based Inspection of Civil Engineering Structures*, Ph.D. Thesis. University of Aalborg: Denmark, 1993.

[2] Todd MD. Optical-based sensing. In *Damage Prognosis*, Inman DJ (ed). John Wiley & Sons, 2004.

[3] Todd MD, Johnson GA, Althouse BL. A novel Bragg grating sensor interrogation system utilizing a scanning filter, a Mach-Zehnder interferometer, and a 3×3 coupler. *Measurement Science and Technology* 2001 **12**(7):771–777.

[4] Cox HL. The elasticity of paper and other fibrous materials. *British Journal of Applied Physics* 1952 **3**:72–79.

[5] Guemes A, Menendez JM. Embedded fibre Bragg gratings for design and manufacturing optimisation of advanced composites. In *Structural Health Monitoring 2005*, Chang F-K (ed). Destech Publications, 2005; pp. 462–469.

[6] Brotzu A, Caponero MA, Colonna D, Felli F, Gasbarro N, Maddaluno G. FBG sensor embedded in pultruded composite bars. In *Structural Health Monitoring 2006*, Guemes A (ed). Destech Publications, 2006; p. 922.

[7] Zhou G-D, Li H, Ren L, Li D. Influencing parameters analysis of strain transfer in optic fiber bragg grating sensors. *Proceedings of SPIE 6179*. SPIE, 2006.

[8] Belarbi A, Watkins SE, Chandrashekhara K, Corra J, Konz B. Smart fiber-reinforced polymer rods featuring improved ductility and health monitoring capabilities. *Smart Materials and Structures* 2001 **10**:427–431.

[9] Kulwanoski G, Schnellinger J. The principles of piezoelectric accelerometers. *Sensors Magazine* 2004 **21**(2):(may be found at http://www.sensorsmag.com).

[10] Farrar CR, Duffey TA, Doebling SW, Nix DA. A statistical pattern recognition paradigm for vibration-based structural health monitoring. In *Structural Health Monitoring 2000*, Chang F-K (ed). Technomic Publications, 1999.

[11] Doebling SW, Farrar CR, Prime MB, Shevitz DW. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. LA-13070-MS, Los Alamos National Laboratory Report. 1996 (may be found at http://www.lanl.gov/projects/ncsd/publications.htm).

[12] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. A review of structural health monitoring literature: 1996–2001. LA-13976-MS, Los Alamos National Laboratory Report. 2004 (may be found at http://www.lanl.gov/projects/ncsd/publications.htm).

[13] Chang F-K (ed). *Structural Health Monitoring 2005*. Destesch Publications, 2005.

[14] Guemes A (ed). *Structural Health Monitoring 2006*. Destesch Publications, 2006.

[15] Ansari F (ed). *Fiber Optic Sensors for Construction Materials and Bridges*. Technomic Publications, 1998.

[16] Ansari F (ed). *Sensing Issues in Civil Structural Health Monitoring*. Springer, 2006.

[17] Tennyson RC, Mufti AA, Rizkalla S, Tadros G, Benmokrane B. Structural health monitoring of innovative bridges in Canada with fiber optic sensors. *Smart Materials and Structures* 2001 **10**:560–573.

[18] Vohra ST, Johnson GA, Todd MD, Danver BA, Althouse BA. Distributed strain monitoring with arrays of fiber Bragg grating sensors on an in-construction steel box-girder bridge. *IEICE Transactions* 2000 **E83-C**:454–461.

[19] Idriss RL, Kodindouma MB, Kersey AD, Davis MA. Multiplexed Bragg grating optical fiber sensors for damage evaluation in highway bridges. *Smart Structures and Materials* 1997 **7**:209–216.

[20] Jones R, Galea S. Health monitoring of composite repairs and joints using optical fibres. *Composite Structures* 2002 **58**:397–403.

[21] Malkin M, Qing XP, Leonard M, Derriso M. Flight demonstration: health monitoring for bonded structural repairs. *Structural Health Monitoring 2006*. Destesch Publications, 2006.

[22] Weis M, Joflin J, Deimel P, Drechsler K. Evaluation of impact tests on the TANGO barrel by means of fibre Bragg grating sensor (FBGS) measurements. *Structural Health Monitoring 2006*. Destesch Publications, 2006.

[23] Staszewski WJ, Boller C, Tomlinson G. Health monitoring of aerospace structures. *Smart Sensors and Signal Processing*. John Wiley & Sons, 2004.

[24] Todd MD, Johnson GA, Vohra ST. Deployment of a fiber Bragg grating-based measurement system in a structural health monitoring application. *Smart Materials and Structures* 2000 **10**:534–539.

[25] Nichols JM, Todd MD, Seaver M, Virgin LN. The use of chaotic excitation and attractor property analysis in structural health monitoring. *Physical Review E* 2003 **67**:016209.

[26] Todd MD, Nichols JM, Trickey ST, Seaver M, Nichols CJ, Virgin LN. Bragg grating-based fibre optic sensors in structural health monitoring. *Philosophical Transactions of the Royal Society of London A* 2007 **365**:317–343.

[27] Kang H-K, Park J-S, Kang D-H, Kim C-U, Hong C-S, Kim C-G. Strain monitoring of a filament wound composite tank using fiber Bragg grating sensors. *Smart Materials and Structures* 2002 **11**:848–853.

[28] Stubbs N, Kim JT, Farrar CR. Field verification of a nondestructive damage localization and severity estimation algorithm. *Proceedings of the 13th International Modal Analysis Conference*. Nashville, TN, 1995; 210–218.

[29] Todd MD, Johnson GA, Chang CC, Malsawma L. Real-time girder deflection reconstruction using a fiber Bragg grating system. *International Modal Analysis Conference XVIII*. San Antonio, TX, 7–10 February 2000.

[30] Worden K, Fieller NRJ. Damage detection using outlier analysis. *Journal of Sound and Vibration* 1999 **229**:647–667.

[31] Nichols JM, Seaver M, Trickey ST, Salvino LW, Pecora DL. Detecting impact damage in experimental composite structures: an information-theoretic approach. *Smart Materials and Structures* 2006 **15**:424–434.

[32] Collins JA. *Failure of Materials in Mechanical Design*. John Wiley & Sons: New York, 1981.

[33] Hudson HL, Little RR. Installation and demonstration of embedded sensors for solid rocket motor health monitoring. In *Structural Health Monitoring 2005*, Chang FK (ed). DEStech Publishing: Lancaster, PA, 2005.

[34] El-Bakry M. Commercial aircraft landing gears operational loads monitoring (OLM) systems engineering requirements. In *Structural Health Monitoring 2005*, Chang FK (ed). DEStech Publishing: Lancaster, PA, 2005.

[35] Liu KC, Wang JA. An energy method for predicting fatigue life, crack orientation, and crack growth under multiaxial loading conditions. *International Journal of Fatigue* 2001 **23**:129–234.

Chapter 11

Modal–Vibration-based Damage Identification

Keith Worden[1] and Michael I. Friswell[2]

[1] *Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*
[2] *Department of Aerospace Engineering, University of Bristol, Bristol, UK*

1 INTRODUCTION

Structural health monitoring (SHM) arguably emerged as a subdiscipline of structural dynamics with an emphasis on determining the health or condition of a given structure on the basis of the sort of measurements that are commonly available from structural dynamic testing. The standard for a structural dynamic test is an application of *modal analysis*, where the measurements correspond to a modal model—an expansion of the system in terms of a sequence of fictitious single degree-of-freedom (SDOF) systems [1]. The *modal parameters* are the SDOF system-damping ratios and natural frequencies, together with the coefficients of the terms in the expansion that form the modeshapes. Vibration-based SHM does not solely use the modal quantities, it also uses the physical parameter matrices and quantities such as frequency response functions (FRFs) and transmissibilities considerably. The objective of this article is to describe the use of modal and general vibration-based quantities for SHM. It is not intended as a comprehensive review by any means. In fact, such a review already exists in the form of the extensive literature surveys carried out by researchers at Los Alamos National Laboratories in the United States [2, 3]. For a much more comprehensive overview, the reader is directed to these documents. This article is intended to illustrate the use of a small number of vibration-based methods by the consideration of case studies. The specific methods for illustration here are chosen simply by the particular tastes of the authors.

The rationale for vibration-based SHM is the long-held belief that any damage occurring within a structure will be reflected in a change in its dynamic properties. The problem of damage identification is then to translate the changes in the measured quantities into a diagnosis. This diagnosis may simply be a yes/no indication of the presence of damage

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

or may give more detailed information like location and severity. In any case, some degree of mathematical processing is required to translate the measurements into a diagnosis. At present, there are two main frameworks for carrying out this translation. The first is based on the linear-algebraic solution of the identification problem considered as an inverse problem. The second is based on concepts of machine learning and pattern recognition. This article concentrates on the former, as the pattern-recognition approach is adequately represented in other chapters of this Encyclopedia (*see* **Statistical Pattern Recognition**; **Machine Learning Techniques**; **Artificial Neural Networks**; and **Novelty Detection**). The omission of pattern-recognition approaches implies that this article does not consider methods based on passing vibration-based features into neural networks and other learning algorithms.

The breakdown of this article is as follows: Section 2 describes the basic use of modal features in detecting and locating damage. Only basic signal processing operations are considered at this stage. In addition to natural frequencies and modeshapes, derived features—curvatures and Yuen functions—are also dealt with. Section 3 expands on the use of curvatures by discussing the modal strain energy method in some detail, the application of the method is illustrated by use of a detailed experimental case study. Section 4 deals with more advanced algorithms for the processing of data based on the consideration of damage as an inverse problem: the predominant methodology is that of linear algebra or matrix analysis. Section 5 takes a more detailed look at the methodology of finite element (FE) updating and Section 6 contains a discussion of some of the issues that arise in developing a credible SHM methodology based on vibration measurements.

2 NATURAL FREQUENCIES AND MODESHAPES

A fundamental requirement for SHM is that measurements that are sensitive to the presence of damage must be available from a structural test. Such measurements are, in the terminology of pattern recognition, called *features*. A basic requirement of any feature for SHM is that it should distinguish between the normal condition of the structure and *any* damaged state. Such a feature allows *level one* damage identification under Rytter's generally accepted hierarchy [4]; this is simply *detection*. More advanced features allow one to advance further and potentially localize and quantify the damage.

The basic features that one can use for damage detection are the natural frequencies. These have the advantage that they can be acquired without recourse to a full modal test; a basic spectral analysis from a single random excitation test with one response sensor can suffice. A further advantage of the natural frequencies is that they can be estimated quite accurately, typically to within 1%. The use of natural frequencies dates back to the seminal paper of Cawley and Adams [5]. A change in the natural frequency can be taken as an indication of damage within the system. If one considers the changes in many natural frequencies, one can potentially deduce further characteristics of the damage.

Modeshapes can be considered rather than frequencies as they are well known to suffer local changes in the presence of localized damage. Thus, they may prove more effective in *locating* damage. The price one pays for the extra flexibility provided by the modeshapes is the expense of carrying out a full modal test with multiple response sensors and the associated instrumentation. A further problem with modeshapes is that it is not possible to measure them with the same accuracy as natural frequencies. Test errors in the range of 5–10% are typical.

The use of natural frequency and modeshape features is demonstrated here using case studies based on comparatively simple structures. The structures under investigation here are FE models of a beam and plate. To investigate the possibilities of further data processing, the curvatures of the modeshapes are also computed and used as features for network training, as are the diagnostic functions proposed by Yuen [6]. It is shown that the features computed by further processing of the basic modal quantities provide higher-level identification for the damaged structures.

2.1 A cantilever beam structure

FE simulation of the beam was carried out using the package LUSAS [7]. To allow experimental

validation of the model, the properties of a cantilever beam currently in use in the laboratory were adopted. This was of aluminum with dimensions $920 \times 50 \times 12.5\,\text{mm}^3$. Fifteen Timoshenko beam elements were used to model the system and the first five natural frequencies and modeshapes of the system were obtained. The value of each modeshape was specified at 16 nodes along the beam. Comparison of the predicted natural frequencies with those from the experimental beam showed almost perfect agreement. Damage to the structure at a given location was simulated by reducing the Young's modulus of the corresponding element and consequently reducing its stiffness. The results presented here are for 25, 50, and 75% reductions in Young's modulus (and hence stiffness) for the appropriate elements.

The first features to consider are the natural frequencies. These are tabulated below (Table 1 for a stiffness reduction in the fifth element from the cantilever root. (The precisions given is a reflection of the fact that the frequencies are derived from simulation.)

It is noted that increasing severity of damage causes a progressive reduction in all the natural frequencies. The extent of the reduction of a specific frequency depends on the position of the damage relative to the nodes of the corresponding modeshape. In this case, the largest relative decrease is for the first mode. The frequencies appear to provide an effective indication of damage. One problem with using natural frequencies as a damage indicator is the fact that the frequencies may also change for a structure under varying environmental or operational conditions [8]. This matter is discussed in **Statistical Pattern Recognition**. Generally speaking, as the natural frequencies are global properties of a structure, it is difficult to use them to progress beyond simple detection. Where this is possible, several frequencies are needed because the position of the damage affects each frequency differently, as described above. Also, if higher frequency, more local, modes are used, there is potential for damage location, if a higher frequency and more local modes are used, there is potential for damage location [5]. However, higher modes are not always available because of the high sensor density required and the associated instrumentation cost.

As a means of progressing beyond detection, one can consider the modeshapes of the structure. When the most severe level of damage (75% reduction in E) is induced in the model here at element 5, again, the resulting modeshapes are shown in Figure 1 (as solid lines) together with the modes for the undamaged structure (dashed). There is only a small distortion, even for this severely damaged condition. While it is true that distortion becomes more marked in the higher modes, these would not usually be available. It is assumed in the following that only the first few modeshapes are available. The low level of distortion suggests that it might be useful to consider further processing of the modeshapes to obtain features that are known to give at least a clear *visual* indication of the damage location. The derived quantities examined below are the Yuen functions [6] and the *mode curvatures*.

2.1.1 Yuen functions

These are functions defined over the length of the beam, and constructed from the eigenvectors and eigenvalues for the damaged and undamaged system. Suppose $\{\psi\}_i$ is the ith eigenvector of the undamaged system and λ_i is the corresponding eigenvalue.

Table 1. First five natural frequencies of the cantilever beam under various damage scenarios

Stiffness reduction (%)	Natural frequencies				
	1	2	3	4	5
0	12.15	76.05	212.7	415.9	686.0
25	11.96	75.71	208.9	413.1	683.6
50	11.62	75.08	202.6	408.7	679.1
75	11.30	74.53	197.6	405.2	675.0
Maximum % change	7.0%	2.0%	7.1%	2.6%	1.6%

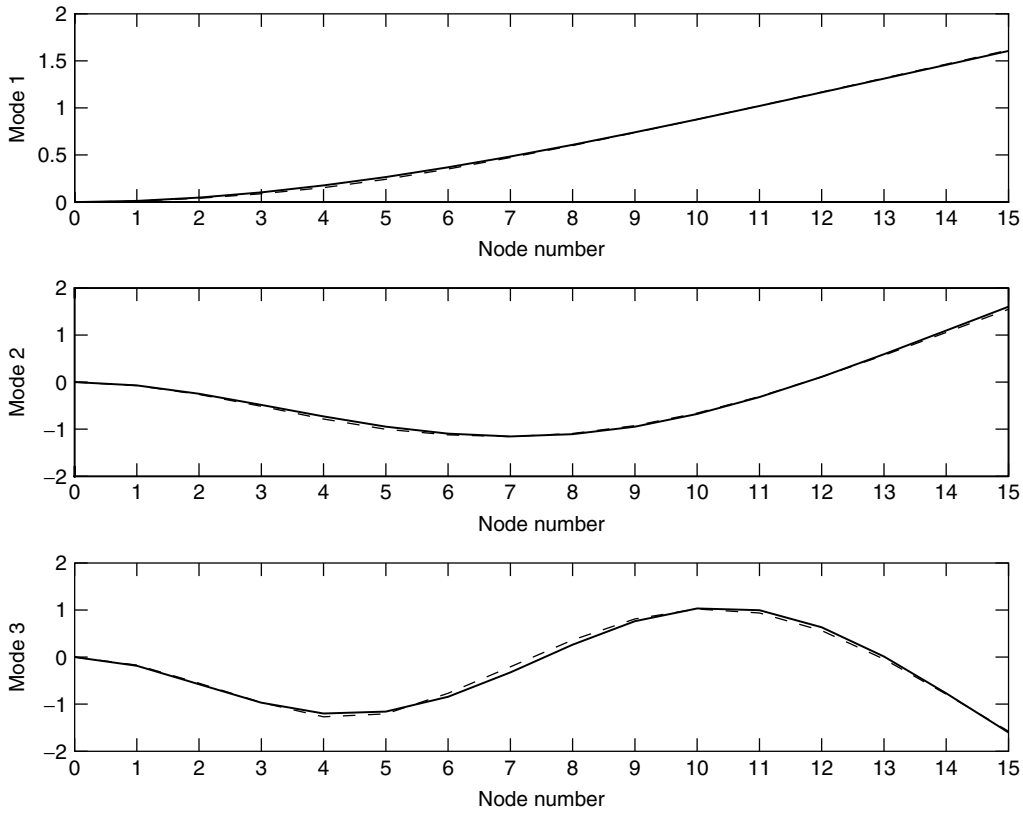


Figure 1. The first three modeshapes for the cantilever beam: undamaged (solid) and damaged (dashed).

Furthermore, suppose $\{\psi'\}_i$ and λ'_i are the corresponding quantities for the damaged system. Then the ith Yuen function is defined by

$$\{Y\}_i = \frac{\{\psi'\}_i}{\lambda'_i} - \frac{\{\psi\}_i}{\lambda_i} \tag{1}$$

The rationale behind these functions is given in [6], and their properties are most easily summarized in a diagrammatic form. Figure 2 shows the function $\{Y\}_1$ (i.e., from the first mode) for each of five damage locations (elements: 2, 4, 6, 8, and 10, with 75% reduction in stiffness). The location of the damage is signaled quite clearly as the node at which the function first becomes nonzero.

Note that the discontinuity in the function becomes less marked toward the free end of the beam; this is expected as very little strain energy is concentrated in this region for vibrations at the first natural frequency. As a consequence, it is difficult to distinguish between the damaged and undamaged modeshape if the fault is in this region.

2.1.2 Modeshape curvatures

Another possibility for a damage-sensitive feature is data obtained by taking derivatives of the modeshapes with respect to the position along the beam. This has the effect of amplifying any discontinuities in the modeshape caused by the localized damage. If the beam stiffness is small in a region, bending is large in that region, and, as a consequence, the beam curvature is high locally. The second derivative with respect to position is used here as an approximation to the curvature. One of the first papers to use curvatures for damage detection was [9]. In this study, the differentiation is carried out by using a centered difference in the main body of the beam with backward and forward differences for the ends.

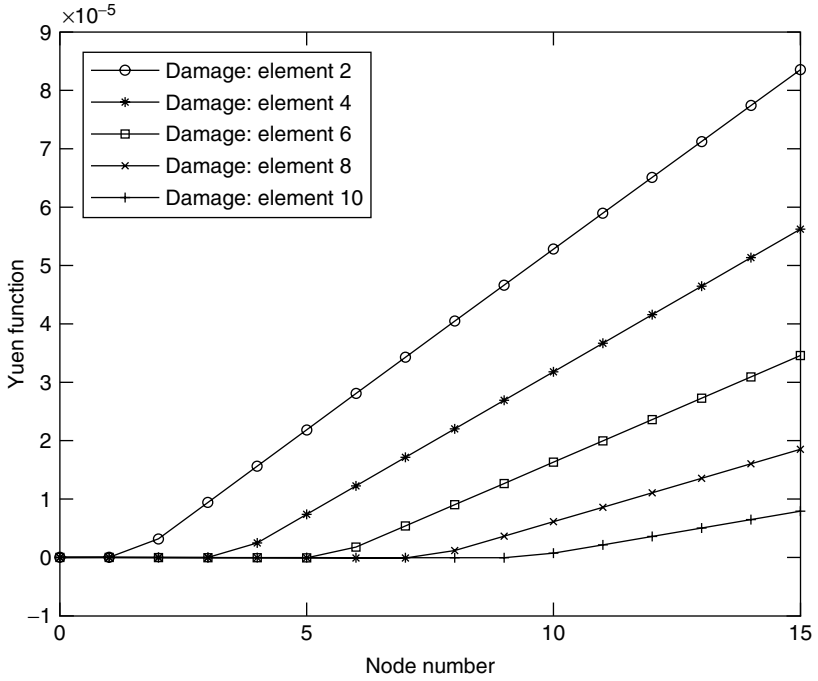


Figure 2. The first Yuen function for the cantilever beam for various different locations of damage.

Figure 3 shows the first mode curvatures when a 75% stiffness reduction is enforced at the five elements considered for the Yuen function. A clear maximum occurs near the location of the damage. This maximum is present for all fault locations. However, as with the Yuen functions, damage in the neighborhood of the free end is poorly indicated.

The effect of the damage can be amplified by subtracting the modeshape curvature of the undamaged structure as in Figure 4.

An interesting approach related to curvature modeshapes was developed by Ratcliffe in [10].

2.2 A cantilever plate structure

To illustrate the use of the modeshape–derived features in a slightly more realistic situation, a two-dimensional cantilever plate structure was considered. The plate was modeled using LUSAS in much the same way as the beam structure. The material constants of aluminum were used and the dimensions chosen were $300 \times 200 \times 2.5\,\text{mm}^3$; the built-in end was taken along one of the short sides. The aspect ratio of 1.5 was chosen to have a check on the results from the FE model; for this value, Leissa [11] provided estimates of the first six natural frequencies, namely, 23.7, 79.5, 147.3, 269.0, 367.0, and 422.3 Hz. The FE mesh was composed of 20×20 elements; however, as such dense sensor networks are not usually available, it was decided to also estimate the modeshapes on a regular 4×4 subgrid. The simulated faults were located within the areas defined by the coarse grid. Faults were simulated by lowering the stiffness of the elements as before; in this case, the Young's Modulus was reduced to 1% of its usual value. To give faults of different severities, different groups of elements were "deleted". Only the most severe case is considered here, where a 3×3 group of elements was deleted to give the fault.

For the FE model, eight-node semiloof elements were used, although this was a little time consuming, the results justified the expenditure. The estimated natural frequencies were 23.5, 79.4, 146.2, 364.6, and 419.2 Hz, agreeing with [11] to well within its stated confidence interval. Figure 5 shows the first modeshape for the undamaged system as a surface and in contour-map form.

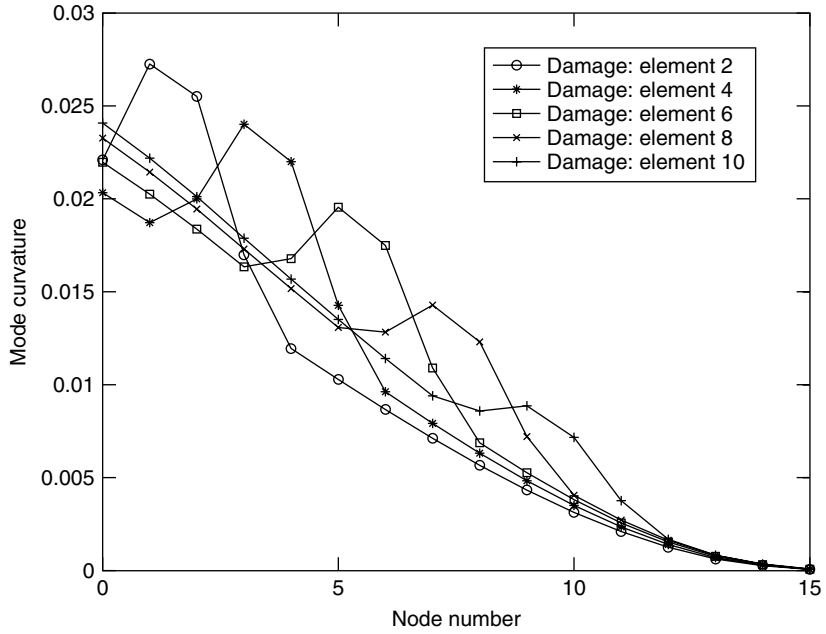


Figure 3. First modeshape curvature for cantilever beam for various different locations of damage.

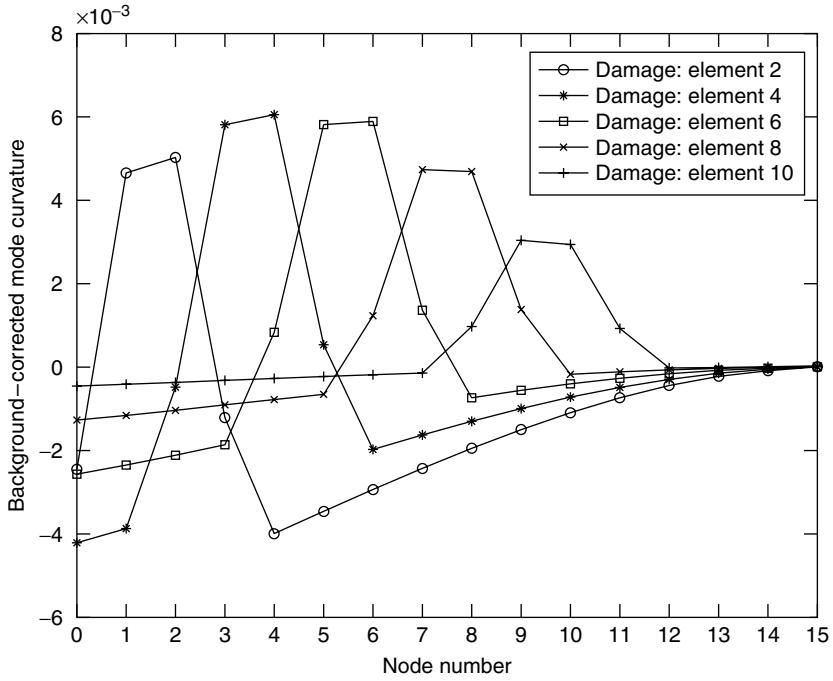


Figure 4. Background-corrected modeshape curvature for cantilever beam for various different locations of damage.

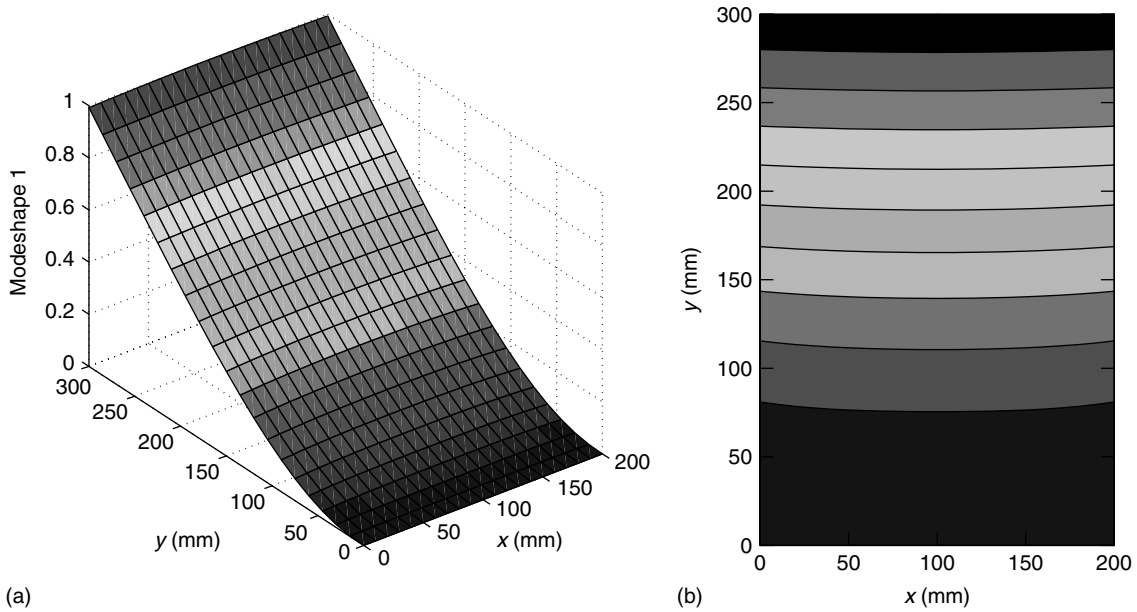


Figure 5. The first modeshape for the undamaged cantilever plate. (a) Surface and (b) contour.

First, the modeshape is considered as a feature for damage detection. When damage is induced in one of the edge cells nearest the clamped edge, there proves to be very little indication of damage in the modeshape. There is some improvement if the modeshape of the undamaged plate is subtracted. The result is shown in Figure 6.

As in the case of the cantilever beam, the modeshape itself gives very little indication of the damage; note that Figure 6 shows the modeshape over the fine mesh. Over the coarse mesh, which would represent a more realistic situation, there is very little discrimination. This is motivation for considering quantities derived from the modeshapes once more. As the modeshape curvatures proved useful in the one-dimensional case, they are considered here.

The curvatures proved a little more involved in this case; as the modal displacement $\{\psi\}$ was now a function of both x and y (convenient coordinates on the plate), the derivative was replaced by the gradient vector,

$$\underline{\nabla}\psi = \left(\frac{\partial \psi}{\partial x}, \frac{\partial \psi}{\partial y}\right) \tag{2}$$

As some scalar analog of the curvature was still required for visualization, the following quantity was chosen:

$$\kappa = \|\underline{\nabla}\|\underline{\nabla}\psi\|\| \tag{3}$$

where

$$\|\underline{\nabla}\psi\| = \sqrt{\left(\frac{\partial \psi}{\partial x}\right)^2 + \left(\frac{\partial \psi}{\partial y}\right)^2} \tag{4}$$

explains the use of the norm $\|\ \|$. A simple centered difference was used to estimate the partial derivatives. To amplify the disturbance due to the fault, the curvatures corresponding to the undamaged state are subtracted throughout. Figure 7 shows the first mode curvature for a fault at the same location as previously. The position of the fault is clearly visible, whereas it was not at all discernable in the modeshape.

A legitimate criticism at this point would be that the curvatures would usually be estimated from the data on the measurement grid and therefore would not be available with the accuracy shown here. In fact, the process of double differentiation to generate the modeshape curvatures is badly influenced by noise, and it was shown in [12] that numerical analysis gives unacceptable errors. However, it was shown

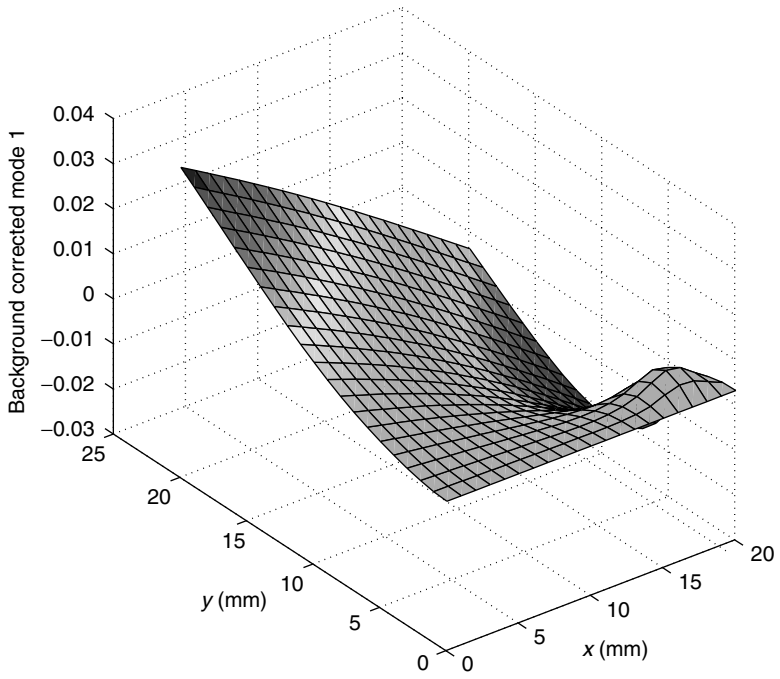


Figure 6. Background-corrected modeshape for the damaged cantilever plate.

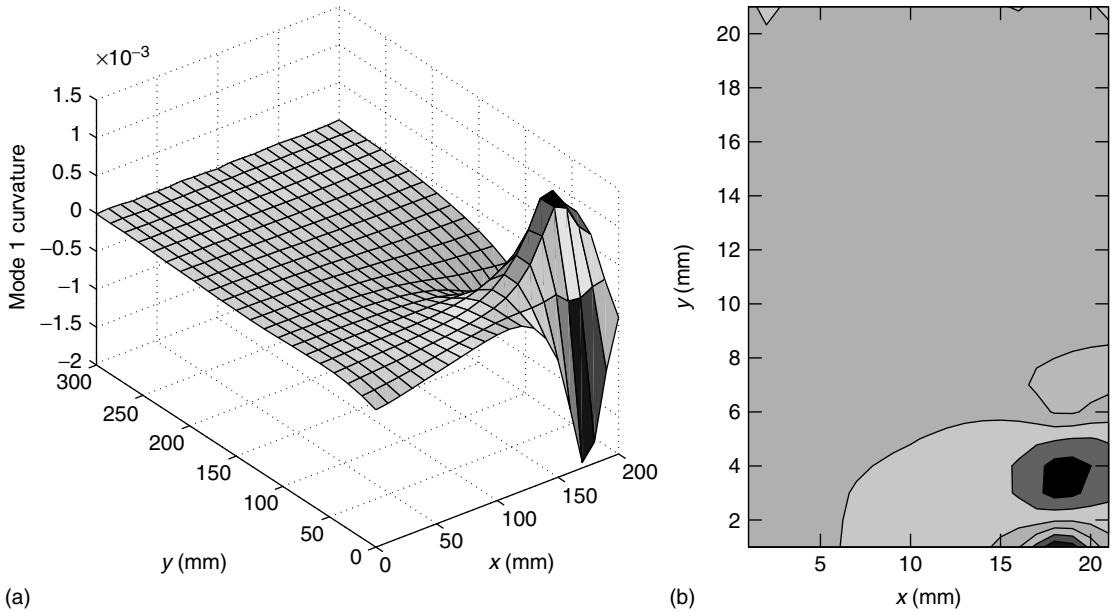


Figure 7. First modeshape curvature for damaged cantilever plate. (a) Surface and (b) contour.

in the same study that strain measurements make an acceptable proxy for the curvatures as long as the sensor density is reasonably high. The feasibility was demonstrated on an experimental cantilever beam with a machined element simulating a crack. Assuming that strain measurements are available, it will still not be possible to use the sort of sensor density implicit in Figure 7. If the coarse 4×4 mesh is used, which would require 16 sensors, the curvatures of Figure 8 are obtained.

The results are less impressive, but the vicinity of the damage is indicated. Before leaving the subject of modeshape features, it should be remarked that the modeshape curvatures play an important part in the *damage index* or *strain energy* method of Stubbs and Kim [13]. In Refs. 14 and 15, the authors compared a number of damage detection methods using the example of a road bridge with a concrete deck and steel supports. The strain energy method proved the most promising, and a detailed discussion is provided in the next section.

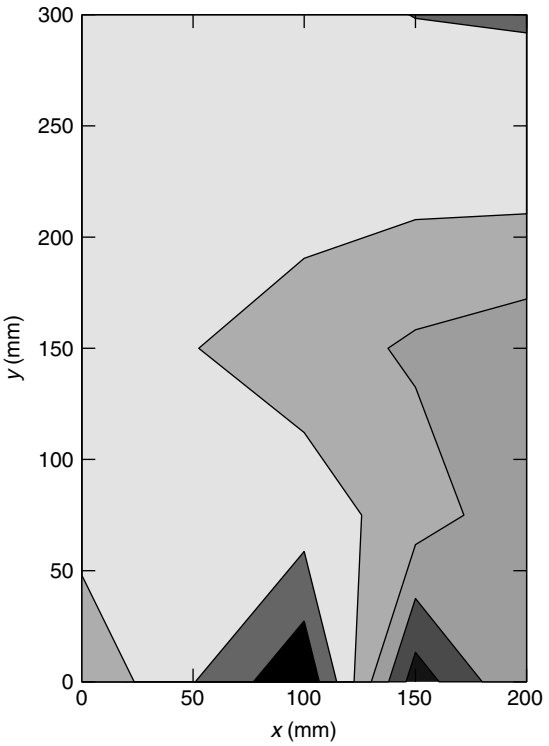


Figure 8. Background-corrected modeshape curvature for the damaged cantilever plate: coarse mesh.

3 THE STRAIN ENERGY METHOD

The approach considered here was developed by Stubbs and Kim [13, 16] and is based on *modal strain energy*. It has been applied in a number of experimental studies over the years, including an analysis of a full-scale bridge structure [17]. The method has mainly been applied to beamlike or one-dimensional structures. In fact, as the object of the exercise here is to identify and localize damage along an aircraft panel stringer, this is sufficient. The extension to two-dimensional structures was made in [18].

3.1 Theory

The analysis here closely follows that of [18]. Consider a beamlike structure discretized into a number of elements labeled by $j = 1, \ldots, N_e$, and suppose that measured modal data is available for N_m modes.

The strain energy of an Euler–Bernoulli beam of length L is given by

$$U = \frac{1}{2}\int_0^L \mathrm{d}x\,(\mathrm{EI})\left(\frac{\mathrm{d}^2 w}{\mathrm{d}x^2}\right)^2 \tag{5}$$

where $w(x)$ is the displacement profile of the beam and EI is the structural rigidity. The "fraction" of the energy associated with a given modeshape $\phi_i(x)$ is simply

$$U_i = \frac{1}{2}\int_0^L \mathrm{d}x\,(\mathrm{EI})\left(\frac{\mathrm{d}^2 \phi_i}{\mathrm{d}x^2}\right)^2 \tag{6}$$

If one further divides the energy among the N_e elements, one finds for element j,

$$U_{ij} = \frac{1}{2}\int_{a_{j-1}}^{a_j} \mathrm{d}x\,(\mathrm{EI})_j\left(\frac{\mathrm{d}^2 \phi_i}{\mathrm{d}x^2}\right)^2 \tag{7}$$

where a_{j-1} and a_j are the endpoints of element j. (The nodes at the ends of the beam are a_0 and a_{N_e}.)

The proportion of modal energy in the jth element is

$$F_{ij} = \frac{U_{ij}}{U_i} \tag{8}$$

and

$$\sum_{j=1}^{N_e} F_{ij} = 1 \tag{9}$$

If the system is damaged, there is a redistribution or scaling of the strain energy and the analogous quantities can be defined as

$$U_i^* = \frac{1}{2}\int_0^L \mathrm{d}x\,(\mathrm{EI})^* \left(\frac{\mathrm{d}^2\phi_i^*}{\mathrm{d}x^2}\right)^2 \tag{10}$$

$$U_{ij}^* = \frac{1}{2}\int_{a_{j-1}}^{a_j} \mathrm{d}x\,(\mathrm{EI})_j^* \left(\frac{\mathrm{d}^2\phi_i^*}{\mathrm{d}x^2}\right)^2 \tag{11}$$

$$F_{ij}^* = \frac{U_{ij}^*}{U_i^*} \tag{12}$$

where again

$$\sum_{j=1}^{N_e} F_{ij}^* = 1 \tag{13}$$

If the elements are small enough, the flexural rigidity is constant on the elements, and, if it is further assumed that (EI) and $(\mathrm{EI})^*$ are constant and equal for the purposes of evaluating U_i and U_i^*, then

$$F_{ij}^* = \frac{(\mathrm{EI})_j^*}{(\mathrm{EI})} \frac{\int_{a_{j-1}}^{a_j} \mathrm{d}x\ \left(\mathrm{d}^2\phi_i^*/\mathrm{d}x^2\right)^2}{\int_0^L \mathrm{d}x\ \left(\mathrm{d}^2\phi_i^*/\mathrm{d}x^2\right)^2} = \frac{(\mathrm{EI})_j^*}{(\mathrm{EI})} f_{ij}^* \tag{14}$$

with a similar definition for the undamaged quantity.

Now, suppose that the damage is only in element k and that all the undamaged elements store the same fraction of modal strain energy before and after damage i.e., $F_{ij}^* = F_{ij} \forall j \neq k$. It follows from (9) and (13) that $F_{ik}^* = F_{ik}$ or

$$\frac{(\mathrm{EI})_k^*}{(\mathrm{EI})} f_{ik}^* = \frac{(\mathrm{EI})_k}{(\mathrm{EI})} f_{ik} \tag{15}$$

so,

$$\frac{(\mathrm{EI})_k}{(\mathrm{EI})_k^*} = \frac{f_{ik}^*}{f_{ik}} = \beta_{ik} \tag{16}$$

The quantity β_{ik} is the *damage index* associated with mode i and element k. It is assumed that damage to an element manifests itself in a decrease in the flexural rigidity and a consequent increase in the index. To eliminate possible problems due to near-zero values of the denominator, a slightly modified version is used here.

$$\beta_{ik} = \frac{f_{ik}^* + 1}{f_{ik} + 1} \tag{17}$$

To obtain a more robust diagnostic, one can integrate the information across all the measured modes and form

$$\beta_k = \frac{\sum_{i=1}^{N_m} f_{ik}^* + 1}{\sum_{i=1}^{N_m} f_{ik} + 1} \tag{18}$$

As in most statistical detection methods, it is important to have a threshold, which asserts that there is a significant departure from unity for the damage index. In [19] and [18], it is assumed that the statistics can be obtained by taking the mean and standard deviation over all the elements. If the index is assumed to have a Gaussian distribution, then one can assert that indices more than two standard deviations above the mean can be associated with possible damage locations. Alternatively, if the normalized index Z_k is used, where

$$Z_k = \frac{\beta_k - \overline{\beta}}{\sigma_\beta} - 2 \tag{19}$$

then potential damage sites are associated with positive Z_k. Note that, if the statistics are estimated from index values on a damaged structure, this threshold is equivalent to using an inclusive discordancy measure to detect outliers of the index [20]. In practice, the statistics of Z_k are unlikely to be Gaussian and this means that the decision rule above will not usually give 97% confidence in the diagnosis as would be expected from the two standard deviations rule. Whether the decision rule is conservative or not will depend on the true statistics of Z_k. A more rigorous approach to determining the threshold could be by means of probability density estimation.

3.2 The experimental panel and data capture

The panel was constructed to the following specifications. The upper surface was $750 \times 500 \times 3\,\text{mm}^3$ aluminum sheet (Figure 9 shows a schematic figure). This was stiffened by the addition of two ribs composed of lengths of C-channel riveted to the short edges. Two stiffening stringers composed of angle section ran along the length of the sheet. The tests were all conducted with free–free boundary conditions for the panel, which was suspended from a substantial frame using springs and nylon line.

Damage was simulated by the introduction of a saw cut in the outside stringer, 125 mm from the edge of the panel (Figure 9). Nine levels of damage were investigated from 10% depth to 90%. As the stringer is 1 in. (25.4 mm) in depth, each level corresponds to 2.5 mm of damage.

The system was excited using a Gearing and Watson electrodynamic shaker driven by broadband white noise amplified by a Gearing and Watson power amplifier. The responses were measured using PCB resonant piezoelectric accelerometers and sampled using a DIFA/Scadas acquisition system running LMS software under the control of an HP computer. The DIFA system was also used to form the random excitation.

The object of the experiment was to carry out a complete top surface modal analysis using 19 measurement points (Figure 10) for each stage of damage (location 1 is the shaker attachment point, points 2–20 are the accelerometer positions). However, for the purposes of this illustration, only the data measured along the line of the stringer is relevant.

3.3 Results from the strain energy method

For the strain energy method, the analysis required the measurement of modeshapes from the structure. This was accomplished by using the LMS modal analysis package. The sensor layout was as shown in Figure 10. To use the 1D formulation of the strain energy method, only the modeshapes along the

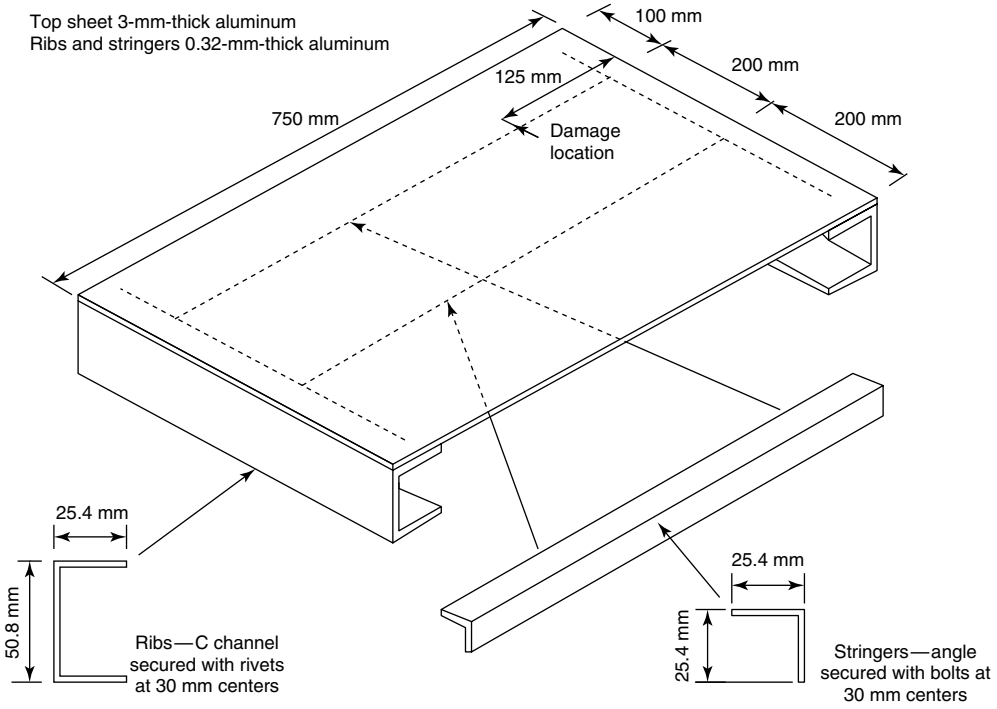


Figure 9. Schematic of the simulated skin panel.

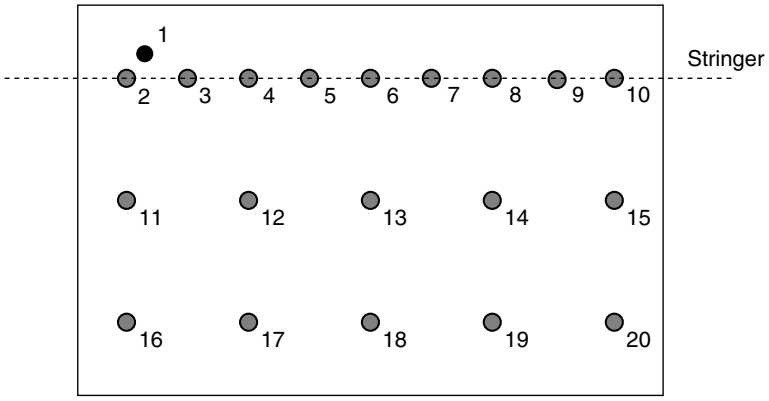


Figure 10. Modal analysis measurement points.

stringer are used in what follows i.e., those at sensors 2–10.

The frequency interval up to 150 Hz contained eight distinguishable modes. These were selected for the analysis. A time-domain curve-fitting technique was used to extract modal parameters from the measured FRFs; broadband excitation was used. A modal analysis was carried out for the normal condition and for all nine levels of damage. The natural frequencies of the eight modes of interest are given in Table 2 as a function of the damage level.

Only modes 6 and 8 show any significant systematic variation in their frequencies as the damage increases. Mode 4 was difficult to identify and in one case (damage level 50%) the curve fitting and subsequent stability analysis failed to identify it at all.

Figures 11 and 12 show some of the modeshapes along the stringer for each level of damage. The modeshapes are normalized so that $||\phi_i|| = 1$. The first mode appeared to be completely insensitive to damage, and the second mode was not much different. Mode 3 showed no variation until the damage level reached 90%, when there was a perturbation in the vicinity of sensor 9, which is closest to the damage site. Mode 4 showed a degree of variation, but, given that this mode was consistently difficult to identify, this may be a manifestation of a repeatability problem. Modes 5 and 6 both displayed some systematic variation with the level of damage. Modes 7 and 8 showed the most marked variation with damage (Figures 11 and 12, the highest levels are highlighted using dotted and dashed lines); in fact, the four highest levels for mode 7 can be distinguished by the eye.

It is rather surprising that there is not more correlation between the extent of the variation for natural frequency and that for modeshape. In particular, mode 7 shows by far the greatest change in modeshape while retaining a constant natural frequency over the different levels of damage.

Table 2. Natural frequencies from modal tests

Mode	Natural frequency (Hz)									
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1	24.2	24.2	24.2	24.2	24.2	24.2	24.2	24.2	24.2	24.2
2	30.9	30.8	31.0	31.0	31.0	30.2	30.9	31.0	30.8	30.7
3	65.1	65.0	64.9	65.0	65.1	65.0	65.0	65.0	65.1	65.1
4	79.7	80.9	79.8	79.8	79.8	—	79.6	79.8	79.6	79.4
5	86.7	86.7	86.7	86.7	86.7	86.7	86.7	86.7	86.6	86.8
6	108.5	108.6	108.5	108.5	108.5	108.3	108.0	107.7	106.4	103.5
7	110.3	110.3	110.3	110.3	110.2	110.3	110.3	110.3	110.3	110.3
8	118.4	118.4	118.3	118.0	117.5	116.6	115.1	114.0	111.9	110.8

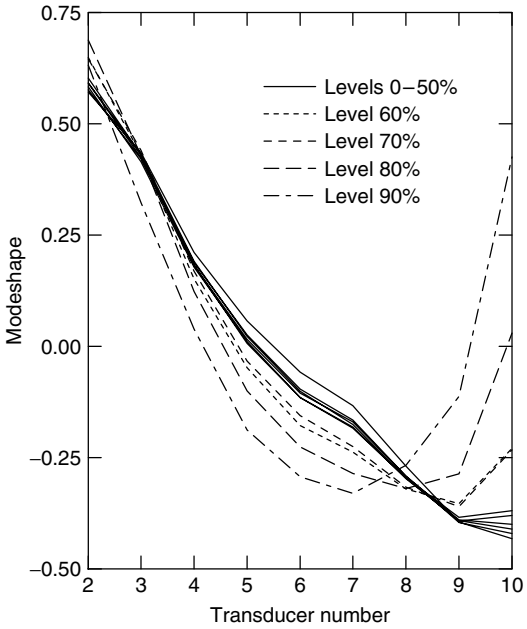


Figure 11. Measured modeshapes for mode 7 for all levels of damage.

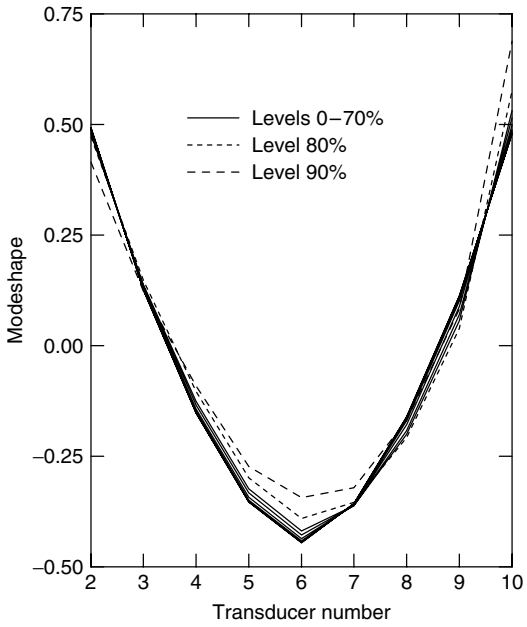


Figure 12. Measured modeshapes for mode 8 for all levels of damage.

The damage indices for all modes were computed using equation (17) and normalized as in equation (19). As the data stands, there are only nine measurement points per modeshape and it is therefore impossible to localize the damage at a nonsensor location. To increase the resolution, a cubic convolution polynomial interpolation was used to estimate the modeshape at 19 regularly spaced points between the sensors. This gave an array of 161 *pseudosensor* values for the modeshape. This, in turn, gave 160 elements. The differentiations that were needed to estimate the curvatures $\mathrm{d}^2\phi_i/\mathrm{d}x^2$ were carried out using a centered difference in the main body of the array and forward and backward differences at the ends. The computation was carried out in Matlab to take advantage of the built-in interpolation routines.

For the modal damage indices, modes 1–5 gave poor results; only the highest level of damage was flagged for modes 1, 3, and 5. (The statistics for normalizing the indices were computed for each mode individually, but averaging over all damage levels. Missing data from mode 4 at the 50% level was approximated by the corresponding 60% level data.) More worryingly, these indices showed many excursions above threshold at non-damaged locations e.g., see Figure 13 for mode 2. This behavior is consistent with the fact that the first five modes showed little or no systematic variation with the damage. The results for the three highest modes were far more encouraging. Each signaled damage: mode 6 from the 80% extent onward and modes 7 and 8 from the 60% level onward. Figure 14 shows the indices for mode 8. More importantly, there were no spurious excursions above the threshold. One of the main reasons for using the strain energy method was to locate the damage. The peaks in the index for mode 8, which is the clearest, are consistently over element 139. Note that this is the location of sensor 9. In reality, the damage is in element 134. As each element is 4.24 mm wide, this amounts to a location error of 21.25 mm. However, note that the index is not above the threshold at a single point; each excursion signals damage over an interval. The extent of the intervals for mode 8 are given in Table 3. A star in Table 3 indicates that the identified interval includes the damage location.

It can be seen from Table 3 that the correct damage element is included from the 80% extent onward. The starred values in the table indicate that the end effects from the interpolation routine have created a plateau,

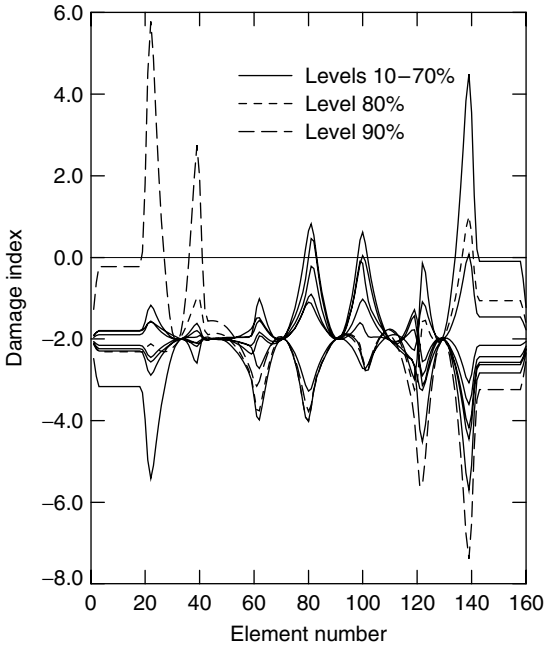


Figure 13. Damage indices for mode 2 for all levels of damage.

Table 3. Damage location intervals for mode 8 index

Level (%)	Interval
60	136–140
70	135–141
80	132–160*
90	132–160*

which causes the index to be above threshold over the whole region between the index maximum and the final element. This appears to be due to the cubic nature of the interpolation and might be avoided by passing to a higher-order interpolant. This plateau is also visible in the results of [21]. The plateaus can be eliminated by using a cubic spline; however, this option reduces the sensitivity to damage.

Multimode damage indices were also computed using equation (18). When all the modes were used, the two highest damage levels were detected and located in the correct vicinity (the plateau was present again). Unfortunately, there was a small indication of damage at a spurious location. Figure 15 shows the multimode index computed using modes 6–8; the

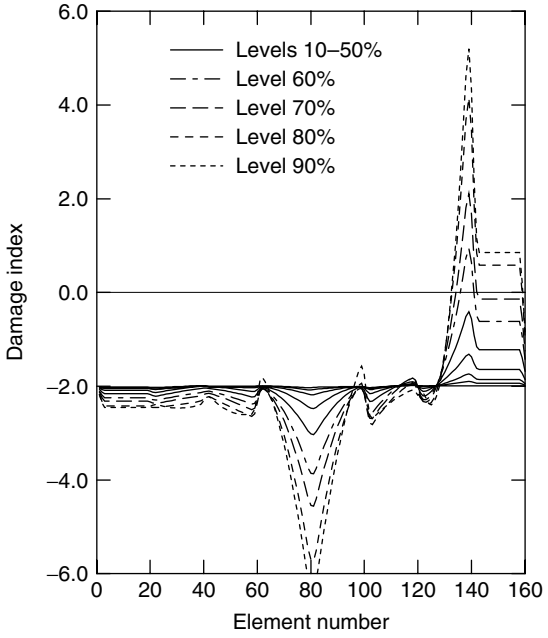


Figure 14. Damage indices for mode 8 for all levels of damage.

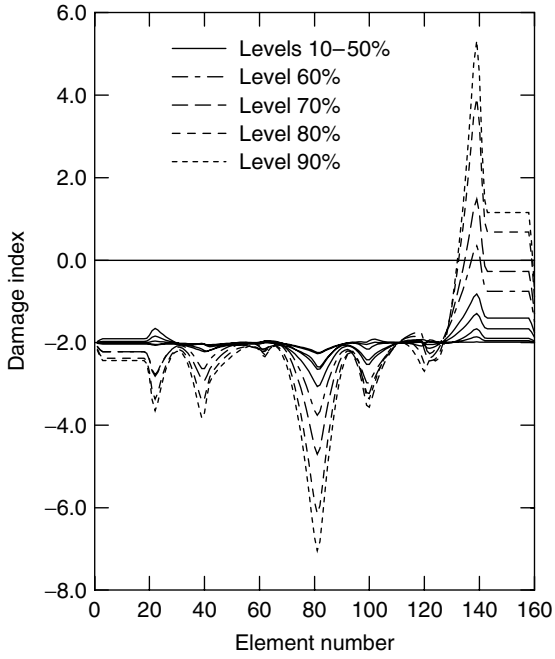


Figure 15. Damage indices for modes 6 to 8 for all levels of damage.

results are much better, damage is detected from the 60% level onward, the location accuracy is the same as before, and the most probable location appears to be at sensor 9.

4 LINEAR-ALGEBRAIC METHODS

The methods described so far are quite basic in the sense that they require only simple numerical operations (simple in nature, if not always in implementation) on single modeshapes. The object of this section is to look at methods based on more sophisticated numerical processing and involving more detailed specifications of the structure. If, for example, the modal and/or physical matrices for the structure are obtainable in the undamaged and (potentially) damaged states, it will be possible to make a comparison and this may allow one to identify the location and severity of the damage. All of the methods described in the section make critical use of the fact that "baseline" data is available for the undamaged structure. There is an extensive literature on the application of linear algebra in the context of SHM, and no attempt is made here to provide anything like a review. Only a small selection of the available techniques is presented here. The interested reader can consult the comprehensive reviews [2, 3] for a more complete coverage of the field. One linear-algebraic approach, the method of FE updating, has received a great deal of attention. Because this method is used extensively, it is described in detail in the following section.

4.1 A direct-update method

To apply this method, one must assume that one or more of the natural frequencies and corresponding modeshapes have been measured. Suppose $[M_u]$, $[C_u]$, and $[K_u]$ are the mass, damping, and stiffness matrices of the undamaged structure and that λ_{ui} and $\{\phi_{ui}\}$ denote the ith natural frequency and corresponding modeshape vector. The eigenvalue equation for the structure gives

$$(\lambda_{ui}^2[M_u] + \lambda_{ui}[C_u] + [K_u])\{\phi_{ui}\} = 0 \qquad (20)$$

For the corresponding quantities of the damaged structure (indicated by a subscript d), one has

$$(\lambda_{di}^2[M_d] + \lambda_{di}[C_d] + [K_d])\{\phi_{di}\} = 0 \qquad (21)$$

If one assumes, for simplicity, that damage only affects the stiffness matrix, then one can summarize the change between the undamaged and damaged systems by the simple expression

$$[K_d] = [K_u] + [\delta K] \qquad (22)$$

and the problem of damage identification is reduced to the problem of computing the "update" matrix $[\delta K]$. One makes use of equation (21), which now takes the form (noting that $[M_d] = [M_u]$ and $[C_d] = [C_u]$),

$$\begin{aligned}\{B_i\} &= [\lambda_{di}^2[M_u] + \lambda_{di}[C_u] + [K_u]]\{\phi_{di}\} \\ &= [\delta K]\{\phi_{di}\} \quad \forall i = 1, \ldots, r \end{aligned} \qquad (23)$$

where r is the number of measured modal properties. The vector $\{B_i\}$ can be tested to see whether it departs significantly from zero and this provides a detection method for damage. If the damage is local, one can further infer the location of the damage from the position of the nonzero entries within the vector $\{B_i\}$. Various methods of computing the physical matrices or their perturbations have been investigated and the reader is referred to [2] for a discussion under the title "Matrix update methods".

To illustrate the method, data were generated from a simple synthetic system. The system was a linear chain of 20 identical masses m, connected in series by identical springs of stiffness k. The two end masses were connected to the ground by two further springs of stiffness k. The damping of the system was assumed to be zero. Values of $m = 1\,\mathrm{kg}$ and $k = 10^4\,\mathrm{N\,m^{-1}}$ were used to generate the system matrices and solve the associated eigenvalue equation for the natural frequencies and modeshapes.

To simulate damage in the system, the stiffness of the spring joining masses 5 and 6 was reduced by 50%. When the residual vector $\{B_i\}$ was computed for the first three modes using equation (23), the results shown in Figure 16 were obtained.

The results are perfect and locate the damage to be between the fifth and sixth masses. However, one should be aware that the data used so far is also perfect as it has been generated from numerical simulation. To investigate the effects of noise, the

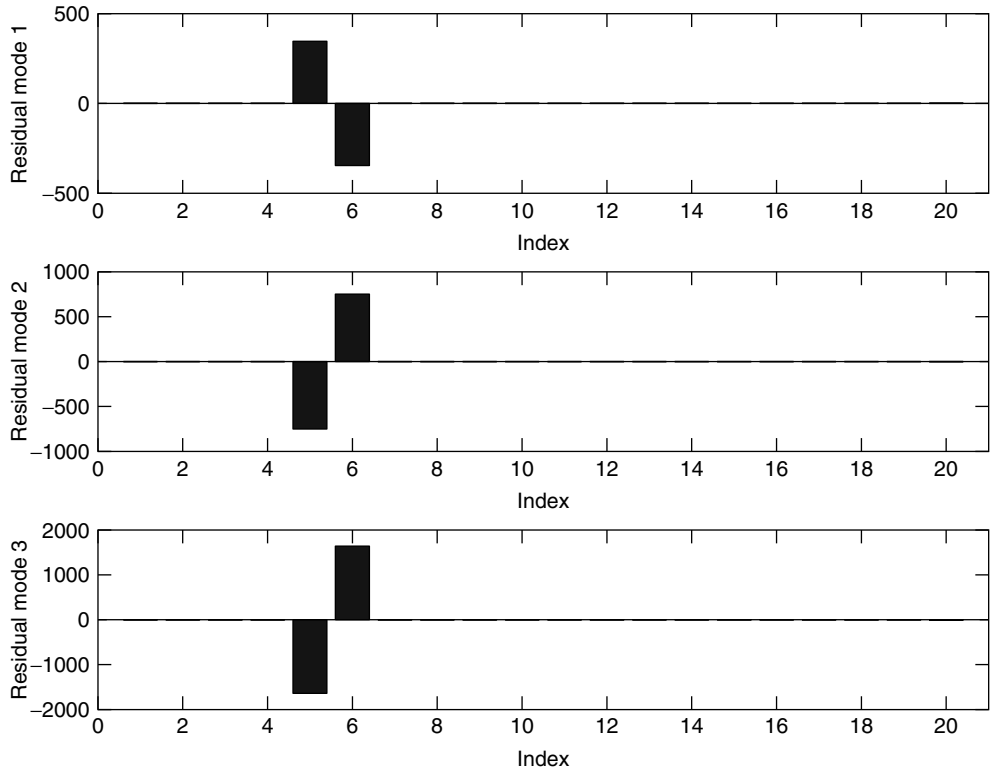


Figure 16. Residual vectors for the first three modes of the damaged system.

modeshapes from the damaged system were perturbed by Gaussian noise. The noise level was taken as 5% of the maximum modeshape value, which is a realistic level. The natural frequencies, which can be measured more accurately, were left clean. The residual vectors for the noisy modes are given in Figure 17. The average residual vector is also given, taken over the three modes.

Although the location is still indicated, the results are not as good. The problem is that the update with the noisy modeshape vectors has smeared the stiffness change over the whole stiffness matrix. One possible solution to this problem is given by the next method discussed.

4.2 A minimum-rank update method

Generally, one would wish to add a constraint to the computation of the residual vectors $\{B_i\}$ or perturbation matrix $[\delta K]$. This is the requirement that the perturbation, if it is to truly reflect a *local* change, must be mainly zero with only a small number of elements being nonzero. Unfortunately, the update method shown above has no such constraint and the result is that the effect of the damage is smeared across the residuals. Zimmerman and Kaouk [22, 23] proposed that one way to enforce locality within the perturbation matrix would be to demand that it had a low rank. The method proceeds by computing

$$[B] = [M_{\mathrm{u}}][\phi_{\mathrm{d}}][\Lambda_{\mathrm{d}}]^2 + [C_{\mathrm{u}}][\phi_{\mathrm{d}}][\Lambda_{\mathrm{d}}] + [K_{\mathrm{d}}][\phi_{\mathrm{d}}] = [\delta K][\phi_{\mathrm{d}}] \tag{24}$$

where $[\phi_{\mathrm{d}}] = [\{\phi_{\mathrm{d}1}\}\{\phi_{\mathrm{d}2}\}\ldots\{\phi_{\mathrm{d}r}\}]$ and $[\Lambda_{\mathrm{d}}] = \mathrm{diag}\,[\lambda_{\mathrm{d}1}, \lambda_{\mathrm{d}2}, \ldots, \lambda_{\mathrm{d}r}]$. Zimmerman and Kaouk showed that the minimum-rank solution to equation (24) is [22, 23],

$$[\delta K] = [B]([B]^{\mathrm{T}}[\phi_{\mathrm{d}}])^{-1}[B]^{\mathrm{T}} \tag{25}$$

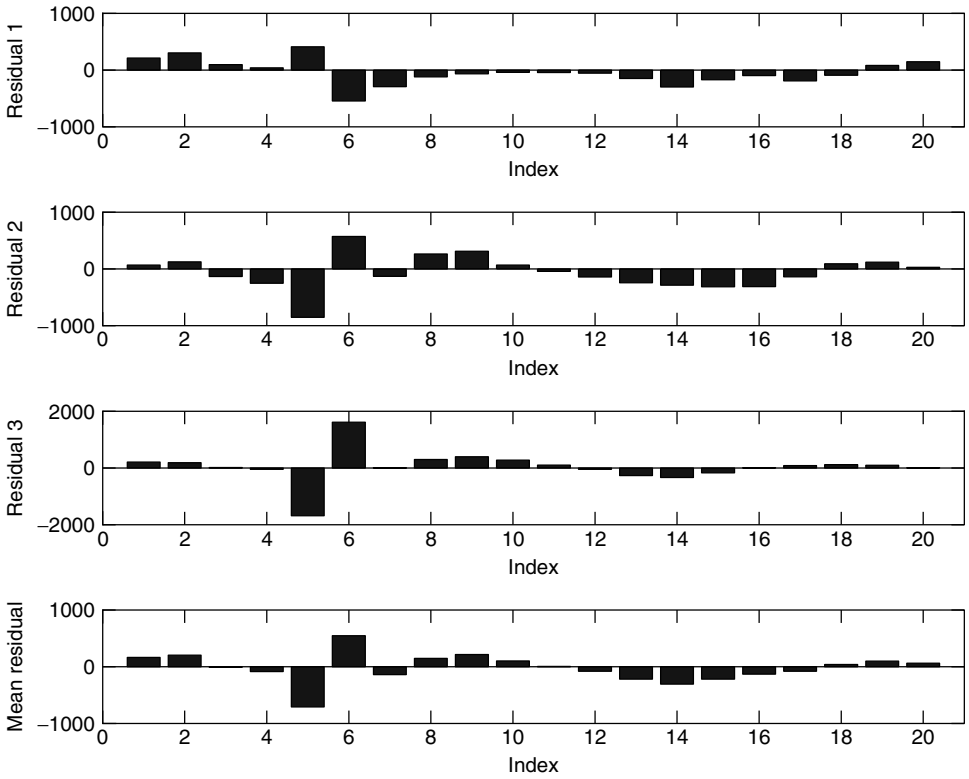


Figure 17. Residual vectors for the first three noisy modes of the damaged system.

To illustrate this approach, data from the same simulated 20-DOF system as before was used. The perturbation matrix for the damaged system in equation (25) was computed using all of the clean modeshapes. The resulting perturbation is shown in Figure 18.

The advantage of the minimum-rank update approach is shown clearly when the method is then applied to the noisy modeshapes. The results are as shown in Figure 19.

The position and size of the damage is shown just as clearly in Figure 19 as in Figure 18. Kaouk and Zimmerman further extended the approach so that it could simultaneously estimate perturbations to the mass, damping, and stiffness matrices [24].

4.3 Flexibilities

Rather than trying to estimate the change in the stiffness matrix, it is sometimes more straightforward to estimate the change in its inverse, namely, the flexibility matrix $[G] = [K]^{-1}$. In terms of modal quantities, the flexibility is given by

$$[G] = [\phi][\Lambda][\phi]^{\mathrm{T}} = \sum_{i=1}^{n} \frac{1}{\omega_i^2} \{\phi_i\}\{\phi_i\}^{\mathrm{T}} \qquad (26)$$

or, if only the first r modes are measured,

$$[G] \approx \sum_{i=1}^{r} \frac{1}{\omega_i^2} \{\phi_i\}\{\phi_i\}^{\mathrm{T}} \qquad (27)$$

It is clear from these expressions that the lowest modes have the most influence on the flexibility matrix. Pandey and Biswas [25] proposed a damage indicator based on the difference between the measured flexibilities of the damaged and undamaged systems,

$$[\delta G] = [G_{\mathrm{u}}] - [G_{\mathrm{d}}] \qquad (28)$$

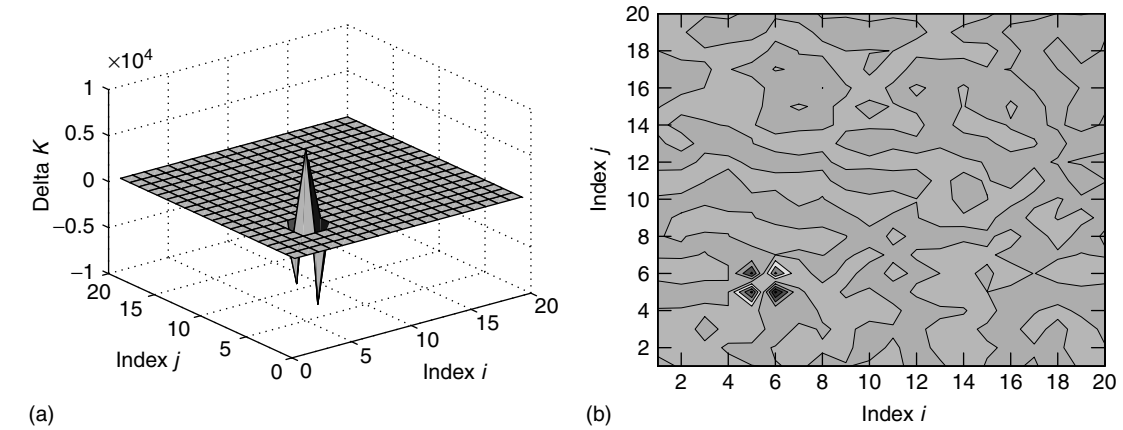


Figure 18. Perturbation matrix from the minimum-rank update for the damaged 20-DOF system: clean data. (a) Surface and (b) contour.

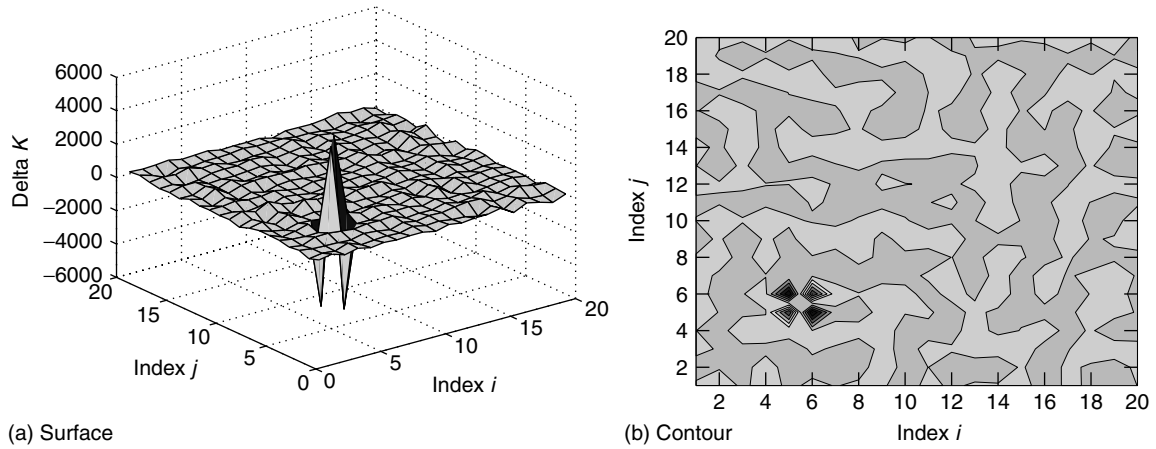


Figure 19. Perturbation matrix from the minimum-rank update for the damaged 20-DOF system: noisy data.

where, of course,

$$[G_{\mathrm{u}}] \approx \sum_{i=1}^{r} \frac{1}{\omega_{\mathrm{u}i}^2} \{\phi_{\mathrm{u}i}\}\{\phi_{\mathrm{u}i}\}^{\mathrm{T}} \tag{29}$$

and

$$[G_{\mathrm{d}}] \approx \sum_{i=1}^{r} \frac{1}{\omega_{\mathrm{d}i}^2} \{\phi_{\mathrm{d}i}\}\{\phi_{\mathrm{d}i}\}^{\mathrm{T}} \tag{30}$$

The use of this measure can be illustrated using the data from the simulated 20-DOF system introduced earlier. The change in flexibility for the damaged system was computed using (all of) the clean modeshapes and is displayed as a contour map in Figure 20.

The results of this method are often presented as a plot of selected columns of the flexibility difference, and the location of the damage is evident from the plot shown in Figure 21.

If the noisy modeshapes from the simulation are used to compute the flexibility change, there is not much degradation in diagnostic performance, as shown in Figures 22 and 23. Although there is some distortion in the plots, the features that indicate the location of damage are still present.

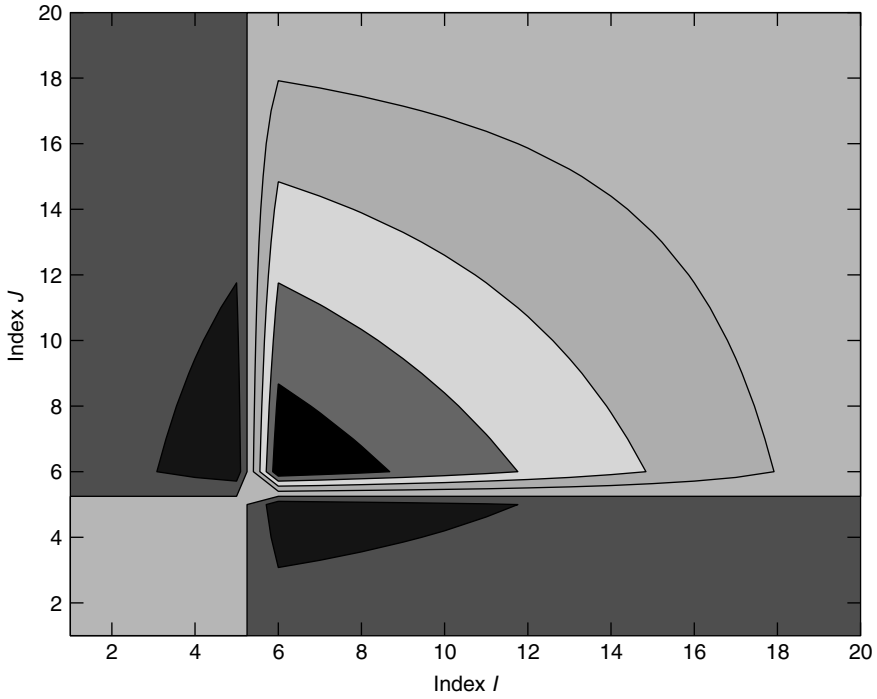


Figure 20. Computed flexibility matrix for the damaged 20-DOF system: clean modes.

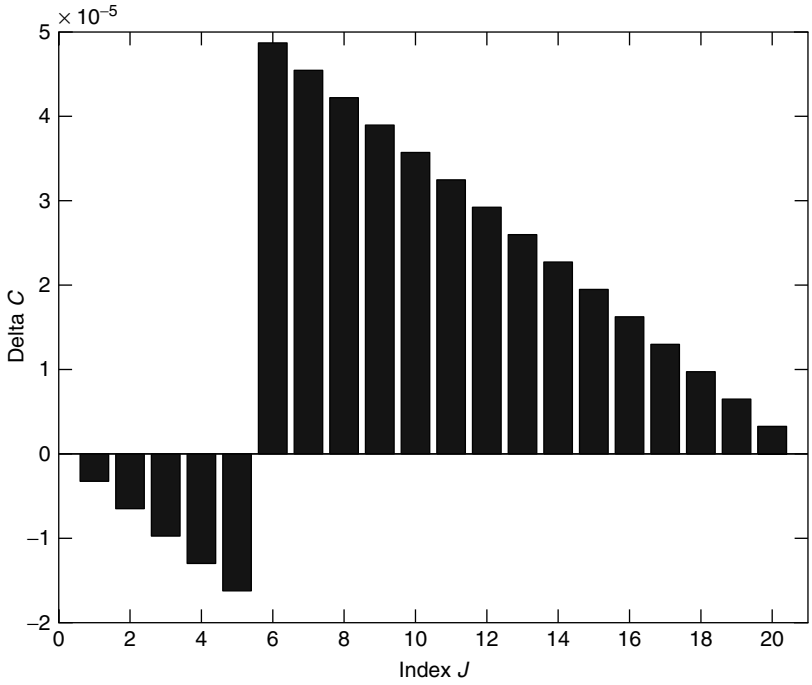


Figure 21. Column 6 of the computed flexibility matrix for the damaged 20-DOF system: clean modes.

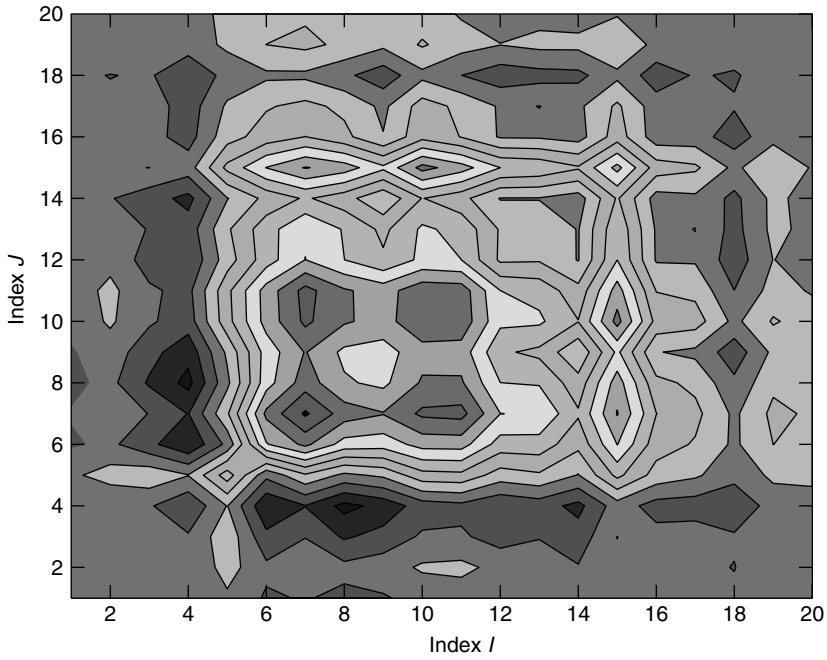


Figure 22. Computed flexibility matrix for the damaged 20-DOF system: noisy modes.

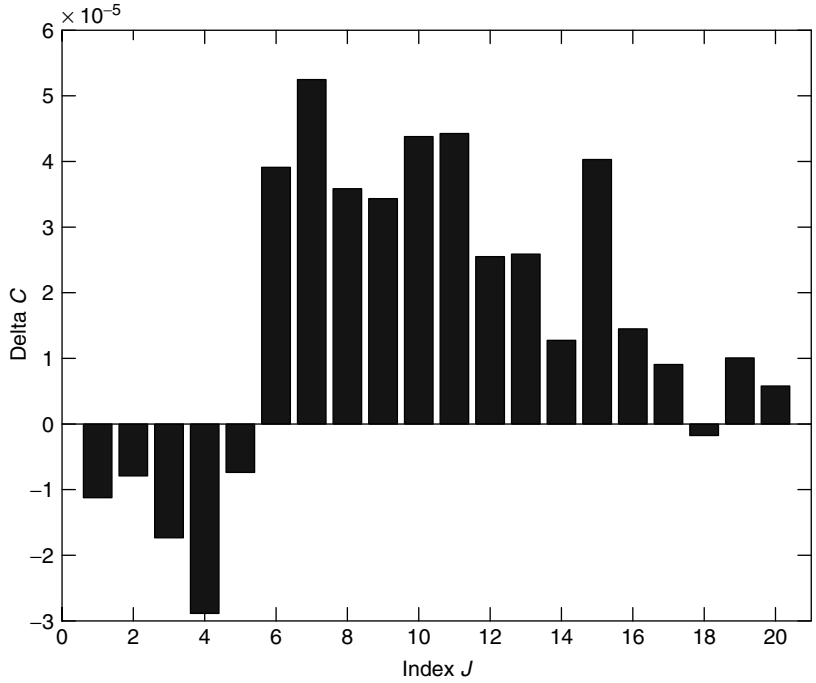


Figure 23. Column 6 of the computed flexibility matrix for the damaged 20-DOF system: noisy modes.

5 PARAMETRIC MODEL UPDATING

The alternative to the algebraic approach to model updating described in the previous section is sensitivity-type methods that rely on a parametric model of the structure and the minimization of some penalty function based on the error between the measured data and the predictions from the model. These methods offer a wide range of parameters to be updated that have physical meaning and allow a degree of control over the optimization process.

5.1 Objective functions

Friswell and Mottershead [26] have discussed sensitivity-based methods in detail. The approach minimizes the difference between modal quantities (usually natural frequencies and less often modeshapes) of the measured data and model predictions. This problem may be expressed as the minimization of J, where

$$J(\theta) = \left\| \{z_m\} - \{z(\{\theta\})\} \right\|^2 = \{\varepsilon\}^{\mathrm{T}}\{\varepsilon\} \qquad (31)$$

and

$$\{\varepsilon\} = \{z_m\} - \{z(\{\theta\})\} \qquad (32)$$

Here $\{z_m\}$ and $\{z(\{\theta\})\}$ are the measured and computed modal vectors, $\{\theta\}$ is a vector of all unknown parameters, and $\{\varepsilon\}$ is the modal residual vector. The modal vectors may consist of both natural frequencies and modeshapes, although often modeshapes are only used to pair individual modes. If modeshapes are included, then they must be carefully normalized, the sensor locations must be carefully matched to the FE degrees of freedom, and weighting should be applied to equation (31).

FRFs may also be used, although a model of damping is required, and the penalty function is often a very complicated function of the parameters with many local minima, making the optimization very difficult. dos Santos *et al.* [27] present an example of such a method for damage in a composite structure.

5.2 Sensitivity methods

Sensitivity-based methods allow a wide choice of physically meaningful parameters, and these advantages have led to their widespread use in model updating. The approach is very general and relies on minimizing a penalty function, which usually consists of the error between the measured quantities and the corresponding predictions from the model. Parameters are then chosen that are assumed uncertain, and these are usually estimated by approximating the penalty function using a truncated Taylor series and iterating to obtain a converged solution. If there are sufficient measurements and a restricted set of parameters, then the identification may be well conditioned. Often, some form of regularization must be applied. This is considered in detail later. Other optimization methods may be used such as quadratic programming and simulated annealing or genetic algorithms, but these are not considered further in this article. Problems also arise if an incorrect or incomplete set of parameters is chosen or, even worse, if the structure of the model is wrong.

The modal residual in equation (31) is a nonlinear function of the parameters and the minimization is solved using a truncated linear Taylor series and iteration. Thus, the Taylor series is

$$\{z_m\} = \{z_j\} + [S_j]\{\delta\theta_j\} + \text{higher-order terms} \qquad (33)$$

where

$$\{z_j\} = \{z(\{\theta_j\})\}, \quad [S_j] = [S(\{\theta_j\})],$$
$$\{\delta\theta_j\} = \{\theta_m\} - \{\theta_j\} \qquad (34)$$

The matrix $[S_j]$ consists of the first derivatives of the modal quantities with respect to the model parameters; index j denotes the jth iteration and $\{\theta_m\}$ is the parameter vector that gives the measured outputs. Standard methods exist to calculate the modal derivatives required [26, 28]. By neglecting higher-order terms in equation (33), an iterative scheme may be derived using the linear approximation

$$\{\delta z_j\} = [S_j]\{\delta\theta_j\} \qquad (35)$$

where $\{\delta z_j\} = \{z_m\} - \{z_j\}$ and $\{\delta\theta_j\} = \{\theta_{j+1}\} - \{\theta_j\}$. Often, for damage location studies, only the residual

and sensitivity matrix for the initial model are used. Avoiding iteration reduces the computation required, particularly where multiple parameter sets have to be estimated. However, particularly if the damage is severe, there is a risk that the wrong location is identified.

As indicated above, one of the problems with sensitivity methods is the need for a parametric model of the damage. Mottershead *et al.* [29] proposed an approach where the system was constrained so that unknown stiffnesses are replaced with rigid connections. The constraint is not imposed physically, but the behavior is inferred from the unconstrained measurements. The best fit between the measured and predicted data is obtained when the damage is located in the substructure that is made rigid.

5.3 Model parameters

One of the key aspects of a model-based identification method is the parameterization of the candidate damage. Since inverse approaches rely on a model of the damage, the success of the estimation is dependent on the quality of the model used. The type of model used depends on the type of structure and the damage mechanism, which leads to an increase either in local or the distributed flexibility. The damage model may be simple or complex. For example, a cracked beam may be modeled as a reduction in stiffness in a large FE or substructure, or alternatively using a very detailed model from fracture mechanics. Whether such a detailed model is justified often depends on the requirements of the estimation procedure and the quality of the measured data. Using a measured modal model consisting of the lower natural frequencies and associated modeshapes means that only a coarse model of the damage may be identified. The simple example used for illustration uses element stiffnesses as the parameters and is the simplest form of an equivalent model for the damage.

5.3.1 Crack models

The modeling of cracks in beam structures and rotating shafts has been a significant topic for research. The models fall into three main categories: local stiffness reduction, discrete spring models, and complex models in two or three dimensions. Dimarogonas [30] and Ostachowicz and Krawczuk [31] gave comprehensive surveys of crack modeling approaches. The simplest methods for FE models reduce the stiffness locally, for example, by reducing a complete element stiffness to simulate a small crack in that element [32]. This approach suffers from problems in matching damage severity to crack depth, and is affected by the mesh density. An improved method introduces local flexibility based on physically based stiffness reductions, where the crack position may be used as a parameter for identification purposes. The second class of methods divides a beam-type structure into two parts that are pinned at the crack location and the crack is simulated by the addition of a rotational spring. These approaches are a gross simplification of the crack dynamics and do not involve the crack size and location directly. The alternative, using beam theory, is to model the dynamics close to the crack more accurately, for example, by producing a closed-form solution giving the natural frequencies and modeshapes of cracked beam directly or using differential equations with compatible boundary conditions satisfying the crack conditions [33–35]. Friswell and Penny [36] compared several of the simple crack models that may be used for health monitoring for both the linear and nonlinear response. Alternatively, two- or three-dimensional FE meshes for beam-type structures with a crack may be used. Meshless approaches may also be used, but are more suited to crack propagation studies. No element connectivity is required and so the task of remeshing as the crack grows is avoided, and a growing crack is modeled by extending the free surfaces corresponding to the crack [37]. However, the computational cost of these meshless methods generally exceeds that of conventional finite element analysis (FEA). Rao and Rahman [38] avoided this difficulty by coupling a meshless region near the crack with an FEA model in the remainder of the structure. The two- and three-dimensional approaches produce detailed and accurate models but are a complicated and computational intensive approaches to model simple structures like beams, and are unlikely to lead to practical algorithms for damage identification.

Although the geometry of a crack can be very complicated, for low frequency vibration only an effective reduction in stiffness is required. Thus, for comparison, a simple model of an open crack, which

is essentially a saw cut, is used. This allows the comparison of models using beam elements with those using plate elements. Only a selection of beam models are used, which illustrates the fact that many beam models are able to model the effect of the crack at low frequencies. In the first approach, the stiffness of a single element is reduced, which requires a fine mesh, and also the derivation of the effect of a crack on the element stiffness. In the second approach, the beam is separated into two halves at the crack location. The beam sections are then pinned together and a rotational spring is used to model the increased flexibility due to the crack. Translational springs may also be used in place of the pinned constraint. The major difficulties with this approach are that an FE node must be placed at the crack location, requiring remeshing for health monitoring applications, and the relationship between the spring stiffness and crack depth needs to be derived. The third method is the model of Sinha *et al.* [34], which is based on the work of Christides and Barr [33]. The fourth model is based on fracture mechanics [35]. In the final model, the open crack is modeled using plate elements, and elements are removed where the crack is located.

Table 4. Natural frequencies (Hz) for the undamaged beam

Number DOF		Beam 40	Plate 13 233
Modes	1	20.709	20.707
	2	129.78	129.39
	3	363.40	360.62
	4	712.16	701.96
	5	1177.4	1150.6

The approaches to crack modeling are compared using a simple example of a steel cantilever beam 1 m long, with cross section $25 \times 50\,\text{mm}^2$. Bending in the more flexible plane is considered. The crack is assumed to be located at a distance of 200 mm from the fixed end, and has a constant depth of 10 mm across the beam width. The beam is modeled using 20 Euler–Bernoulli beam elements, and gives the natural frequencies shown in Table 4. For the plate elements, the length is split into 401 elements and the depth into 10 elements. Thus, the elements are approximately $2.5\,\text{mm}^2$. A large number of elements is required because an element with linear shape functions is used. Table 4 shows the estimated natural frequencies using the Quad4 element in the structural dynamics toolbox [39]. The damaged beam was modeled using the approaches discussed earlier, and the results are shown in Table 5. All the beam models contain 20 elements, and the nodes are arranged such that the crack occurs in the middle of an element. Of course, in the case of the discrete rotational spring, a node is placed at the crack location. The reduction in the element stiffness is adjusted so that the percentage change in the first natural frequency is the same as that for the plate model. The other beam models are adjusted in a similar way. In the plate model, the crack is simulated by removing four elements and thus represents a saw cut 10 mm deep. The row of elements below the crack is also made thinner, so that the crack has negligible width. The differences in the lower natural frequencies are very similar for all models, and these differences are smaller than the changes that would occur due to small modeling errors, or changes due to environmental effects. Of course, the accuracy at higher frequencies becomes less since the modes are influenced more by local stiffness variations.

Table 5. The percentage changes in the natural frequencies for the damaged beam

		Beam				Plate
		Element stiffness reduction	Discrete spring	Sinha *et al.* [34]	Lee and Chung [35]	
Modes	1	4.18	4.18	4.18	4.18	4.18
	2	0.07	0.04	0.08	0.04	0.04
	3	1.24	1.23	1.24	1.20	1.22
	4	2.99	3.08	2.98	2.99	3.07
	5	2.37	2.45	2.37	2.34	2.69

5.3.2 Composite structures

Composite structures have an excellent performance, although this deteriorates significantly with damage. Unfortunately, damage, due to impact events for example, is difficult to detect visually, and hence some method of nondestructive testing of these structures is required. Zou *et al.* [40] reviewed the vibration-based methods that are available to monitor composite structures. Since this article considers inverse methods for damage estimation, this section only considers the parameterization of the damage in composite structures, and, in particular, the modeling of delaminations. Although composite structures have other modes of failure, such as matrix cracking, fiber breakage, or fiber-matrix debonding [31], these damage mechanisms produce similar changes in the vibration response to that obtained for damage in metallic structures. However, delamination is a serious problem in composite structures, and has no parallel to damage mechanisms in other materials. Once the damage is parameterized, then inverse methods, such as sensitivity analysis, may be applied.

Zou *et al.* [40] reviewed methods to model delaminations. Here the focus is on simple models. For example, if a structure is modeled with beam or plate elements, then only beam or plate elements should be used to model the structure with delaminations. Delamination occurs when adjacent plies in a laminated composite debond. For beam structures, the simplest case of a through-width delamination, parallel to the beam surface, was modeled using four beam segments [41, 42]. Separate beam elements were used above and below the delamination, and the constraints to join these elements to those of the undamaged parts of the beam needed to be applied carefully. Zou *et al.* [40] detailed further development of these models. One difficulty with using these models for parameter-based identification is that changing the length and position of a delamination requires the model to be remeshed, and care must be exercised in calculating the associated sensitivity matrices. The techniques detailed by Sinha *et al.* [34] for the position of cracks might be extended to this case. Paolozzi and Peroni [43] highlighted that the most sensitive modes are those whose wavelength is approximately the same size as the delamination. Luo and Hanagud [44] used a sensitivity-based method to detect delaminations, and they also discovered that some modes split to give two closely spaced natural frequencies.

5.3.3 Joint models and generic elements

One major difficulty in parametric approaches is that a model is required that accurately reflects the effect of damage on the mass and stiffness matrices. To some extent, using low-frequency vibration measurements is of help because any local stiffness reduction will have a very similar effect on the dynamic response. Thus, it is possible to use equivalent parameters, such as element stiffnesses, to model the damage. Generic elements [45, 46] take this approach further by allowing changes to the eigenvalues and eigenvectors of the stiffness matrices of structural elements or substructures. These changes are usually constrained so that properties such as the rigid body modes and the geometric symmetry are retained.

Generic elements introduce flexibility into the joint in a controlled way. Other equivalent models, such as discrete rotational springs, offset parameters, or changing element properties may also be used, although generic parameters do have advantages [46]. In particular, all models prejudge how the damage will affect the full model of the structure, whereas the generic element approach automatically finds the likely low-frequency motion of the joint. Consider a two-dimensional T-joint constructed from three beam elements. Each node has three degrees of freedom and, since the substructure has four nodes, the substructure stiffness matrix has three rigid body eigenvectors and nine flexible eigenvectors [47]. The lower eigenvectors have much simpler deformation shapes that are more likely to represent the motion the substructure would undergo in many of the global modes of the structure. Thus, reducing the eigenvalues corresponding to these eigenvectors makes the joint substructure more flexible in the frequency range of the global dynamics, and may be used to model damage. Higher frequency eigenvectors of the substructure may also be included if the motion of the joint is more complex; however, the lower eigenvectors of the joint are likely to adequately characterize the low-frequency dynamics of the structure.

Generic elements have been developed for use in model updating and may be considered as equivalent models of elements or substructures [45]. Law *et al.* [48] applied generic elements to the FE model

updating of the Tsing Ma bridge in Hong Kong. Wang *et al.* [49] used generic elements in damage detection, dealing with the simulated problem of damage detection in a frame structure with flexible L-shaped and T-shaped structural joints. Titurus *et al.* [47, 50] used generic elements and subset selection for model updating and damage location of an experimental frame structure.

5.3.4 Distributed damage

Teughels *et al.* [51] presented a sensitivity-based FE updating method for damage assessment that minimized differences between the experimental and predicted modal data. The parameterization of the damage (both localization and quantification) was represented by a reduction factor of the element bending stiffness. The number of unknown variables was reduced to obtain a physically meaningful result by using a set of damage functions to determine the spatial bending stiffness distribution. The updating parameters were then the multiplication factors of the damage functions. The procedure was illustrated on a reinforced concrete beam and on a highway bridge [52].

5.4 Optimization procedures and ill-conditioning

When the parameters of a model are unknown, they must be estimated using measured data. Usually, the measured response is a nonlinear function of the parameters. In these cases, minimizing the error between the measured and the predicted response produces a nonlinear optimization problem, with the usual questions about convergence and local minima. The most common approach is to linearize the residuals, obtain a least squares solution, and iterate. If the identification problem is well posed, then this simple approach is adequate. The usual response to problems encountered in the optimization is to try more advanced algorithms, but often the issue is that the estimation problem has not been posed correctly, and including some physical insight into the problem provides a much better solution.

Probably the most important difficulty in parameter estimation is ill-conditioning. In the worst case, this can mean that there is no unique solution to the estimation problem, and many sets of parameters are able to fit the data. Many optimization procedures result in the solution of linear equations for the unknown parameters. The use of the singular value decomposition (SVD) [53] for these linear equations enables ill-conditioning to be identified and quantified. The options are then to increase the available data, which is often difficult and costly, or to provide extra conditions on the parameters. These can take the form of smoothness conditions (for example, the truncated SVD), minimum norm parameter values (Tikhonov regularization), or minimum changes from the initial estimates of the parameters [54, 55].

However, there are significant differences in the application of these methods in model updating and damage location, which necessitates different methods of regularization. In both cases, the number of potential parameters is very large and the estimation process is likely to be ill-conditioned unless the physical understanding can be used to introduce extra information. In model updating, the number of parameters may be reduced by only including those parameters that are likely to be in error. Thus, if a frame structure is updated, the beams are likely to be modeled accurately but the joints are more difficult to model. It would therefore be sensible to concentrate the uncertain parameters to those associated with the joints. Even so, a large number of potential parameters may be generated, the measurements may still be reproduced, and the parameters are unlikely to be identified uniquely. In this situation, all the parameters are changed, and regularization must be applied to generate a unique solution [46]. Regularization generally applies extra constraints to the parameter estimation problem to ensure a unique solution. Applying the standard Moore–Penrose pseudoinverse is a type of regularization where the parameter vector with the minimum norm is chosen. The parameter changes may be weighted separately to give a weighted least squares problem, where the penalty function is a weighted sum of squares of the measurement errors and the parameter changes. Such weighting may also be extended to include minimizing the difference between equivalent parameters that are nominally equal in different substructures such as joints. Although using parametric models can reduce the number of parameters considerably, for damage location there will still be a large number of parameters. Most regularization techniques rely on minimum-norm-type solutions that tend to spread the

identified damage over a large number of parameters. Using subset selection, where only the optimum subset of the parameters are used for the estimation [56], has been used for model updating and also for damage location.

Black box methods are often not considered as model-based approaches. However, any simulation of an input–output relationship must make some assumptions about the underlying process, and hence essentially has an underlying model. For example, a neural network is essentially a very sophisticated curve-fitting algorithm, and ill-conditioning is a major problem, evidenced by overfitting and a lack of generalization. The advantage of neural networks is that the class of input–output relationships that may be fitted is huge. However, better results are always obtained if physical insight is used to guide the modeling and estimation process. Indeed, there is often a need to reduce the number of input nodes to present to a neural network, and understanding is vital to obtain the correct feature extraction and data reduction. Another use of physical models is the generation of training or test data for these identification schemes. Typically, experimental data for a sufficient range of events is difficult or expensive to obtain. Since running a model many times is relatively easy and cheap, these simulations may be used to increase the quantity of the test data. However, it is vital that this simulated data correctly reproduces the important features of the real structure, and hence requires a validated and, if necessary, updated model.

Neural networks and genetic algorithms have been viewed as potential saviors for the solution of the difficult problems in damage location. Although these methods may be useful in some circumstances, they do not deal with the root cause of the problem. Genetic algorithms have some advantage in finding a global minimum in very difficult optimization problems, particularly where there are many local minima as is often the case in damage location. That said, the method still requires that the dynamics of the structure changes sufficiently and predictably enough for the optimization to be meaningful. The crucial decision and difficulty is often regarding what to optimize, not the optimization method used. Neural networks are able to treat damage mechanisms implicitly, so that it is not necessary to model the structure in so much detail. The method can also deal with nonlinear damage mechanisms easily. Models are still required to provide the training cases for the networks, and this is a major problem. There are always systematic errors between the model used for training and the actual structure. For success, neural networks require that the essential features in the damaged structure are represented in the training data.

5.4.1 Regularization with side constraints

Model updating often leads to an ill-conditioned parameter estimation problem, and an effective form of regularization is to place constraints on the parameters. This could be that the deviation between the parameters of the updated and the initial model are minimized, or differences between parameters could be minimized. For example, in a frame structure, a number of "T" joints may exist that are nominally identical. Owing to manufacturing tolerances, the parameters of these joints are slightly different, although these differences are small. Therefore, a side constraint is placed on the parameters, so that both the residual and the differences between nominally identical parameters are minimized. Thus, if equation (31) generates the residual, the parameter that minimizes the quadratic cost function is sought,

$$J(\{\theta\}) = \left\|\{z_m\} - \{z(\{\theta\})\}\right\|^2 + \lambda^2 \left\|[C]\{\theta\} - \{d\}\right\|^2 \tag{36}$$

for some matrix $[C]$, vector $\{d\}$, and regularization parameter λ. The regularization parameter is chosen to give a suitable balance between the residual and the side constraint. An example of the choice of $[C]$ is when there are only two parameters, which are nominally equal, then

$$[C] = [1 \ \ -1] \tag{37}$$

Equation (36) is usually linearized using the sensitivity matrix to give the regularized penalty function corresponding to equation (35). The resulting set of linear equations is often well conditioned. Indeed, the constraints should be chosen to satisfy Morozov's complementation condition,

$$\text{Rank}\begin{bmatrix}[S_j] \\ [C]\end{bmatrix} = p \tag{38}$$

where p is the number of unknown parameters, which ensures that the coefficient matrix is full rank.

5.4.2 Regularization using subset selection

One solution to the problem of ill-conditioning is to select only a subset of the parameters for updating [56]. The parameters that are chosen are those to which the response data is sensitive, but the parameters must also be able to correct the errors in the model. Parameter subset selection is a technique that selects the best subset of parameters from a candidate set, utilizing some application-dependent cost function that provides a measure of goodness of each subset. Often, these techniques only obtain a suboptimal estimate of the best subsets in some sense due to the excessive computational burden posed by the original problem. These techniques are firmly rooted in statistics and related fields [57], although recently applications in structural mechanics have appeared. Friswell *et al.* [56] gave an overview of subset selection and also proposed the use of this technique for damage detection. They suggested an approach based on forward parameter subset selection, which is especially suited to local damage, and applied the method to a simulated cantilever beam example with physical parameters corresponding to either element or node properties. Different selection and iteration strategies were evaluated, and the case where multiple measurement sets are available was handled by computing the principal angles between two vector subspaces. Fritzen *al.* [58] used a orthogonalization scheme for subset selection.

In damage location, statistical methods and performance measures have been used that work on a similar principle [5, 59, 60]. Only a limited number of sites are assumed to be damaged, and the model is updated based on the reduced number of parameters. This process is repeated for all possible combinations of damage sites, and possibly even damage mechanisms. The results from all the updated models are compared and the one that best matches the measured data is chosen.

The major problem with both subset selection and the statistical type approach is that many smaller model-updating exercises have to be performed. To optimally derive the best set of parameters, or the best damage location, requires the evaluation of many subsets of parameters. With a large number of parameters evaluating all subsets of even two or three parameters can become daunting. Thus, suboptimal methods must be used to derive good, but not necessarily the best, subsets of the parameters. In the forward approach, parameters are chosen one at a time, and the parameters selected previously are retained. However, there is no guarantee that the optimal subset will be found. The number of candidate damage locations may be controlled on the basis of the expected reduction in the residual [57]. The addition of a parameter to a previously selected subset inevitably reduces the residual terms, and thus there is a trade-off between the number of parameters selected and the magnitude of the residual. Often, only a single damage location is required, in which case the optimal parameter may be determined. Often, a reasonable number of parameter subsets (say between 3 and 20) are selected for a more detailed study [57]. Friswell *et al.* [56] reviewed the relationship between subset selection and matrix decomposition, and also expanded the methods to parameter groups using subspace angles.

The process of subset selection is now described. It should be highlighted that the standard approach to subset selection is not iterative, but only uses the sensitivity equation (35) evaluated at the initial parameter values. For ease of notation this equation is written as

$$[A]\{x\} = \{b\} \tag{39}$$

It would be possible to update each candidate parameter set until convergence, and then compare the performance of the different subsets, although, in practice, the computational cost is prohibitive. As the model parameters are usually local in nature and may also allow for different damage mechanisms, parameter subset selection selects parameters from $\{x\}$ that identify both the damage location and mechanism. This formulation requires the selection of the optimum parameter subset from $\{x\}$. The most straightforward approach is to use an exhaustive search where all $(2^p - 1)$ possible cases have to be searched. The number of cases renders this approach computationally intensive and thus impractical in many real situations. Consequently, suboptimal schemes have to be used. An additional problem is that the addition of a parameter to a previously

selected subset inevitably reduces the residual generated by equation (39). Thus, there is a trade-off between the number of parameters selected and the magnitude of the residual.

Equation (39) may be written as

$$[A]\{x\} = [\{a_1\}, \{a_2\}, \ldots, \{a_p\}]\{x\} = \{b\} \quad (40)$$

The case of a single damage location leads to a simplified version of the above philosophy. When only one parameter is selected, the optimum parameter is that which best fits the changes due to damage characterized by the vector $\{b\}$ in equation (40). Thus, the goal is to find the column $\{a_j\}$ of matrix $[A]$ that minimizes

$$J = \left\| \{b\} - \{a_j\}\hat{x}_j \right\|^2 \quad (41)$$

where $\hat{x}_j$ is the least squares estimate of the jth parameter in $\{x\}$. Friswell *et al.* [56] showed that minimizing equation (41) is equivalent to finding the column of $[A]$ that minimizes the angle with $\{b\}$. Hence, the best parameter is the jth and is found by

$$\min\left(\left\{\psi_1, \psi_2, \ldots, \psi_p\right\}\right) \Longrightarrow \hat{x}_j, \{a_j\} \quad (42)$$

where

$$\cos^2 \psi_i = \frac{\left(\{a_i\}^{\mathrm{T}}\{b\}\right)^2}{\left(\{a_i\}^{\mathrm{T}}\{a_i\}\right)\left(\{b\}^{\top}\{b\}\right)}, \quad i = 1, 2, \ldots, p \quad (43)$$

and ψ_i is the angle between vectors $\{a_i\}$ and $\{b\}$. This step is part of a general technique used in damage detection [56] and is called *forward parameter subset selection*. This is a suboptimal technique of subset selection, starting with the above step and continuing by additional parameter searches where the already selected parameters are retained. For subsequent steps, a new modified problem is created, respecting the previous parameter selections. Suppose the single parameter with index j_1 has been chosen, then the parameter estimate $\hat{x}_{j_1}$ and residual $\{\varepsilon\}$ are

$$\hat{x}_{j_1} = \frac{\{a_{j_1}\}^{\mathrm{T}}\{b\}}{\{a_{j_1}\}^{\mathrm{T}}\{a_{j_1}\}} \Longrightarrow \{\varepsilon\} = \{b\} - \hat{x}_{j_1}\{a_{j_1}\} \quad (44)$$

Note that $\{\varepsilon\}$ is orthogonal to $\{a_{j_1}\}$. A new parameter is then sought by considering the subspace defined by columns of $[A]$, but orthogonal to $\{a_{j_1}\}$. The modified problem is defined as [56]

$$\{a_j\} \rightarrow \{a_j\} - \alpha_j\{a_{j_1}\}, \qquad \{b\} \rightarrow \{b\} - \hat{x}_{j_1}\{a_{j_1}\} \quad (45)$$

where

$$\alpha_j = \frac{\left(\{a_{j_1}\}^{\mathrm{T}}\{a_j\}\right)}{\left(\{a_{j_1}\}^{\mathrm{T}}\{a_{j_1}\}\right)} \quad (46)$$

A second parameter may now be selected by means of the modified problem defined by equation (45), where $j \neq j_1$. Further parameters may be selected in the same way. An algorithm is thus created to search for the best parameter subset, denoted by $[\hat{x}_{j_1}, \hat{x}_{j_2}, \ldots, \hat{x}_{j_r}]$, that minimizes the cost function,

$$J_r = \left\| \{b\} - \sum_{i=1}^{r} \hat{x}_{j_i}\{a_{j_i}\} \right\|^2 \quad (47)$$

This cost function is also employed in Efroymson's algorithm for forward subset selection that focuses on adding or removing parameter selections from chosen subsets. Thus, the number of candidate damage locations may be controlled on the basis of the expected reduction in the residual [56, 57]. Since only single damage location cases are examined in detail here, this subject is not considered further.

5.5 A simple cantilever beam example

A simulated cantilever beam example is used to demonstrate some of the problems. Although the example is somewhat artificial, it highlights how easily methods fail even on very simple structures. Any practical method would have to be robust and should therefore succeed on simple structures, even though some systematic errors are included. This example also demonstrates the use of simulation in damage identification. Not all of the methods are tried and this example is not supposed to represent an extensive scientific evaluation of the methods. It is used here for the purpose of illustration.

The beam has a cross section of $25 \times 50\,\mathrm{mm}^2$, a length of 1m, and is assumed to be rigidly clamped at one end. Only motion in the plane of the thinner

beam dimension is considered. The beam has a Young's modulus of $210\,\mathrm{GN\,m^{-2}}$ and a mass density of $7800\,\mathrm{kg\,m^{-3}}$.

The first test of any method is its application to a simulated example with no noise or systematic errors. Any parameter changes in the model should be identified exactly. The simulated *measurements* are assumed to be the relative changes in the lower natural frequencies of the beam and are taken from a model with 20 elements. The undamaged natural frequencies are taken from the uniform beam, while the damaged frequencies are derived from a model where the stiffness of element 4 has been reduced by 30%. Table 6 gives the damaged and undamaged natural frequencies, showing that the 30% damage only results in a 2.4% change in natural frequency at most. These small frequency changes are typical in damage location and are one of the major difficulties in the identification of the location of damage. Measurement noise, environmental factors, and structure nonstationarity can easily lead to incorrect conclusions on damage location.

The standard sensitivity approach based on modal data is now used to identify the damage. The set of candidate parameters is chosen to be relatively large and consists of the stiffness of each of the 20 elements. If the relative changes to the first eight natural frequencies are used as the measurements, then the identification of the parameters is underdetermined. In this case, some form of regularization must be employed. Figure 24 shows the change in element stiffness required to reproduce the damaged natural frequencies, using a minimum norm constraint on the parameter changes. Although the largest stiffness change occurs at element 4, the identified damage is spread over the whole beam, and there are some significant increases in stiffness. Note that this is the ideal case with no measurement noise or modeling errors. Suppose, by some means, the damage is known to be somewhere in the eight elements closest to the fixed end. The number of parameters is now reduced to eight, the same as the number of natural frequencies. There is now a unique solution to the estimation problem, and this solution is given in Figure 25. Note that the stiffness of elements 9–20 cannot change, but are included in Figure 25 for easy comparison. The damage has clearly been correctly located to element 4. However, the magnitude of the damage is incorrect because the estimation is based on the sensitivity matrix, which is a linear approximation to the residual. The other seven parameters are nonzero for the same reason.

This simple cantilever beam is now used to demonstrate some of the properties of the subset selection

Table 6. Natural frequencies of the simulated undamaged and damaged beam

Mode number	Undamaged (Hz)	Damaged (Hz)	Difference (%)
1	20.96	20.45	2.39
2	131.3	131.1	0.15
3	367.7	366.6	0.31
4	720.6	711.3	1.29
5	1191	1172	1.61
6	1780	1762	1.02
7	2487	2479	0.32
8	3313	3303	0.30

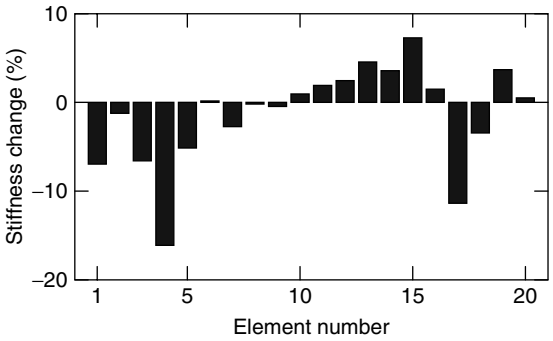


Figure 24. The change in element stiffness estimated for the cantilever beam example with no noise and a minimum norm constraint.

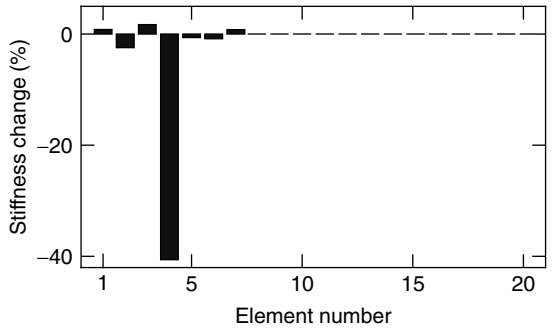


Figure 25. The change in element stiffness estimated for the cantilever beam example with no noise, but only eight nonzero parameters.

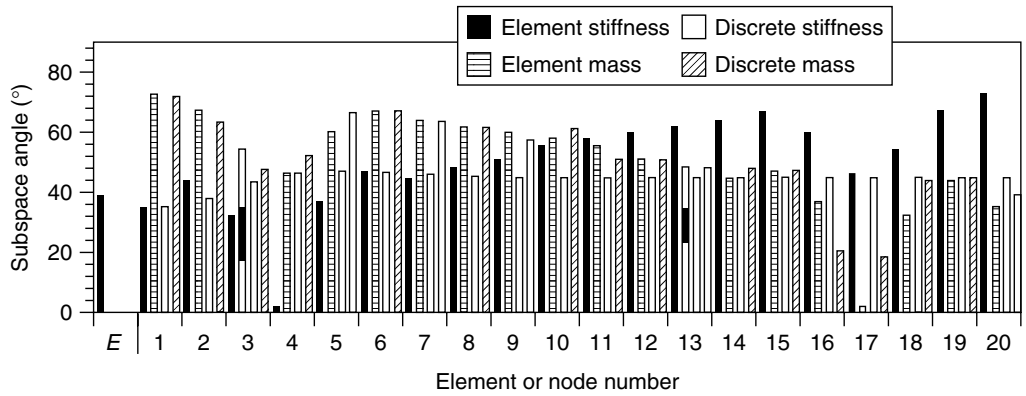


Figure 26. Subspace angles for a 30% change in the stiffness of element 4.

methods. Candidate parameters include element mass, and discrete mass and springs, as well as the element stiffness. Figure 26 shows the angles between the columns of the sensitivity matrix of the initial FE model and the vector of the relative changes in the first eight natural frequencies due to the damage. Clearly, the column relating to the stiffness of element 4 has a small angle, although it is not zero because the method is based on a first-order approximation and the extent of the damage (30%) is large. Changing the mass of element 17 also enables the modeling of the *measured* changes accurately. This is a problem that relates to the symmetry of the beam, and the fact than no spatial information is incorporated into the measurements. Modeshapes could also be incorporated into the measurement vector, although the accuracy with which they could be measured may be insufficient to show a change in modeshape due to damage. This is an example of the more general problem, where damage or changes in parameters at more than one location cause the same changes in the lower natural frequencies.

Subset selection is next demonstrated in an example where damage is introduced at two locations. A 0.1 kg mass is added to node 12, in addition to the 30% stiffness change to element 4. This example does not include any measurement noise or modeling errors. Table 7 shows the results when the best subsets of 1, 2, and 3 parameters are chosen. The parameters are specified by type and element or node number. Thus, $(\rho A)_{17}$ is the mass/unit length of element 17, $(\mathrm{EI})_4$ is the stiffness of element 4, k_1 is a discrete spring at node 1, and m_{12} is a discrete mass at node 12. At each stage, the two best parameters are chosen. The residuals under the first two parameters relate to the values when a subset of size 1 or 2 is selected. Also shown are the residuals after convergence based on optimizing the values of the chosen parameters. From the values of the residuals, it is clear that the two correct parameters should be selected. A number of

Table 7. The selection of three parameters for the beam example

Parameter 1			Parameter 2			Parameter 3		
	Residual	Converged		Residual	Converged		Residual	Converged
$(\rho A)_{17}$	154.6	160.4	m_{12}	1.49	0.782	m_8	1.21	0.701
						$(\rho A)_{12}$	1.48	0.286
			m_8	8.70	4.76	m_{12}	1.21	0.701
						$(\rho A)_{12}$	8.68	4.75
$(\mathrm{EI})_4$	154.7	160.5	m_{12}	1.49	0.000	m_8	1.20	0.000
						$(\rho A)_{12}$	1.48	0.000
			m_8	8.70	5.40	m_{12}	1.20	0.000
						$(\rho A)_{12}$	8.69	5.35

other parameter subsets have small residuals and the addition of random noise would make the selection of the best subset more difficult.

6 POTENTIAL PROBLEMS IN DAMAGE IDENTIFICATION USING VIBRATION DATA

The discussion thus far has indicated some of the problems with damage identification. There are always errors in the measured data and the numerical model that affect all of the algorithms. These errors, and the adequacy of the data, are now discussed. Damage identification algorithms should always be tested on realistic experimental examples, as many methods that work well on simulated data often fail due to the problems highlighted in this section. As a first step, methods may be tested using simulated data, but even then realistic systematic errors should be incorporated.

6.1 Modeling errors

One of the major problems in damage location is the reliance on the FE model. This model is also important because the incomplete set of measured data requires extra information from the model to enable identification of the damage location. There will undoubtedly be errors even in the model of the undamaged structure. Thus, if the measurements on the damaged structure are used to identify damage locations, the methods will have great difficulty in distinguishing between the actual damage sites and the location of errors in the original model. If suitable parameters are not included to allow for the undamaged model errors, then this will the result in a systematic error between the model and the data. Identification schemes generally have considerable difficulty with systematic errors. It is very likely that the original errors in the model will produce frequency changes that are far greater than those produced by the damage. There are two basic approaches to reducing this problem, although both rely on having measured data from an undamaged structure. The first is to update the FE model of the undamaged structure to produce a reliable model [26]. Obviously, the quality of the damage location assessment is critically dependent upon the updated model being physically meaningful [46, 61]. Generally, this requires model validation using a control set of data not used for the updating. The second alternative uses differences between the damaged and undamaged response data in the damage location algorithm [50, 62]. To first order, any error in the undamaged model of the structure that is also present in the damaged structure will be removed. This relies on the structure remaining unchanged, except for the damage, between the two sets of measurements.

Another potential source of error is the mismatch between the measurement locations and the model degrees of freedom. Such a mismatch makes the direct comparison of FRFs and modeshapes impossible, and the generation of residuals inaccurate. The magnitude of the errors involved depends on the mesh density in the sensor region and the complexity of the modeshapes. The best solution is to ensure that nodes in the model exist at the sensor locations. Alternatively, interpolation techniques may be used.

6.2 Environmental and other nonstationary effects

One very difficult aspect of damage assessment is the change in the measured data due to environmental effects. This is one undesirable nonstationary effect and makes damage location very difficult. Of course, progressive damage is also a nonstationary phenomenon, and damage can be difficult to identify if other nonstationary effects are also present. Typical environmental effects are demonstrated by highway bridges, especially those constructed using concrete, which have been the subject of many studies in damage location. For example, temperature changes can cause the stiffness properties of a bridge to change significantly, and the difficulty is to predict the effects of temperature from readily available measurements. Peeters and de Roeck [63] reported on measurements of the Z24 bridge over a whole year and suggested a *black box* model to predict the temperature variation. Sohn *et al.* [64] considered the effect of temperature on the Alamosa Canyon Bridge. Sohn *et al.* [65] used a combination of time series analysis, neural networks, and statistical inference to determine

the damage state for structures affected by environmental conditions. Mickens *et al.* [66] corrected FRF measurements by assuming that the temperature affected the global stiffness of the structure. On a highway bridge, the changing traffic conditions cause different mass loading effects that can change the natural frequencies by as much as 1% [67]. There are further difficulties with highway bridges because they are highly damped with low natural frequencies. They are in a noisy environment and are difficult to excite. The frequency resolution in the measurements is invariably quite low, leading to considerable difficulties in detecting small frequency changes due to damage.

Typical of environmental effects are those on highway bridges. These bridges have been the subject of many studies in damage location, but, in the United Kingdom, where most bridges are constructed using concrete, such identification has considerable problems with changes due to environmental factors [68]. For example, concrete absorbs considerable moisture during damp weather, which considerably increases the mass of the bridge. Temperature changes the stiffness properties of the road surface, known as the *black-top*, significantly. On a hot summer's day in the United Kingdom, the road surface provides little stiffness, but, on a cold winter's day, the stiffness contribution is considerable. The difficulty is trying to predict the effects of temperature and moisture absorption from readily available measurements.

6.3 The effect of frequency range

The range of frequencies employed in damage location has a great influence on the resolution of the results and also the physical range of application. The great advantage in using low-frequency vibration measurements is that the low-frequency modes are generally global and so the vibration sensors may be mounted remotely from the damage site; equally, fewer sensors may be used. The problem with low-frequency modes is that the spatial wavelengths of the modes are large, and are typically far larger than the extent of the damage. The spatial resolution of the damage identification scheme requires that there is a significant change in response between two adjacent potential damage sites. If low-frequency modes are used, then this resolution is closely related to the spatial wavelengths of the modes. Using high-frequency excitation uses very local modes, which are able to accurately locate damage, but only very close to the sensor and actuator position. Estimating accurate models at these high-frequency ranges is also very difficult, and often changes in the response are used for damage identification. For example, Park *et al.* [69, 70] used changes in measured impedance to identify damage in civil structures and pipeline systems. Schulz *et al.* [71] used high-frequency transmissibility to detect delaminations in composite structures.

Moving to even higher frequencies can also yield good results. Acoustic emission [72] is a transient elastic wave, typically in the region of 50–500 kHz, and is able, for example, to detect the energy released when cracks propagate. One approach to damage assessment is to use physical models to deduce quantitative relationships between measured acoustic emission signals and the damage mechanism that cause them. Significant research has been undertaken to obtain a physical understanding of various source mechanisms [73] and the radiation pattern of bulk shear and longitudinal acoustic waves that they produce [74]. The difficulty in using these models in inverse estimation procedures is the accuracy of these high-frequency models, and the huge computational requirements. In recent years, a pattern-recognition philosophy has dominated, that relies on using large databases of empirical data from which correlations between measured acoustic emission signals and damage mechanisms are inferred. Many advanced signal-processing algorithms have been employed to interpret experimental data. Damage location is often determined using time of flight methods, which require the events to be well separated in time, the wave speed to be approximately equal in all parts of the structure, and the effects of reflection and refraction to be insignificant.

6.4 Damage magnitude

A frequent problem that arises in model- and vibration-based damage detection, whether parametric or nonparametric, is the need for a very accurate mathematical model, so that it correctly captures the actual structural dynamic behavior in some predetermined frequency range. Often, in SHM, the changes

in the measured quantities caused by structural damage are smaller than those observed between the healthy (i.e., undamaged) structure and the mathematical model. Consequently, it becomes almost impossible to discern between inadequate modeling and actual changes due to damage. There are two alternative approaches to this problem. The first is to update the healthy model so that the correlation between the model and the measured data is improved. This approach requires that the errors that remain after updating are smaller in magnitude than the changes due to the damage. Furthermore, the changes to the model should be physically meaningful, so that the updating process corrects actual model errors, and does not merely reproduce the measured data. The second approach is based on the use of (relative) differences between data measured on healthy and potentially damaged structure. In this case, assuming that the only changes in the structure are due to damage, the problem may be reduced to finding those parameters that reproduce the measured changes.

6.5 Nonlinearity

Many forms of damage cause a change in the stiffness nonlinearity that qualitatively and quantitatively affects the dynamic response of a structure. For example, Nichols *et al.* [75, 76] used the features of the chaotic response of a structure to detect changes in a joint. Adams and Nataraju [77] gave a variety of features based on the nonlinear dynamic response. Kerschen *et al.* [78] considered model-based estimation methods and identified the form of nonlinearity that is most likely present in the measured data. Meyer and Link [79] identified a parametric nonlinear model using harmonic balance and a model-updating approach. A breathing crack, which opens and closes, can produce interesting and complicated nonlinear dynamics. Brandon [80] and Kisa and Brandon [81] gave an overview of some of the techniques that may be applied. Many techniques to analyze the resulting nonlinear dynamics are based on approximating the bilinear stiffness when the crack opens and closes. Linear approaches to damage estimation approximates a local reduction in the stiffness matrix of the beam. Since the nonlinearity introduced by a crack is often weak, many of the common testing techniques tend to linearize the response [36]. Sinusoidal forcing tends to emphasize the nonlinearity, thus damage detection methods based on detecting harmonics of the forcing frequency have been proposed [82]. In rotor dynamic applications, these approaches are useful because the forcing is inherently sinusoidal [30]. However, in SHM applications, this approach requires considerable hardware and software to implement, and also requires a lengthy experiment. Johnson *et al.* [83] used a transmissibility approach that was insensitive to boundary condition nonlinearities. Neild *et al.* [84] investigated the potential of a time–frequency analysis procedure to identify damage in concrete beams.

Although using the nonlinear response has a huge potential in health monitoring, model-based inverse approaches have a number of difficulties because of the high number of degrees of freedom required, and therefore the computational burden imposed. In practice, any realistic multiple degree of freedom nonlinear analysis would have to be based on a reduced order model of the structure. Furthermore, many of the difficulties outlined in this section for linear systems are also a problem for nonlinear systems.

6.6 Prognosis

Rytter [4] gave prognosis as the fourth level of damage estimation and Farrar *et al.* [85] gave a summary of the state of the art. The philosophy of damage detection using measured vibration data is based on the premise that the damage will change the stiffness of the structure. In some instances, there is a significant difference between strength and stiffness. Indeed, estimating the remaining useful life of a component based on conclusions from a dynamic analysis is very difficult. For example, a concrete highway bridge has steel reinforcement cables running in channels in the concrete. The cables are tensioned, either before or after the concrete has set, to ensure that the concrete remains in compression. One major failure mechanism is by the corrosion of these cables. Once the cables have failed, the concrete has no strength in tension and therefore the bridge is liable to collapse. Unfortunately, the stiffness of the bridge is mainly due to the concrete, and therefore the progressive corrosion of the cables is very difficult to identify from stiffness changes.

Essentially, the dynamics of the bridge changes very little until it collapses. Even for metallic structures, the estimation of remaining life requires an estimate of the damage present, an assessment of the probable future loads, and an accurate model of how the damage may develop and the structure might fail. Although this process is very difficult, the use of inverse methods to generate a physically meaningful model offers a route to prognosis.

6.7 The role of simulation and physical testing

Many of the algorithms suggested for damage location are tested on simulated data. It is necessary to fully test any method on both simulated and real data. The simulated tests are able to fully exercise the location methods, with the benefit that the answer is known. In simulation, far more damage cases may be used and the effect of errors may be fully investigated. The need for real testing arises because experimental work always produces errors and problems that are unexpected. For simulation to be useful, the errors that might be expected in real structures must be simulated. Thus, adding random noise to a model of the structure and then using the same model to identify the damage is not enough. Most identification schemes are able to cope very well with random noise, and although such simulations are important parts of the overall performance assessment of an algorithm, they are not sufficient. It is vital that systematic-type errors are included in the simulation. Thus, discretization errors may be included by generating the simulated measurements using a fine FE model; the damage mechanism introduced to generate the measurements may be different from those modeled for the identification; or boundary conditions on the structure could be changed between the measured data set and the identification.

7 CONCLUSIONS

Given the nature of the last section, this article does not require a detailed conclusions section. The objective has been to illustrate a number of modal vibration-based approaches to the SHM problem and to discuss the merits and demerits of such approaches. The overall judgment must surely be that vibration-based approaches have their place in the SHM canon, but that they have limitations. Arguably, the main limitation is the resolution or minimum level of detectable damage. If the damage is small compared to the spatial wavelength of the modal vibrations, then there is likely to be a problem. Another remark worth making here is that vibration-based damage identification arguably works best when the damage causes a change in the load path. Therefore, it would likely be good for the failure of a member in a truss, but not so good for a crack in a plate. (The authors would like to thank Dr Chuck Farrar of the Los Alamos National Laboratories in the United States for this observation.) In the situation in which one is effectively below the "diffraction limit" for detection by vibration-based methods, one would always have recourse to higher frequency wave-based approaches as used commonly by the NDT (nondestructive testing) community. For example, Lamb wave inspection has proved a powerful technique in platelike areas of structures. Such approaches are discussed in some detail elsewhere in this encyclopedia.

REFERENCES

[1] Ewins D. *Modal Testing Theory, Practice and Application*, *Second Edition*. Research Studies Press, 1999.

[2] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*, Los Alamos National Laboratories Report LA-13070-MS, 1996.

[3] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates SW, Nadler BR, Czarnecki JJ. *A Review of Structural Health Monitoring Literature from 1996–2001*, Los Alamos National Laboratory Report, LA-13976-MS, 2004.

[4] Rytter A. *Vibration Based Inspection of Civil Engineering Structures*, Ph.D. Dissertation. Aalborg University: Aalborg, 1993.

[5] Cawley P, Adams RD. The location of defects in structures from measurements of natural frequencies. *Journal of Strain Analysis* 1979 **14**:49–57.

[6] Yuen MMF. A numerical study of the eigenparameters of a damaged cantilever. *Journal of Sound and Vibration* 1985 **103**:301–310.

[7] *LUSAS User Manual*, Version 10.0. Finite Element Analysis Ltd, 1990.

[8] Sohn H. Effects of environmental and operational variability on structural health monitoring. *Transactions of the Royal Society, Series A* 2007 **365**:538–560.

[9] Pandey AK, Biswas M, Samman MM. Damage detection from changes in curvature modeshape. *Journal of Sound and Vibration* 1991 **145**:321–332.

[10] Ratcliffe CP. Damage detection using a modified Laplacian operator on mode shape data. *Journal of Sound and Vibration* 1997 **204**:505–517.

[11] Leissa AW. The free vibration of rectangular plates. *Journal of Sound and Vibration* 1973 **31**:257–293.

[12] Chance J, Tomlinson GR, Worden K. A simplified approach to the numerical and experimental modelling of the dynamics of a cracked beam. *Proceedings of 12th International Modal Analysis Conference (IMAC)*. Honolulu, HI, 1994; pp. 778–785.

[13] Stubbs N, Kim J-T. Damage localization in structures without baseline modal parameters. *AIAA Journal* 1996 **34**:1644–1649.

[14] Farrar CR, Jauregui DA. Comparative study of damage identification algorithms applied to a bridge: I. experiment. *Smart Materials and Structures* 1998 **7**:704–719.

[15] Farrar CR, Jauregui DA. Comparative study of damage identification algorithms applied to a bridge: 2. Numerical study. *Smart Materials and Structures* 1998 **7**:720–731.

[16] Stubbs N, Kim J-T, Topole K. An efficient and robust algorithm for damage localisation in offshore structures. *Proceedings of 10th ASCE Structures Conference*, 1992; pp. 543–546.

[17] Kim J-T, Stubbs N. Nondestructive crack detection algorithm for full-scale bridges. *Journal of Structural Engineering: ASCE* 2003 **129**:1358–1366.

[18] Cornwell P, Doebling SW, Farrar CR. Application of the strain energy damage detection method to plate-like structures. *Journal of Sound and Vibration* 1999 **224**:359–374.

[19] Stubbs N, Park S, Sikorsky C, Choi S. A level IV global nondestructive damage assessment methodology for civil engineering structures. *International Journal of System Science* 2000 **31**:1361–1373.

[20] Worden K, Manson G, Fieller NR. Damage detection using outlier analysis. *Journal of Sound and Vibration* 1999 **229**:647–667.

[21] Farrar CR, Jauregui D. *Damage Detection Algorithms Applied to Experimental and Numerical Model Data from the I-40 Bridge*, Los Alamos National Laboratories Report, LA-13074-MS, 1996.

[22] Zimmerman DC, Kaouk M. Structural damage detection using a minimum rank update theory. *Journal of Vibration and Acoustics* 1994 **116**: 220–230.

[23] Kaouk M, Zimmerman DC. Structural damage assessment using a generalised minimum rank perturbation theory. *AIAA Journal* 1994 **23**:1431–1436.

[24] Kaouk M, Zimmerman DC. Assessment of damage affecting all structural properties. *Proceedings of the 9th VPI & SU Symposium on Dynamics and Control of Large Structures*, 1994; pp. 445–455.

[25] Pandey AK, Biswas M. Damage detection in structures using changes in flexibility. *Journal of Sound and Vibration* 1994 **169**:3–17.

[26] Friswell MI, Mottershead JE. *Finite Element Model Updating in Structural Dynamics*. Kluwer Academic Publishers, 1995.

[27] dos Santos JVA, Soares CMM, Soares CAM, Maia NMM. Structural damage identification in laminated structures using FRF data. *Composite Structures* 2005 **67**:239–249.

[28] Adhikari S, Friswell MI. Eigenderivative analysis of asymmetric non-conservative systems. *International Journal for Numerical Methods in Engineering* 2001 **51**:709–733.

[29] Mottershead JE, Friswell MI, Mares C. A method for determining model-structure errors and for locating damage in vibrating systems. *Meccanica* 1999 **34**:153–166.

[30] Dimarogonas AD. Vibration of cracked structures: a state of the art review. *Engineering Fracture Mechanics* 1996 **55**:831–857.

[31] Ostachowicz W, Krawczuk M. On modelling of structural stiffness loss due to damage. *DAMAS 2001, 4th International Conference on Damage Assessment of Structures*. Cardiff, 2001; pp. 185–199.

[32] Mayes IW, Davies WGR. Analysis of the response of a multi-rotor-bearing system containing a transverse crack in a rotor. *Journal of Vibration, Acoustics, Stress and Reliability in Design* 1984 **106**:139–145.

[33] Christides S, Barr ADS. One dimensional theory of cracked Bernoulli-Euler beams. *International Journal of Mechanical Sciences* 1984 **26**:639–648.

[34] Sinha JK, Friswell MI, Edwards S. Simplified models for the location of cracks in beam structures

using measured vibration data. *Journal of Sound and Vibration* 2002 **251**:13–38.

[35] Lee Y, Chung M. A study on crack detection using eigenfrequency test data. *Computers and Structures* 2001 **77**:327–342.

[36] Friswell MI, Penny JET. Crack modelling for structural health monitoring. *Structural Health Monitoring: An International Journal* 2002 **1**:139–148.

[37] Belytschko T, Lu YY, Gu L. Crack propagation by element-free Galerkin methods. *Engineering Fracture Mechanics* 1995 **51**:295–315.

[38] Rao BN, Rahman S. A coupled meshless—finite element method for fracture analysis of cracks. *International Journal of Pressure Vessels and Piping* 2001 **78**:647–657.

[39] Balmes E. *Structural Dynamics Toolbox: For Use with MATLAB User's Guide*, Version 4, 2004.

[40] Zou Y, Tong L, Steven GP. Vibration-based model-dependent damage (delamination) identification and health monitoring for composite structures—a review. *Journal of Sound and Vibration* 2002 **230**:357–378.

[41] Majumdar PM, Suryanarayan S. Flexural vibrations of beams with delaminations. *Journal of Sound and Vibration* 1988 **125**:441–461.

[42] Tracy JJ, Pardoen GC. Effect of delamination on the natural frequencies of composite laminates. *Journal of Composite Materials* 1989 **23**:1200–1215.

[43] Paolozzi A, Peroni I. Detection of debonding damage in a composite plate through natural frequency variations. *Journal of Reinforced Plastics and Composites* 1990 **9**:369–389.

[44] Luo H, Hanagud S. Delamination detection using dynamic characteristics of composite plates. *Proceedings of the AIAA/ASME/ASCE/AHS Structures, Structural Dynamics & Materials Conference*, New York, 1995; pp. 129–139.

[45] Gladwell GML, Ahmadian H. Generic element matrices suitable for finite element model updating. *Mechanical Systems and Signal Processing* 1995 **9**:601–614.

[46] Friswell MI, Mottershead JE, Ahmadian H. Finite element model updating using experimental test data: parameterisation and regularisation. *Transactions of the Royal Society of London, Series A* 2001 **359**:169–186.

[47] Titurus B, Friswell MI, Starek L. Damage detection using generic elements: part I, model updating. *Computers and Structures* 2003 **81**:2273–2286.

[48] Law SS, Chan THT, Wu D. Efficient numerical model for the damage detection of a large scale complex structure. *Engineering Structures* 2001 **23**:436–451.

[49] Wang D, Friswell MI, Nikravesh P, Kuo EY. Damage detection in structural joints using generic joint elements. *Proceedings of the 17th International Modal Analysis Conference (IMAC)*, 1999; pp. 792–798.

[50] Titurus B, Friswell MI, Starek L. Damage detection using generic elements: part II, damage detection. *Computers and Structures* 2003 **81**:2287–2299.

[51] Teughels A, Maeck J, Roeck GD. Damage assessment by FE model updating using damage functions. *Computers and Structures* 2002 **80**:1869–1879.

[52] Teughels A, Roeck GD. Structural damage identification of the highway bridge Z24 by FE model updating. *Journal of Sound and Vibration* 2004 **278**:589–610.

[53] Golub G, van Loan CF. *Matrix Computations*. The John Hopkins University Press, 1996.

[54] Hansen PC. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review* 1992 **34**:561–580.

[55] Hansen PC. Regularisation tools: a MATLAB package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms* 1994 **6**:1–35.

[56] Friswell MI, Penny JET, Garvey SD. Parameter subset selection in damage location. *Inverse Problems in Engineering* 1997 **5**:189–215.

[57] Millar AJ. *Subset Selection in Regression*, *Monographs on Statistics and Applied Probability, Vol. 40*. Chapman & Hall, 1990.

[58] Fritzen CP, Jennewein D, Kiefer T. Damage detection based on model updating methods. *Mechanical Systems and Signal Processing* 1998 **12**:163–186.

[59] Cawley P, Adams RD, Pye CJ, Stone BJ. A vibration technique for non-destructively assessing the integrity of structures. *Journal of Mechanical Engineering Science* 1978 **20**:93–100.

[60] Friswell MI, Penny JET, Wilson DA. Using vibration data and statistical measures to locate damage in structures. *Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis* 1994 **9**:239–254.

[61] Link M, Friswell MI. Working group 1. Generation of validated structural dynamic models—results of a benchmark study utilising the GARTEUR SM-AG19 testbed. *Mechanical Systems and Signal Processing, COST Action Special Issue* 2003 **17**:9–20.

[62] Parloo E, Guillaume P, van Overmeire M. Damage assessment using mode shape sensitivities. *Mechanical Systems and Signal Processing* 2003 **17**: 499–518.

[63] Peeters B, de Roeck G. One-year monitoring of the Z24-bridge: environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics* 2001 **30**:149–171.

[64] Sohn H, Dzwonczyk M, Straser EG, Kiremidjian A, Law KH, Meng T. An experimental study of temperature effects on modal parameters of the Alamosa Canyon Bridge. *Earthquake Engineering and Structural Dynamics* 1999 **28**:879–897.

[65] Sohn H, Worden K, Farrar CR. Statistical damage classification under changing environmental and operational conditions. *Journal of Intelligent Material Systems and Structures* 2002 **13**:561–574.

[66] Mickens T, Schulz M, Sundaresan M, Ghoshal A, Naser AS, Reichmeider R. Structural health monitoring of an aircraft joint. *Mechanical Systems and Signal Processing* 2003 **17**:285–303.

[67] Zhang QW, Fan LC, Yuan WC. Traffic induced variability in dynamic properties of cable-stayed bridge. *Earthquake Engineering and Structural Dynamics* 2002 **31**:2015–2021.

[68] Wood MG. *Damage Analysis of Bridge Structures Using Vibrational Techniques*, Ph.D. Thesis. Aston University, 1992.

[69] Park G, Cudney H, Inman DJ. Impedance-based health monitoring of civil structural components. *ASCE Journal of Infrastructure Systems* 2000 **6**:153–160.

[70] Park G, Cudney H, Inman DJ. Feasibility of using impedance-based damage assessment for pipeline systems. *Journal of Earthquake Engineering and Structural Dynamics* 2001 **30**:1463–1474.

[71] Schulz MJ, Pai MJ, Inman DJ. Health monitoring and active control of composite structures using piezoceramic patches. *Composites Part B: Engineering* 1999 **30**:713–725.

[72] Rogers LM. *Structural and Engineering Monitoring by Acoustic Emission Methods—Fundamentals and Applications*, Technical Report. Lloyd's Register of Shipping, London, 2001.

[73] Scruby CB, Buttle DJ. Quantitative fatigue crack measurement by acoustic emission. In *Fatigue Crack Measurement: Techniques and Applications*, Marsh KJ, Smith RA, Ritchie RO (eds). Engineering Materials Advisory Service Ltd., 1991; 207–287.

[74] Ono K. Acoustic emission. In *Fatigue Crack Measurement: Techniques and Applications*, Marsh KJ, Smith RA, Ritchie RO (eds). Engineering Materials Advisory Service Ltd., 1991; pp. 207–287.

[75] Nichols JM, Todd MD, Wait JR. Using state-space predictive modeling with chaotic interrogation in detecting joint preload loss in a frame structure. *Smart Materials and Structures* 2003 **12**: 580–601.

[76] Nichols JM, Virgin LN, Todd MD, Nichols JD. On the use of attrator dimension as a feature in structural health monitoring. *Mechanical Systems and Signal Processing* 2003 **17**:1305–1320.

[77] Adams DE, Nataraju M. A nonlinear dynamical systems framework for structural diagnosis and prognosis. *International Journal of Engineering Science* 2002 **40**:1919–1941.

[78] Kerschen G, Golinval J-C, Hemez FM. Bayesian model screening for the identification of nonlinear mechanical structures. *Journal of Vibration and Acoustics* 2003 **125**:389–397.

[79] Meyer S, Link M. Modelling and updating of local non-linearities using frequency response residuals. *Mechanical Systems and Signal Processing* 2003 **17**:219–226.

[80] Brandon JA. Some insights into the dynamics of defective structures. *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science* 1998 **212**:441–454.

[81] Kisa M, Brandon JA. The effects of closure of cracks on the dynamics of a cracked cantilever beam. *Journal of Sound and Vibration* 2000 **238**:1–18.

[82] Shen MH, Guran A, Tzou H. On-line structural damage detection. In *Structronic Systems: Smart Structures, Devices, and Systems, Vol. 1: Smart Materials and Structures*, Anderson GL, Natori M (eds). World Scientific, 1998; pp. 271–332.

[83] Johnson TJ, Brown RL, Adams DE, Schiefer M. Distributed structural health monitoring with a smart sensor array. *Mechanical Systems and Signal Processing* 2004 **18**:555–572.

[84] Neild SA, Williams MS, McFadden PD. Nonlinear vibration characteristics of damaged concrete beams. *Journal of Structural Engineering: ASCE* 2003 **129**:260–268.

[85] Farrar CR, *et al. Damage Prognosis: Current Status and Future Needs*, Los Alamos National Laboratory Report, LA-14051-MS, 2003.

Chapter 15
Nonlinear Acoustic Methods

Dimitri M. Donskoy
Stevens Institute of Technology, Hoboken, NJ, USA

1 INTRODUCTION

Traditional active acoustic/ultrasonic methods utilize the linear effects of reflection, scattering, transmission, and attenuation of the elastic waves by structural inhomogeneities to identify damage (*see* **Ultrasonic Methods**). Despite the practical utility and success of these methods, there are inherent limitations. One of the major limitations includes difficulties in detecting a small-scale incipient damage of the size comparable to or smaller than the wavelength of the transmitted wave. Another difficulty is distinguishing between the actual damage and structural features of comparable or greater size, such as notches, holes, borders, and other structural features, which produce multiple reflections effectively masking the signals associated with the damage. One possible approach to circumvent these limitations is to explore the nonlinear nature of the material damage by utilizing nonlinear acoustic methods for damage detection.

Numerous experimental and theoretical studies [1–58] demonstrate that damaged or fatigued materials exhibit noticeable nonlinear properties (nonlinear stress–strain relationship) because of micro/meso and macro defects. Strong nonlinearities have been observed for defects having contact interfaces such as cracks, delaminations, and disbonds. This type of contact acoustic nonlinearity is explained by "clapping" and "rubbing" of the interfacial rough surfaces subjected to vibrations [2, 5–10]. The physical nature of the contact nonlinearity can be understood by considering a contact defect (crack) as a planar interface that separates two elastic materials with the surfaces in intimate contact but with no traction forces across the interface [2]. As a longitudinally polarized elastic wave (vibration) is applied to the interface, the surfaces are pressed together during the compression phase of the wave and then separate under the tensile phase (Figure 1a). The elastic deformation of the medium containing such a defect will be different for tension and compression leading to a bilinear (nonlinear) stress–strain relationship, as illustrated in Figure 1(b).

A more realistic model considers a defect's interface as two rough elastic surfaces with Hertzian contact at surface microasperities (Figure 2a). The

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

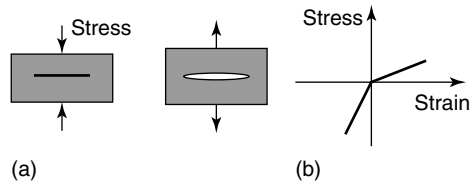


Figure 1. (a) Clapping defect and (b) its stress–strain bilinear dependence.

applied stress varies the asperity contact area leading to nonlinear elastic deformation of the defect (Figure 2b) [10, 11, 50]. Microasperity contacts may exhibit even more complicated nonlinear behavior such as hysteresis (Figure 2c) [12, 13], nonlinear thermal dissipation [14, 15], and nonlinear friction [33].

Another type of material degradation associated with increased nonlinearity is micro- and mesoscopic fatigue damage accumulation. Here the stronger nonlinearity is due to dislocations, hysteresis, formation of slip planes, and microcrack development and clustering [16–24].

It has to be understood, however, that not all types of material or structural flaws are characterized by strong nonlinearity. Large voids and porosity, dents, cuts and scratches, loss of material due to corrosion, breakage of structural components (cable wires, for example), and other flaws may not exhibit appreciable (or any) nonlinearity and, therefore, cannot be detected and characterized with nonlinear acoustic technique (NAT).

There are various acoustic manifestations of material and structural nonlinearities: generation of second- and higher-order harmonics [1–6, 24, 32], dependence of time-of-flight on applied external stress [25, 26], amplitude dependence of the resonance spectrum [27, 28], and the frequency mixing and modulation of high-frequency ultrasound by low-frequency vibration [7, 19–31]. Very intensive acoustic excitation may lead to the generation of subharmonics, intermodulation, and even chaotic behavior of highly nonlinear contact interfaces [33–35].

All of these phenomena have been explored for nondestructive testing and evaluation (NDT&E) of different materials and structures. The resulting nonlinear methods have advantages and disadvantages depending on the particular application. The principal advantage of all these methods is their highly *selective* sensitivity to flaws with nonlinear properties. Another important feature of the nonlinear techniques is their ability to detect flaws in highly nonhomogeneous and complicated geometries/structures, because structural inhomogeneities and features (holes, voids, channels, bonded laminations, boundaries, etc.) are linear and have no or little effect on the nonlinear readouts.

The NATs are not without limitations. One of the primary problems with their practical implementations for NDT&E is the need for a well-established reference. There are structural (nonflaw) sources of contact nonlinearity: structural supports and connections, inserts, etc.; all of these nonlinearities contribute to the nonlinear response. There could also be instrumentation and measurement nonlinearities, which may contribute to the "background" nonlinear readouts as well. Therefore, the background nonlinearity must be determined for a reference structure and particular measurement setup and then compared with the structure undergoing NDT&E. For many applications it is not easy to characterize this background nonlinearity. This drawback is, perhaps, one of the primary reasons that the nonlinear methods most reported till date are still experimental and are not yet established as practical and reliable defect detection and characterization tools.

In contrast to NDT&E, structural health monitoring (SHM) detects (monitors) changes in the materials/structure over time, so the initial measurements could be used as a reference for the very same structure. This approach makes SHM applications

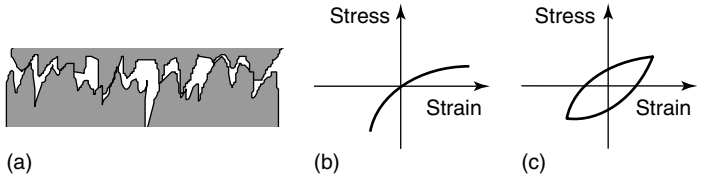


Figure 2. A defect with rough contact surfaces (a) and its stress–strain dependencies: Hertzian (b) and hysteretic (c).

extremely attractive for nonlinear techniques offering high sensitivity to the initiation and development of defects with a built-in nonlinear reference obtained from initial measurements on the healthy structure.

2 NONLINEAR ACOUSTIC METHODS FOR DAMAGE DETECTION

Among the number of different nonlinear methods mentioned above, there are two that are practically viable and most widely used: harmonic distortion and modulation methods. A simplified graphical illustration of these two methods is presented in Figure 3.

These methods rely on the experimental observations that the nonlinear response is proportional to the degree of damage regardless of the particular nonlinear mechanism involved: classical nonlinear elasticity, contact nonlinearity, hysteretic, thermal, frictional, slow dynamics, or any other yet-to-be-discovered nonlinearity.

2.1 Harmonic distortion methods

Historically, one of the first methods to characterize the acoustic nonlinearity is to measure the degree of the nonlinear (harmonic) distortion of a sinusoidal acoustic (vibration) signal. This approach has been widely used for the characterization of nonlinearity in fluids, biological media, electromechanical systems, and material nonlinearity of solids. The essence of the method is illustrated in Figure 3. An input signal is a sinusoidal waveform with frequency f_1 and

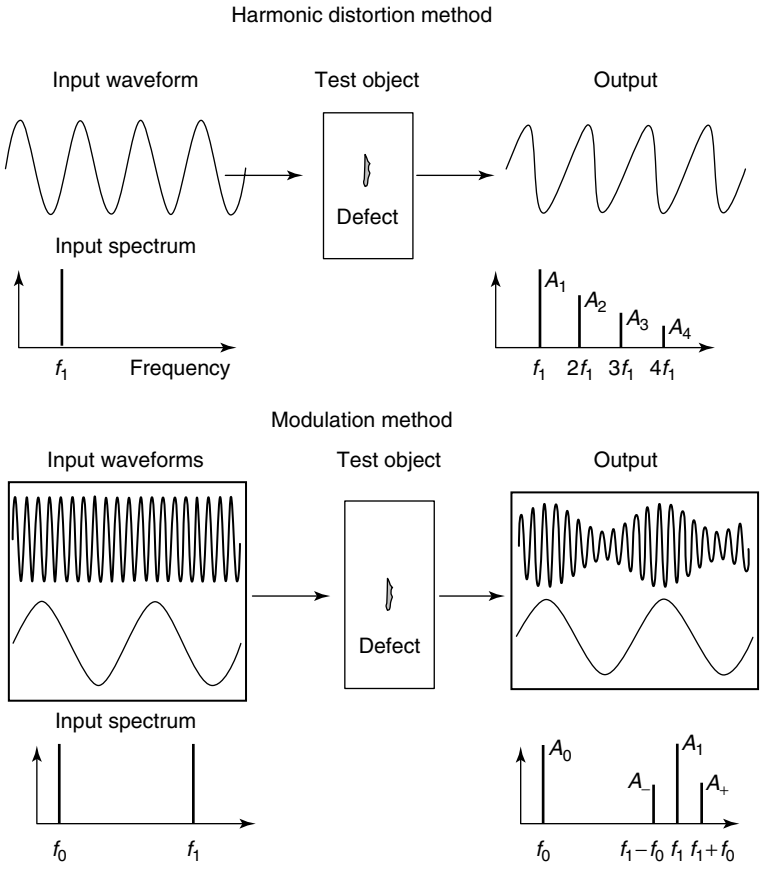


Figure 3. Illustrative diagrams of the harmonic distortion and modulation methods.

amplitude A_1. The nonlinearity distorts the waveform, so its spectrum contains additional harmonics. Typically, these are higher harmonics with frequencies $2f_1, 3f_1, \ldots$ and, respectively, diminishing amplitudes $A_1 > A_2 > A_3 > \cdots$. Because of this decrease in amplitude, most of the studies consider only the second harmonic for characterization of the defect's nonlinearity. The second-harmonic approach has been used for evaluation of fatigue cracks [1–3, 36], adhesive joints [5, 25, 26, 37–42], diffusion-bonded interfaces [43, 44], cracks in concrete [45], dislocations, and other fatigue damage [18, 24, 46–48, 56–58].

Very strong contact nonlinearity coupled with very large acoustic excitation may lead to much more pronounced nonlinear distortion with an appreciable level of higher-order harmonics and intermodulation and subharmonics [32–35]. Some of these high-order distortions could be localized in the vicinity of a defect, so the authors used a scanning laser vibrometer to visualize the location of the defect.

The range of frequencies and type of acoustic/vibration waves vary significantly depending on the specific applications: type of material, size of structure, and type and size of flaws. Thus, the reported frequencies used for the nonlinear detection span from hundreds of hertz to tens of megahertz. Flexural and torsional vibrations, longitudinal, shear, surface, and guided acoustic waves were utilized.

Two examples of investigations utilizing the second-harmonic method are presented here. Both investigations deal with the detection and characterization of disbonds in adhesive joints using very low-frequency vibrations, $f_1 = 190\,\text{Hz}$ [5], and high-frequency ultrasound $f_1 = 1.85\,\text{MHz}$ [37]. In both examples, the nonlinear measurements provided more sensitive evaluations when compared to the linear acoustic techniques.

The first exemplary study investigates the bonding quality of a thermoprotective coating (tiles) of a space shuttle [5]. The coating consists of the relatively rigid tiles attached to the ship's metal skin through an elastic bonding layer. The tiles (ceramic fiber) and the elastic bonding layer are so absorptive for the acoustic (ultrasonic) signals that only very low-frequency vibrations could be used. Then harmonic vibration was applied to the structure (and its level exceeds the static stress at the disbonded interface initiating its clapping), the tile's vibration was noticeably distorted as is evident in the vibration spectrum shown in Figure 4.

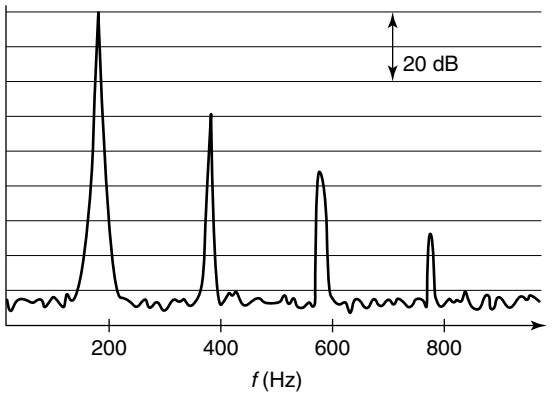


Figure 4. Harmonic distortion of sinusoidal vibration due to disbond.

The stiffness of the elastic layer with the clapping disbond is modeled as a bilinear spring (Figure 1b):

$$K(\varepsilon) = \begin{cases} \kappa S & \text{if } \varepsilon < 0 \\ \kappa\left(S - S_0\right) & \text{if } \varepsilon \geq 0 \end{cases} \tag{1}$$

where κ is the stiffness per unit area of the elastic layer without the defect, S is the full area of the layer, S_0 is the area of a disbond, and ε is the local strain at the elastic layer: negative ε corresponds to compression (full closure of the disbond) and positive ε is for tension (opening of the disbond).

Using a simple spring-mass oscillator model with the nonlinear stiffness (equation 1), the steady-state nonlinear solution (higher harmonics) of the oscillator's equation with nonlinear spring $K(\varepsilon)$ was found [5], using Bogolyubov–Mitropolsky's perturbation method. The results showed that the amplitude of the second harmonic A_2 is proportional to the area S_0 of the disbond:

$$A_2 \sim \frac{S_0}{S} A_1 \tag{2}$$

where A_1 is the amplitude of the fundamental (first) harmonic oscillation of the solid layer. This result is in good agreement with the experiment showing that the ratio of the amplitude of the second harmonic to the first may serve as a quantitative indicator of the defect.

It is interesting to note that the unique feature of the bilinear system is a linear dependence between

the second and the fundamental harmonic amplitudes (equation 2). For systems with a quadratic nonlinearity, the second harmonic is proportional to $A_1{}^2$.

The challenging problem of evaluating kissing bonds in aluminum–aluminum adhesive joints using a second-harmonic measurement is reported in [37]. Kissing bonds are disbonds between two compressively loaded disbonded surfaces creating contact, which is difficult to detect with conventional linear ultrasonic techniques. The result of this study demonstrated a high sensitivity of the nonlinear distortion technique to the disbond at a relatively low contact load. At higher compressive load, the disbond contact nonlinearity was reduced and "overwhelmed by the inherent system nonlinearity" [37] from the equipment and measurement setup.

The latter is a typical example highlighting the challenges in implementing the harmonic measurement approach: system nonlinearities from electronic and electromechanical equipment, such as signal generators, amplifiers, and transducers, generate a certain level of the harmonic distortion in the first place. This background level in the nonlinear signal limits the sensitivity of the method to defects with smaller nonlinearities.

2.2 Modulation methods

The modulation methods utilize the effect of the nonlinear interaction of acoustic/vibration waves in the presence of the nonlinear defects. There are two modifications of the method: vibromodulation (VM) and impact modulation (IM). The VM method utilizes two sinusoidal waves with the frequencies f_0 and f_1. The defect's nonlinearity leads to mixing of these two signals resulting in a new signal with the combination frequencies $f_1 \pm f_0$. Typically, the VM method employs lower frequency modulating and higher frequency probing signals: $f_0 \ll f_1$ [8–10, 28–31, 51]. Applied lower frequency vibration varies the contact area within a defect or damaged area (Figure 2a), effectively modulating the amplitude of the higher frequency probing wave passing through the varying contacts. In the frequency domain, the result of this modulation manifests itself as the sideband spectral components, $f_1 \pm f_0$, as shown in Figures 3 and 5. The defect or damage can be detected and characterized by the amplitude of the sideband components or, better, the modulation index (MI) (in decibel scale):

$$MI = 20\log_{10}\left(\frac{A_- + A_+}{2A_1}\right) = 20\log_{10}\left(\frac{A_- + A_+}{2}\right) - 20\log A_1 \quad (3)$$

In reality, the modulation phenomenon is more complicated because of the complexity of the nonlinear mechanisms at the interface as mentioned in the previous section of this article. Strong defect nonlinearities may lead to the occurrence of numerous sideband components with the frequencies $f_1 \pm mf_0$, where $m = 1, 2, \ldots$ as evident from Figure 5(b) and other experimental observations [8, 9, 15]. In practice, however, only the first sidebands ($m = 1$) are used as a reliable indicator of a damage.

The IM method uses impact-excited low-frequency vibration modes as the modulating signal for the

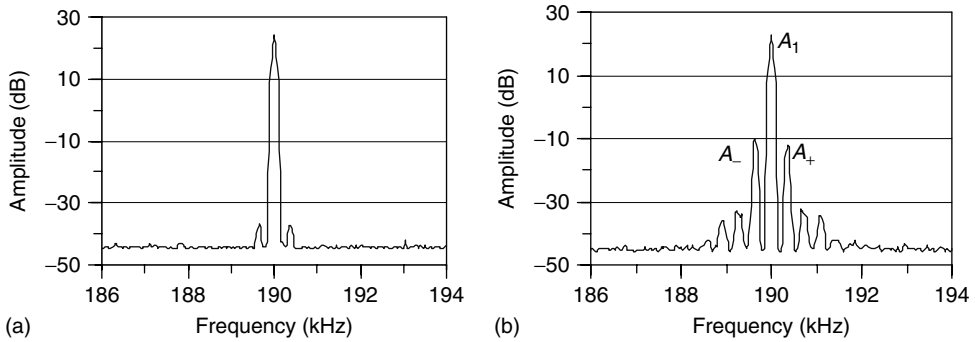


Figure 5. Spectra of a high-frequency ultrasonic signal modulated by a low-frequency vibration: (a) undamaged section of the steel pipe and (b) the same pipe section with stress-corrosion cracks.

high-frequency ultrasound [8–10, 49, 51]. The main advantage of IM over the VM approach is the ease of excitation of the low-frequency signal: a simple hammer can be used instead of an electronically controlled low-frequency vibration/acoustic source. IM works well, however, only for structures with low vibration damping. The impact-excited vibration modes should "ring" long enough, so that the resulting sideband component of the modulated ultrasonic signal can be resolved using Fourier spectrum analysis. At a minimum, the quality factor of the resonating vibration modes should be greater than 4 to resolve the sidebands. Calibration of the impact-excited vibration could also be a problem for comparative testing.

The modulation methods could be implemented using a continuous wave (CW) or a sequence of burst ultrasonic signals [51]. CW implementations of the VM (CW-VM) and IM (CW-IM) methods showed that the choice of the ultrasonic frequency, f_1, may have a significant impact on the MI, often leading to the erroneous interpretation of the test result. As seen from the recorded structural frequency responses of the probing ultrasonic signal and its sidebands (Figure 6), MI could vary as much as 40 dB depending on the choice of the primary frequency, f_1. This variation is due to resonances and antiresonances of the structure. Theoretical analysis [9] and numerous tests with various structures and materials [52] demonstrated that reliable damage detection and characterization could be achieved with frequency averaging as follows:

$$MI = 20\log_{10} \times \left(\frac{1}{N} \sum_{n=1}^{N} \frac{A_+ \left(f_n + f_0\right) + A_- \left(f_n - f_0\right)}{2A_1 \left(f_n\right)} \right) \tag{4}$$

where $f_n = F_{\text{start}} + n \cdot \Delta F$ is the fundamental ultrasonic frequency swept in steps n over the frequency range $F_{\text{start}} + N \cdot \Delta F$, with F_{start} being the starting frequency, ΔF the frequency step, and N the total number of steps. The choice of ΔF is determined by the density of the resonances of the particular structure's frequency response in the chosen frequency range. For proper averaging, ΔF should be less than the frequency separations between the resonances. The number of frequency steps should be at least 30, preferably 100.

In the burst implementation of the vibromodulation method (B-VM), a sufficiently long sequence of bursts with the central frequency f_1 for each burst and the repetition frequency $F_R > 2f_0$ is used instead of a CW ultrasonic signal. The bursts are effectively "sampling" the modulating envelope, as shown in Figure 7. The B-VM method does not require frequency averaging. Another significant

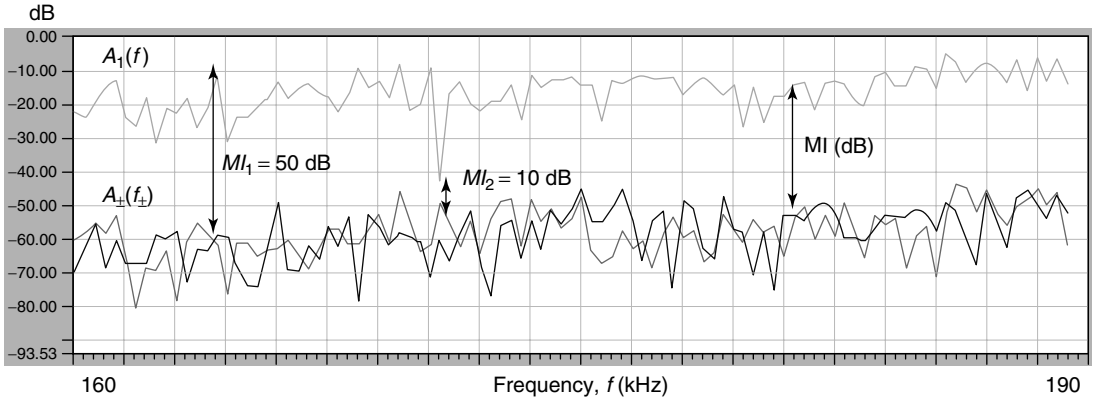


Figure 6. Frequency response ($A_1(f)$ upper curve, in decibels) of 0.5-m-long steel beam for the high-frequency ultrasonic signal f_1 swept the frequency range f: 160–190 kHz. Lower two curves are the corresponding frequency responses of the sidebands $A_\pm(f_\pm)$ (also in decibel scale) at the frequencies $f_\pm = f_1 \pm 250$ Hz recorded as f_1 is swept. MI is the modulation index (in decibels). It is graphically defined as a difference between the linearly averaged value of two lower curves and the upper curve. As can be seen, MI could vary as much 40 dB (MI_1–MI_2) simply due to resonances and antiresonances of the frequency response.

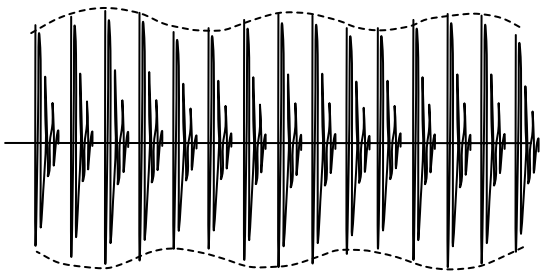


Figure 7. Burst VM method: sequence of ultrasonic bursts modulated by low-frequency vibration.

advantage of the B-VM method is an ability to localize a damage area and, because of localization, reduce the "background" nonlinearity contributed by structural connections and supports. However, the B-VM method may not work at all if there is no sufficient reflection from the damage (kissing bonds, for example) or there are significant reflections (reverberation) from the test object's structural elements, inhomogeneities and boundaries. The B-VM method is more complicated to implement, requiring more elaborate signal collection and processing.

There are numerous examples of damage detection and characterization using modulation methods. Examples of CW-VM and CW-IM methods (using longitudinal ultrasonic waves with frequencies up to 200 kHz) to detect fatigue cracks and characterize their size in steel pipes and beams can be found in [29] and [52]. The B-VM method utilizing 5 MHz surface acoustic waves has been used to characterize fatigue crack growth during a fatigue test of Al 2024-T3 alloy plate [31]. Some other examples of the VM tests include detection of fatigue cracks in automotive aluminum alloy casting parts [52], airframe [22], combustion engine parts [28], cracks in Plexiglas and other polymer materials [28, 34], stress-corrosion cracks in steel pipes, and cracks in concrete initiated by rebar corrosion [54].

Compared to the harmonic distortion method, the modulation methods are not influenced by nonlinearities in the signal-generating electronic equipment because the sources of the probe and modulating signals are independent. The use of low-frequency vibrations, especially flexural resonance vibrations, significantly amplifies nonlinearity of the defect, thus increasing the sensitivity of the measurements.

One of the problems implementing VM and, especially IM methods, for NDT&E screening for damage in multiple parts of the same kind is the calibration of the modulating vibration. In order to achieve repeatability and high sensitivity, the modulating stress applied to the damaged area should be constant or measured so that the MI can be normalized with respect to the modulating stress. It is not usually possible to measure this stress. Thus, if the modulating vibration utilizes resonating structural modes (this is always the case for IM methods and the most effective modulating signal for VM), the mode's frequencies and their quality factors could vary from part to part respectively varying the stress. If the position of the defect also varies from part to part and is unknown, there is no way to calibrate the modulating stress applied to the defect. This limitation, however, may not be so critical for SHM applications, which monitor changes in the same structure.

It should be mentioned that fatigue defects are developed in stress concentration areas most often caused by structural vibrations. Therefore, dynamic stresses provide the best modulation conditions for the probing ultrasonic signal because the stresses are concentrated exactly in the damaged area. In effect, this stress concentration enhances the nonlinear interaction between high- and low-frequency waves and improves damage detectability. This aspect, which is often overlooked, creates an interesting opportunity for SHM applications: VM techniques can utilize the structure's own vibration during its normal operation as a modulating signal. For example, aircraft in-flight VM SHM could use engine and fuselage vibrations; VM monitoring of a bridge could utilize vibration due to traffic and wind, etc.

3 FATIGUE DAMAGE ACCUMULATION MONITORING

The most challenging task for assessment and monitoring of structural health is the characterization of damage at the smallest possible scale often even before macrofracture is developed (*see* **Fatigue Life Assessment of Structures**). The sensitivity of linear ultrasonic testing (UT) significantly degrades as the damage size gets smaller. Being orders of magnitudes more sensitive to micro- and mesoscopic damages, [12, 17–19, 24] nonlinear acoustics offers a unique opportunity to monitor and characterize the damage accumulation at these scales.

There are growing numbers of studies devoted to characterization of precracking fatigue damage using nonlinear acoustics. Thus, a substantial increase of the nonlinear parameter during transition from micro- to macrodamage was reported in [55] using the VM technique. The harmonic distortion method was utilized to study the nonlinear acoustic response of metals (steel, aluminum, and titanium alloys) during fatigue tests [24, 47, 48, 56–58]. These investigations clearly demonstrated strong quantitative relationships between the nonlinear parameter measured with the second harmonic (and even third [58]) and fatigue damage.

Cantrell and Yost [24] investigated the effect of the dislocation dipoles on the nonlinear response of the aluminum 2024-T4 alloy. The authors measured the nonlinear parameter of the specimens subjected to an increasing number of fatigue cycles and suggested a relationship between the nonlinear acoustic parameter and the volume fraction of the persistent slip band consisting of dislocation dipoles. A model describing the contribution of the dislocation dipoles to the nonlinear acoustic parameter was proposed, and the theoretical predictions indicated reasonable agreement with the experimental data.

Jhang and Kim [56] measured the second harmonic of 5 MHz longitudinal ultrasonic wave to assess the damage accumulation in stainless steel samples subjected to static and dynamic tensile load. The test demonstrated proportionality between the measured nonlinear parameter and the fatigue load be it static or dynamic.

Frouin *et al.* [57] discussed the real-time finite amplitude measurements of the nonlinear acoustic parameter in titanium alloys subjected to low-cycle and the high-cycle fatigue. The authors used special grips and a transducer holder to enable acoustic measurements during the fatigue test. High-frequency transducers operating at 10 and 20 MHz were used to transmit a fundamental frequency signal and measure the material response at the second harmonic. Results obtained for both high-cycle and low-cycle regimes show stable growth of the nonlinear acoustic parameter at initial stages of the material deterioration.

Campos-Pozuelo *et al.* [58] developed a testing procedure for assessment of the fatigue behavior of titanium alloys and duraluminum using a very high intensity 22-kHz ultrasonic signal. Tuned to the test article's flexural or longitudinal resonances, the same excitation signal was used to fatigue the article as well as to measure the degree of nonlinear distortion. The setup enabled fatigue stresses up to 450 MPa. A strong nonlinear behavior was observed showing a 900 and 800% increase in the nonlinear parameter of the fatigued titanium alloy and duraluminum samples, respectively. These are compared to only a 0.8 and 1.4% change in linear response (resonance frequency shift) for the respective samples.

Donskoy *et al.* [19–23] conducted a series of CW-VM tests to characterize micro- and mesoscale damage accumulation during dynamic fatigue of Al 2024 T-4. The samples were fatigued at 10 Hz using the strain-control three-point bending test. The same dynamic load was used as a modulating stress. A specially developed correlation algorithm allowed for measuring and calculating an averaged MI (equation 4) every 18 s during the fatigue loading enabling essentially real-time continuous monitoring of the damage index (DI in decibels), defined as $DI = MI_{\text{damaged}} - MI_{\text{intact}}$. Over 100 samples were tested under different load condition demonstrating near linear dependency of the damage index versus number of fatigue cycles to failure. This observation suggests that the linear damage accumulation rule, a familiar concept in fracture mechanics, may be reformulated in terms of the nonlinear acoustic data to enable prediction of the specimen's remaining fatigue life. To confirm that the nonlinear acoustic damage index is responsive to the micro- and mesoscale structural changes, a microscopic analysis of the fatigue samples using a scanning acoustic microscope (SAM) and a scanning electron microscope (SEM) were conducted. The images depicting material in the damage accumulation area and the corresponding nonlinear damage index as a function of the number of fatigue cycles are shown in Figure 8. Under the cyclic fatigue load, the material transitions from a consistent microstructure to zones of extensive plastic deformation. During various stages of this transformation, SEM micrographs reveal nucleation of microdefects, formation of slip planes and striation marks, and development and clustering of microcracks. The coalescence of microfailures finally results in propagation of one or several macrocracks, which cause ultimate material fracture. The damage index presented in Figure 8 shows a gradual increase at all stages of the fracture

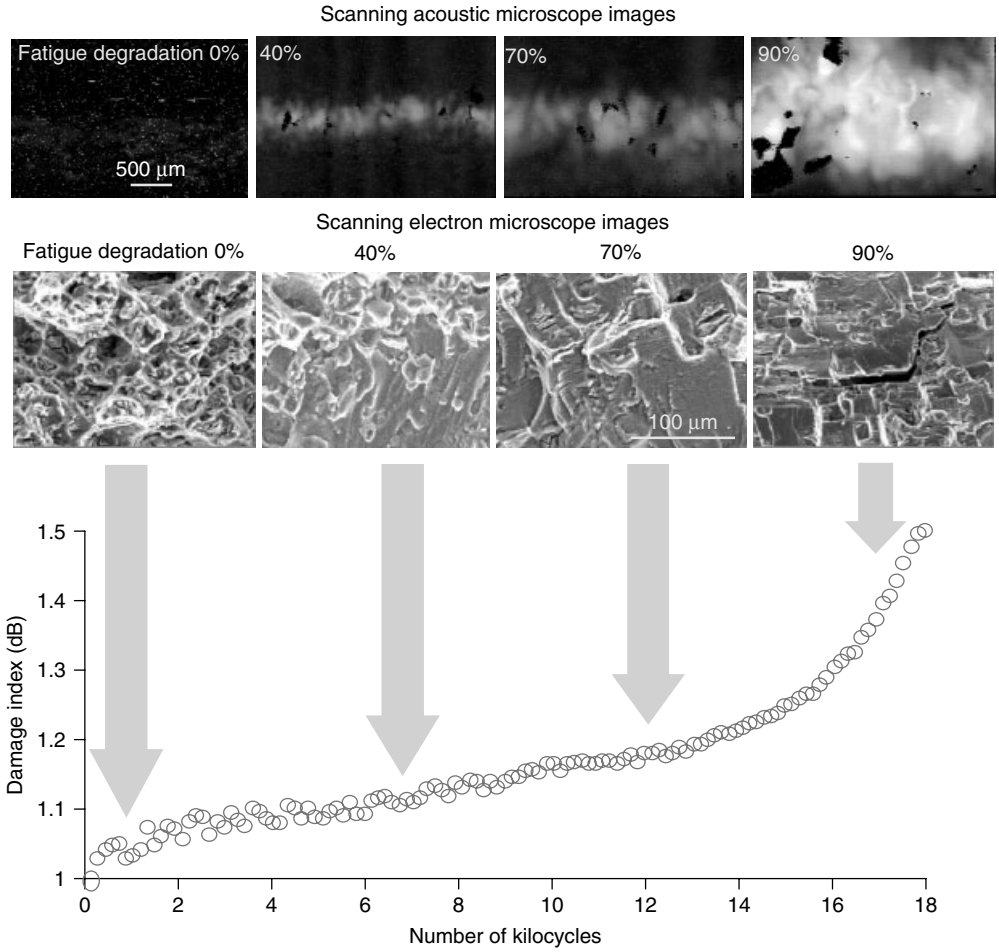


Figure 8. Nonlinear acoustic damage index versus number of fatigue cycles and corresponding sequential stages of the fracture process—SAM and SEM micrographs.

process and rapid growth just before the macroscale cracking.

4 CONCLUSION

Investigations of the nonlinear dynamics of materials with contact-type macrodefects (cracks, disbonds, delaminations) as well as fatigued materials with micro- and mesoscale damages reveal their unusually high acoustic nonlinearities, often orders of magnitude greater than found in undamaged materials. This phenomenon is utilized for detection and characterization of the nonlinear damages in a broad spectrum of materials: different metals, composites, wood, and concrete. There are two primary and most widely used NATs: harmonic distortion and modulation. The first correlates the amplitude of the second harmonic (sometimes higher-order harmonics and even intermodulation and subharmonics) with the presence and severity of the damage. The modulation methods employ the nonlinear interaction of two signals: typically lower frequency modulating vibration and higher frequency probing ultrasound. The strong low-frequency signal varies interface conditions at the damaged area. These variations effectively modulate the high-frequency signal passing through: the greater the damage, the greater the modulation effect.

Advantages of the NAT include much higher nonlinear response contrast between damaged/undamaged materials: studies report hundreds of percents change in the nonlinear response versus only a fraction of a percent in the linear response for the same damage. Being responsive to only nonlinear defects, the NAT can be used in structures with complicated geometries in which multiple reflections (reverberation) often preclude the use of the linear UT.

One of the difficulties in implementing NAT for many NDT&E applications is the requirement for a well-established "nonlinear background" reference for a particular structure. However, because SHM detects (monitors) changes in the materials/structure over time, the initial measurements could be used as a reference for the very same structure. This reference, coupled with the extremely high responsiveness to changes due to damage, makes the NAT highly attractive for SHM applications. Additionally, many implementations of the NAT are perfectly suited for monitoring of large portions of a structure using just a few sensors in fixed locations not requiring sensor spatial scanning. We believe these advantages are the primary reasons for selecting NAT over competing techniques in some SHM applications.

REFERENCES

[1] Buck O, Morris WL, Richardson JN. Acoustic harmonic generation at unbonded interfaces and fatigue cracks. *Applied Physics Letters* 1978 **33**(5):371–373.

[2] Richardson M. Harmonic generation at an unbonded interface. I. Planar interface between semi-infinite elastic media. *International Journal of Engineering Science* 1979 **17**:73–75.

[3] Morris WL, Buck O, Inman RV. Acoustic harmonic generation due to fatigue damage in high-strength aluminum. *Journal of Applied Physics* 1979 **50**(11):6737–6741.

[4] Yost WT, Cantrell JH. Materials characterization using acoustic nonlinearity parameters and harmonic generation. In *Engineering Materials, Review of Progress in Quantitative Nondestructive Evaluation*, Thompson DO, Chimenti DE (eds). Plenum Press: New York, 1990; Vol. 9B, pp. 1669–1676.

[5] Antonets VA, Donskoy DM, Sutin AM. Nonlinear vibro-diagnostics of flaws in multilayered structures. *Mechanics of Composite Materials* 1986 **15**:934–937.

[6] Solodov Yu. Ultrasonics of non-linear contacts: propagation, reflection and NDE applications. *Ultrasonics* 1998 **36**:383–390.

[7] Tohjirna E, Sato T. Nonlinear acoustical sounding system for detection of defects in pipelike objects. *The Journal of the Acoustical Society of America* 1988 **83**(4):1661–1666.

[8] Donskoy DM, Sutin AM. Vibro-acoustic modulation nondestructive technique. *Journal of Intelligent Material Systems and Structures* 1999 **9**:765–771.

[9] Donskoy DM, Sutin AM, Ekimov A. Nonlinear acoustic interaction on contact interfaces and its use for nondestructive testing. *NDT International* 2001 **34**:231–238.

[10] Zaitsev VY, Sutin AM, Belyaeva IY, Nazarov VE. Nonlinear interaction of acoustical waves due to cracks and its possible usage for cracks detection. *Journal of Vibration and Control* 1995 **1**:335–344.

[11] Baltazar A, Rokhlin SI, Pecorari C. On the relationship between ultrasonic and micromechanical properties of contacting rough surfaces. *Journal of the Mechanics and Physics of Solids* 2002 **50**:1397–1416.

[12] Nazarov VE, Ostrovsky LA, Soustova IA, Sutin AM. Nonlinear acoustics of micro-inhomogeneous media. *Physics of the Earth and Planetary Interiors* 1988 **50**:65–70.

[13] McCall KR, Guyer RA. Equation of state and wave propagation in hysteretic nonlinear elastic materials. *Journal of Geophysical Research* 1994 **99**(23):887–897.

[14] Gusev V, Mandelis A, Bleiss R. Theory of strong photothermal nonlinearity from sub-surface non-stationary (breathing) cracks in solids. *Applied Physics A: Material Science Process* 1993 **A57**:229.

[15] Zaitsev VYu, Sas P. Dissipation in microinhomogeneous solids: inherent amplitude-dependent attenuation of a non-hysteretical and non-frictional type. *Acta Acustica* 2000 **86**:216–228.

[16] Van Den Abeele K, Johnson PA, Guyer RA, McCall KR. On the quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation. *The Journal of the Acoustical Society of America* 1997 **101**(4):1885–1898.

[17] Nicata A, Elbaum C. Generation of ultrasonic second and third harmonics due to dislocations. *Physical Review* 1966 **144**(2):469–477.

[18] Buck O. Harmonic generation for measurement of internal stresses as produced by dislocations. *IEEE*

Transactions on Sonics and Ultrasonics 1976 **SU-23**(5):346–350.

[19] Zagrai A, Donskoy D, Chudnovsky A, Golovin E, Agarwala E. Micro/meso scale fatigue damage accumulation monitoring using nonlinear acoustic vibro-modulation measurements. *SPIE Proceeding* 2006 **6175**.

[20] Donskoy D, Zagrai AN, Chudnovsky A, Wu H. Nonlinear acoustic vibro-modulation technique for materials damage diagnostics and prognostics. Presented at *AEROMAT 2005 Conference and Exposition*. Orlando, FL, 6–9 June 2005; Also published in *Advanced Materials and Processes* **163**(4): 34, 2005, www.asminternational.org/AMP (2005).

[21] Zagrai A, Donskoy D, Chudnovsky A, Wu H. Assessment of material degradation using nonlinear acoustic vibro-modulation technique. *3rd US-Japan Symposium on Advancing Applications and Capabilities in NDE*. Maui, HI, 20–24 June 2005.

[22] Zagrai AN, Donskoy D, Sedunov N, Chudnovsky A, Wu H. Nonlinear acoustic assessment of material fatigue damage. *Greater Philadelphia AIAA/ASME Inaugural Aerospace/Mechanical Engineering Mini-Symposium, Plymouth Meeting*. Pennsylvania, PA, 29 January 2005.

[23] Donskoy D, Zagrai A, Chudnovsky A, Golovin E, Agarwala V. Nonlinear acoustic vibro-modulation technique for life prediction of aging aircraft components. In *Proceedings of the Third European Workshop on Structural Health Monitoring*, Guemes A (ed). DEStech Publications: Pennsylvania, PA, 2006, pp. 251–258.

[24] Cantrell JH, Yost WT. Nonlinear ultrasonic characterization of fatigue microstructures. *International Journal of Fatigue* 2001 **23**(S1):S487–S490.

[25] Nagy PB, McGowan P, Adler L. Acoustic nonlinearities in adhesive joints. In *Review of Progress in Quantitative Nondestructive Evaluation*, Thompson DO, Chimenti DE (eds). Plenum Press: New York, 1990; Vol. 10B, pp. 1685–1692.

[26] Adler L, Nagy PB. Second order nonlinearities and their application in NDE. In *Review of Progress in Quantitative Nondestructive Evaluation*, Thompson DO, Chimenti DE (eds). Plenum Press: New York, 1991; Vol. 10B, pp. 1813–1820.

[27] Van Den Abeele K, Carmeliet J, TenCate JA, Johnson PA. Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. Part II: single mode nonlinear resonance acoustic spectroscopy. *Research in Nondestructive Evaluation* 2000 **12**(1):31–42.

[28] Van Den Abeele K, Johnson PA, Sutin A. Nonlinear Elastic Wave Spectroscopy (NEWS) techniques to discern material damage. Part I: Nonlinear Wave Modulation Spectroscopy (NWMS). *Research in Nondestructive Evaluation* 2000 **12**(1):17–30.

[29] Duffor P, Morbidini M, Cawley P. A study of the vibro-acoustic modulation technique for the detection of cracks in metals. *The Journal of the Acoustical Society of America* 2006 **119**(3):1463–1475.

[30] Kim J-Y, Yakovlev VA, Rokhlin SI. Parametric modulation mechanism of surface acoustic wave on a partially closed crack. *Applied Physics Letters* 2003 **82**:3203–3205.

[31] Kim J-Y, Yakovlev VA, Rokhlin SI. Surface acoustic wave modulation on a partially closed fatigue crack. *The Journal of the Acoustical Society of America* 2004 **115**:1961–1972.

[32] Krohn N, Stoessel R, Busse G. Acoustic nonlinearity for defect selective imaging. *Ultrasonics* 2002 **40**:633–637.

[33] Solodov I, Wackerl J, Pfleiderer K, Busse G. Nonlinear self-modulation and subharmonic acoustic spectroscopy for damage detection and location. *Applied Physics Letters* 2004 **84**:5386–5388.

[34] Ballad EM, Vesirov SYu, Pfleiderer K, Solodov IYu, Busse G. Nonlinear modulation technique for NDE with air-coupled ultrasound. *Ultrasonics* 2004 **42**:1031–1036.

[35] Solodov IYu, Korshak BA. Instability, chaos, and "memory" in acoustic wave-crack interaction. *Physical Review Letters* 2002 **88**:014303.

[36] Ouahabi A, Thomas M, Lakis AA. Detection of damages in beams and composite plates by harmonic excitation and time-frequency analysis. In *Proceedings of the Third European Workshop on Structural Health Monitoring 2006*, Guemes A (ed). DEStech Publications: Pennsylvania, PA, 2006, pp. 775–782.

[37] Brotherhood CJ, Drinkwater BW, Dixon S. The detectability of kissing bonds in adhesive joints using ultrasonic techniques. *Ultrasonics* 2003 **41**: 521–529.

[38] Fassbender SU, Kroning M, Arnold W. Measurement of adhesion strength using nonlinear acoustics. *Materials Science Forum* 1996 **210–213**: 783–790.

[39] Tang Z, Cheng A, Achenbach JD. Ultrasonic evaluation of adhesive bond degradation by detection of

the onset of nonlinear behavior. *Journal of Adhesion Science and Technology* 1999 **13**(7):837–854.

[40] Liu G, Qu J, Jacobs LJ, Li J. Characterizing the curing of adhesive joints by a nonlinear ultrasonic technique. In *Review of Progress in QNDE*, Thompson DO, Chimenti DE (eds). Plenum Press: New York, 1999; Vol. 18, pp. 2191–2199.

[41] Rothenfusser M, Mayr M, Baumann J. Acoustic nonlinearities in adhesive joints. *Ultrasonics* 2000 **38**:322–326.

[42] Achenbach JD, Parikh OK. Ultrasonic analysis of nonlinear response and strength of adhesive bonds. *Journal of Adhesion Science* 1991 **5**(8):601–618.

[43] Barnard DJ, Dace GE, Rehbein DK, Buck O. Acoustic harmonic generation at diffusion bonds. *Journal of Nondestructive Evaluation* 1997 **16**(2):77–89.

[44] Kawashimaa K, Muraseb M, Yamadac R, Matsushimad M, Uematsue M, Fujitaf F. Nonlinear ultrasonic imaging of imperfectly bonded interfaces. *Ultrasonics* 2006 **44**(Suppl. 1:):e1329–e1333.

[45] Stauffer JD, Woodward CB, White KR. Nonlinear ultrasonic testing with resonant and pulse velocity parameters for early damage in concrete. *Materials Journal* 2005 **102**(2):118–121.

[46] Hikata A, Chick BB, Elbaum C. Effect of dislocations on finite amplitude ultrasonic waves in aluminum. *Applied Physics Letters* 1963 **3**:195–197.

[47] Cantrell JH, Yost WT. Acoustic harmonic generation from fatigue-induced dislocation dipoles. *Philosophical Magazine A* 1994 **69**(2):315–326.

[48] Cantrell JH. Dependence of microelastic-plastic nonlinearity of martensitic stainless steel on fatigue damage accumulation. *Journal of Applied Physics* 2006 **100**(6):063508–063508-7.

[49] Ryles M, McDonald I, Ngau FH, Staszewski WJ. Ultrasonic wave modulations for damage detection in metallic structures. In *Proceedings of the Third European Workshop on Structural Health Monitoring 2006*, Guemes A (ed). DEStech Publications: Pennsylvania, PA, 2006, pp. 275–282.

[50] Rudenko O, Chin AV. Nonlinear acoustic properties of a rough surface contact and acousto-diagnostics of a roughness height distribution. *Acoustical Physics* 1994 **40**(4):593–596.

[51] Donskoy DM, Sutin AM. *Method and Apparatus for Acoustic Detection and Location of Defects in Structures or Ice on Structures*, US Patent #6,301,967, 2001.

[52] Donskoy DM, Ekimov A, Luzzato E, Lottiaux J-L, Stoupin S, Zagrai A. N-SCAN®: new vibro-modulation system for detection and monitoring of cracks and other contact-type defects. In *Smart Systems and Nondestructive Evaluation for Civil Infrastructures*. SPIE Proceedings, 2003; Vol. 5057, pp. 400–409.

[53] Sigworth GK, DeHart F, Major JF, Donskoy D. Bulking up aluminum alloys. *Modern Casting* 2003 **93**(5):40–41.

[54] Donskoy DM, Ferroni K, Sutin AM, Sheppard K. A nonlinear acoustic technique for crack and corrosion detection in reinforced concrete. *Proceeding of 8th International Nondestructive Testing Conference*. Boulder, CO, 1997.

[55] Van Den Abeele KE-A, Sutin A, Carmeliet J, Johnson PA. Micro-damage diagnostics using nonlinear elastic wave spectroscopy (NEWS). *NDT and E International* 2001 **34**:30–248.

[56] Jhang K-Y, Kim K-C. Evaluation of material degradation using nonlinear acoustic effect. *Ultrasonics* 1999 **37**:39–44.

[57] Frouin J, Sathish S, Na JK. Real-time monitoring of acoustic linear and nonlinear behavior of titanium alloys during low-cycle fatigue and high-cycle fatigue. *Proceedings of the SPIE's 5th International Symposium on Nondestructive Evaluation and Health Monitoring of Aging Infrastructure*, Newport Beach, CA, 5–9 March 2000.

[58] Campos-Pozuelo C, Vanhille C, Gallego-Juarez JA. Comparative study of the nonlinear behavior of fatigued and intact samples of metallic. *IEEE Transactions on Ultrasonics Ferroelectrics, and Frequency Control* 2006 **53**(1):175–184.

Chapter 14
Ultrasonic Methods

Wolfgang Hillger
German Aerospace Center (DLR), Braunschweig, Germany

1 INTRODUCTION

Nondestructive testing (NDT) means to test materials and components in a way so that its later usage is not affected. NDT methods do not directly indicate flaws but only the variation of physical values. Therefore, it is the challenge to interpret flaws out of these variations. Reference standards (i.e., a test specimen with artificially inserted defects like flat bottom bore holes) are required for a calibration and a system check. Applications of NDT are quality control, determination of material properties, and defect detection.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

NDT started in 1900 when Roentgen discovered his X rays. Another old NDT method is the testing with sound; internal defects in forgings can be detected by changes in the sound when they are hammered. The beginning of ultrasonic testing (UT) started in 1929. Sokolow [1, 2] detected defects in components by applying UT-through-transmission techniques.

The UT is the most used NDT method without any negative impact on the environment. UT is a true NDT method because no loading and no chemical agents are necessary. There is no limit to the number of times that the UT testing can be carried out (recurring testing of components).

2 FUNDAMENTALS OF ULTRASONIC TESTING

2.1 Wave propagation and wave modes

2.1.1 Wave propagation

This article presents the basic principles of UT. Additional information can be found in the literature [3–6] and on websites [7, 8]. UT is based on elastic waves contrary to X-ray testing that uses electromagnetic waves. Contrary to electromagnetic waves, acoustic waves do not propagate in a vacuum.

A mechanical wave consists of mechanical vibrations of the particles in a material. This periodic movement is identified as an oscillation. The

recording of the movement versus time indicates a sinus curve:

$$x(t) = r \sin \alpha \tag{1a}$$

where

$$\alpha = \omega t \tag{1b}$$

$x(t)$ identifies the oscillation motion, t the time, r the amplitude, $\alpha = \omega t$ the phase, and ω the angular speed.

T is the oscillation period and indicates the time between two zero crossings. The number of oscillations per seconds is called *frequency f*

$$f = 1/T \tag{2}$$

The angular frequency ω is given by

$$\omega = 2\pi f = 2\pi/T \tag{3}$$

If all particles of a plan in a component (by excitation of an actuator) are vibrating sinuslike with the same clock, the outcome is an elastic wave. This excitation is propagating with the characteristic velocity v of the material. This velocity depends on the elastic properties of the material and can be determined out of measurements of the sound path s and its time of flight t:

$$v = s/t \tag{4}$$

An important factor of material testing is the wavelength λ. The wavelength is given by the regular spaces of the particles with the same vibration phases (Figure 1) and can be calculated by equation (5):

$$\lambda = v/f \tag{5}$$

The wavelength can be selected by the exciting (test) frequency f. For testing of homogeneous materials, frequencies in a range of 0.5–20 MHz

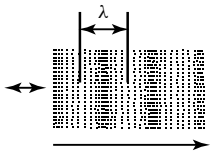
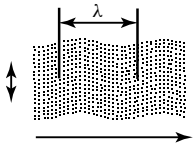
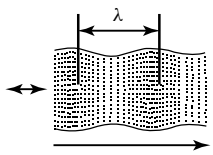
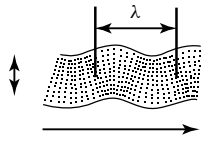
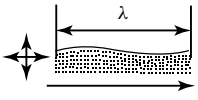
Mode	Vibration of particles	Velocity
Longitudinal wave		$v_L = \sqrt{\frac{E}{\rho} \cdot \frac{(1-\mu)}{(1+\mu)\cdot(1-2\mu)}}$
Shear wave		$v_T = \sqrt{\frac{E}{\rho} \cdot \frac{1}{2(1+\mu)}}$
Dilatational wave		$v_D = \sqrt{\frac{E}{\rho}}$
Bending wave		$v_B = \frac{2w}{\lambda} \sqrt{\frac{E}{\rho} \cdot \frac{I}{F}}$
Surface wave		$v_O = \frac{0.87 + 1.12\,\mu}{1-\mu} \cdot v_T$

Figure 1. Elastic wave modes in solid materials.

are mostly used. The wavelength is situated in the millimeter range.

2.1.2 Wave modes

The most used wave mode of the UT testing is the longitudinal wave (Figure 1).

It is characterized by the vibration of the particles in propagation direction. The longitudinal wave has the largest velocity of all elastic wave modes. The sound waves in air are longitudinal waves. Also, propagation in solids and in water is possible.

The equations in Figure 1 present the correlation between the velocities of the different wave modes and the elastic properties: the dynamic E modulus E_{dyn}, Poisson's ratio μ_{dyn}, as well as the density ρ.

In solids, a propagation of shear (transverse) waves is possible, too. The oscillation of the particles is perpendicular to the wave propagation. An excitation is possible by a periodic shearing force on the surface of the component so that the particles are periodically moved up and down. In solids, a shear force can be propagated to the next layer so that propagation in the material is possible.

In air and in liquids, no thrusts power can be carried so that propagation of shear waves is not possible. Using the same test frequency, the wavelength of shear wave is half of those ones of longitudinal waves because $v_{\mathrm{T}} \cong 0.5\, v_{\mathrm{L}}$.

Bending wave propagate in bars if the length is not larger than the wavelength, the diameter smaller than λ. In dependence of the excitation with longitudinal or shear waves, dilatational waves or bending waves are generated.

Surface or Rayleigh waves only propagate at the interfaces of components. These waves are very sensitive to surface defects.

There are different plate waves such as Lamb waves (*see* **Modeling of Lamb Waves in Composite Structures**) which are used for damage detection. Guided waves like Lamb waves can penetrate large areas and interact with defects and structural inhomogeneities. In contrary to UT, the sensors and actuators are permanently applied to the structure. This "long-range ultrasonics" enable a damage detection and SHM without time-consuming scanning. The most used guided waves are surface waves like Rayleigh waves and Lamb waves. These types of waves interact with defects and material discontinuities.

Lamb discovered these dispersive plate waves, which are propagating in solid plate with free boundaries [9]. For a plate with a thickness of 2 h, the symmetrical and the antisymmetrical modes can be expressed by [10]

$$(k^2 + s^2)^2 \cosh(qh) \sinh(sh) - 4k^2 qs \sinh(qh) \cosh(sh) = 0 \quad (6)$$

$$(k^2 + s^2)^2 \sinh(qh) \cosh(sh) - 4k^2 qs \cosh(qh) \sinh(sh) = 0 \quad (7)$$

with $q^2 = k^2 - k_{\mathrm{l}}{}^2$ and $s^2 = k^2 - k_{\mathrm{t}}{}^2$.

k indicates the wavenumber, k_{l} and k_{t} are wavenumbers for longitudinal and shear waves.

It is important to know that there exist multiple wave modes that satisfy equations (1a) and (1b).

A dispersion diagram (*see* **Modeling of Lamb Waves in Composite Structures**) presents a plot of the phase velocities versus the product of frequency and plate thickness. Using a broadband excitation, the different frequency components travel with different speeds through the material. As a result, the shape of the wave package changes during propagation and the receiver signals are difficult to evaluate.

In a defined frequency-thickness product only two fundamental modes exist; a symmetrical mode and an antisymmetrical one, which are widely separated by different phase velocities. Lamb wave testing should be carried out in this low-frequency range.

2.1.3 Properties of the sound field

An important factor of the sound field is the acoustic pressure (excess pressure) p. Bergmann [6] described the acoustic pressure and deflection of particles ξ in the case of plane waves and spherical waves in the following correlation:

$$p = \rho v \omega \xi \quad (8)$$

The product of density ρ and velocity v refers to the acoustical impedance z:

$$z = \rho v \quad (9)$$

Materials having high acoustic impedance are called *reverberant*.

Example For copper, the acoustic impedance z_c can be calculated as follows:

$\rho_c = 8.9 \times 10^3\,\mathrm{kg\,m^{-3}}$, $v_L = 4700\,\mathrm{m\,s^{-1}}$, $z_c = 8.9\,\mathrm{kg\,m^{-3}} \times 10^3 \times 4.7 \times 10^3\,\mathrm{m\,s^{-1}} = 41.8 \times 10^6\mathrm{kg\,m^{-2}\,s^{-1}} = 41.8 \times 10^6\,\mathrm{Ns\,m^{-3}} = 41.8\,\mathrm{MRayl}$

The intensity J is proportional to the square of the acoustic pressure.

$$J = 1/2p^2/z = 1/2z\omega^2\xi^2 \tag{10}$$

UT transducers deliver an output voltage proportional to the acoustic pressure.

For the determination of the intensity, the output voltage has to be squared.

2.2 Sound waves at plane boundaries

2.2.1 Normal incidence

For the testing of materials, the behavior of sound waves at boundaries and interfaces with different acoustic impedances is very important. Boundaries can be external surfaces of components and defects such as cavities in the component. A sleek boundary delivers a reflection, whereas a rough one delivers a scattering. The scale for sleek and rough is the wavelength.

If a plane wave with a sound pressure p_e perpendicularly reaches a sleek plane boundary, one part of the plane wave propagates the boundary (p_d), whereas the other one (p_r) is being reflected (Figure 2). The reflected part is contrarily propagating to the invading one.

$$R = p_r/p_e \tag{11}$$

and

$$D = p_d/p_e \tag{12}$$

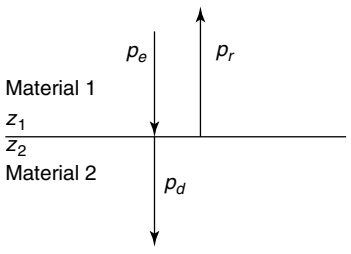


Figure 2. Perpendicular incidence on a flat interface.

R (reflection coefficient) and D (transmission coefficient) can be calculated out of the impedances z_1 and z_2:

$$R = (z_2 - z_1)/(z_2 + z_1) \tag{13}$$

$$D = 2z_2/(z_2 + z_1) \tag{14}$$

Reflection coefficient R and transmission coefficient D are referred to the sound pressure, which is measured by transducers. The following equations describe the relationships between the different sound pressures and R and D:

$$p_e + p_r = p_d \tag{15}$$

or

$$1 + R = \mathrm{D} \tag{16}$$

For the different intensities J, equation (15) calculated the balance

$$J_e = J_r + J_d \tag{17}$$

Equations (16) and (17) calculate the percentile parts of the reflected (r: reflection factor) and transmitted sound intensities (d: transmission factor) at a plane interface:

$$r = [(z_2 - z_1)/(z_2 + z_1)]^2 \tag{18}$$

$$d = 4z_2z_1/(z_2 + z_1)^2 \tag{19}$$

Because of the energy balance,

$$r + d = 1 \tag{20}$$

Note that in the literature there are different definitions for transmission factor d, transmission coefficient D, and reflection coefficient R, and reflection factor r. In order to prevent confusions, both possibilities are listed here.

The transducers deliver an electrical voltage, which is proportional to the sound pressure; therefore, it is more useful to use equations (11) and (12). The information about the phase (prefix of R) is loosening by calculating with intensities because $J \sim p^2$.

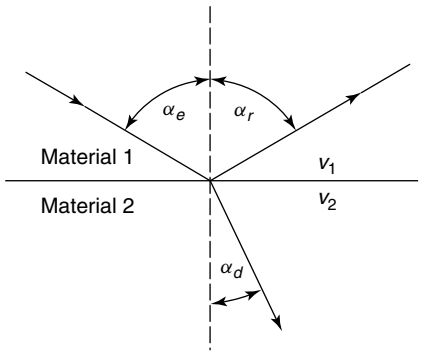


Figure 3. Reflection and refraction of a plane wave with angular incidence at a boundary between two materials.

2.2.2 Angular incidence

The incoming wave with the angle α_e to the perpendicular is reflected with the same angle. A part is penetrating material 2 with the angle α_d (Figure 3). The relationship between these angles and the velocities describes Snell's law:

$$\sin\alpha_e / \sin\alpha_d = v_1/v_2 \tag{21}$$

Using equation (21), it is possible to provide a conversion from one wave mode with a velocity v_1 to another one with v_2. Also, it is possible to select a defined mode.

2.3 Sound propagation

The amplitude of the spherical wave is inversely proportional to the distance from the transmitter. Besides this material-independent sound attenuation, there are also material-dependant influences like scattering and absorption. Both terms are summarized in attenuation. The sound pressure decreases exponentially as

$$p(x) = p_0 \mathrm{e}^{-\alpha x} \tag{22}$$

where p_0 identifies the sound pressure at the initial location, $p(x)$ the sound pressure in a distance x from the source, and α the sound attenuation.

It is also possible to refer α to the sound intensity. Because of $J \sim p^2$, equation (21) delivers

$$\alpha_J = 2\alpha \tag{23}$$

The sound attenuation can be divided into two parts:

$$\alpha = \alpha_\mathrm{a} + \alpha_\mathrm{s} \tag{24}$$

where α_a identifies the absorption and α_s the scattering coefficient.

The absorption increases nearly proportionally to the test frequency. The reason is the conversion of acoustical energy into heat. The absorption can be settled by a higher transmitter voltage and/or a higher gain of the receiver amplifier and also by a lower test frequency.

Contrary to the absorption, the scattering cannot be settled by a higher amplification. It is useful to compare scattering to the effect of fog on a car driver who is dazzled by his own headlamps. Scattering does not only provide attenuation but also several small additional echoes with different times of flight. Scattering is performed by inhomogeneities in the material, which are smaller than the wavelength. Inhomogeneities are interfaces at which the sound impedance is excursively changing.

Inhomogeneities that have a diameter below 1% of the wavelength do not produce scattering. With increasing diameter, the scattering increases with the third power of the diameter and the fourth power of the frequency [11]:

$$\alpha_\mathrm{s} \sim {d_k}^3 f^4 \tag{25}$$

In order to reach a high-resolution UT, the test frequency spectrum (which defines the band of wavelengths) has to be optimized carefully, especially for inhomogeneous materials. Using UT, it is not possible to distinguish between α_s and α_a because the resulting sound attenuation defines the sound pressure at the receiver.

3 ULTRASONIC TESTING

3.1 Transducers

3.1.1 Normal incidence

Transducers are used as a sender and a receiver for UT of materials. They are manufactured for a wide range of applications with different element diameters, center frequencies, bandwidths, and couple

techniques like immersion technique, water split, and squirter [6, 7]. Mostly, transducers provide normal incidence to the surface of the component. Such a contact transducer is shown schematically in Figure 4. These types of transducers are used for hand manipulating with hand-held ultrasonic flaw detectors. The active element consists, in most cases, of a piezoelectric crystal (*see* **Piezoceramic Materials—Phenomena and Modeling**; **Piezoelectricity Principles and Materials**). It is well protected in a rugged casing and with a wear plate, which is also used as a matching layer to the material.

In order to get a short pulse response, the active element is damped by a backing that reduces the ringing and increases the bandwidth. The bandwidth of the transducer (up to 100% of the center frequency) is inversely proportional to the pulse duration, which means that a high bandwidth provides a short pulse response of the transducer.

3.1.2 Sound field

The sound field of a transducer consists of a near field, which is characterized by interferences and a far field. The near-field length is given by (26) [3]

$$N = \frac{D^2 - \lambda^2}{4\lambda} \approx \frac{D^2 \cdot f}{4c} \tag{26}$$

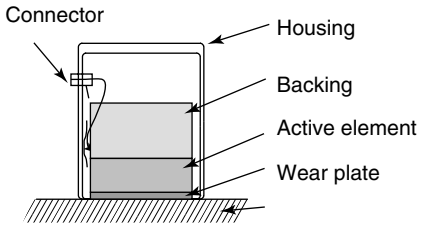


Figure 4. Construction of a normal incidence contact transducer.

where D is the element diameter, λ is the wavelength in the material, and c is the velocity in the material.

The sound pressure decreases inversely to the distance z, because the sound energy is distributed to even larger areas. Figure 5 shows the principle characteristic of the sound field of a flat cirque transducer, which is simultaneously used as a sender and receiver (echo technique). The smallest beam diameter can be found at the near-field distance N; for large distances the field is opened by the divergence angle γ.

3.1.3 Focused transducer

By bending the active element (polyvinylidene fluoride (PVDF) piezoelectric polymer film) or by applying a length the sound field can be focused. Focusing increases the lateral resolution and provides the detection of small flaws. A focusing always reduces the near-field length.

Figure 6 shows the focal distances of the same transducer in water without test specimen (left-hand side) and with test specimen (right-hand side). The velocity in the test specimen is higher than that in water; as a result, refraction reduces the focal distance.

The length of the water path calculates equation (27) [12, 13]

$$l_w = F - l_m \cdot \frac{c_m}{c_w} \tag{27}$$

where l_w is the water path, l_m is the focal distance from the surface, c_m is the material sound velocity, c_w is the water sound velocity, and F is the focus in water.

3.1.4 Phased array transducers

Contrary to single-element transducers, phased array transducers (Figure 7) consist of 8–256 small

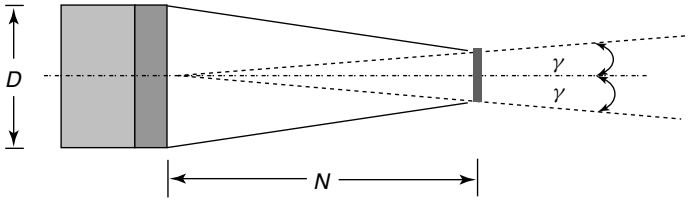


Figure 5. Sound field of a flat transducer.

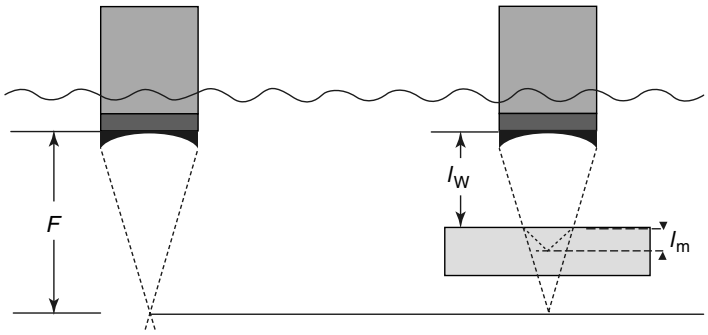


Figure 6. Focal distances in water and in material.

Figure 7. Phased array transducer (linear array with 64 elements, product of R/D-Tech/Olympus).

individual elements. Each element can be pulsed and received separately with sophisticated computer-based instruments.

The elements can be fired in groups where individual delays provide beam steering, sector scan, and linear scanning. The first application of these software-controlled transducers was in medical ultrasonic diagnostics. Because of the fast-developing electronic devices, computer systems, and the big drop of prices for hardware and software, the phased array technique is largely used for material testing [12–14]. More articles concerning phased array techniques can be found in the online journal NDT.net [8].

3.1.5 Coupling techniques

For a reproducible UT, it is important to have a constant coupling between the transducer and the component. Because of the high-impedance mismatch, air in the sound path would produce a reflection back to the transducer (see Section 6). For manual testing, a coupling paste or gel is used.

Immersion technique means that the component and the transducer are in a water tank (Figure 8). The transducer is moved by an XY scanning system, which provides automatic testing. An additional Z axis can be used for adjusting the water delay between the transducer and the component in order to set the focal point into the component. For curved components, additional axis is necessary for a normal incidence.

Squirter technique [15] is characterized by a plastic housing with a built-in transducer and a nozzle (Figure 9). The nozzle forms a guide for the water jet, which carries the ultrasound to the component.

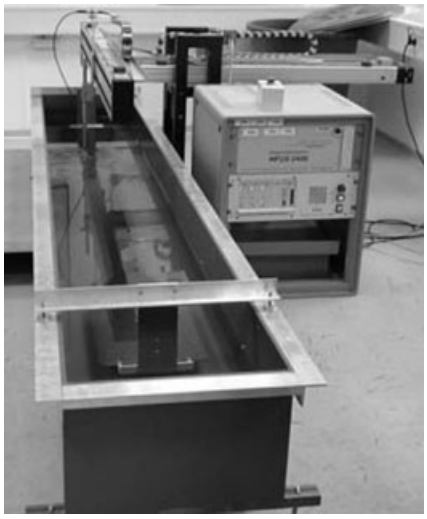


Figure 8. Immersion technique (DLR FA, Germany).

Figure 9. Inspection of a CFRP cylinder using squirter technique. (DLR FA, Germany).

Air bubbles and turbulences in the jet decrease the signal-to-noise ratio.

Water gap coupling also provides a reproducible coupling for automatic testing. The transducer is mounted on a carrier, which is guided by rollers on the surface of the component so that simply curved components can be tested with two-axis scanners. Figure 10 shows the testing of a curved carbon fibre reinforced plastic (CFRP) component with a water gap coupling device with rollers and a cardanic fixture.

3.2 Ultrasonic testing methods

3.2.1 Through-transmission technique

Through-transmission technique is useful for materials with high sound attenuation. This method is characterized by two transducers (one used as a transmitter, the other one used as a receiver) on opposite sides of the component. Contrary to the echo technique, the penetration of the material is only on time, but a double side's access of the component is necessary. An evaluation of the amplitude of the transmitted signal and a time-of-flight measurement are possible. The amplitude is dependant on the sound attenuation of the material.

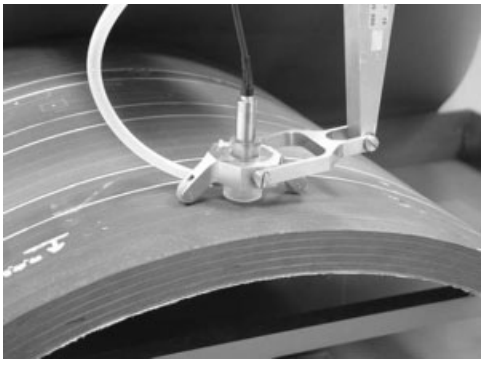


Figure 10. Water gap coupling for curved CFRP components, the transducer is mounted on a carrier with rollers and cardanic adapter to the scanning axis (DLR FA, Germany).

Defects principally produce an amplitude decrease and an increase in time of flight.

Measurements of the relative changes of amplitudes (A_1, A_2) caused by flaws or different attenuations can be calculated in dB (decibel):

$$\mathrm{dB} = 20 \log_{10}(A_1/A_2) \tag{28}$$

3.2.2 Pulse-echo technique

The most used testing method is pulse-echo technique, which provides a single-sided access. Only one transducer is necessary, which is simultaneously used as a transmitter and a receiver. This method is based on radar techniques and was a breakthrough for UT in the late 1940s.

Figure 11 demonstrates the echo technique. After a short excitation, the transducer converts the electrical signal to a longitudinal wave pulse, which propagates into the test object. The wave is reflected at the backwall of the transducer, which converts the acoustic wave back to an electrical signal, which is amplified and displayed. A defect in the test objects causes an additional reflection. Its time of flight (t_2) is smaller than that of the backwall (t_1). Time-of-flight measurements can be used for the determination of the material velocity (l: length of the test object):

$$v = 2l/t \tag{29}$$

and/or of the measurement of the distance d of the defect from the surface ("defect depth").

$$d = vt/2 \tag{30}$$

In dependence of the size and reflectivity of the defect, the amplitude of the backwall echo decreases.

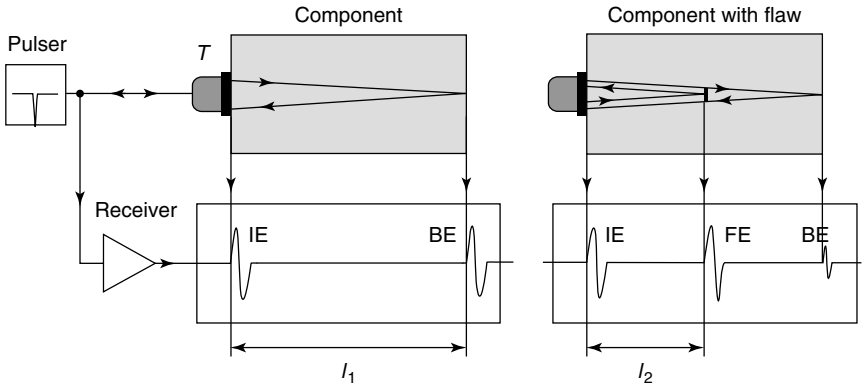


Figure 11. Principle of echo technique.

3.2.3 Immersion technique

Immersion technique is often combined with echo technique and a well-known method for automatic testing with ultrasonic imaging. The transducer and the test object are situated in the water tank (Figure 12).

The receiver signal displayed on the screen of the ultrasonic system consists of

1. interface echo
2. defect echo
3. backwall echo and
4. tank echo.

It should be noted that there are multiple reflections from the interface and, in some cases, form the defect(s), too. Therefore, the time of flight of the second reflection of the interface echo must be larger than the time of flight of the tank bottom echo.

For automatic testing and ultrasonic imaging, the echoes have to be evaluated automatically.

Therefore, different gates are used: gate A for the interface echo, B for the defect echo, C for the backwall echo, and D for the tank echo. The methods of imaging are described in Section 4.3.

It is also possible to use through-transmission technique in immersion technique. For this application, a u-shaped fixture for the two transducers is necessary.

4 AUTOMATIC ULTRASONIC SYSTEMS

4.1 Survey

Figure 13 shows the block diagram of an ultrasonic imaging system [16], with immersion technique

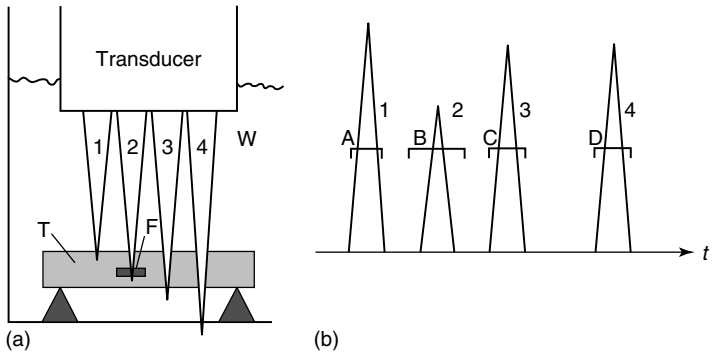


Figure 12. Principle of immersion technique: (a) test object (T) with transducer in water, F: defect and (b) echoes and gates versus time.

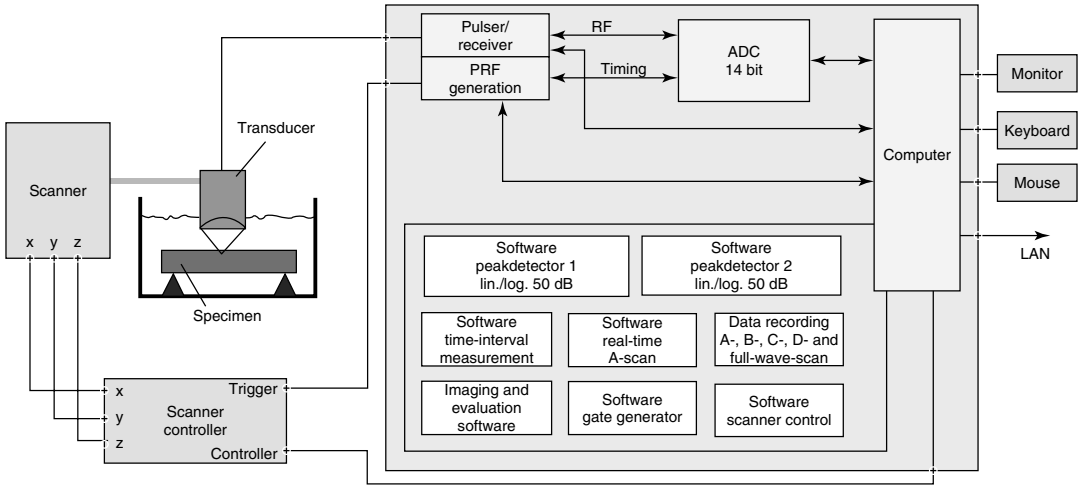


Figure 13. Block diagram of an ultrasonic imaging system.

consisting of a scanner with controller, immersion tank, a transducer with fixture, a pulser/receiver unit, an analog-to-digital converter (ADC), a timing unit for the trigger generation and a computer with ultrasonic software, which provides the control of the system, the data capture, procession, evaluation, and imaging.

4.2 Hardware

4.2.1 Pulser/receiver unit

The pulser/receiver unit consists of a pulser that generates the excitation for the transducer, an amplifier, a high- and a low-pass filter section, and a timing unit for the system trigger. This unit can be built on a Personal Computer PC-board [17]. Spike pulsers, which provide a broadband excitation and the shortest pulse responses, are the most used type of pulsers in ultrasonic systems. A capacity C is loaded and discharged by an electronic switch such as a field-effect transistor (FET) via a resistor R. With a short switching time of the FET, the frequency spectrum starts at 0 Hz and decreases (−3 dB) at

$$f = 1/2\Pi RC \tag{31}$$

Usually, the resistor is chosen near the impedance of the cable to the transducer (50−75 Ω), the capacity determines the maximum frequency. For high-frequency applications ($f > 20$ MHz), the FET should be replaced by an Avalanche transistor, which provides switching times lower than 1 ns. Also, for high-frequency application, the pulser/preamplifier unit should be built in a separate case and mounted nearby the transducer in order to get short cable to the transducer and to avoid electromagnetic noise (HFUS 2000 system [18]).

If a high-power excitation is required (for materials with a high sound attenuation, thick components, and others), a square pulser or a tone-burst pulser is the best solution. Such a device can also be a PC-board [18]. The energy of such a pulse is more concentrated on a smaller frequency range. Such kind of pulsers is useful for applications in a frequency range from 20 kHz up to 3 MHz.

The receiver consists of an amplifier, a gain setting and a high- and low-pass filter unit. Because of the large dynamic range of the input signals (>100 dB), the amplifier is separated in an input attenuator, a preamplifier, a fine-stepped attenuator and a main amplifier and a high- and low-pass filter unit. In order to get a constant sensitivity, for the inspection of thick components, a distance amplitude control unit (DAC) is intended.

The amplified signal is given to a high- and low-pass filter unit, which suppresses unwanted frequency parts and provides an antialiasing filter for the ADC board. The filter unit eliminates electromagnetic noise as well as the noise produced by ultrasonic scattering.

4.2.2 Analog-to-digital converter

An ADC converts the preprocessed analog receiver signal to a digital one (to a digital number). The digital signal cannot be influenced by noise and can easily be stored, digital processed, evaluated, and displayed by a computer system. There are two most important ADC features: the sampling frequency f_s (MHz) or sample rate (M samples per second) and the amplitude resolution (number of bits). The Shannon theorem says that for continuous signals, the sample frequency f_s must be larger than twice the input bandwidth B.

For ultrasonic applications, the sampling frequency is used up to 10 times larger than the maximum ultrasonic frequency in order to get a more precise amplitude and time-of-flight measurements. A low-cost ADC provides 8-bit amplitude resolution, 2^8 gives 256 levels of vertical resolution. Because of the bipolar ultrasonic (RF) signal, 128 steps are used for the positive amplitudes and 128 for the negative ones. An input range of 1 V_{pp} (peak to peak) provides a resolution of 3.906 mV.

Ultrasonic systems like USPC 3040 typically comprehend ADCs with 14 bits, which give 16 284 steps and 200-MHz sampling frequency. For example, if the maximum input voltage is 1 V_{ss}, the lowest step is 0.125 m $V_{ss} = 125\,\mu V_{ss}$, which means a dynamic range of about 78 dB. Therefore, it makes no sense to use high-gain preamplifiers, which only generate noise levels at their outputs much greater than the lowest amplitude step of the ADC. The higher the dynamic range, the better are the possibilities for digital signal processing after data capturing.

4.3 Software

The software not only controls the scanner and the ultrasonic hardware and displays the results but also replaces a lot of hardware. Examples are the gated peak detector (linear or logarithmic mode, 70-dB single-shot dynamic range) and the time-interval measurement system (resolution of 5 ns). These software solutions are more flexible than those of the hardware. The accuracy of these parts is given by the ADC. The user interface of the software is shown in Figure 14 [18]. The tap sheets on the left hand side provide the instrumentation settings; a large A-scan display is useful during the system setup. It can be separated into two parts in order to have an additional fast Fourier transform (FFT) display.

Figure 15 shows the possibilities for ultrasonic imaging. The A-scan presents the amplitude as a function of time (oscilloscope display of the echo signals), a B-scan gives a cross-sectional view of the component, whereas the amplitude is displayed along the vertical axis and the position of the transducer is displayed along the horizontal axis, a C-scan is a plan-type view or top view of the component with all flaws (two-dimensional amplitude presentation). For a C-scan, the amplitude of the flaw echo or the amplitude of the backwall echo has to be evaluated. Therefore, a precise gate setting is required in which the amplitude measurement is carried out. A D-scan also provides a plan-type view; however, contrary to a C-scan, the time of flight is displayed. A D-scan is able to display the depths of the flaws or the thickness of the component.

Figure 16 demonstrates the possibilities of ultrasonic imaging using an impacted CFRP panel. The C-scan of the defect echo is presented in Figure 16(a). For this presentation, the amplitude in gate B situated between the interface and the backwall echo is evaluated in a logarithmic scale (dB). The highest amplitude (100% in an A-scan) is calibrated to 0 dB and mapped to the color black in the C-scan. A defect is characterized by a higher amplitude than the "natural noise" caused by scattering at inhomogeneities. The black area indicates the delaminated area. A C-scan of the backwall echo is given in Figure 16(b). This one presents the "acoustic shadow" of the defect. Therefore, a defect causes smaller echo amplitudes than its surroundings. The delaminated area is indicated by the lighter region in the C-scan.

A D-scan of the defect echo shown in Figure 16(c) presents the different distances of the delaminations from the surface. The black area presents a distance of 0 mm (no defect) and white areas nearly 4 mm. It is shown that the damaged area consists of different delaminations in different depths, which means that they are situated between different layers. This information can also be received from a B-scan (Figure 16d), which presents a view into the material along a selected line out of a C-scan. Not only the different depths of the delaminations are indicated but also horizontal lines, which present layer echoes.

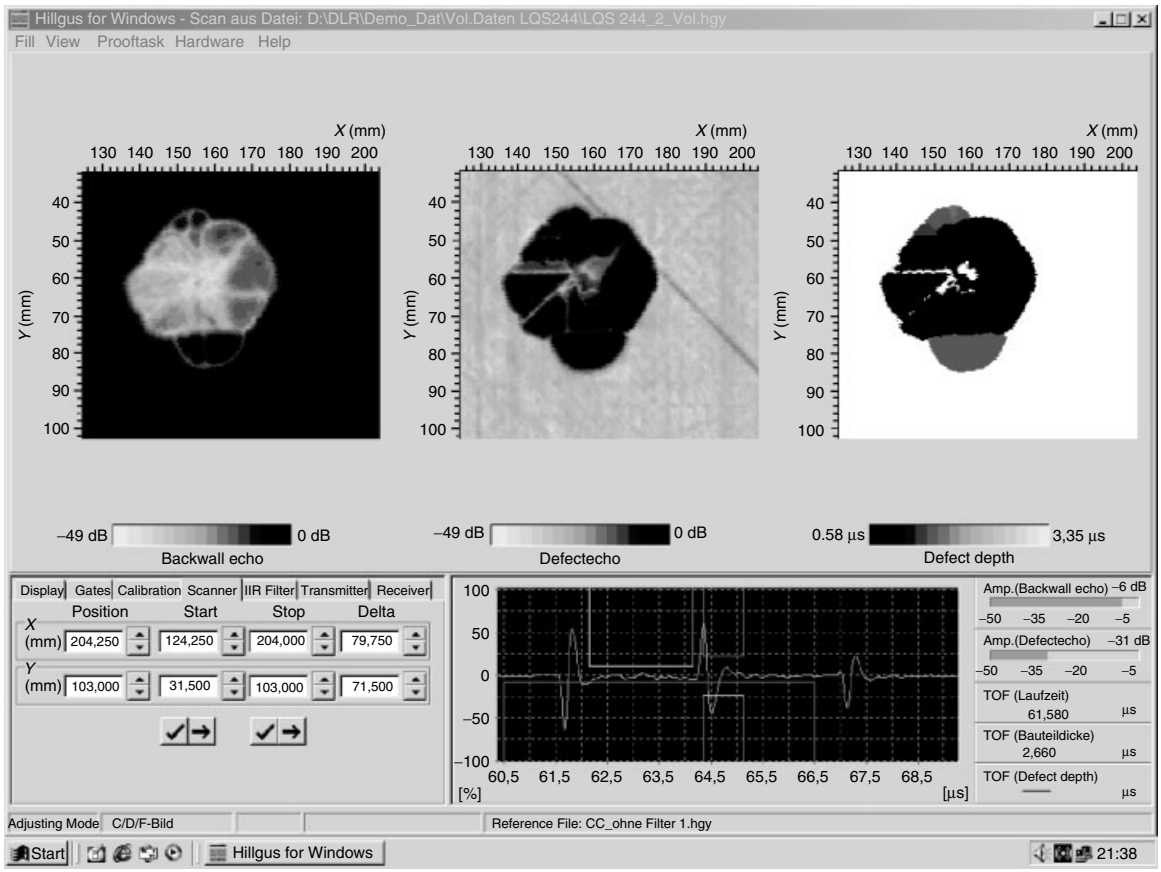


Figure 14. User interface of the software Hillgus for windows.

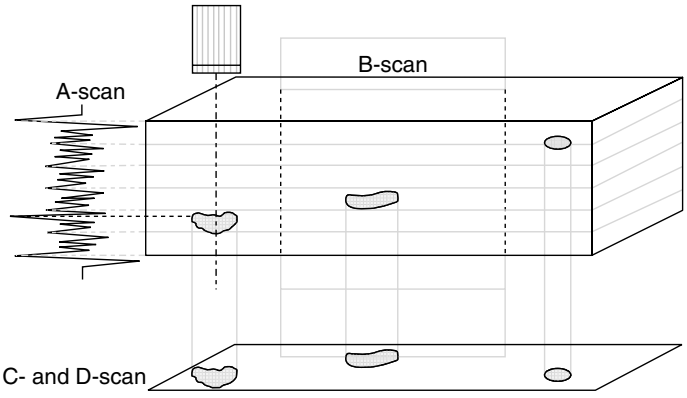


Figure 15. Ultrasonic imaging techniques: A-, B-, C-, and D-scans.

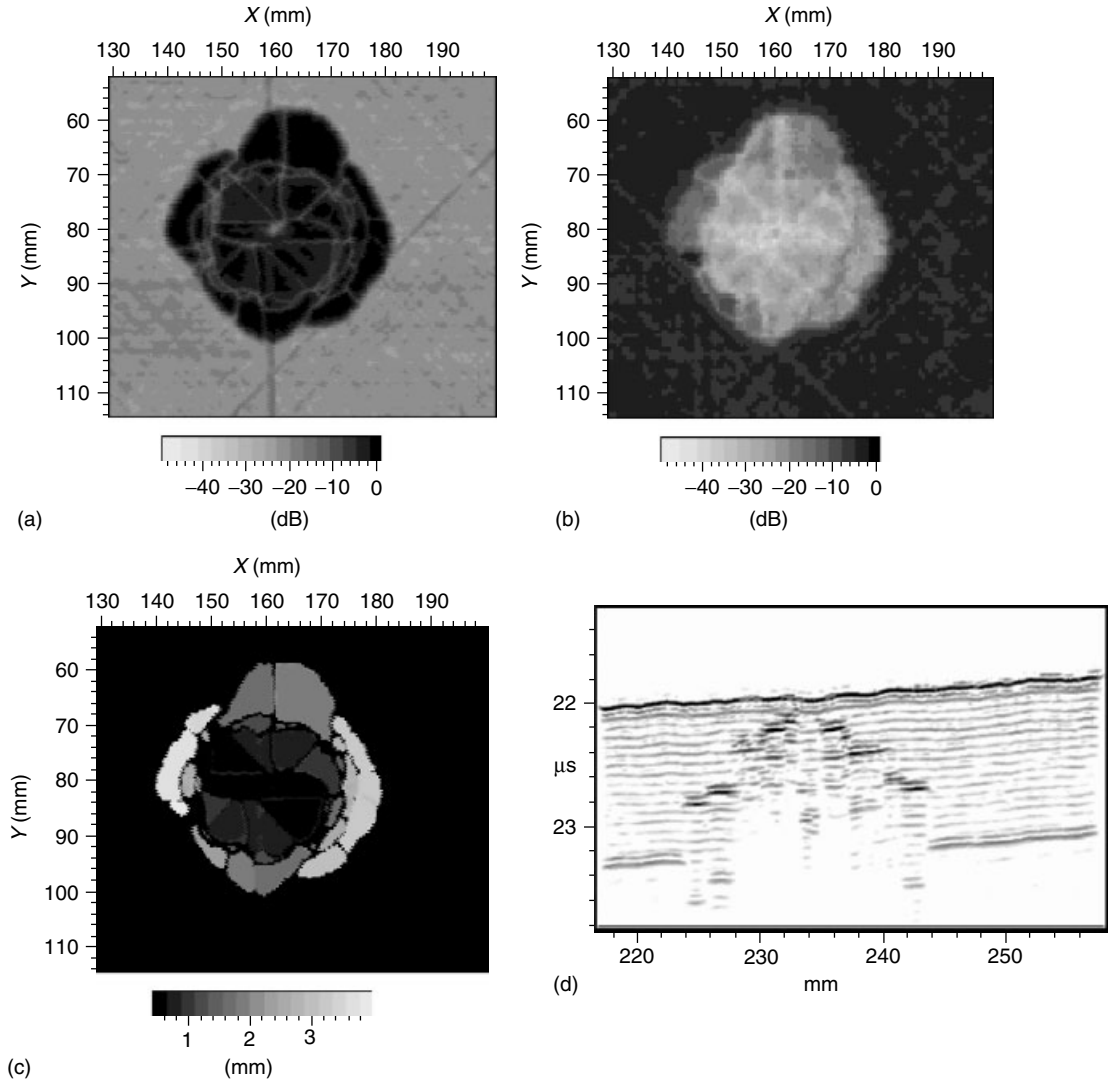


Figure 16. Ultrasonic imaging of an impacted CFRP panel. (a) C-scan defect echo, (b) C-scan backwall echo, (c) D-scan defect echo, and (d) B-scan.

These indications require a high-quality laminate and a well-optimized UT.

5 MOBILE ULTRASONIC INSPECTION SYSTEM "MUSE"

For in-service inspections of built-in aerospace components, a mobile equipment is necessary.

The manual scanning is time consuming and a reproducible coupling is difficult to achieve. Therefore, the mobile ultrasonic inspection system "mobile ultrasonic equipment (MUSE)" has been developed [19]. The manipulator is attached to the component with sucking devices. This system shown in Figure 17 consists of a motor-driven manipulation system, a water-circulation system for the coupling, and the PC-board-based ultrasonic system USPC built into a portable computer. The manipulator is attached to

MUSE manipulator with adapter for transducer | Water circulation system | Flaw detector with built-in motor controller

Figure 17. MUSE system for in-field inspections.

the component with sucking devices and provides an inspection area of 200–300 mm^2, a resolution of 0.020 mm and a maximum speed of 200 mm s^{-1}. The test frequency range for the MUSE is 0.05–35 MHz. Ultrasonic imaging is possible in A-, B-, C-, and D-scans.

The coupling technique developed for the MUSE ("local immersion technique") provides the application of focused transducers without problems with air bubbles so that the high resolution is also available for field inspections. The local immersion technique is also useful for laboratory investigations because no water tanks are required and the components stay dry.

6 AIR-COUPLED TECHNIQUES

Air-coupled ultrasonic testing (AC-UT) is a very attractive noncontact technique, which avoids the disadvantages of the coupling liquid (water, squirter, and immersion technique) or coupling paste [20–22]. However, the large impedance mismatch between solids and gas (air) produces an amplitude loss of more than 150 dB using through-transmission technique with separate standard transducers as a transmitter and a receiver on the opposite sides of a CFRP component [19].

Figure 18 shows the amplitude differences between water and air coupling using through-transmission technique with different transmitter and receiver transducers on the opposite sides of a thin CFRP specimen. We set the amplitude on the transmitter to 0 dB. Using water coupling, the sound pressure decreases to −21 dB in the water because of the differences in the acoustic impedance of piezoelectric material ($z_{\text{piezo}} = 35$ MRayl) and water ($z_{\text{water}} = 1.5$ MRayl). The amplitude on the CFRP surface ($z_{\text{CFRP}} = 4.5$ MRayl) increases to −18 dB and decreases again to −24 dB in the water. At least the amplitude on the receiver becomes −18 dB. This small amplitude difference between transducer and receiver can easily be compensated by amplification.

In the case of air coupling, the amplitude difference between transmitter and receiver reaches −156 dB. The extreme high differences between the impedance of air ($z_{\text{air}} = 0.0004$ MRayl) and of the lead zirconate titanate (PZT) material cause a decrease of −93 dB and also the interface between CFRP and air (−75 dB). In this very simple model, only the transmission factors are calculated and no losses of sound attenuation and sound divergences have been regarded.

Therefore, standard flaw detectors with standard transducers cannot be used for air coupling. Special transducers, transmitter and receiver electronics are required. Since 15 years, more papers have been describing the air-coupled technique [6]. Normally, the investigations are carried out in through-transmission technique.

In order to get better signal-to-noise ratio special transducers, an optimized excitation, ultralow noise preamplifier, and a signal processing are necessary. An example of a modular system is the USPC 4000 AirTech [15], which was used for the recording of the right hand side C-scan in Figure 6.

Figure 19 shows two C-scans of an impacted sandwich component consisting of two CFRP skins each 0.5 and 15-mm-thick honeycomb cores. The C-scan in Figure 19(a) was recorded in echo technique using a broadband transducer (2.2 MHz) with water gab coupling. In order to get a high signal-to-noise ratio, a square-pulse transmitter is used. The software receiver filter provides a center frequency of the back-wall echo (back skin) of 700 kHz. This frequency enables the detection of defects in the skins and core, demonstrated by the impact indication in the C-scan [8].

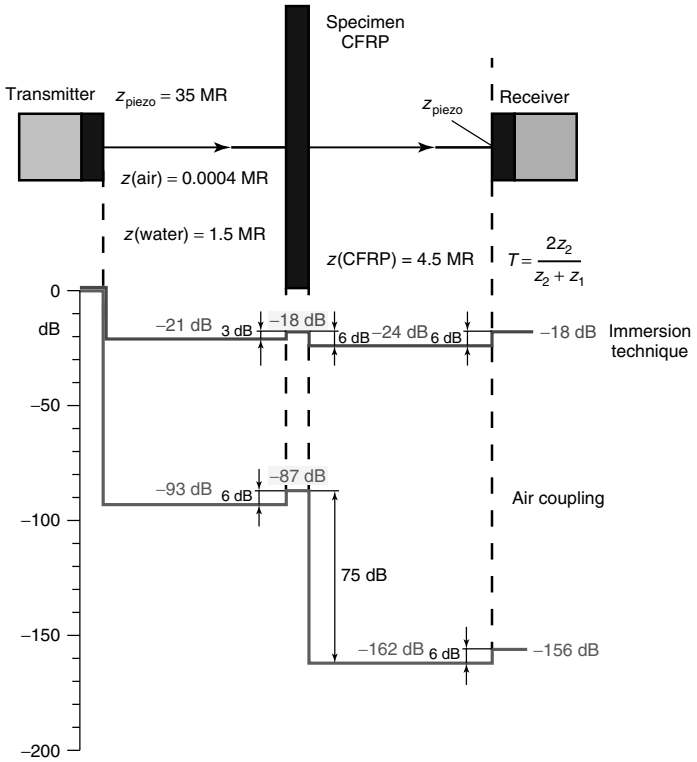


Figure 18. Sound pressure curves between transmitter and receiver probe using water and air coupling.

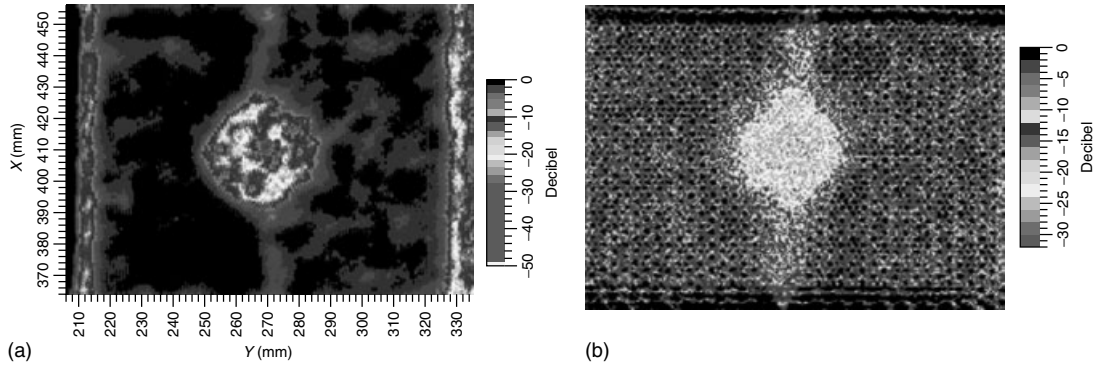


Figure 19. C-scans of an impacted CFRP sandwich component. (a) Echo technique with water gab coupling and (b) air-coupled ultrasonic testing.

The C-scan in Figure 19(b) of the same component is carried out by air coupling and gives a much clearer presentation. The reason for the better resolution using air coupling is the smaller velocity in air in comparison with water, which provides a smaller wavelength and—using focused transducers—a smaller beam size.

7 VISUALIZATION OF LAMB WAVE PROPAGATION

The knowledge of the propagation of different Lamb wave modes and their complex interactions with structure elements, discontinuities, and defects is a requirement for interpreting the complex receiver

signals and for optimizing all the parameters of Lamb wave system such as test frequency, design, number, and distribution of the transducers, required signal procession and evaluation. Only for simple geometries, the Lamb wave propagation can be calculated. Therefore, experimental methods are used in order to get this important information.

One actuator at a fixed position on the bottom of the specimen has been used as a transmitter (Figure 20) [23, 24]. The excitation is carried out by a burst generator. Because of the harmonics a filter on the receiver side is used, which provides a desired narrow band signal.

A second PZT patch is moved by an XY scanner in a meander track. Using a two-axis scanning system, the curve of the component can be compensated by a cardanic sensor holder. The scanning grid of 1.5 mm is used, which is smaller than a tenth of the Lamb wavelength. In order to get a reproducible acoustic coupling between the sensor and the component, a water split coupling is used.

During scanning at each point of the scanning grid, a full-wave Lamb wave A-scan is recorded. This 3-D recorded data file consists of a number of B-scans along the y axis (scanning axis). In x direction the index axis is situated, the time of flight in z direction. A scanning area of 400 mm $\times$ 500 mm with a scanning grid of 2 mm delivers the information of more than 50 000 virtual sensors.

Out of the 3-D- data files several presentations can be calculated and presented [23]:

- 2-D amplitude images, showing the maximal amplitude during a selected gate $[t_1, t_2]$ (C-scan);

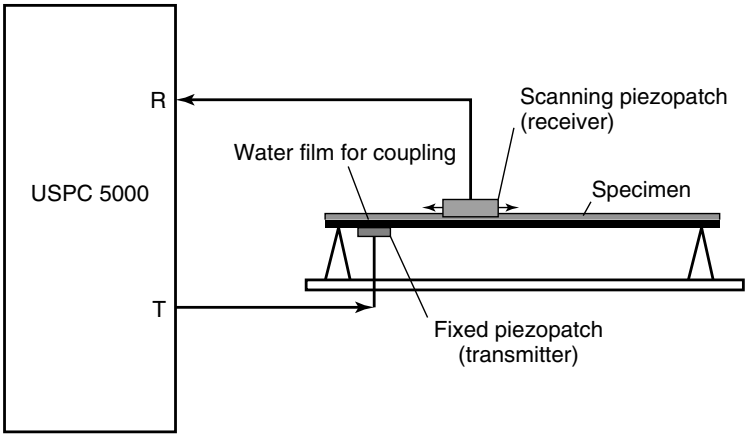


Figure 20. Experimental arrangement for the visualization of the Lamb wave propagation.

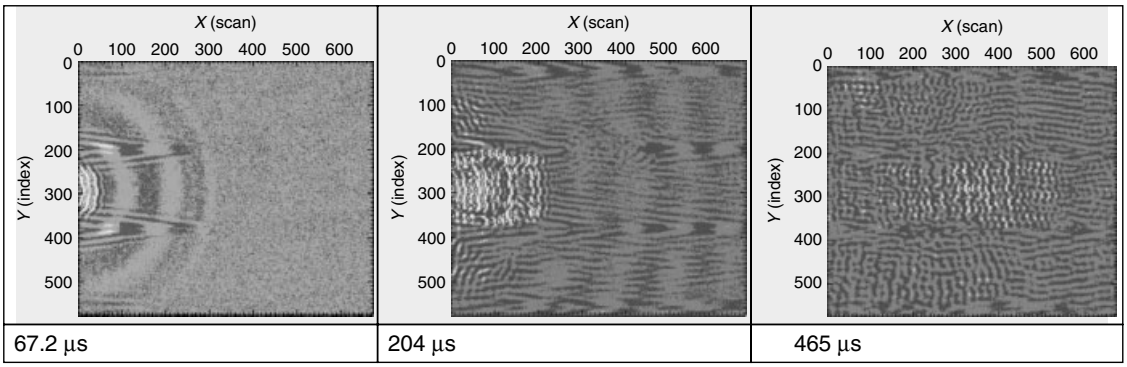


Figure 21. Video snap shots with excitation of PZT T4 situated on the skin between two stringers.

- 2-D time-of-flight images within a certain gate $[t_1, t_2]$ ("D-scan");
- Lamb wave A-scans at any recorded position together with spectral analysis;
- B-scans $[x, t]$, $[y, t]$; and
- Video animations showing the wave propagation.

Figure 21 presents three snap shots of a video animation with excitation of T4 situated on the skin. There are two wave modes generated on fast mode with a wavelength of 105 mm (s_0) propagating mainly on the stringers with a phase velocity of 5250 m s^{-1}. The wavelength of the slower mode (a_0) propagating between the stringers was measured to 15 mm with a phase velocity of 750 m s^{-1}.

The circular propagation in the skin is influenced by reflections from the stringers, so that after a time of flight of about 100 μs more interferences are indicated. After 465 μs, a chaotic wave field is visible.

REFERENCES

[1] Sokolow SJ. Zur Frage der Fortpflanzung ultraakustischer Schwingungen in verschiedenen Körpern (Propagation of ultrasonic waves in different components). *Elektrische Nachrichten-Technik* 1929 **6**:454–461.

[2] Mclntire R. *Akustische Zeitschrift* 1938 **3**:132–136.

[3] Krautkraemer J, Krautkraemer H. *Ultrasonic Testing of Materials*. Springer, 1990.

[4] American Society for Nondestructive Testing. *Nondestructive Testing Handbook*, *Vol. 7, Ultrasonic Testing*, 1991.

[5] Lynnworth LC. *Ultrasonic Measurements for Process Control*. Academic Press, 1989.

[6] Bergmann L. *Der Ultraschall (The Ultrasound)*, 6. Auflage. Hirzel: Stuttgart, 1954.

[7] http://www.ndt-ed.org/index_flash.htm.

[8] http://www.ndt.net/.

[9] Lamb H. On waves in an elastic plate. *Proceedings of the Royal Society of London, Series A* containing papers of mathematical and physical character 1917 **39**(651):114–129.

[10] Park HW, Sohn H, Law KH, Farrar CR. Time reversal active sensing for health monitoring of a composite plate. *Journal of Sound and Vibration* 2007 **302**:50–56.

[11] Akashi T. On the measurement of loss in ultrasonic pulse in concrete. *Proceedings of 2nd Japan Congress Testing Materials*. STM, Kyoto, 1959, pp. 165–168.

[12] http://www.olympusndt.com/data/File/panametrics/panametrics-UT.en.pdf.

[13] Wüstenberg H, Erhard A, Schenk G. Scanning modes at the application of ultrasonic phased array inspection systems. *WCNDT Roma 2000*. Roma, 2000, Conference Proceedings, also available on: http://www.ndt.net/article/wcndt00/papers/idn193/idn193.htm.

[14] Maurer A, Haase W, De Odorico W. Phased array application in industrial scanning systems. *ECNDT 2006*. Berlin, 2006, paper Mo.2.1.3, Conference Proceedings on CD, also available on: http://www.ndt.net/article/ecndt2006/doc/Mo.2.1.3.pdf.

[15] Deutsch WAK, Schulte P, Joswig M, Kattwinkel R. Automatic inspection of welded pipes with ultrasound. *ECNDT 2006*. Berlin, 2006, paper Tu.2.3.1, Conference Proceedings on CD, also available on: http://www.ndt.net/article/ecndt2006/doc/Tu.2.3.1.pdf.

[16] Hillger W. Ultrasonic systems for imaging and detection. *19th International Congress on Acoustics (ICA 2007)*. Madrid, 2–7 September 2007, Special Issue of the journal *Revista de Acoustica*, 2007, **38**, ISBN: 84-87985-12-2.

[17] Hillger W. Ultrasonic PC- boards for different applications. *7th ECNDT 1998, Conference Proceedings*. Copenhagen, 26–29 May 1998; pp. 3113–3119.

[18] http://www.Dr-Hillger.de.

[19] Hillger W. Ultrasonic testing of composites—from laboratory research to in-field inspections. *15th World Conference on Non-destructive Testing*. Roma, 15–21 October 2000, Conference Proceedings on CD WCNDT Rom 2000.

[20] Grandia WA, Fortunko SM. NDE applications of air-coupled ultrasonic transducers. *1995 IEEE International Ultrasonic Symposium, Conference Proceedings*. Seattle, WA, 1995; pp. 697–709.

[21] Bhardhwaj MC. High efficiency non-contact transducers and very high coupling piezoelectric composite. *16th World Conference on Non-destructive Testing*. Montréal, 30 August–3 September 2004, Conference Proceedings on CD.

[22] Hillger W, Meier R, Henrich R. Inspection of CFRP components by ultrasonic imaging with air coupling. *Insight* 2004 **46**(3):147–150.

[23] Hillger W, Pfeiffer U. Structural health monitoring using lamb waves. *9th European Conference on Non-destructive Testing*. Berlin, 25–29 September 2006, published on CD.

[24] Hillger W. Visualisation and animation of the lamb wave propagation in composites. *Proceedings of the 6th International Workshop on Structural Health Monitoring 2007*. Stanford University, Stanford, CA, 11–13 September 2007; Vol. 1, pp. 145–152.

Chapter 16
Guided-wave Array Methods

Paul D. Wilcox
Department of Mechanical Engineering, University of Bristol, Bristol, UK

1 INTRODUCTION

Ultrasonic array transducers for bulk waves were developed originally for medical imaging applications in the 1950s. In the last few decades, this technology has been taken up in the nondestructive evaluation (NDE) community as a superior alternative to conventional ultrasonic techniques using monolithic transducers. The enabling technology that has lead to the rapid growth in array-based systems for NDE has been in digital electronics and data processing in the supporting instrumentation, rather than the fabrication of the arrays themselves. In NDE, there are a number of different reasons for using arrays. First, they enable multiple conventional inspections (e.g., with ultrasonic probes at different angles) to be performed simultaneously, hence reducing inspection time and possibly inspection complexity. Second, they enable some form of scanning to take place without any physical movement of the device. This can be exploited to scan the interior of a complex component from a range of angles from one transducer position or to reduce the number of mechanical scanning dimensions in a raster scanning system. These two reasons for using arrays are essentially concerned with reducing the cost of NDE by performing an inspection more efficiently with an array. The third reason is somewhat different and is only just starting to be realized. This is the fact that arrays can potentially provide much richer information about a component than conventional NDE. A comprehensive review of the current state of the art in array-based NDE may be found in [1].

Digital electronics and computational power have also been key enabling technologies for long-range guided-wave NDE systems using arrays. The most commercially successful of these have been for the inspection of pipelines. Such systems allow the rapid inspection of long lengths of pipeline from a single access point and are now in widespread industrial use throughout the petrochemical industry and elsewhere. A review of applications may be found in [2]. Despite the success on pipeline inspection, there has been

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

very little commercial use of guided-wave NDE systems for large area inspection to date, although the technology has been demonstrated in research laboratories. Guided-wave NDE systems are based on the concept of a device that can be deployed on a structure when inspection is required. In principle, the same devices could be permanently attached to a structure to form the basis of a structural health monitoring (SHM) system, but this is not necessarily an economic solution because of the complexity of the arrays used. It is fair to say that guided-wave arrays specifically designed to be permanently attached for use in an SHM system are still a long way from widespread industrial usage.

This article begins with a brief overview of the salient features of guided waves and their application to damage detection. It then examines the various types of guided-wave arrays in use in NDE applications, beginning with those for one-dimensional (1D) waveguides such as pipes. The article compares and contrasts these with devices to perform inspection on two-dimensional (2D) platelike structures. The final section is on sparse distributed arrays specifically for SHM applications.

2 BACKGROUND

2.1 Guided-wave terminology

Guided waves are elastic waves that propagate along structures, their energy guided by the boundaries of the structure. In guided-wave terminology, the structure is referred to as the *waveguide*, and guided waves include waves on free surfaces (Rayleigh waves), waves in plates (Lamb waves and shear-horizontal or SH waves), waves in rods, and many others. Guided-wave phenomena exist over a wide frequency range from $<$1-Hz seismic waves generated by earthquakes to $>$100-MHz surface waves used in acoustic microscopes. There is sometimes debate about whether guided waves or bulk waves exist in a particular situation, but this is a misleading argument. The actual wavefield that exists is the solution that satisfies the governing partial differential equations and boundary conditions. The guided or bulk wave distinction is about the most appropriate way to formulate the solution. If the wavelength of bulk waves is small compared with the dimensions of the structure then the bulk wave formulation is most appropriate, but if the wavelength is comparable or greater than the dimensions of the waveguide a guided-wave solution is generally more useful. The ratio of waveguide dimension to bulk wavelength is therefore an important parameter, and for this reason, the various properties of guided waves in a particular shape of waveguide are often considered as functions of the product of a characteristic waveguide cross-section dimension and frequency (since this is inversely proportional to the bulk wavelength). Further details on the physics of guided waves may be found in **Fundamentals of Guided Elastic Waves in Solids** and **Modeling of Lamb Waves in Composite Structures**.

Key characteristics of guided waves are the presence of multiple modes of propagation, which generally have frequency-dependent (i.e., dispersive) velocities. Associated with each mode at each frequency is a characteristic pattern of particle displacements and stresses through the cross section of the waveguide that is referred to as the *mode shape*. By definition, the mode shape of a particular guided-wave mode at a particular frequency is invariant as the wave propagates along the waveguide, although the amplitude may diminish because of attenuation.

The interest in the use of guided waves for long-range NDE and SHM is their potential for detecting damage remote from the location of a sensor, an essential attribute if a limited number of sensors are required to detect damage at any location in a large structure. From a guided-wave perspective, structures can be divided into 2D waveguides where the structural elements are platelike (e.g., aircraft fuselages, ships' hulls, storage vessels, etc.) and 1D waveguides (e.g., pipes, rails, etc.). The two types are shown schematically in Figure 1.

In 2D platelike structures, the guided waves are Lamb and SH modes, and the salient dimension is the plate thickness. To give an idea of the numbers involved for typical engineering materials (steel, aluminum, many composites), a typical operating point for an SHM or NDE system is around 1 MHz mm (i.e., 1 MHz in a 1-mm-thick plate, 500 kHz in a 2-mm-thick plate, etc.). Up to around 2 MHz mm, only three fundamental modes (the A_0 and S_0 Lamb modes and the fundamental SH mode, SH_0) exist in platelike structures of such materials; above this value, higher order guided-wave

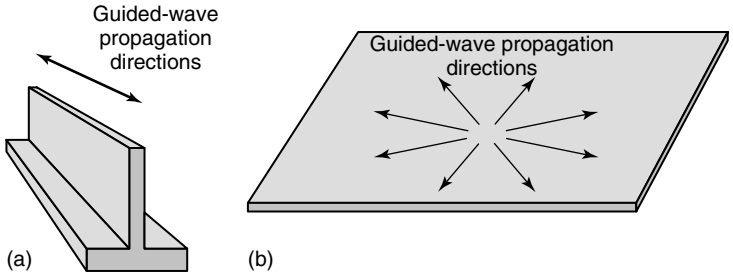


Figure 1. Schematic diagrams of typical (a) 1D and (b) 2D waveguides.

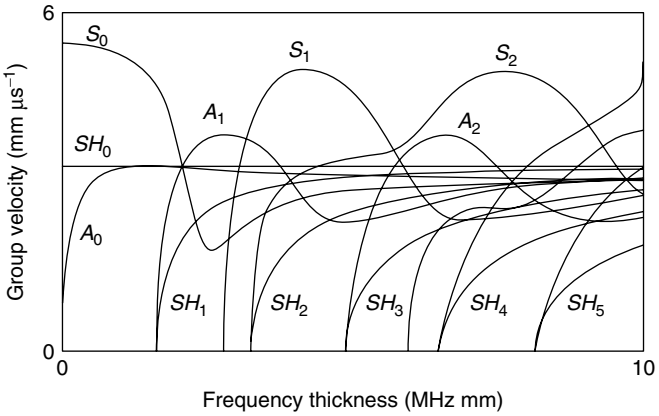


Figure 2. Group velocity dispersion curves for steel plate.

modes begin to appear and by 10 MHz mm there are some 15 possible modes. Example group velocity dispersion curves for steel are shown in Figure 2.

The usable propagation distances of guided waves for SHM or NDE in 2D waveguide structures are of the order of meters. The waves will actually propagate much further than this, but the complexity introduced by interactions with structural features make signal interpretation increasingly difficult. In 1D waveguide structures, such as pipes and rails, long lengths of uninterrupted structure may be present and interpretable guided-wave signals may be obtained even after propagation over distances of tens or even hundreds of meters.

2.2 Overview of guided-wave arrays

2.2.1 Monolithic devices versus arrays

The reason for using arrays for guided-wave testing is to allow waves from different modes and different directions to be separated, with the ultimate goal of localizing and characterizing damage by the guided waves scattered from that damage. Modal and directional selectivity can be achieved by various monolithic devices, including wedge transducers [3], electromagnetic acoustic transducers (EMATs) [4, 5], interdigital transducers (IDTs) [6], and comb transducers [7]. Examples of some of these devices are illustrated in Figure 3. The limitation of these devices is that they must be mechanically moved to obtain selectivity in different directions and need to be physically altered to obtain selectivity to different modes. The advantage of a correctly designed array is that this functionality can be obtained in postprocessing without any physical change to the array or its position.

2.2.2 Compact versus sparse distributed arrays

A distinction should be made between compact arrays where all the elements are in close proximity to

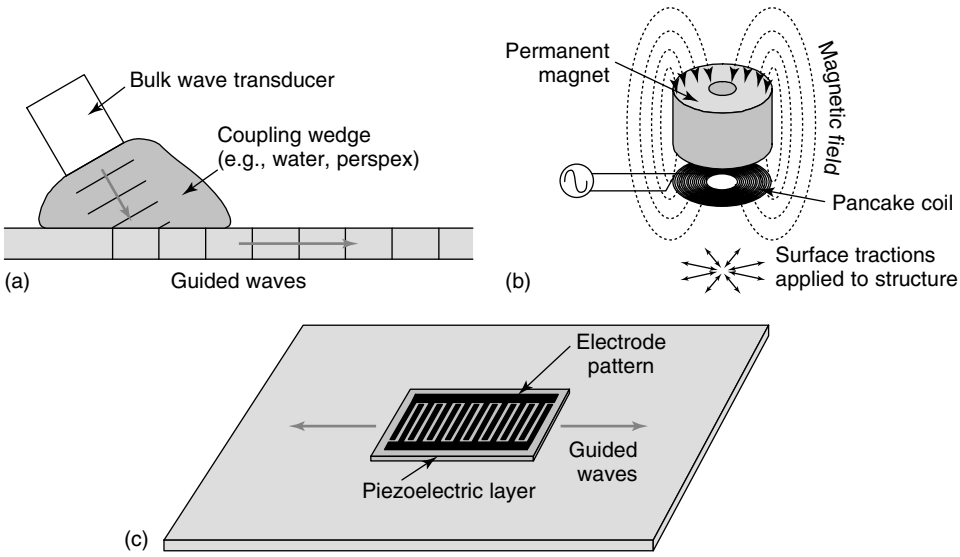


Figure 3. Monolithic transducers for guided waves: (a) wedge transducer, (b) EMAT [Reproduced from Ref. 5. © IEEE, 2005.] and (c) IDT. [Reproduced from Ref. 6. © Elsevier, 1997.]

one another, often in a single unit, and sparse distributed arrays. Deployable systems for NDE are generally compact arrays, while permanently attached systems may be either. The reason for using a compact array in a deployable system is to allow the reflected signals from multiple structural features to be resolved spatially. Without this attribute, it is generally impossible to reliably separate signals from damage, other than in structures with very low feature density or in laboratory experiments, where the location of damage can be controlled. In the case of a permanently attached system, the situation is potentially somewhat different, since subsequent data can be directly compared with the reference data. For this reason, the use of sparse distributed arrays is possible and may be a more economic solution than permanent attachment of compact arrays. Sparse distributed arrays are discussed separately in Section 5.

2.2.3 General requirements for compact arrays

To localize and characterize damage with a compact array, it is necessary to separate different guided-wave modes propagating in different directions. To achieve modal or directional selectivity, it is necessary to find a distinguishing guided-wave property that can be measured. Fundamentally, this must be one or more of the phase velocity, group velocity, or mode-shape properties. Differences in group velocity correspond to differences in arrival time and have limited use in deployable systems where there is often no *a priori* information about the location of scatterers. Measurements of differences in mode shape require access to different points in the cross section of the waveguide and are therefore generally not applicable to 2D waveguides, but are used extensively for mode separation in 1D waveguides. The measurement of differences in phase velocity (or wavenumber) is the fundamental basis of directional selectivity in guided-wave arrays. Phase velocity differences can also be used for modal selectivity if the phase velocities of the guided-wave modes are well separated.

The requirements and means of achieving modal and directional selectivity with compact arrays are somewhat different in the case of 1D and 2D waveguides as will be discussed in the following sections.

3 COMPACT ARRAYS FOR 1D WAVEGUIDE STRUCTURES

In a 1D waveguide, the number of possible propagation directions for guided waves is precisely two.

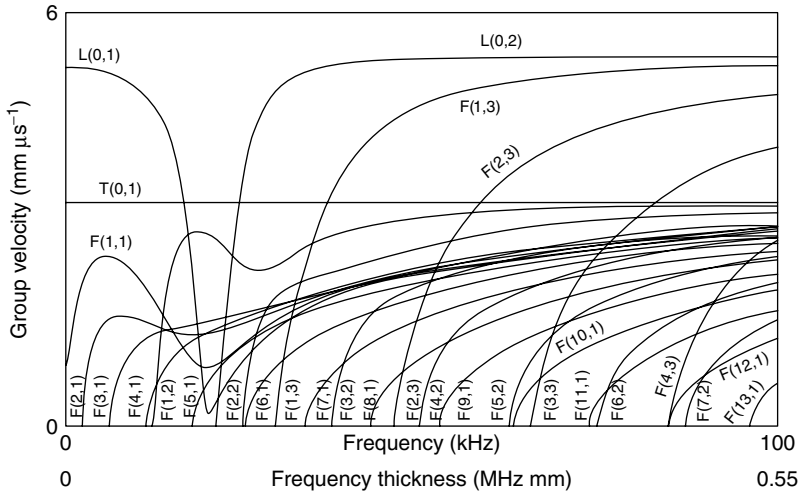


Figure 4. Group velocity dispersion curves for a 5.5-mm-thick steel pipe with internal diameter 76 mm, and also showing the equivalent frequency-thickness axis. [Reproduced from Ref. 9. © Elsevier, 1998.]

However, at useful operating frequencies for damage detection the number of possible guided-wave modes is high. For example, Figure 4 shows the dispersion curves for a 5.5-mm-thick steel pipe with an internal diameter of 76 mm. Although the frequency-thickness range in Figure 4 is much lower than that of the dispersion curves for steel plate shown in Figure 2, there are now many more modes. This is because of the additional constraint of periodicity around the circumference of the pipe, leading to the concept of angular order. A zero angular order mode has an axisymmetric and a first angular order mode has mode shape that is a sinusoidal distribution of displacements around the circumference with one period. In this article, the naming convention used for guided-wave modes in a pipe follows that described in [8]. The first numerical index refers to the order of the periodicity of the mode shape in the angular direction and the second refers to the through-thickness order of the mode. The prefix letters L and T are used to highlight axisymmetric (i.e., zero angular order) modes with through-thickness mode shapes similar to Lamb (L) or SH (T) type waves in a flat plate and the prefix letter F is used for all nonaxisymmetric modes.

The key to producing signals that can be readily interpreted is to achieve both modal and directional selectivity and this attribute is shared by all commercially successful guided-wave systems for NDE. In a 1D waveguide, directional selectivity is achieved by exploiting differences in phase velocity and modal selectivity by exploiting differences in phase velocity and/or mode shapes. In a 1D waveguide, differences in mode shape can be easily exploited as an array can contain multiple elements distributed around the perimeter of the waveguide cross section. In fact, the task of obtaining modal and directional selectivity can be generalized, but in the first instance the two are considered separately for the example case of a pipe.

3.1 Modal selectivity

The mode shapes of guided waves in pipes are characterized by various parameters, including the angular order. Of great importance in guided-wave NDE is the separation of modes with different angular orders, most commonly zero and one, as this provides valuable information about the nature of reflectors.

Consider a pipe instrumented with an array of four point transducers, identified by subscripts $j = 1-4$, evenly distributed around the outer perimeter as shown in Figure 5(a).

An angular datum is defined at the 12 o'clock position. As an example, it is assumed that under a particular set of operating conditions (i.e., frequency, transducer element type) only two modes of guided-wave propagation are possible, the first M_0, denoted

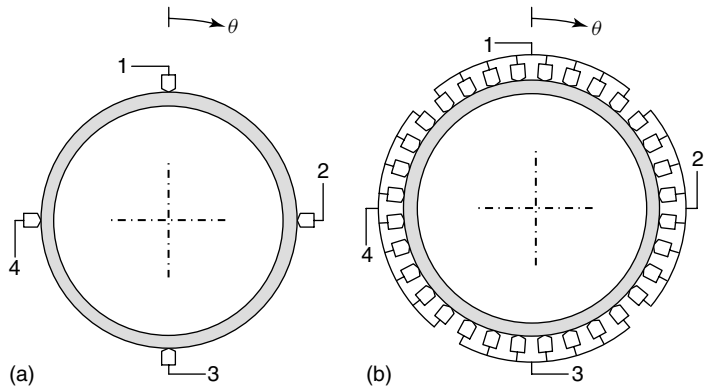


Figure 5. Schematic diagrams of pipe transducer arrays for modal selectivity based on angular order: (a) array of four transducers and (b) array of 32 transducers connected in quadrants.

by (0) superscripts, with angular order zero and the second, M_1, denoted by (1) superscripts, of angular order one. At a particular instant, the incident field on the array consists of quantities $A^{(0)}$ and $A^{(1)}$ of the two modes. The modal amplitudes are assumed to be defined in such a way that the particle displacement associated with each of the modes at a particular angular position, θ, around the circumference is expressed as

$$\begin{aligned} u^{(0)} &= A^{(0)} \\ u^{(1)} &= A^{(1)} \cos\left(\theta + \phi^{(1)}\right) \end{aligned} \tag{1}$$

where $\phi^{(1)}$ is the polarization angle of the mode M_1 (i.e., the angle at which its peak displacement occurs) with respect to the angular datum. Hence the associated particle displacement seen by the jth transducer in the array is

$$u_j = A^{(0)} + A^{(1)} \cos\left(\theta_j + \phi^{(1)}\right) \tag{2}$$

where $\theta_j = (j-1)\pi/2$ is the angle of the jth transducer. The objective of mode separation is to deduce the mode amplitudes $A^{(0)}$ and $A^{(1)}$ from the measured displacements u_j. In this case, the natural orthogonality of the different angular periodicity of M_0 and M_1 can be exploited. This means that the mode amplitudes can be extracted effectively from the first two terms of the discrete Fourier transform performed around the circumference with four sampling points at the transducer locations:

$$\begin{aligned} A^{(0)} &= \frac{1}{4}\sum_{j=1}^{4} u_j \\ A^{(1)} &= \frac{1}{2}\sqrt{\left(\sum_{j=1}^{4} u_j \cos\theta_j\right)^2 + \left(\sum_{j=1}^{4} u_j \sin\theta_j\right)^2} \end{aligned} \tag{3}$$

The two terms under the square root sign in the second expression relate to two perpendicular polarizations of the M_1 mode, the first term polarized in the vertical (0–180°) direction and the second polarized in the horizontal (90–270°) direction. For axisymmetric structures such as pipes, this type of Fourier approach can be extended to separating higher angular orders of modes in a straightforward manner. However, for waveguides with more complex cross sections, a more general approach is necessary. With this in mind, situation in the previous example can be described in matrix form:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} A^{(0)} \\ A^{(1v)} \\ A^{(1h)} \end{bmatrix} \tag{4}$$

where $A^{(1v)}$ and $A^{(1h)}$ refer to the components of the M_1 mode in the vertical and horizontal polarization directions and hence its overall amplitude is

$$A^{(1)} = \sqrt{\left[A^{(1h)}\right]^2 + \left[A^{(1v)}\right]^2} \tag{5}$$

The matrix equation (4) describes an overdetermined set of linear equations, and the matrix is referred to as the *mode-shape matrix*. Each column in the mode-shape matrix refers to the displacement of a mode at each of the transducer locations. The overdetermined system of equations can be solved in a manner that minimizes the least squares error by the application of the Moore–Penrose inversion technique [10]. This technique yields

$$\begin{bmatrix} A^{(0)} \\ A^{(1v)} \\ A^{(1h)} \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & -2 & 0 \\ 0 & 2 & 0 & -2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{6}$$

It is easy to see that this equation is equivalent to equation (3). However, this is a more general formulation that can be applied to 1D waveguides of arbitrary cross section. The requirement is that there must be more transducer positions than there are modes, and that the resulting mode-shape matrix is not ill-conditioned. Note that in the case of the pipe example, the mode-shape matrix is frequency independent and the mode separation technique could be applied in either the time or the frequency domain. However, this is not the case in general and if the mode-shape matrix is frequency dependent it should strictly be recalculated for each frequency present in the signals and applied in the frequency domain.

It should be noted that in 1D waveguides such as pipes the number of modes is potentially quite high because of the large number of angular orders possible. However, if higher order modes are not to be used for testing then they can be suppressed by the array design itself. For example, an array with 32 physical elements around the perimeter of a pipe could have groups of eight adjacent elements connected in parallel to form four quadrants as shown in Figure 5(b). Although this would only allow zero and first-order modes to be used for testing it would still suppress modes up to 16th order.

3.2 Directional selectivity

Point measurements at a single axial location along a 1D waveguide can be used to separate modes but not directions.[a] To separate directions of propagation, it is necessary to measure the relative phase of a wave at two different axial positions. As an example, consider the situation where harmonic waves of frequency ω of the same mode are propagating in both directions along a 1D waveguide. Let x be the distance along the waveguide, and $A^{(+)}$ and $A^{(-)}$ be the (complex) amplitudes of the wave propagating in the positive and negative x directions respectively. The measured displacement, U_j, at an axial position x_j along the waveguide is therefore

$$U_j = A^{(+)}\mathrm{e}^{\mathrm{i}(kx_j - \omega t)} + A^{(-)}\mathrm{e}^{\mathrm{i}(-kx_j - \omega t)} \tag{7}$$

where k is the wavenumber of the mode at the frequency ω. The problem is to deduce the values of $A^{(+)}$ and $A^{(-)}$ from the measured displacements, of which there must clearly be at least two. For the case of two transducers at locations x_1 and x_2 ($x_2 > x_1$), the previous expression can be written in matrix form as

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-\mathrm{i}k\delta} & \mathrm{e}^{\mathrm{i}k\delta} \\ \mathrm{e}^{\mathrm{i}k\delta} & \mathrm{e}^{-\mathrm{i}k\delta} \end{bmatrix}\begin{bmatrix} \mathrm{e}^{\mathrm{i}kx_0} & A^{(+)} \\ \mathrm{e}^{-\mathrm{i}kx_0} & A^{(-)} \end{bmatrix} \tag{8}$$

where $2x_0 = x_1 + x_2$ and $2\delta = x_2 - x_1$ and the time-harmonic term has been omitted for clarity. The complex exponential terms containing x_0 are grouped with the unknown amplitudes $A^{(+)}$ and $A^{(-)}$, since these are the unknown phase shifts due to propagation from the source of the waves (i.e., a reflector) and the center of the array. Once again, the solution is obtained by inversion of the matrix

$$\begin{bmatrix} \mathrm{e}^{\mathrm{i}kx_0} & A^{(+)} \\ \mathrm{e}^{-\mathrm{i}kx_0} & A^{(-)} \end{bmatrix} = \frac{-1}{2i\sin 2k\delta}\begin{bmatrix} \mathrm{e}^{-ik\delta} & -\mathrm{e}^{\mathrm{i}k\delta} \\ -\mathrm{e}^{ik\delta} & \mathrm{e}^{-\mathrm{i}k\delta} \end{bmatrix}\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \tag{9}$$

If $2k\delta = n\pi$ (n, an integer), then the matrix is singular, and the closer $2k\delta$ is to $n\pi + \pi/2$ the more stable the inversion. Practically, this means that the optimum spacing between transducers is 1/4λ, 3/4λ, etc. For nondispersive modes, the complex exponential terms in the matrix act as time shifts, but for accurate directional selectivity of dispersive modes, the procedure should again be applied in the frequency domain.

3.3 Generalized problem

From the preceding sections, it can be seen that the mathematical basis for separating received modes and

directions are very similar, and both can be combined in general form:

$$\mathbf{U}(\omega) = \mathbf{M}(\omega)\mathbf{A}(\omega) \tag{10}$$

where the vector **U** represents the measured displacements, **M** is the mode-shape matrix and **A** is the vector of unknown amplitudes. In this general formulation **U** now contains measurements both around the perimeter of the waveguide and along its axis. Similarly **M** contains both mode-shape information and phase according to the axial position of the associated transducer and the coefficients in **A** are for each possible mode in each direction. Again, the matrix **M** needs to be inverted using an appropriate technique.

The discussion thus far has considered reception only. The same arguments can be applied in transmission to determine what amplitude and phasing of transducers is required to transmit a particular mode in a particular direction. By virtue of reciprocity, the quantities turn out to be precisely the same as those required to receive the same mode in that direction. The most general form of the combined transmit and receive situation can be written in the same way as equation (10), except that the displacement vector **U** now has N^2 elements (N being the total number of elements in the array) and each element in **U** now corresponds to the signal transmitted and received between that pair of elements. This type of approach has been used to perform the processing in 1D waveguides such as rails [11] that support multiple modes.

3.4 Example—pipe inspection

Pipe inspection is the most successful example of any type of long-range guided-wave inspection. At least three companies, Guided Ultrasonics Ltd [12], Pipeline Integrity Ltd [13], and MKCNDT [14], supply commercial guided-wave systems for pipe inspection, primarily targeted at the petrochemical industry. The fundamental principles of these systems is similar, the main differences lie in the type of transducer elements used with guided ultrasonics and pipeline integrity favoring piezoelectric devices, while MKCNDT use magnetostrictive sensors. A typical result obtained using the guided ultrasonics system is shown in Figure 6 as an example, which illustrates the key aspects of long-range guided-wave NDE. The data is presented in the form of an A-scan, with the abscissa indicating range measured from the test location and the ordinate axis representing reflected signal amplitude. An array is used to transmit an axisymmetric torsional mode, T(0,1), into the pipe and then to separate received signals into those that return as the same mode, and those that have been mode converted to a mode with angular order one, F(1,2). The directly reflected T(0,1) signals are indicated by the black line and the mode converted T(0,1) to F(1,2) signals are indicated by the gray line. The dominant features in the black signal are a series of regularly spaced reflections from butt welds between pipe sections. The challenge is discriminating between signals from benign structural

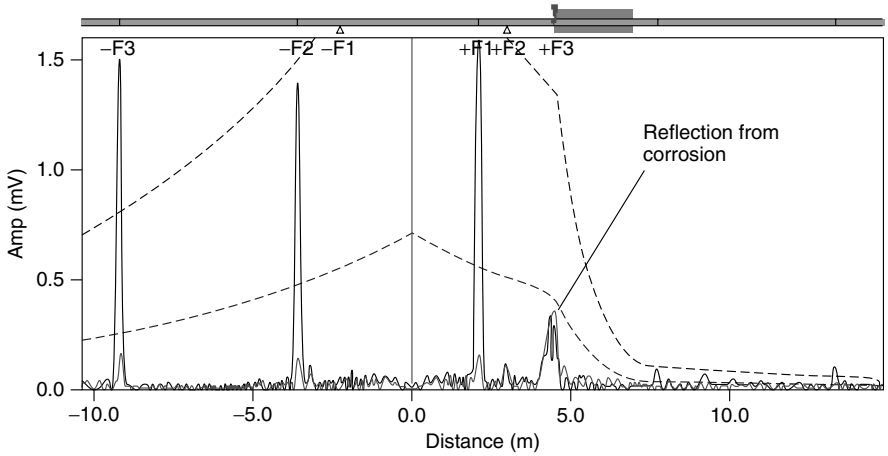


Figure 6. Example of results obtained from guided-wave pipe inspection system. [Reproduced with permission from Guided Ultrasonics Ltd.]

features such as these and defects. In pipe inspection, this is the reason for using multiple modes. The axisymmetric butt welds do not cause any mode conversion of the incident T(0,1) wave and hence there are no corresponding peaks in the gray mode converted signal. However a patch of localized corrosion at a distance of approximately 5 m produces a peak in both the reflected T(0,1) mode and the F(1,2) mode, indicating a nonaxisymmetric feature and possible defect.

4 COMPACT ARRAYS FOR 2D WAVEGUIDE STRUCTURES

In a 2D waveguide structure, access is generally restricted to one surface of the waveguide and differences in mode shape cannot be used to separate different modes.[b] Instead, modal and directional selectivity must be achieved by exploiting differences in the phase of waves at different points on the surface. This is much more similar to the case of a bulk wave ultrasonic array transducer, but with two important differences.

First, there are the usual complicating properties of guided waves, such as dispersion and multiple modes. However, the issue of modal selectivity is often overstated in the case of 2D waveguides, where there are only three possible modes (A_0, S_0, and SH_0) at typical operating frequencies below 2 MHz mm. These modes have significantly different properties and unwanted modes can generally be suppressed by the careful design of array elements themselves. For example, a piezoelectric disc operating in through-thickness mode is an order of magnitude more sensitive to A_0 than S_0 and completely insensitive to SH_0. Similarly, EMAT elements can be designed to be much more sensitive to S_0 than A_0 [5]. A general procedure for modeling the sensitivity of guided-wave transducers to different guided-wave modes in isotropic and anisotropic 2D waveguides using a Green's function formulation is described in [15]. The problem of dispersion is actually not a problem; as long as the dispersive properties of the waveguide are known, all processing can be performed in the frequency domain and dispersion compensation [16] applied.

The second and more important complication for guided-wave arrays over bulk wave arrays is the fact that a guided-wave array is in the middle of the wavefield rather than being at a boundary. This means that any consideration of directionality for guided-wave arrays must consider the wavefield in all 360° about the array, even if the array is only designed to detect damage over a limited range of angles. For example, the wavefield in front of a linear array comprising a single row of elements on the surface of a unimodal 2D waveguide will be very similar to the wavefield from an equivalent linear array for bulk waves. The crucial difference is that the guided-wave array will also emit an identical wavefield behind the array as well and it is impossible to subsequently distinguish which side of the array a reflector is on. This means that linear guided-wave arrays with a single row of elements have little practical use with 2D waveguides, although they have been used as a research tool in controlled laboratory experiments [17–19].

4.1 Omni-directional inspection

A generic application for 2D waveguide arrays is to interrogate the surrounding area of a structure in all directions. This suggests a layout of array elements that is approximately (i.e., as far as the discretization into elements will allow) axisymmetric. In such a design, the elements themselves should exhibit omnidirectional selectivity. For such arrays, it is useful to describe performance in terms of a polar coordinate system with its origin at the center of the array. Figure 7(a–c) shows some candidate designs for 2D arrays, including two that are axisymmetric and one gridlike design that is not.

The expression for the intensity, I, of a point in an image using data from a bulk wave array is given by the delay and sum equation:

$$I(\mathbf{r}) = \sum_{m=1}^{N} \sum_{n=1}^{N} A\left(\mathbf{r}, \mathbf{d}_{(\mathbf{m})}, \mathbf{d}_{\mathbf{n}}\right) h_{(m)n} \times \left(t = \tau\left(\mathbf{r}, \mathbf{d}_{(\mathbf{m})}, \mathbf{d}_{\mathbf{n}}\right)\right) \tag{11}$$

where $h_{(m)n}(t)$ is the time (t) domain recorded when the mth element acts as a transmitter and the nth element acts as a receiver, $\mathbf{r}$ is the point in the image, A is the amplitude factor, τ is the delay, $\mathbf{d}_{(\mathbf{m})}$ is the position of the transmitter element, and

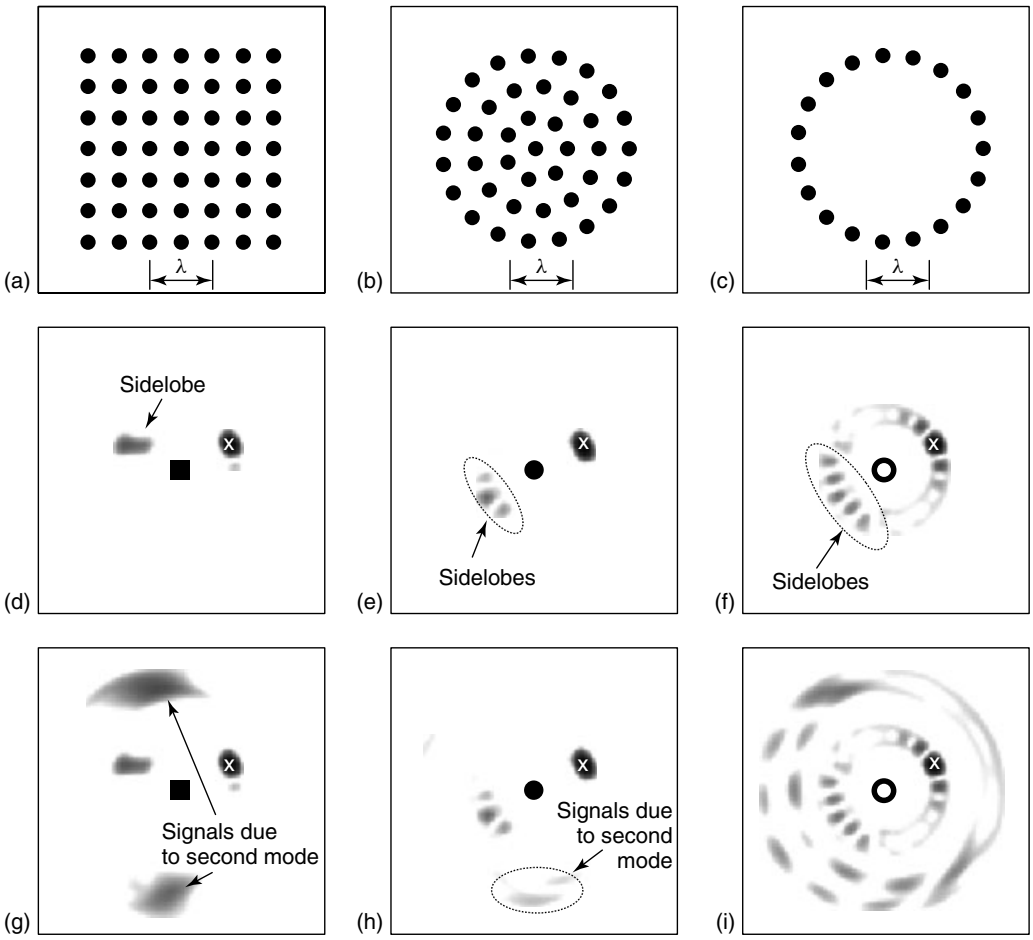


Figure 7. (a–c) Possible 2D array designs for omni-directional inspection of 2D waveguides showing the target wavelength; (d–f) associated PSFs for a reflector at the location marked x assuming a nondispersive unimodal system; (g–i) associated PSFs assuming a nondispersive bimodal system where the unwanted second mode has half the wavelength of the first mode. In (d–i) the signal used is a Gaussian windowed toneburst with an 80% bandwidth at −40 dB and the gray scale is from 0 dB (black) to −40 dB (white).

$\mathbf{d_n}$ is the position vector of the receiver element. The summations are performed over the N elements in the array, assuming in this case that each element acts both as transmitter and receiver, although this is not a necessity. This expression is applicable to all linear imaging algorithms, the difference being in the choice of coefficients, A, and delays, τ [1].

Equation (11) is also valid for a guided-wave array if the waves are nondispersive, but is more usefully written in the frequency domain, as this allows for frequency-dependent time delays to be applied (as phase shifts in the frequency domain) if necessary:

$$I(\mathbf{r}) = \int \sum_{m=1}^{N} \sum_{n=1}^{N} A\left(\mathbf{r}, \mathbf{d_{(m)}}, \mathbf{d_n}, \omega\right) H_{(m)n}(\omega)\, \mathrm{d}\omega \tag{12}$$

where $H_{(m)n}$ is the Fourier transform of the signal received when the mth element is transmitting and the nth element is receiving, and A is now a complex coefficient that includes phase and amplitude components. The challenge is the selection of coefficients, A, to give the best imaging performance. In terms of characterizing the imaging performance, both the

approaches of beam profile and point spread function (PSF) modeling may be employed [1].

The most conceptually simple choice of coefficients, A, is to select those which make the mode of interest arrive in phase when the array is focused at each point, $\mathbf{r}$, in the image. In this case,

$$A\left(\mathbf{r}, \mathbf{d}_{(\mathbf{m})}, \mathbf{d}_{\mathbf{n}}, \omega\right) = \mathrm{e}^{\mathrm{i}k(\omega)\left(|\mathbf{r}-\mathbf{d}_{(\mathbf{m})}|+|\mathbf{r}-\mathbf{d}_{\mathbf{n}}|\right)} \qquad (13)$$

where $k(\omega)$ is the wavenumber of the guided-wave mode. This is analogous to the total focusing algorithm for bulk wave arrays [20]. The larger the ratio of the overall dimension of the array to the wavelength of the guided-wave mode, the better the resolution. PSFs calculated using this approach for the arrays shown in Figure 7(a–c) are shown in Figure 7(d–f). In all cases, it is assumed that there is one nondispersive mode present with the wavelength indicated in the figure.

There are two reasons why this approach does not necessarily give good results in the case of a 2D guided-wave array. First, there is the potential problem of multiple modes. If the elements are sensitive to more than one mode, then application of equations (12) and (13) to one mode may cause an unwanted mode to be steered in a different direction. If this happens, phantom signals appear at incorrect positions in the PSF at different radial distances (because of the velocity differences between modes) and different angular positions. Whether this occurs depends on both the element configuration and the relative wavenumbers of the guided-wave modes involved. Example PSFs for the arrays shown in Figure 7(a–c) operating in a bimodal environment are shown in Figure 7(g–i). The second potential problem, that occurs even in a unimodal environment is the appearance of large sidelobes in the PSF over a wide range of angular positions, which can be seen to varying degrees in all of Figure 7(d–f). A fully populated circular array, Figure 7(b), comes close to providing optimum performance as shown in Figure 7(e). If the array only consists of the perimeter elements, Figure 7(c), then there is a severe degradation of the PSF as shown in Figure 7(f). The reason why the design shown in Figure 7(c) is of interest is because of the significantly reduced number of elements compared with a fully populated array of the same diameter.

To improve the situation with an array that is not fully populated such as that shown in Figure 7(c), it is necessary to choose different coefficients, A, than those suggested by equation (13). The problem with equation (13) is that the coefficients are chosen simply to maximize the signal due to the desired mode at each point in the image, $\mathbf{r}$, in turn; what occurs elsewhere in the image as a consequence of this is ignored. Conceptually this can be written as

$$\frac{|I(\mathbf{r})|^2}{\sum_{m,n} |A\left(\mathbf{r}, \mathbf{d}_{(\mathbf{m})}, \mathbf{d}_{\mathbf{n}}, \omega\right)|^2} \rightarrow \max^{\mathrm{c}} \qquad (14)$$

A more sophisticated choice of coefficients must in some way consider the overall image. One possibility that can be applied to axisymmetric element configurations is to use angular deconvolution [21]. This is because the PSF of such an array is angularly invariant, hence the angular content of an image obtained using equations (12) and (13) is the angular convolution of a number of delta functions with the angular PSF of the array. Conventional deconvolution techniques can therefore be applied and this is most easily performed in the Fourier domain. Here, the angular Fourier transform of an image obtained using equations (12) and (13) is divided by the angular Fourier transform of the PSF and the final image is the inverse angular Fourier transform of the result. The deconvolution technique does not allow perfect delta functions to be observed in the final image, as the angular Fourier transform of the PSF from a finite sized array only contains information up to a limited number of angular orders. Division of angular spectra beyond this point results in extreme sensitivity to noise in the original data and hence some form of angular order filter must be applied. However, as long as the angular spectrum of the PSF from the array does not contain any zeros up to this point, the final result is a PSF that resembles that which would be obtained by direct application of equations (12) and (13) to data from a fully populated array of the same size.

A more general approach that can be applied to any array configuration is the maximization of contrast technique [22]. This seeks to maximize the signal at the focal point relative to the total signal elsewhere in the image and can be written as

$$\frac{|I(\mathbf{r})|^2}{\int |I(\mathbf{r}')|^2 \, d\mathbf{r}'} \rightarrow \max \tag{15}$$

Unfortunately, direct application of this technique leads to an unstable solution, where a small amount of random noise in data leads to a major deterioration in image quality. This is because equation (15) implies that the array can be focused at a single point that cannot be physically achieved with an array of finite dimensions. This is analogous to performing the previously described angular deconvolution without a filter. An enhancement that improves stability is to perform the following optimization instead:

$$\frac{\int_{\mathbf{r}+\boldsymbol{\varepsilon}} |I(\mathbf{r}')|^2 \, d\mathbf{r}'}{\int |I(\mathbf{r}')|^2 \, d\mathbf{r}'} \rightarrow \max \tag{16}$$

where $\mathbf{r} + \boldsymbol{\varepsilon}$ is a region in the image in the vicinity of the focusing point. The choice of the size of $\boldsymbol{\varepsilon}$ is made with knowledge of the best resolution that the array can achieve. Again, the final result is a PSF close to that which would be obtained from a fully populated array with the same overall dimensions.

If the integrals in equations (15) and (16) are instead considered as summations over finite numbers of angular directions, it can also be shown that the maximization of contrast technique is the same as the Moore–Penrose solution to the generalized problem for 1D waveguide arrays, described in Section 3.3. Finally, it should be noted that the maximization of contrast techniques can, in principle, be readily extended to dealing with the suppression of unwanted modes. This is performed by extending the denominator of equation (16) to include the integrations over the images from unwanted mode combinations.

4.2 Examples

4.2.1 Deployable arrays

In 2000, Wilcox *et al.* [17] demonstrated a prototype design for a fully populated 2D array with 16 elements, as shown in Figure 8. The array used 5-mm-diameter piezoelectric elements operating in the through-thickness mode to excite and detect the A_0 Lamb wave mode. The array was able to achieve a signal to coherent noise ratio of more than 30 dB when deployed on a large 5-mm-thick aluminum plate and was able to detect the signal scattered from a bonded mass simulating the presence of a localized defect. It was recognized that in many practical situations the A_0 mode is not a suitable choice because of its high attenuation, if the plate is in contact with liquid on one or both sides.

To widen the application scope, the same authors subsequently developed a deployable omnidirection array using EMAT elements to perform 2D inspection using the S_0 Lamb wave mode [23]. The EMAT

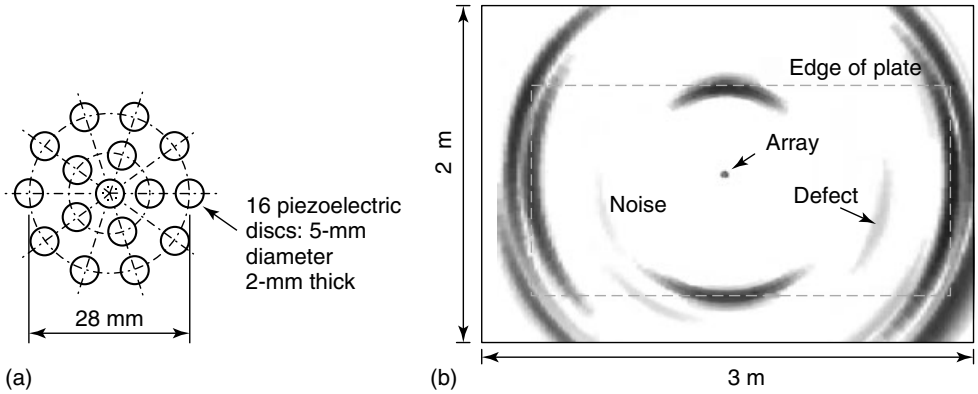


Figure 8. Example of use of deployable circular array from [17] showing (a) layout of array and (b) example results from a 2.5 m by 1.25 m by 5-mm-thick aluminum plate with a bonded mass simulating a defect. The gray scale in (b) is from 0 dB (black) to −30 dB (white). [Reproduced from Ref. 17. © American Institute of Physics, 2000.]

elements have the unique property of acting as omnidirectional sources and receivers of this mode. They also have the advantage of being noncontact and so do not require a couplant. It was shown [5] that the EMAT diameter required to optimize the S_0 sensitivity was considerably more than the maximum interelement spacing needed in an array to prevent spatial aliasing. As a result, it was necessary to overlap adjacent elements, an obstacle that was only surmountable because the EMAT coils could be printed onto a multilayer printed circuit board. The layout of elements is shown in Figure 9(a) and example results in Figure 9(b).

4.2.2 Permanently attached arrays

Subsequently, Fromme *et al.* [24] refined the A_0 array concept of [17] to make a compact self-contained device designed for permanent installation on offshore structures. The device is shown in Figure 10(a) and (b) and some example results from a 5-mm-thick steel plate are shown in Figure 10(c). The device achieved a superior angular resolution to that described in [17] by first, using 32 rather than 16 elements, and second, by arranging these in a ring and using the angular deconvolution algorithm to suppress sidelobes.

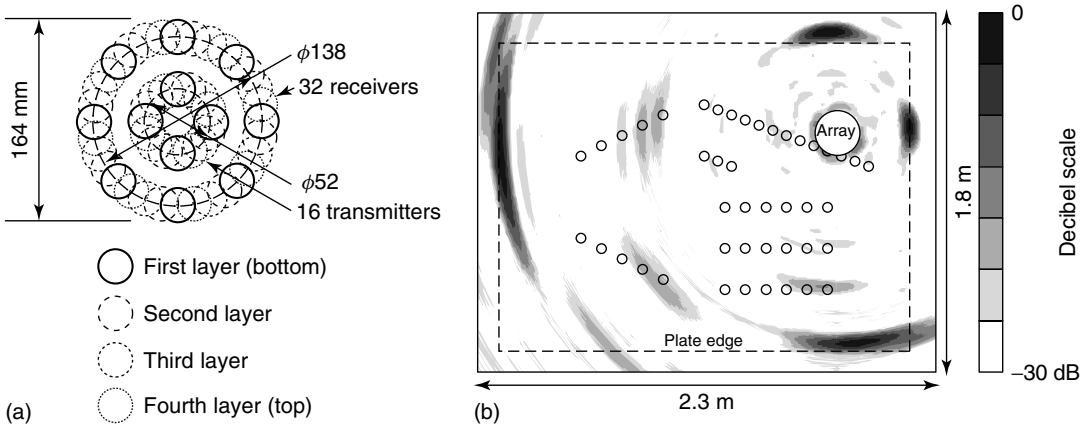


Figure 9. Example of deployable array from [23] with EMAT elements showing (a) array layout and (b) example results from a 6-mm-thick steel plate containing multiple artificial defects. [Reproduced from Ref. 23. © IEEE, 2005.]

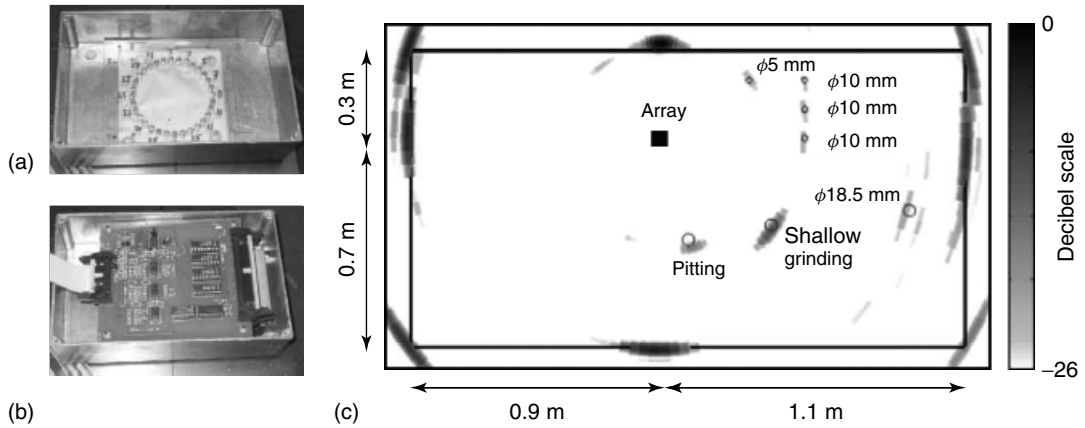


Figure 10. Prototype circular array from [24] showing (a) array, (b) multiplexing circuitry, and (c) example results from a 2 m × 1 m × 5 mm thick steel plate containing a number of artificially introduced defects. [Reproduced from Ref. 24. © IEEE, 2006.]

Giurgiutiu *et al.* [18] have also built permanently attached arrays, this time using so-called piezo wafer active sensors (PWASs) as the array elements (*see* **Piezoelectric Wafer Active Sensors**). The thin PWAS devices provide shear tractions on the surface of the waveguide and are hence particularly sensitive to S_0 Lamb waves, if bonded to the surface. Figure 11(a) shows the layout of a simple linear array of nine PWASs and Figure 11(b) shows example results where a crack was detected within the field of view of the array. Note that the linear array has no directionality about its centerline, hence the clarity of the image shown relies on the absence of any reflectors behind the array. Diamanti *et al.* [19] have also published results from a linear array, this time using the A_0 mode at low frequency (20 kHz) in a composite plate to detect impact damage in a composite plate. In this case, the array was mounted close to an edge of the plate, hence eliminating the problem of backward propagating waves.

The previous examples have all considered the plates as isotropic. In principle, many of the concepts can be applied to highly anisotropic composite plates, but these introduce some additional complexity. First, the choice of guided-wave mode in composites becomes more critical, as the attenuation of certain modes at certain frequencies can be prohibitively high. Secondly, the anisotropic material properties mean that guided-wave propagation is different in different directions. Rajagopalan *et al.* [25] have attempted to solve this problem through the adaptation of the basic imaging equation to account for directionally dependent wavenumbers.

Reference should also be made to the work of Sicard *et al.* [26] who have exploited array-based processing techniques with scanned guided-wave probes to perform defect characterization and Rose *et al.* [7] who have used comb type transducers for mode selection.

5 SPARSE DISTRIBUTED ARRAYS

For a permanently attached guided-wave system for damage detection, it is possible to exploit differences in the response of the structure over the course of time due to the appearance of damage. This approach is referred to as *reference signal comparison*. In the context of reference signal comparison, the compact arrays described previously are not necessarily the best solution, since the ability to detect damage by spatially resolving it from structural features is no longer required. Instead, the array needs only to be able to locate the signal from a single scatterer (the damage). The assumption of only one damage location occurring at any one time is a perfectly valid starting point in most practical situations where the occurrence of any damage is a fundamentally infrequent event.

5.1 Localization

If it is assumed *a priori* that there is only one scatterer present then localization is a trivial problem that can be achieved with as few as three sensing elements, by direct application of equations (12) and (13). The data

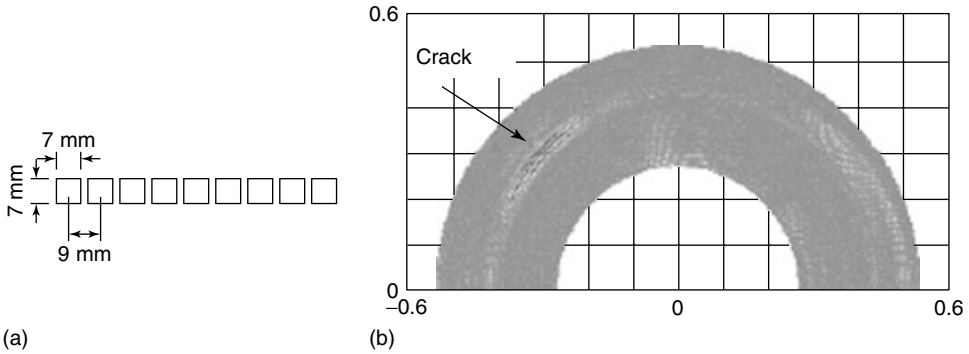


Figure 11. Permanently attached linear array using piezo wafer active sensors (PWASs) from [18] showing (a) array, (b) example results obtained on a 1-mm-thick aluminum plate containing a 19-mm-long machined crack. [Reproduced from Ref. 18. © Sage Publications, 2005.]

from a pair of sensors in the array produces an ellipse in the image with the sensors at the foci [27]. The intersection point of the ellipses from three or more pairs of sensors (i.e., at least three separate sensors acting in pitch–catch mode) defines the damage location. This is illustrated with simulated data in Figure 12(a). Note that the success of the approach is predicated on there being only one scatterer present, hence its use in conjunction with reference signal comparison. If there are multiple scatterers present, then the image rapidly becomes too complex to interpret, as shown in the example in Figure 12(b) where reflections from eight additional structural features as well as the direct transmission between transducer pairs have been added. The reflectors represent the edge and corner reflections from a rectangular plate.

5.2 Reference signal comparison

The far bigger challenge for any permanently installed system that relies on reference signal comparison is robustly performing the comparison itself, whether this is performed by simple subtraction or by more complicated methods. The problem is that the reference signal is likely to contain multiple large signals due to reflections from structural features, while the size of potential signals scattered from damage is much smaller. Whatever method of comparison is used, the signals from structural features must be suppressed by a significant amount, the remnants of these signals constituting the noise of the system [28]. If the noise is greater than the size of signals from damage, then the damage cannot be reliably detected. The following simple calculation gives an idea of the numbers involved and the degree of noise suppression necessary. A straight edge of a structure is the largest reflector likely to be encountered in practice, and has a 100% reflection coefficient for normally incident S_0 and A_0 guided waves. If a pointlike transducer emits a guided wave, then in the absence of material losses, the amplitude of the wave decreases in proportion to the square root of the propagation distance. The amplitude, A_{refl}, of

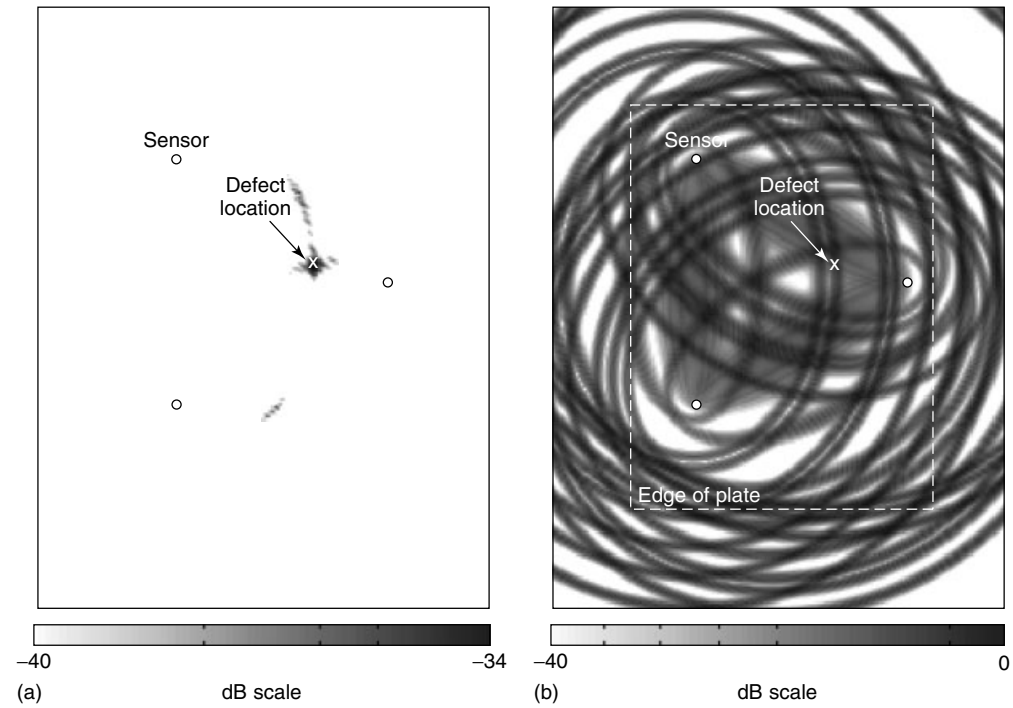


Figure 12. Localization using pitch–catch data from a sparse distributed array of three sensors in (a) the presence of a single scatterer and (b) multiple scatterers from the structural features indicated.

a received signal reflected back from an edge of the structure at a distance d from the transducer is therefore

$$A_{\text{refl}} = \frac{A}{\sqrt{2d}} \tag{17}$$

where A is the amplitude of the outgoing wave at unit distance. The fact that these signals decay with root distance dependence is very important. A signal reflected from damage will undergo secondary scattering and the received amplitude, A_{damage}, will be of the form:

$$A_{\text{damage}} = \frac{AR}{d} \tag{18}$$

where R is the backscattering reflection coefficient of the damage. Although at this stage R is not known, it is already clear from equations (17) and (18) that the anticipated signal amplitude as a function of distance from a particular type of damage decreases at a faster rate than that from large structural features. In other words, the ratio of signals from damage to signals from features is expected to decrease with $d^{-1/2}$. This in itself would not be a problem if the comparison process perfectly suppressed all signals due to structural features. Unfortunately, the signals from structural features do not remain precisely the same due to environmental changes such as temperature, which perturb the velocity of guided waves.

For signal comparison based on subtraction, it is straightforward to show that, if the temperature changes, the fraction of a signal from a structural feature remaining after subtraction increases in proportion to the propagation distance, d. This, combined with equations (17) and (18), means that the overall signal-to-noise ratio (SNR) decreases with $d^{-3/2}$. Without performing some sort of temperature compensation, it can be shown [28] that the inter-sensor spacing needed to reliably monitor a structure is very small indeed. As a consequence of the $d^{-3/2}$ signal to noise dependence, it can further be shown that each 15 dB improvement in the SNR corresponds to an order of magnitude reduction in the number of sensors required to monitor a given area of the structure. Hence, any improvement in the SNR is very valuable. Croxford *et al.* [29] demonstrated two candidate temperature compensation techniques that both gave SNR improvements of around 30 dB under varying temperature conditions, as shown in Figure 13. One technique was inspired by the work

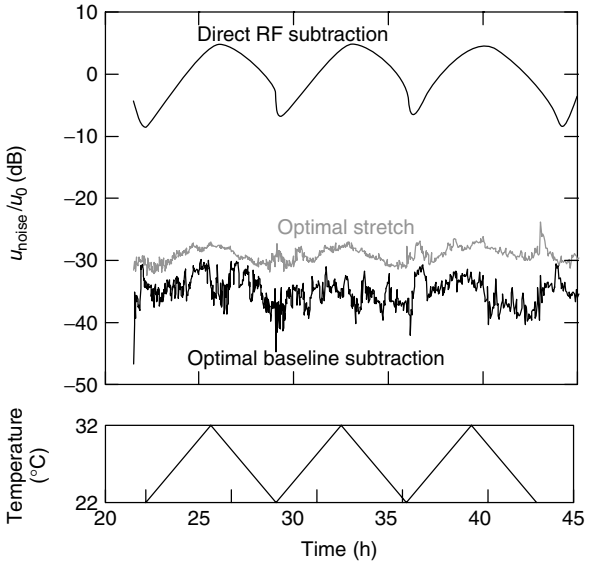


Figure 13. Measurement of remnant signal (u_{noise}) amplitude after performing reference signal comparison by subtraction under cyclic temperature variations of 10 °C. The reductions in noise level obtained by the application of two candidate compensation techniques (optimal stretch and optimal reference subtraction) are also shown. [Reproduced with permission from Ref. 29. © SPIE, 2007.]

of Lu and Michaels [30], where the best matched reference signal from an ensemble of signals recorded at different ambient temperatures was selected. The second is based on the optimization of a simple mathematical dilation operation applied to the reference signal that mimics the effect of the change in guided-wave velocity due to temperature variations.

5.3 Examples

Acellent Technologies Inc. [31] manufacture one of the few commercially available guided-wave SHM systems (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**). Their product utilizes a distributed array of piezoelectric devices (the "SMART layer") in a grid pattern mounted on a flexible sheet that can be either embedded or mounted to the surface of a component. Guided-wave signals are transmitted between pairs of transducers. Proprietary software is used to detect changes between reference and subsequent signals and uses this information to produce a map of potential damage locations in a structure. As the sensor pitch is around 100 mm, the number of sensors required to instrument the whole of a large structure is very large and the system is presumably designed for damage detection in known "hot-spot" regions. A picture of the transducer layer is shown in Figure 14. The advantage of a small sensor pitch is, for the reasons described previously, a combination of increased sensitivity to damage (since the damage will necessarily be nearer to a sensor) and reduced sensitivity to temperature. Demonstration results have been obtained on a variety of components, including metals [32] and composites [33], but routine deployment in industrial situations has not yet occurred.

Konstantinidis *et al.* [27] and Lu and Michaels [30] both attempt to use widely spaced arrays of sensors for damage detection, although the subsequent

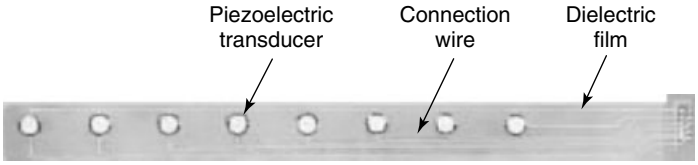


Figure 14. Picture of the Acellent SMART layer device. [Reproduced from Ref. 27. © Institute of Physics Publishing, 2006.]

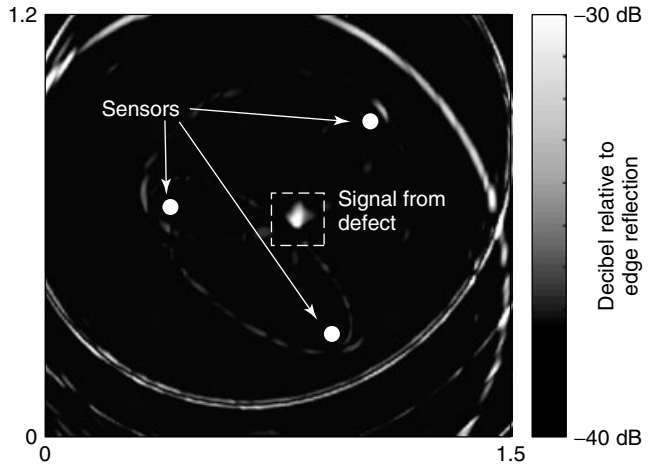


Figure 15. Results obtained by Konstantinidis *et al.* [27] on a 1.5 m × 1.25 m × 3 mm thick aluminum plate using an array of three piezoelectric transducers and optimal reference subtraction. [Reproduced from Ref. 32. © Institute of Physics Publishing, 2006.]

processing methodology is different. The former simply use the earlier part of the time–domain signal recorded from different sensor pairs and effectively isolate the signal directly scattered from a defect. Lu and Michaels use information in the diffuse field that occurs over a much longer time period. Both groups of workers use reference signal comparison to identify changes in the structural response, with Konstantinidis *et al.* opting for a direct subtraction and Lu and Michaels using an integrated root mean square approach. The reference signal comparison methods are both sensitive to temperature fluctuations and similar temperature compensation schemes based on ensembles of reference signals are used. Example results obtained by Konstantinidis *et al.* are shown in Figure 15.

It is appropriate to mention the use fiber Bragg gratings (FBGs) as detectors of guided-wave signals. The advantage is the relative ease with which a distributed network of such sensors can be embedded in a composite structure. The principle is that multiple FBGs can be etched onto the same fiber and interrogated separately either using temporal or (optical) wavenumber multiplexing. Betz *et al.* [34, 35] have studied the sensitivity of such sensors to Lamb waves, and the sensors appear to be ideal for use as a distributed array of guided-wave receivers. It should be stressed that FBGs do not overcome any of the other difficulties associated with distributed guided-wave arrays, and still require a different device, such as a piezoelectric one to act as a transmitter. The latter point is a rather major limitation, since if the transmitters in an array still need to be piezoelectric elements, the advantage of using FBG receivers is somewhat reduced.

6 CONCLUSIONS

The physical principles of array operation and processing have been reviewed with specific reference to the use of guided-wave arrays for NDE and SHM applications. The reason for using arrays in these applications is to improve the detection, localization, and characterization of defects. In all cases, the factor that ultimately limits detectability is coherent noise because of interaction between guided waves and benign features in the structure. In the case of deployable arrays for NDE, the primary purpose of using an array is to enable signals from defects to be resolved from signals due to benign features. This resolution is generally achieved spatially, but can also be achieved by separating different types of interaction with guided waves, using the information in multiple modes. However, even with state-of-the-art spatial resolution, long-range guided-wave NDE is fundamentally limited to structures with low feature density such as pipes. For this reason, widespread use of guided-wave NDE devices for complex structures, such as aircraft fuselages, is unlikely to ever occur, as interpretation of signals is too difficult.

Demonstrator SHM systems operating on the same principles as the NDE devices have been built, but there has been little or no commercial take up to date. The reason for this is that such systems remain limited to geometrically simple structures with low feature density, and the cost and complexity of such systems is prohibitively high for most suitable applications. However, within the SHM paradigm of a permanently installed system, the obstacle of structural complexity may be surmountable in the future. This is because comparison with reference signals can be performed, that, in principle, enables signals due to structural features to be eliminated. If this can be achieved, the requirements for array processing become trivial, since the *a priori* assumption that there is only one scatterer present (the defect) can be made.

Unfortunately, the suppression of signals from structural features using a reference signal is a very difficult problem, because of environmental fluctuations. Even temperature changes of a few degrees have been shown to cause changes in signals that are much larger than those due to defects. Robustly discriminating between environmental changes and signals from defects is one of the grand challenges for guided-wave SHM that must be surmounted before such systems can be used commercially. Recently, quantitative experiments and modeling are providing a better understanding of environmental effects and initial concepts for temperature compensation schemes appear promising.

END NOTES

[a.] Directional selectivity by mode shape is theoretically possible from a single point if the relative phase of displacements in different directions can

be measured. For example, the in-plane and out-of-plane surface displacements of a Rayleigh wave are in quadrature, with the particles moving in clockwise elliptical trajectories for waves traveling right to left and counterclockwise for waves traveling left to right.

[b.] Differences in mode shape between the top and bottom surfaces of a plate can be used to a limited extent in a laboratory where both surfaces can be accessed to separate symmetric and antisymmetric modes. Separation by mode shape is also used in finite element (FE) [15] where displacements can be monitored at any point through the thickness of the waveguide.

[c.] The normalization by the amplitude of the coefficients is necessary because otherwise the absolute amplitude of the PSF is unlimited and its maximization is meaningless.

REFERENCES

[1] Drinkwater BW, Wilcox PD. Ultrasonic arrays for non-destructive evaluation: a review. *NDT and E International* 2006 **39**(7):525–541.

[2] Cawley P, Lowe MJS, Alleyne DN, Pavlakovic BN, Wilcox PD. Practical long range guided wave testing: applications to pipes and rails. *Materials Evaluation* 2003 **61**:66–74.

[3] Jia X. Modal analysis of Lamb wave generation in elastic plates by liquid wedge transducers. *The Journal of the Acoustical Society of America* 1996 **101**(2):834–842.

[4] Alers GA, Burns LR. EMAT designs for special applications. *Materials Evaluation* 1987 **45**: 1184–1189.

[5] Wilcox PD, Lowe MJS, Cawley P. The excitation and detection of Lamb waves with planar coil electromagnetic acoustic transducers. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2005 **52**(12):2370–2383.

[6] Monkhouse RSC, Wilcox PD, Cawley P, Interdigital PVDF. Flexible transducers for the generation of Lamb waves in structures. *Ultrasonics* 1997 **35**:489–498.

[7] Rose JL, Pelts SP, Quarry MJ. A comb transducer model for guided wave NDE. *Ultrasonics* 1998 **36**(1):163–169.

[8] Silk MG, Bainton KP. The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves. *Ultrasonics* 1979 **17**(1):11–19.

[9] Lowe MJS, Alleyne DN, Cawley P. Defect detection in pipes using guided waves. *Ultrasonics* 1998 **36**(1–5):147–154.

[10] Penrose R. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society* 1955 **51**:406–413.

[11] Wilcox PD, Evans MJ, Pavlakovic BN, Alleyne DN, Vine KA, Cawley P, Lowe MJS. Guided wave testing of rail. *Insight* 2003 **45**(6):413–420.

[12] Guided Ultrasonics Ltd. http://www.guided-ultrasonics.com/ (accessed Jul 2007).

[13] Plant Integrity Ltd. http://www.plantintegrity.com (accessed Jul 2007).

[14] MKC NDT. http://www.mkckorea.com/english.htm (accessed Jul 2007).

[15] Velichko A, Wilcox PD. Modelling the excitation of guided waves in generally anisotropic multi-layered media. *The Journal of the Acoustical Society of America* 2007 **121**(1):60–69.

[16] Wilcox PD. A signal processing technique to remove the effect of dispersion from guided wave signals. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2003 **50**(4):419–427.

[17] Wilcox PD, Lowe MJS, Cawley P. Lamb and SH wave transducer arrays for the inspection of large areas of thick plates. In *Annual Review of Progress in QNDE*, Chimenti DE, Thompson DO (eds). American Institute of Physics: Melville, New York, 2000 Vol. 19A, pp. 1049–1056.

[18] Giurgiutiu V, Cuc A. Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention. *Shock and Vibration Digest* 2005 **37**(2):83–105.

[19] Diamanti K, Hodgkinson JM, Soutis C. Detection of low-velocity impact damage in composite plates using lamb waves. *Structural Health Monitoring* 2004 **3**:33–41.

[20] Holmes C, Drinkwater BW, Wilcox PD. Post-processing of the full matrix of ultrasonic transmit receive array data for non-destructive evaluation. *NDT and E International* 2005 **38**(8):701–711.

[21] Wilcox PD. Omni-directional guided wave transducer arrays for the rapid inspection of large areas of plate structures. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2003 **50**(6):699–709.

[22] Velichko A, Wilcox PD. Guided wave arrays for high resolution inspection. *Journal of the Acoustical Society of America* 2008 **123**(1):186–196.

[23] Wilcox PD, Lowe MJS, Cawley P. Omni-directional guided wave inspection of large metallic plate structures using an EMAT array. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2005 **52**(4):653–665.

[24] Fromme P, Wilcox PD, Lowe MJS, Cawley P. On the development and testing of a guided ultrasonic wave array for structural integrity monitoring. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2006 **53**(4):777–785.

[25] Rajagopalan J, Balasubramaniam K, Krishnamurthy CV. A phase reconstruction algorithm for Lamb wave based structural health monitoring of aniso tropic multilayered composite plates. *The Journal of the Acoustical Society of America* 2006 **119**(2): 872–878.

[26] Sicard R, Chahbaz A, Goyette J. Guided Lamb waves and L-SAFT processing technique for enhanced detection and imaging of corrosion defects in plates with small depth-to-wavelength ratio. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2004 **51**(10):1287–1297.

[27] Konstantinidis G, Drinkwater BW, Wilcox PD. The long-term stability of guided wave structural health monitoring systems. *Smart Materials and Structures* 2006 **15**(4):967–976.

[28] Croxford AJ, Wilcox PD, Drinkwater BW, Konstantinidis G. Strategies for guided wave structural health monitoring. *Proceedings—Royal Society of Edinburgh. Section B: Natural Environment* 2007 **463**:2961–2981.

[29] Croxford AJ, Wilcox PD, Konstantinidis G, Drinkwater BW. Strategies for overcoming the effect of temperature on guided wave structural health monitoring. *Proceedings of SPIE: Health Monitoring of Structural and Biological Systems 2007*. SPIE: Bellingham, Washington, DC, 2007 Vol. 6532, pp. 65321T1–65321T10.

[30] Lu Y, Michaels JE. A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations. *Ultrasonics* 2005 **43**:717–731.

[31] Lin M, Qing X, Kumar A, Beard S. SMART layer and SMART suitcase for structural health monitoring applications. *Proceedings of SPIE, Smart Structures and Materials*. Newport Beach, CA, March 2001 Vol. 4332, pp. 98–106.

[32] Qing XP, Chan H-L, Beard SJ, Kumar A. An active diagnostic system for structural health monitoring of rocket engines. *Journal of Intelligent Material Systems and Structures* 2006 **17**(7):619–628.

[33] Qing XP, Beard SJ, Kumar A, Ooi TK, Chang FK. A built-in sensor network for structural health monitoring of composite structure. *Journal of Intelligent Material Systems and Structures* 2007 **18**(1):39–49.

[34] Betz DC, Staszewski WJ, Thursby G, Culshaw B. Structural damage identification using multifunctional Bragg grating sensors: II. Damage detection results and analysis. *Smart Materials and Structures* 2006 **15**:1313–1322.

[35] Betz DC, Thursby G, Culshaw B, Staszewski WJ. Identification of structural damage using multifunctional Bragg grating sensors: I. Theory and implementation. *Smart Materials and Structures* 2006 **15**:1305–1312.

Chapter 18

Piezoelectric Impedance Methods for Damage Detection and Sensor Validation

Gyuhae Park and Charles R. Farrar

Engineering Institute, Los Alamos National Laboratory, Los Alamos, NM, USA

1 INTRODUCTION

Piezoelectric (PZT) materials produce an electric charge when mechanically stressed and a strain when an electric field is applied across them. This electromechanical coupling property is extremely useful because it can be used to develop a transducer that serves both as a sensor and an actuator. An important characteristic of PZT materials is that they provide an unobtrusive, integrated, and distributed way to add actuation and sensing to a structure. This characteristic makes PZT devices ideal for use in structural health monitoring (SHM) for systems that do not contain natural excitation forces or for diagnostic algorithms that require a known, well-controlled excitation. An SHM technique, which utilizes the benefits of PZT materials and shows great promise for SHM systems, is the impedance-based health monitoring method.

The basic principle of the impedance method is to use high-frequency vibrations to monitor a local area of a structure for changes in mechanical impedance that would indicate damage or imminent damage. This concept is implemented using PZT sensors/actuators whose electrical impedance is directly related to the mechanical impedance of the structure. The impedance measurements are correlated with changes in mechanical parameters, such as resonant frequencies or damping, which enables damage to be detected and located. The impedance method also has applications in the field of sensor self-diagnostics in determining the operational status of PZT active sensors used in SHM.

This chapter provides a brief overview on research work on the impedance method. The first part of this section deals with theoretical background and experimental investigations of impedance-based SHM. A sensor diagnostics and validation process based on impedance measurements that performs *in situ* monitoring of the operational status of PZT active sensors

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

in SHM applications is then presented. This section also includes an investigation into the effects of the sensor/structure bonding defects on high-frequency SHM techniques, including Lamb wave propagation and impedance methods. It has been found that the effects are remarkable, modifying the phase and amplitude of propagated waves and changing the measured impedance spectrum, which could lead to false indications of structural conditions.

2 PRINCIPLE OF IMPEDANCE METHOD

The application of impedance measurements to SHM has its theoretical development first proposed by Liang *et al.* [1] and substantially developed by Chaudhry *et al.* [2, 3], Sun *et al.* [4], Park *et al.* [5–9], Peairs *et al.* [10], Giurgiutiu *et al.* [11–13], Giurgiutiu and Zagrai [14, 15], Soh *et al.* [16], Bhalla and Soh [17–20], Park *et al.* [21–23] and their coworkers. The reader is referred to (*see* **Electromechanical Impedance Modeling**) of the Encyclopedia for more details on the underlying theory. This chapter provides an overview of the experimental implementation of impedance-based damage detection.

When measurements from a damaged structure are compared with baseline measurements from the undamaged structure, the damaged structure will exhibit changes in its stiffness and damping characteristics that affect the mechanical impedance. Because direct measurements of the mechanical impedance of a structure are difficult to obtain, the electromechanical coupling effect of PZT materials is utilized to estimate this impedance in the impedance method. Any damage to a host structure will result in changes to its mechanical impedance, and these changes will be observed through changes in the electrical impedance of the PZT materials. In order to ensure high sensitivity of the electrical impedance to small defects in a structure, the impedance measurements are usually made at higher frequency ranges, typically greater than 30 kHz. At such high-frequency ranges, the dynamic response of the host structure is generally concentrated in a local area near the PZT device. PZT patches require very low level voltages, typically less than 1 V, to exert high-frequency excitations on the host structure.

An electromechanical model, which quantitatively describes the impedance measurement process, is presented in Figure 1. Assuming that an axial PZT actuator is attached to one end of a single degree-of-freedom mass-spring system, whereas the other end is fixed, Liang *et al.* [1] showed that the electrical admittance $Y(\omega)$, which is the inverse of the electrical impedance, of the PZT actuator is a combined function of the mechanical impedance of the PZT actuator $Z_a(\omega)$ and that of the host structure $Z(\omega)$:

$$Y(\omega) = \mathrm{i}\omega \frac{wl}{t_c}\left(\varepsilon_{33}{}^{T} - d_{31}{}^{2}Y_p^E + \frac{Z_a(\omega)}{Z_a(\omega)+Z_s(\omega)} \times d_{31}{}^{2}\hat{Y}_p^E\left(\frac{\tan kl}{kl}\right)\right) \quad (1)$$

where V is the input voltage, and I is the output current from the PZT transducer. $a, d_{3x}, Y_{xx}{}^{E}$, and $\bar{\varepsilon}_{33}^{T}$ are the geometry constant, PZT coupling constant, Young's modulus, and complex dielectric constant of the PZT transducer, respectively. The wavenumber of the PZT patch, k, is defined as

$$k = \omega\sqrt{\frac{\rho}{\hat{Y}_p^E}} \quad (2)$$

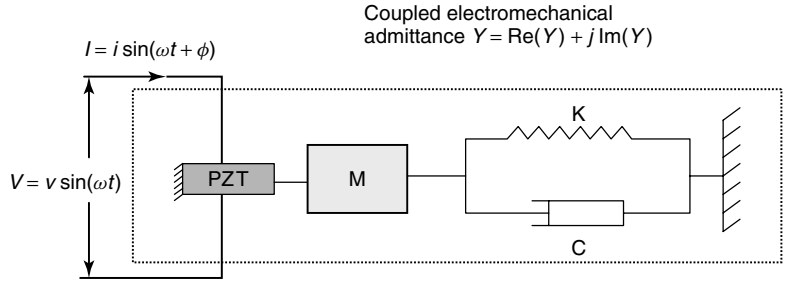


Figure 1. Single degree-of-freedom model used to represent a PZT-driven dynamic structural system.

where ρ is the mass density of the PZT material. The term $\frac{\tan kl}{kl}$ is close to 1 when the frequency range of excitation is much smaller than the first resonant frequency of the PZT patch.

Because all parameters except Z_s, the mechanical impedance of the host structure, are properties of the PZT material, only the mechanical impedance of a structure uniquely defines the electrical impedance of the PZT transducer. As a result, the electrical impedance signature of the PZT transducer is affected by changes in the mechanical impedance of the host structure. Therefore, by monitoring the electrical impedance and comparing this impedance to a baseline impedance measurement, one can determine when structural damage has either occurred or is imminent. The variation in the PZT electrical impedance over a range of frequencies is analogous to that of the frequency response functions of a structure, which contains vital information regarding the health of the structure. Bhalla and Soh [18] extended the 1D impedance model, which is expressed in equation (1), to the 2D governing equation for the PZT electrical impedance.

The impedance method tracks changes in the mechanical properties of structures as in global vibration-based methods. However, impedance measurements are made at much higher frequencies than those used in global vibration-based methods, which are described in (*see* **Free and Forced Vibration Models** and **Modal–Vibration-based Damage Identification**) leading to greater sensitivity to localized damage. Ultrasonic methods, acoustic emission, eddy current, or any other high-frequency testing method can provide detailed information regarding anomalies in some structures, but these methods usually require complex instrumentation and expertise to interpret the measured data. Most of these methods must also be applied when structures are out of service, or can be applied only at certain time intervals, which may not be suitable for autonomous on-line SHM. The sensitivity of the impedance method to minor defects is comparable to that of ultrasonic methods, but the method does not require experienced technicians to discern the details. With the development of low-cost impedance measuring circuits, the cost required for impedance hardware and sensors/actuators is also less than the cost for equipment needed to implement other non-destructive evaluation (NDE) techniques. The sensing regions of impedance sensors are much larger than those of local ultrasonic or eddy current sensors, which are usually moved to scan over certain areas to detect damage in a structure. Detailed comparisons between the impedance method and other NDE approaches can be found in the literature [6, 9, 13].

3 SIGNAL PROCESSING OF IMPEDANCE METHOD

Although the impedance response plots provide a qualitative approach for damage identification, the quantitative assessment of damage is traditionally made using a scalar damage metric. In [4], a simple statistical algorithm is used that is based on frequency-by-frequency comparisons, referred to as the *root mean square deviation* (RMSD),

$$M = \sum_{i=1}^{n} \sqrt{\frac{\left[\mathrm{Re}\left(Z_{i,1}\right) - \mathrm{Re}\left(Z_{i,2}\right)\right]^2}{\left[\mathrm{Re}\left(Z_{i,1}\right)\right]^2}} \qquad (3)$$

where M represents the damage metric, $Z_{i,1}$ is the impedance of the PZT measured in a healthy structural condition, and $Z_{i,2}$ is the impedance in a new structural condition for comparison with the baseline measurement in frequency interval i. In a RMSD damage metric chart, larger numerical values of the metric correspond to larger differences between the baseline reading and the comparison reading indicating the presence of damage in a structure. Another scalar damage metric, referred to as the *cross-correlation* metric, can be used to interpret and quantify the information from different data sets. The correlation coefficient between two data sets determines the relationship between two impedance signatures, and provides an aesthetic metric chart. In most cases, the results with the correlating metric are consistent with those of RMSD, in which the metric values increase when there is an increase in the severity of damage.

Temperature changes, among all other ambient conditions, significantly affect the electric impedance signatures. Some PZT material parameters, such as the dielectric and strain constants are strongly dependent on temperature. Generally speaking, an increase in temperature causes a decrease in the magnitude of electric impedance, and leftward shifting of the

real part of the electric impedances. The RMSD and cross-correlation based damage metrics do not account for these variations. Park *et al.* [5] used a modified RMSD metric, which compensates for horizontal and vertical shifts of the impedance in order to minimize the impedance signature drifts caused by the temperature or other normal variations.

Lopes *et al.* [24] incorporated neural network features with the impedance method for somewhat quantitative damage analysis. The authors proposed a two-step damage identification scheme. In the first step, the impedance-based method detects and locates structural damage and provides an indication of damage in a green/red light form with the use of the modified RMSD. When damage is identified, the neural networks, which are trained for specific damage levels, are then used to estimate the severity of damage. Zagrai and Giurgiutiu [25] investigated several statistics-based damage metrics, including RMSD, mean absolute percentage deviation (MAPD), covariance change, and correlation coefficient deviation. It was found that the third power of the correlation coefficient deviation, $(1 - R^2)^3$, was the most successful damage indicator, which tends to linearly decrease as the location of a crack in a thin plate moves further away from the sensor. Tseng and Naidu [26] also investigated the performance of RMSD, MAPD, covariance and correlation coefficients as indicators of damage. The RMSD and the MAPD were found to be suitable for characterizing the growth and the location of damage, whereas the covariance and the correlation coefficient were efficient in quantifying the increase in damage size at a fixed location.

A debonding detection method for composite patches was proposed based on the coupled use of the impedance method and hybrid genetic algorithms [27]. The impedance methods were also used with an outlier detection framework [28], principle component analysis [29], and frequency-domain autoregressive model for nonlinear feature identification in a structure [30] for improved damage identification performance.

4 EXPERIMENTAL STUDIES

Impedance-based SHM has been successfully implemented on several complex structures; a four bay space truss [4], an aircraft structure [2], complex precision parts [31], temperature varying applications [5], spot-welded structural joints [11], civil structural components [6], concrete structures [17], a reinforced concrete bridge [16], composite reinforced polymers [32], civil pipelines [8], prestress monitoring [33], and corrosion detection [34, 35]. A complete review summarizing experimental results using the impedance method can be found in [9]. Only a few case studies are described here in order to illustrate how the impedance method can be used to monitor the conditions of structures.

A built-in pipeline structure has been investigated by Park *et al.* [8], as shown in Figure 2. The objective of this investigation was to utilize the impedance method for identifying structural damage in areas where rapid condition monitoring is urgently needed, such as in a postearthquake analysis. One PZT sensor/actuator ($15\,\text{mm} \times 15\,\text{mm} \times 0.2\,\text{mm}$) is bonded on each joint to monitor the conditions of this structure. The impedance measurements of a PZT transducer at the joint with three levels of local damage are shown in Figure 2. It can be seen that, with increasing the level of damage, the impedance signature shows a relatively large change in shape and is clearly indicative of imminent damage. For the first level of damage (loosening two bolts), only a small variation along the original signal (curve for undamaged condition) was observed. When four bolts have been loosened, the impedance showed more pronounced variations as compared to previous readings, and finally, when eight bolts have been loosened, it showed a distinct change in the signature pattern, i.e., new peaks and valleys appear in the entire frequency range. This change occurs because the damage modifies the apparent stiffness and damping of the joint, which results in changes in electric impedance of the PZT transducer.

Giurgiutiu *et al.* [11] presented health monitoring results of spot-welded structural joints. The impedance signature was recorded up to 1100 kHz. The initiation and propagation of damage were successfully correlated with the impedance measurements. In addition, through the use of multisite impedance measurements, the sensitivity to minor cracks, localization of damage, and rejection properties to far-field changes have been observed.

Monitoring of a massive built-in bridge structure using impedance sensors are shown in Figure 3. The

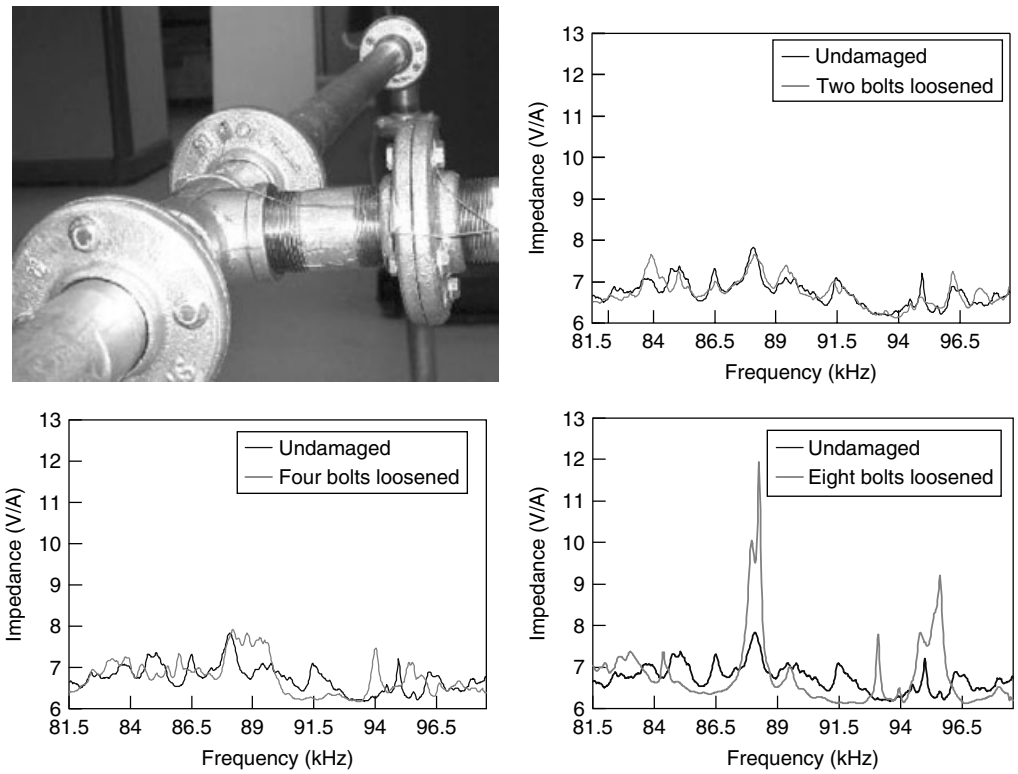


Figure 2. Pipeline experimental results. The impedance variation became more pronounced as the extent of damage increased.

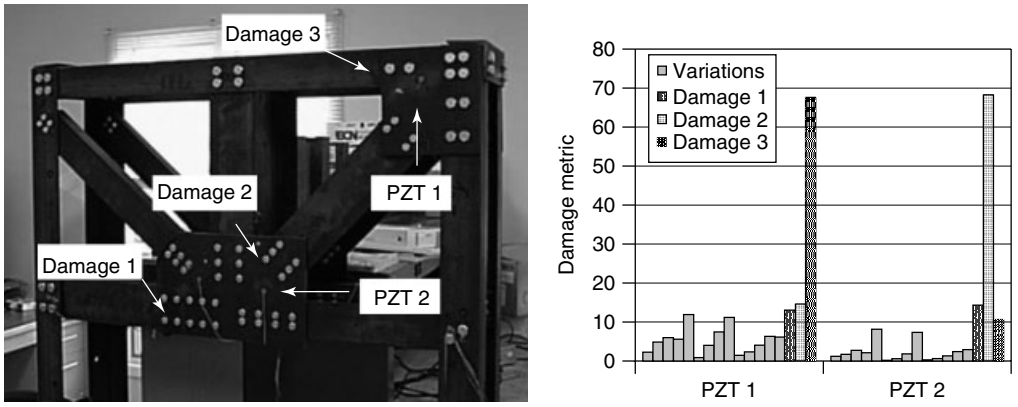


Figure 3. A 1/4 scale steel bridge section. The damage metric shows the localized effect of the impedance method.

bridge model consists of steel angles, channels, plates, and joints connected by over 200 bolts. The structure was 1.8-m tall and it weighed over 250 kg. Some environmental condition changes were imposed on this structure including changes in temperature, addition of a mass, and application of low-frequency vibrations. As can be seen in the damage metric values of Figure 3, these conditions resulted in relatively small variations and would be considered as minor changes. The first 14 variations are those from

the change in environmental conditions. Damage was simulated by loosening connection bolts at the selected locations, as shown in the figure. Damage 1 is believed to be well out of sensing range of both PZT 1 and PZT 2. Hence, only a small increase in damage metric is shown for both PZT transducers. Damage 2 is located close to the PZT 2 and Damage 3 is within sensing region of the PZT 1. Hence the increase in damage metric values are the highest for both PZTs. Note that the effect of loosening a single bolt on the entire structure is minor, and thus damage can be detected in its early stage [6].

Yang *et al.* [36] investigated the use of PZT impedance sensors with fiber-optic sensors for damage assessment and load history monitoring of underground structures, including rock samples. A fuzzy probabilistic damage model was proposed to quantify damage in rock based on the signal captured by piezo-impedance transducers. Soh and Balla [37] also investigated the possibility of monitoring concrete strength gain during its curing process. The stiffness of the concrete was well correlated with variations in impedance peaks, where the peaks shifted rightward and became sharper progressively with time, suggesting that the stiffness was increasing.

The integration of impedance methods with wave propagation-based SHM techniques was also investigated by many researchers [12, 13, 38–40]. Most of the wave propagation approaches utilize the sensing and actuation capability of the PZT patches, which can be nicely integrated with the impedance method. In addition, the waves usually travel long distances and cover a relatively large area, which complement the strengths of impedance-based health monitoring. Some studies include damage detection in the near-field with impedance methods and in the far-field using wave propagation [12, 13], while others utilize different sensitivities to different types of damage [39]. For instance, in pipeline monitoring, the impedance methods are used to detect structural damage occurring at pipeline connection joints, while the Lamb wave propagation measurements identify cracks and corrosion along the surface and through the wall thickness of the pipeline structure [40].

As described in this chapter, the impedance-based health monitoring technique has been successfully applied to various structures ranging from aerospace to civil structures. Because of the high frequencies employed, the method is very sensitive to minor defects in a structure and not affected by any far-field changes. Recognizing that damage is a local phenomenon and that one often knows where to expect damage (either from previous testing or from geometry, such as joints), the impedance method is ideal for tracking and on-line monitoring of critical sections in various structures.

5 SENSOR SELF-DIAGNOSTICS AND VALIDATION USING IMPEDANCE MEASUREMENTS

One critical aspect of PZT active-sensing technologies is that large numbers of distributed sensors and actuators are usually needed to monitor the condition of a structure. In addition, the structures in question are usually subjected to various external loading and environmental condition changes that may adversely affect the functionality of SHM sensors and actuators. The PZT active sensor diagnostic and validation process, where the sensors/actuators are confirmed to be operational, is therefore a critical component to successfully implement SHM. Because piezoceramic materials are brittle, sensor fracture and subsequent degradation of mechanical/electrical properties are the most common types of sensor/actuator failures. In addition, the integrity of bonds between a PZT patch and the host structure must be maintained and monitored throughout their service lives because degradation in bonds modifies the strain and stress transfer mechanism.

For PZT sensors, Saint-Pierre *et al.* [41], Pacou *et al.* [42], and Giurgiutiu *et al.* [15] proposed a debonding identification algorithm by monitoring the resonance of a PZT sensor based on the electrical impedances. As the debonding area between the PZT wafer and the host increased, the shape of the resonance of the PZT wafer became sharper and more distinctive, and the magnitudes of the host resonances decreased. This method requires a high-frequency data-acquisition system because even the first resonance of PZT wafers in SHM applications usually lies in the range of hundreds of kilohertz. In addition, this method is not able to account for sensor fracture, which may simultaneously occur with debonding, as the sensor breakage would apparently generate changes in the resonances of a PZT sensor. Bhalla and Soh [18] investigated the effect of shear-lag loss

on electromechanical impedance measurements. They found that the bond layer can significantly modify the measured admittance signatures, and suggested the use of adhesives with high-shear modulus, the smallest practicable bond thickness, and small-sized PZT transducers in order to minimize the influence of the bond layer. They also suggested that the imaginary part of the electrical admittance of PZT transducers may play a meaningful role in detecting deterioration of the bond layer. However, by concentrating on the effects of bond layers on the electromechanical impedance spectrums only, the metrics that can be used for bond quality assessment was not clearly identified, and the ability to discriminate bond failures from structural damage was not thoroughly investigated.

In general, a completely broken PZT active sensor can be easily identified if a sensor does not produce any meaningful output or an actuator does not reasonably respond to applied signals. However, if only a small fracture or debonding occurs, the sensors/actuators are still able to perform adequately (with distorted signals after the sensor fracture), potentially leading to false indications of the structural condition. In order to fully implement current active-sensing systems into SHM practice beyond the proof-of-concept demonstration, an efficient sensor self-diagnostic procedure should be adopted in the SHM process.

5.1 Sensor diagnostics and validation process

The premise of the sensor self-diagnostic process is to track changes in the capacitive value of PZT materials, which is manifested in the imaginary part of the measured electrical admittances of the PZT material. It has been found that the degradation of the mechanical/electrical properties of the PZT sensor and its attachment to the external structure produces measurable and distinct changes in the capacitance of PZT materials [43–45]. This section briefly describes the sensor diagnostic procedure.

The electrical admittance of a PZT transducer, which is defined as the ratio of the energizing voltage to the resulting current, under a free–free boundary condition is given in the following relation,

$$Y_{\text{free}}(\omega) = \frac{I}{V} = \mathrm{i}\omega\frac{wl}{t_c}\left(\varepsilon_{33}{}^{T}\,(1-\mathrm{i}\delta)\right) \qquad (4)$$

where w, l, t_c is the width, length, and thickness of a PZT transducer, respectively, and δ is the dielectric loss tangent of the PZT wafer. When a PZT patch is surface bonded to a structure, it was shown in equation (1) that the electrical admittance of the PZT transducer is a combined function of the mechanical impedance of the host structure and that of the PZT wafer.

The sensor diagnostic process is based on equations (1) and (4). The electrical admittance is clearly a function of its geometry constants (w, l, t_c), mechanical properties (Y_p^E), and electrical properties $(\varepsilon_{33}{}^{T}, d_{31}, \delta)$ of the PZT transducer. It is also obvious from the equations that the changes in these properties are manifested more distinctly in the imaginary part of the electrical admittance. Therefore, the breakage of the sensor and the degradation of the sensor's quality can be identified by monitoring the imaginary part of the electrical admittance. The breakage/degradation of the sensor quality would cause a downward shift in the slope of the admittance (i.e., decrease in the capacitive value) because the effective size of the sensor would decrease with breakage and the values of dielectric constants and PZT coupling constants would decrease with increasing degradation.

Another significant observation that can be made from equations (1) and (4) is that one can identify the effect of the bonding layer on the measured electrical admittance. The effect of the bonding layer obtained by assuming the mechanical impedance of a structure is much larger than that of the PZT transducer in equation (1), which makes the last term in equation (1) close to zero. The result is as follows:

$$\begin{aligned} Y_b(\omega) &= \mathrm{i}\omega\frac{wl}{t_c}\left(\varepsilon_{33}{}^{T}(1-\mathrm{i}\delta) - d_{31}{}^{2}Y_p^E\right) \\ &= Y_{\text{free}}(\omega) - \mathrm{i}\omega\frac{wl}{t_c}\left(d_{31}{}^{2}Y_p^E\right) \end{aligned} \qquad (5)$$

It is clear from equations (4) and (5) that the electrical admittance of the same PZT patch is different if it is under a free–free or surface-bonded (or commonly referred to as *blocked*) condition. The blocked condition would cause a downward shift in

the slope of the electrical admittance (decrease in the capacitive value) of a free PZT with the factor of $\frac{wl}{t_c} d_{31}{}^2 Y_p^E$. The assumptions that led this result are usually valid, especially at lower frequencies (<20 kHz), because the mechanical impedance of the structure is usually several orders of magnitude greater than that of a PZT transducer. Even though this derivation does not explicitly consider the parameters of bonding materials (such as thickness or shear modulus), it is obvious from equation (5) that the use of a PZT transducer with lower Y_p^E and smaller dimensions, such as a small PVDF patch, will reduce the effect of the bonding layer on the measured admittance, which is consistent with the shear-lag analysis available in the literature [18, 46]. The importance of equation (5) is that the bonding layer also contributes to the overall admittance (capacitance) of PZT patches bonded to a structure.

Figure 4 illustrates the measured admittance of free and surface-bonded PZT patches. The 5A PZT materials with dimensions of 20 mm × 20 mm × 0.25 mm are used. The admittances of three free PZT patches were measured in the frequency range of 40–20 000 Hz using an Agilent 4294A impedance analyzer. These PZT patches were then surface-mounted to a thick aluminum beam and plate using superglue. The admittance measurements were then repeated.

As can be seen in Figure 4, the downward shifting effect of the bonding layer is remarkable. The slope of the imaginary part, which is the capacitive

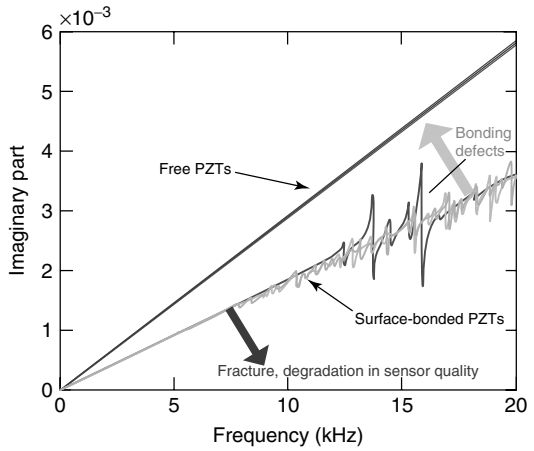


Figure 4. Electrical admittance measurement from PZT patches under free and surface-bonded conditions.

value of the PZT material, was changed from 29.1 to 18 nF resulting in a 38% reduction. Therefore, the sensor functionality including sensor breakage and the degradation of the bonding condition can be assessed by monitoring the imaginary part of the admittances (capacitance) of the PZT materials. Figure 4 also graphically describes the effects of sensor failures, both the sensor fracture and the bonding defects, on the measured electrical admittance.

5.2 Operational status validation of sensing network and the effects of bonding defects on Lamb wave propagations

A series of experimental results are presented to demonstrate the performance of the sensor diagnostics and validation process. This sensor diagnostic process is applied to confirm the operational status of a PZT sensing network right after the installation on an aluminum plate and a bolted washer. The effects of faulty sensors, in particular the bonding defects, on Lamb wave propagations and electromechanical impedance measurements are also presented.

For SHM approaches based on the use of Lamb wave propagation, a single sensor failure creates major problems including failure of signal processing algorithms, or at a minimum, the lack of inspection in a large area of a structure. Sensors can fail owing to extreme operational condition, such as an impact by foreign objects. However, the sensor failure can also occur due to improper installation of the sensors, including imperfect bonding or accidental sensor breakage during handling. The sensor diagnostic process described in the previous section can be efficiently used to determine the operational status of the sensing network immediately after the installation, as shown in the following example.

The test structure shown in Figure 5 is an aluminum plate (1200 mm × 1200 mm × 2 mm). Nine circular PZT patches are mounted using superglue on one surface with a 3 × 3 array with equal distances in all directions. The locations and the numbering schemes of these actuators/sensors are also shown in Figure 5. The circular PZT patch is 5.5 mm in diameter with 0.2-mm thickness. Admittance

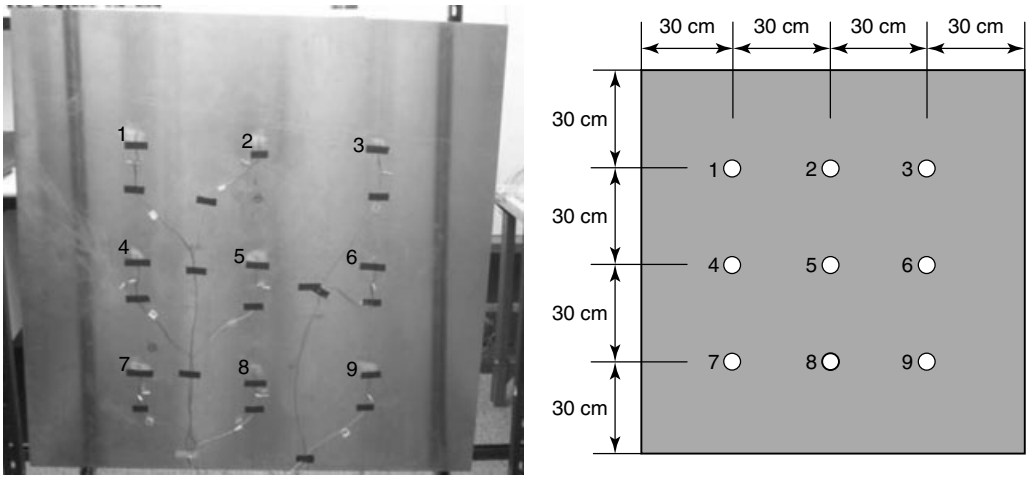


Figure 5. An aluminum plate with PZT transducers attached.

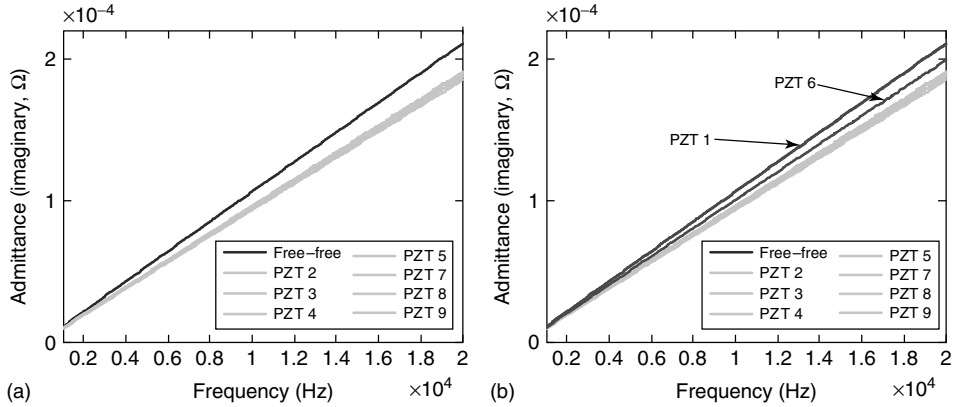


Figure 6. Admittance measurements (imaginary part) from PZT transducers. (a) PZT patches under free–free and healthy conditions and (b) PZT patches under healthy and faulty conditions.

measurements in the frequency range of 1–20 kHz were made at each PZT patch before and after the installation using an Agilent 4294A impedance analyzer.

The admittance measurements of PZT transducers before and after the installation are shown in Figure 6(a). For most PZT patches, the downward shifting effect caused by bonding is evident. The slope of the imaginary part, analogous to the capacitive value of the PZT wafer, results in a 4.5% reduction. However, for PZT 1, virtually no change in the slope of the imaginary part of admittance was observed as shown in Figure 6(b). Further, for PZT 6, only a 2% change in the slope was observed. As described, the small or almost no change in the slope could be an indication of imperfect bonding condition. Therefore, these two sensors are classified as faulty by the sensor diagnostic method. This poor bonding will be problematic because the condition can be easily degraded during the operation.

The effect of bonding defects on the Lamb wave propagation was also investigated. First, the wave propagation (at 100 kHz) in the paths consisting of sensors with the good bonding condition were measured (between PZT 2 and 3, PZT 3 and 5, PZT 4 and 8, and PZT 5 and 9) and are plotted in Figure 7. Because the paths are of equal distance and traverse the same materials, the propagated waves show the

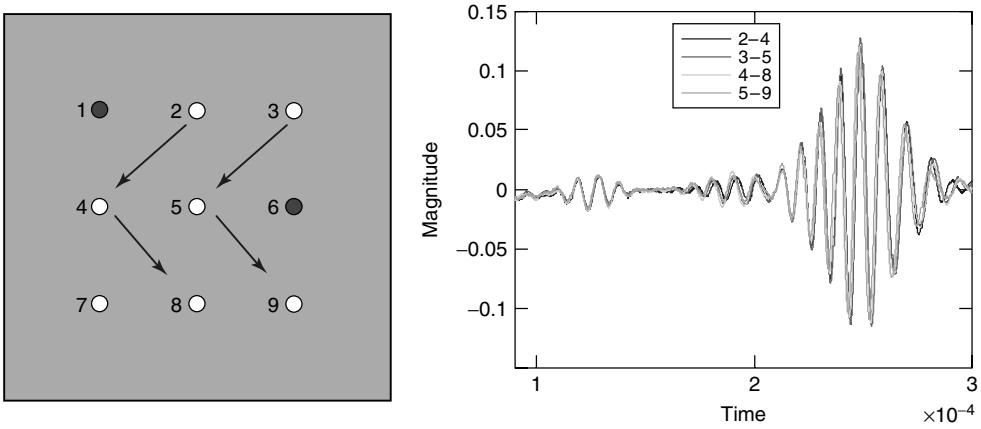


Figure 7. Lamb wave propagations between healthy sensors.

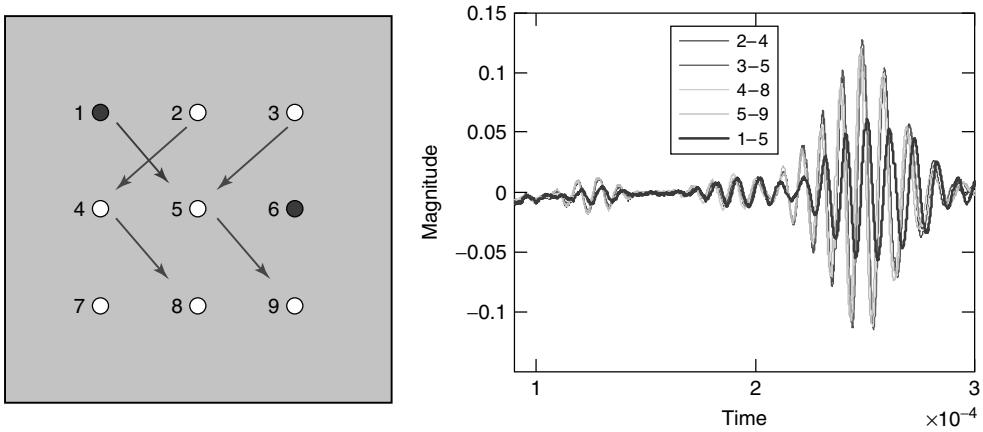


Figure 8. Lamb wave propagations between healthy sensors and faulty sensors.

same characteristics. The arrival times of the first S_o and A_o mode are evident and they also exhibit similar magnitudes. Figure 8 shows the overlap of propagating waves in the path between PZT 1 and 5. PZT 1 was identified as a faulty sensor by the sensor diagnostic process. There is a remarkable change in the measured response. There is a clear change in the magnitude, and further, the arrival time changes significantly. It should be noted that the effects of bonding defects are very similar to those caused by structural damage. If signal processing methods based on the wave attenuation or time of flight are used and the bonding defect occurs during operation, these changes can be mistakenly considered as structural damage, which points to the importance of sensor diagnostics in real-world operation. As shown in this example, the sensor diagnostic method can be used to monitor the operational status of the PZT transducers immediately after their installation. Furthermore, the degradation of bonding conditions over the service life of the PZT sensors could also be identified by tracking changes in the admittance measurements.

5.3 Bonding condition assessment and effects of bonding defects on electromechanical impedance measurements

In lower vibration frequency ranges, several studies [47, 48] have shown that the bonding layer plays a critical role in determining the measured dynamics,

resulting in changes to the measured modal frequencies and damping ratios. Because the frequency range of the impedance method is much higher, the effects of the bonding layer could be much more significant and should be well understood for this method to be successfully applied in practical applications.

A set of experimental tests were performed using different bonding conditions of the PZT patches. In order to assess the effects of different sensor bonding conditions on measured impedance, an aluminum hexagonal washer (with geometry 25.4-mm high, and 31.75 mm across the flats with a 12.7-mm diameter hole) was instrumented with two circular PZT patches (12.7-mm diameter, 0.2-mm thickness), each on a different face, as shown in Figure 9. The patches were attached to the plates using two different bonding techniques to simulate both good and poor bonding conditions. One bonding condition was achieved by using 3M thermo bond film for a temporary (and poor) bond, whereas the other was achieved using superglue to produce a more permanent bonding condition.

Figure 10 shows the imaginary part of the admittance of each of the sensors bonded to the aluminum hexagonal washer in the frequency range up to 20 kHz. This plot also includes an admittance measurement for an unbonded 11-mm-diameter PZT patch. The quality of the bond can be determined based on the changes in the slope steepness of the admittance measurement. From Figure 10, one can see that the superglued sensor demonstrates the lower slope, and therefore, can be considered to have a better bonding condition.

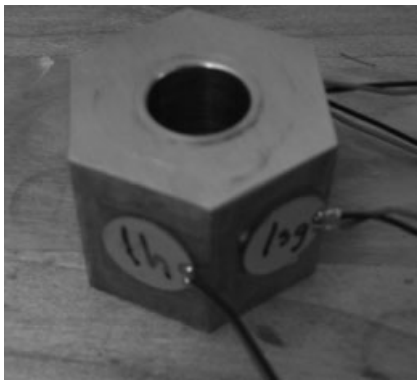


Figure 9. An aluminum washer with two PZT patches installed with different bonding conditions.

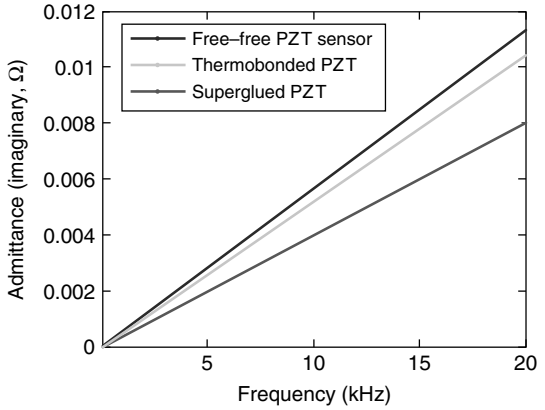


Figure 10. Admittance measurements from different bonding conditions from an aluminum washer.

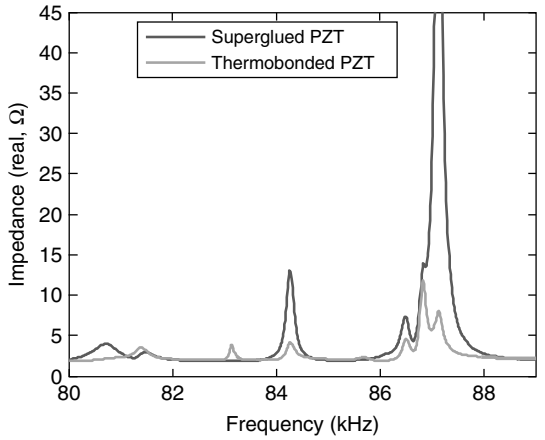


Figure 11. Impedance measurement (real part) under different bonding condition.

The impedance measurements (the real part) under the different conditions are shown in Figure 11. It is clear from the figure that the PZT bonded with the superglue shows resonances near 84 and 87 kHz to be at least an order of magnitude greater than the sensor bonded with the thermo film. From the measurements, it once again confirms that the magnitude of the resonance is affected by the bonding conditions. The poor bonding condition may result in erroneous or misleading results as the stress/strain transfer mechanism is modified by the changed electromechanical interaction between the structure and PZT patches. Furthermore, near the frequency range

at 83 kHz, the PZT active sensor with the thermo-bonded condition exhibits a new resonance of the structure. Therefore, it can be concluded that, as the bonding condition changes, the identified resonances and associated magnitudes are significantly affected. It is very clear from this figure that the bonding condition significantly influences structural identification and health assessment if degradation in bonds is not carefully accounted for. Further, it is obvious that, in order to minimize false indications of the structural condition, the sensor diagnostic process needs to be implemented to discriminate changes due to sensor degradation from structural damage in SHM field applications.

6 SUMMARY

With continual advances in sensor/actuator technology, signal processing techniques, and damage prognosis algorithms, impedance-based SHM methods will continue to attract the attention of researchers and field engineers for monitoring of various structural components. The impedance method provides several advantages over traditional approaches. First, because of the high-frequency range employed, the method is very sensitive to incipient damage in a structure and unaffected by changes in boundary conditions, loading, or operational vibrations. The method is very suitable for an autonomous monitoring system because the data-acquisition procedure can be automated and requires minimal user interference. An analytical model of the structure is also not required, making the method attractive for use on complex structures. Furthermore, the method can be efficiently used for sensor self-diagnostics and validation processes.

Although several successful examples of the use of impedance methods have been presented, it is important to note that extensive research efforts are being devoted to studying the practical implementation issues in real-world field applications. These issues include: (i) developing miniaturized and portable impedance measurement equipment with wireless communication capabilities [49–51]; (ii) packaging of the sensors to facilitate installation; and (iii) integrating these methods with other techniques, such as wave propagation methods and acoustic emission.

REFERENCES

[1] Liang C, Sun FP, Rogers CA. Coupled electromechanical analysis of adaptive material system—determination of actuator power consumption and system energy transfer. *Journal of Intelligent Material Systems and Structures* 1994 **5**:12–20.

[2] Chaudhry Z, Joseph T, Sun F, Rogers CA. Local-area health monitoring of aircraft via piezoelectric actuator/sensor patches. *Proceedings of the SPIE* 1995 **2443**:268–276.

[3] Chaudhry Z, Lalande F, Ganino A, Rogers CA, Chung J. Monitoring the Integrity of Composite Patch Structural Repair via Piezoelectric Actuators/Sensors. *Proceedings of the 36th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and AIAA/ASME Adaptive Structures Forum, 4*. New Orleans, LA, 1995; 2243–2248.

[4] Sun FP, Chaudhry Z, Liang C, Rogers CA. Truss structure integrity identification using PZT sensor-actuator. *Journal of Intelligent Material Systems and Structures* 1995 **6**:134–139.

[5] Park G, Kabeya K, Cudney HH, Inman DJ. Impedance-based structural health monitoring for temperature varying applications. *JSME International Journal, Series A* 1999 **42**(2):249–258.

[6] Park G, Cudney H, Inman DJ. Impedance-based health monitoring of civil structural components. *ASCE Journal of Infrastructure Systems* 2000 **6**(4):153–160.

[7] Park G, Cudney H, Inman DJ. An integrated health monitoring technique using structural impedance sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**(6):448–455.

[8] Park G, Cudney H, Inman DJ. Feasibility of using impedance-based damage assessment for pipeline systems. *Earthquake Engineering and Structural Dynamics Journal* 2001 **30**(10):1463–1474.

[9] Park G, Sohn H, Farrar CR, Inman DJ. Overview of piezoelectric impedance-based health monitoring and path forward. *The Shock and Vibration Digest* 2003 **35**:451–463.

[10] Peairs D, Park G, Inman DJ. Improving accessibility of the impedance-based structural health monitoring method. *Journal of Intelligent Material Systems and Structures* 2004 **15**(2):129–140.

[11] Giurgiutiu V, Reynolds A, Rogers CA. Experimental investigation of E/M impedance health monitoring of spot-welded structural joints. *Journal of*

Intelligent Material Systems and Structures 1999 **10**:802–812.

[12] Giurgiutiu V, Zagrai A, Bao JJ. Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring. *International Journal of Structural Health Monitoring* 2002 **1**:41–61.

[13] Giurgiutiu V, Zagrai AN, Bao J, Redmond J, Roach D, Rackow K. Active sensors for health monitoring of aging aerospace structures. *International Journal of the Condition Monitoring and Diagnostic Engineering Management* 2003 **6**(1):3–21.

[14] Giurgiutiu V, Zagrai A. Characterization of piezoelectric wafer active sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**:959–976.

[15] Giurgiutiu V, Zagrai AN. Embedded self-sensing piezoelectric active sensors for online structural identification. *ASME Journal of Vibration and Acoustics* 2002 **124**:116–125.

[16] Soh CK, Tseng K, Bhalla S, Gupta A. Performance of smart piezoceramic patches in health monitoring of a RC Bridge. *Smart Materials and Structures* 2000 **9**:533–542.

[17] Bhalla S, Soh CK. Structural impedance based damage diagnosis by piezo-transducers. *Earthquake Engineering and Structural Dynamics* 2003 **32**:1897–1916.

[18] Bhalla S, Soh CK. Electromechanical impedance modeling for adhesively bonded piezo-transducers. *Journal of Intelligent Material Systems and Structures* 2004 **15**:955–972.

[19] Bhalla S, Soh CK. High frequency piezoelectric signatures for diagnosis of seismic/blast induced structural damages. *NDT&E International* 2004 **37**:23–33.

[20] Bhalla S, Soh CK. Structural health monitoring by piezo-impedance transducers. I: modeling. *Journal of Aerospace Engineering* 2004 **17**:154–165.

[21] Park S, Yun CB, Roh Y, Lee JJ. Health monitoring of steel structures using impedance of thickness modes at PZT patches. *Smart Structures and Systems* 2005 **1**:339–353.

[22] Park S, Ahmad S, Yun CB, Roh Y. Multiple crack detection of concrete structures using impedance-based structural health monitoring techniques. *Experimental Mechanics* 2006 **46**:609–618.

[23] Park S, Grisso BL, Inman DJ, Yun CB. MFC-based structural health monitoring using a miniaturized impedance measuring chip for corrosion detection. *Research in Nondestructive Evaluation* 2007 **18**:139–150.

[24] Lopes V, Park G, Cudney H, Inman DJ. A structural health monitoring technique using artificial neural network and structural impedance sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**:206–214.

[25] Zagrai AN, Giurgiutiu V. Electro-mechanical impedance method for crack detection in thin plates. *Journal of Intelligent Material Systems and Structures* 2001 **12**:709–718.

[26] Tseng K, Naidu A. Non-parametric damage detection and characterization using smart piezoceramic materials. *Smart Materials and Structures* 2002 **11**:317–329.

[27] Xu YG, Liu GR. A modified electro-mechanical impedance model of piezoelectric actuator-sensors for debonding detection of composite repair patches. *Journal of Intelligent Material Systems and Structures* 2002 **13**:389–405.

[28] Park G, Rutherford AC, Sohn H, Farrar CR. An outlier analysis framework for impedance-based structural health monitoring. *Journal of Sound and Vibration* 2005 **286**:229–250.

[29] Park S, Lee JJ, Yun CB, Inman DJ. A built-in active sensing system-based structural health monitoring technique using statistical pattern recognition. *Journal of Mechanical Science and Technology* 2007 **21**:896–902.

[30] Rutherford CA, Park G, Farrar CR. Nonlinear feature identifications based on self-sensing impedance measurement for structural health assessment. *Mechanical Systems and Signal Processing* 2007 **21**:322–333.

[31] Lalande F, Childs B, Chaudhry Z, Rogers CA. High-frequency impedance analysis for nde of complex precision parts. *Proceedings of the SPIE* 1996 **2717**:237–245.

[32] Pohl J, Herold S, Mook G, Michel F. Damage detection in smart CFRP composites using impedance spectroscopy. *Smart Materials and Structures* 2001 **10**:834–842.

[33] Gopal V, Annamdas M, Yang Y, Soh CK. Influence of loading on the electromechanical admittance of piezoceramic transducers. *Smart Materials and Structures* 2007 **16**:1888–1897.

[34] Simmers GE, Sodano HA, Park G, Inman DJ. Detection of corrosion using piezoelectric impedance based structural health monitoring. *AIAA Journal* 2006 **44**:2800–2803.

[35] Park S, Grisso BL, Inman DJ, Yun CB. MFC-based structural health monitoring using a miniaturized

impedance measuring chip for corrosion detection. *Research in Nondestructive Evaluation* 2007 **18**:139–150.

[36] Yang YW, Bhalla S, Wang C, Soh CK, Zhao J. Monitoring of rocks using smart sensors. *Tunneling and Underground Space Technology* 2007 **22**: 206–221.

[37] Soh CK, Bhalla S. Calibration of piezo-impedance transducers for strength prediction and damage assessment of concrete. *Smart Materials and Structures* 2005 **14**:671–684.

[38] Kabeya K, Jiang Z, Cudney H. Structural health monitoring by impedance and wave propagation measurements. *Proceedings of International Motion and Vibration Control*. Switzerland, 25–28 August 1998; pp. 207–212.

[39] Wait JR, Park G, Farrar CR. Integrated structural health assessment using piezoelectric active sensors. *Shock and Vibration* 2005 **12**:389–405.

[40] Thien AB, Chiamori HC, Ching JT, Wait JR, Park G. The use of macro-fiber composites for pipeline structural health assessment. *Structural Control and Health Monitoring*, 2008 **15**:43–63.

[41] Saint-Pierre N, Jayet Y, Perrissin-Fabert I, Baboux JC. The influence of bonding defects on the electric impedance of piezoelectric embedded element. *Journal of Physics D (Applied Physics)* 1996 **29**:2976–2982.

[42] Pacou D, Pernice M, Dupont M, Osmont D. Study of the interaction between bonded piezoelectric devices and plates. *Proceedings of 1st European Workshop on Structural Health Monitoring*. Paris, 2002; pp. 406–413.

[43] Park G, Farrar CR, Rutherford CA, Robertson AN. Piezoelectric active sensor self-diagnostics using electrical admittance measurements. *ASME Journal of Vibrations and Acoustics* 2006 **128**:469–476.

[44] Park G, Farrar CR, Lanza di Scalea F, Coccia S. Performance assessment and validation of piezoelectric active sensors in structural health monitoring. *Smart Materials and Structures* 2006 **15**: 1673–1683.

[45] Overly TG, Park G, Farrar CR. Development of signal processing tools for piezoelectric sensor diagnostic processes. *Proceedings of the SPIE* 2007 **6530**:1–12.

[46] Sirohi J, Chopra I. Fundamental understanding of piezoelectric strain sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**:246–257.

[47] Seeley CE, Chattopadhyay A. Experimental investigation of composite beams with piezoelectric actuation and debonding. *Smart Materials and Structures* 1998 **7**:502–511.

[48] Sun D, Tong L, Atluri SN. Effects of piezoelectric sensor/actuator debonding on vibration control of smart beams. *International Journal of Solids and Structures* 2001 **38**:9033–9051.

[49] Grisso BL, Martin LA, Inman DJ. A wireless active sensing system of impedance-based structural health monitoring. *Proceedings of the IMAC-XXIII*. Orlando, FL, 2005.

[50] Mascarenas DL, Todd MD, Park G, Farrar CR. Development of an impedance-based wireless sensor node for structural health monitoring. *Smart Materials and Structures* 2007, **16**:2137–2145.

[51] Overly TG, Park G, Farrar CR. Development of impedance-based wireless active-sensor node for structural health monitoring. *Proceedings of 6*th *International Workshop on Structural Health Monitoring*. Stanford, CA, 2007; pp. 1660–1667.

Chapter 19
Thermal Imaging Methods

Daniel L. Balageas
Structure and Damage Mechanics Department, ONERA (The French Aerospace Lab), Châtillon, France

1 INTRODUCTION

The analysis of the space distribution and time evolution of the surface temperature of a structure can permit the establishment of a diagnostic of the structural health of the structure. This analysis is based on the fact that all mechanical phenomena are accompanied by correlated thermal effects (thermomechanical coupling). These effects can be reversible or irreversible. The reversible thermal phenomena are linked to the strain (and stress) state of the structure. Mapping the thermal state of the structure is a means to detect abnormalities in its mechanical behavior. Generally, irreversible phenomena are the signature of the occurrence of a damage. Until recently, the thermomechanical approach was scarcely applied in experimental mechanics, partly due to the lack of sensitivity of the temperature detection hardware. In recent years, infrared (IR) detectors and data processing techniques have achieved considerable improvements, opening new possibilities in this field. The other approach for structural health monitoring (SHM) comes from classical nondestructive evaluation (NDE). Damages are detected, localized, and characterized by the disturbance they introduce in the local physical properties of the structure. These properties are not mechanical, but their evolution can be correlated to a loss of mechanical performance. The variety of physical phenomena used in NDE to detect these damages is very wide. Among them, the thermal properties (heat conduction, heat diffusion, heat capacity, effusivity, interface thermal resistance) are currently used.

Thermal NDE techniques are essentially applied with the help of IR radiometry, although there are several other means of measuring temperature: thermoelectric effect (thermocouples), thermoresistive effect [1], thermochromic liquid crystals [2, 3], thermocolor or heat-sensitive paints [4], fluorescence [5], etc. The present article focuses on IR radiometry, more particularly on IR thermography, the most practical way to easily map surface temperature fields of structures, without any contact, with very high bandpass and sensitivity, allowing a detailed analysis of space and time distributions.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

In the present article, the physical principle of IR radiometry and state of the art of IR thermography are recalled. The analysis of the thermographic process shows how to perform a quantitative measurement, a requisite for the identification of structural health and characterization of damages. Different ways to perform SHM with IR thermography are finally detailed: diagnostic elaboration based on a thermomechanical analysis (strain mapping and fatigue detection), and damage detection and characterization by stimulated thermography.

2 RECALL ON TEMPERATURE MEASUREMENT BY RADIOMETRY

All materials emit and absorb energy by radiation. The amount of energy and the spectral domain of the emitted radiations strongly depend on temperature. This thermal emission may also vary with the direction of emission. Thermal emission has been modeled by introducing the concept of the "blackbody" or perfect radiator. The radiations emitted by the blackbody are a function of its absolute temperature, T, and their magnitudes are a function of the wavelength, λ, but are independent of the direction (isotropically diffuse or Lambertian surface). For a given temperature and wavelength, no other surface can emit more energy. This ideal concept is approached by real blackbodies: isothermal cavities and corrugated blackened isothermal surfaces.

Planck's law describes blackbody emission. It expresses the power irradiated per unit surface, unit solid angle, and unit wavelength domain ($\mathrm{W\,m^{-2}\,\mu m^{-1}\,sr^{-1}}$). This power is called the *spectral radiance*, L_λ, and is given by the following formula:

$$L_\lambda(\lambda, T) = 2hc_0^2\lambda^{-5}\left[\exp\left(\frac{hc_0}{\lambda KT}\right) - 1\right]^{-1} \qquad (1)$$

where c_0 is the speed of light in vacuum ($2.9979 \times 10^8\,\mathrm{m\,s^{-1}}$), h the Planck constant (6.626076×10^{-34} Js), and K the Boltzmann constant ($1.380658 \times 10^{-23}\,\mathrm{J\,K^{-1}}$).

Wien's law expresses the wavelength of the maximum emission as a function of the temperature: $\lambda_{\max} = 2897.7/T$. For near-ambient temperatures, the spectral domain for which the thermal emission is maximum is the IR between 5 and 15 μm.

For real materials, the thermal emission can be different from the blackbody radiation. The ratio of this thermal emission to the blackbody emission is the *spectral emissivity*. A first approximation consists in supposing that the emissivity is independent of the wavelength and direction of emission. A body that satisfies this condition is called a *gray body*. In practice, when measuring the thermal emission by radiometry, very often this assumption is assumed to have been verified.

If the real body differs strongly from a gray body, to be quantitative, radiometry requires knowledge of the value of the spectral emissivity in the spectral domain of measurement. With regard to the influence of the direction of emission, in general, radiometric measurements are performed following lines of sight corresponding to angles $\theta < 50^\circ$, making the directional effect negligible and validating the Lambertian assumption.

Radiometry consists of detecting and measuring radiant electromagnetic energy. These tasks are achieved by conjugating an element in the region of thermal activity with a radiation detector and with the help of lenses. The energy received can be related to the temperature of the emitter, at least in the simpler case (see Section 3). If a calibration is made, the radiometer signal can measure the emitter temperature. Generally, the measurement is performed in the IR domain (see above). For more details on the subject, please refer to DeWitt and Nutter [6].

3 INFRARED THERMOGRAPHY

If a scanning system is included between the lenses and the detector, or if there is an array of detectors instead of a unique detector, it is possible to measure the temperature in a certain field of view and to build a thermal image. Such an instrument is a thermographic imager, also called *infrared camera*. Figure 1 presents two types of cameras. The first type is the single detector camera with an optomechanical scanning system allowing to successively image all the points of the region and to build an image with the help of video-type scanning. This type of camera is less in use and is now being replaced by the focal plane array (FPA) cameras. In the FPA camera,

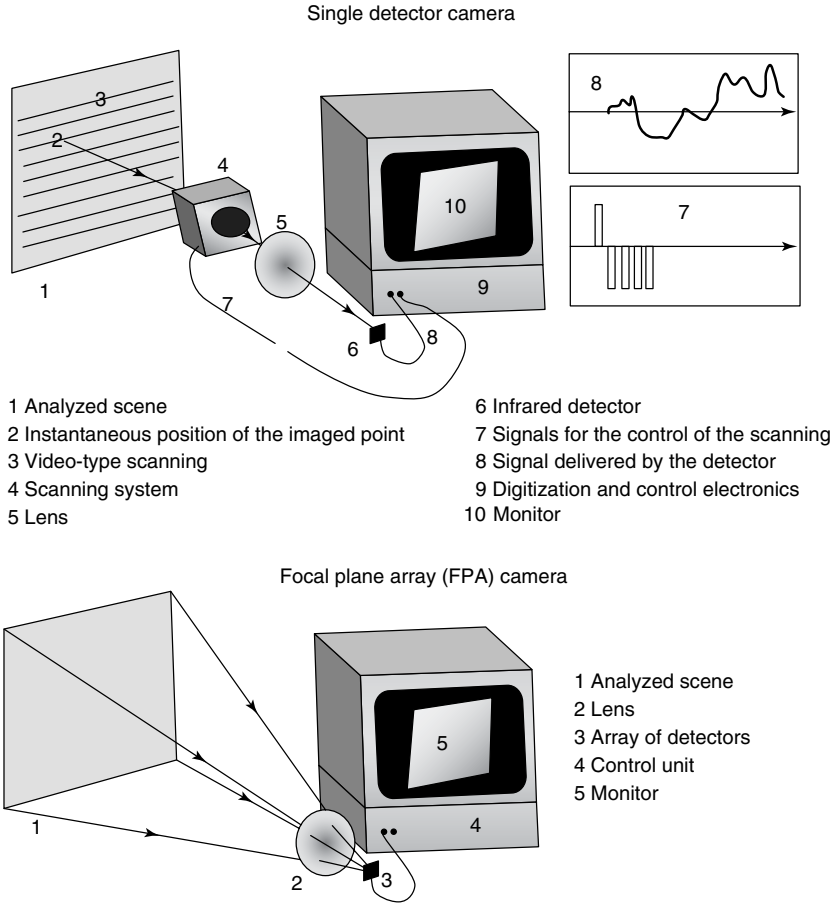


Figure 1. The two types of infrared cameras: single detector cameras and focal plane array (FPA) cameras.

there is no scanning device. Using lenses, the region is imaged onto an array of detectors. This type of camera is much simpler, and the electronics does not need a high bandpass as in the single detector cameras; this leads to a higher signal-to-noise ratio and consequently a higher sensitivity. More details can be found in the books by Gaussorge [7] and by Maldague [8].

3.1 Complexity of a quantitative absolute thermographic measurement

Measuring the distribution of temperature in space and its evolution with time is then possible with a thermographic instrument. Nevertheless, the quantitative measurement of the absolute value of the temperature remains difficult. This result is shown in Figure 2.

The radiant energy received by the camera is a complex function of the temperatures of the object, T_o, surroundings, T_{amb}, and atmosphere, T_{atm}, of the emissivity of the object, ε_o, and of the transmissivity of the atmosphere, τ_{atm}. The simplest case—radiometric signal only depending on the temperature of the object—is obtained when ε_o and τ_{atm} are both equal to 1. The second assumption is generally satisfied when the camera–object distance is of the order of a few meters. To satisfy the first assumption, black coatings are often applied on the object. Another realistic condition can lead to simple interpretation: the object is assimilable to a gray body with a temperature notably higher than the ambient temperature. Then the radiometric

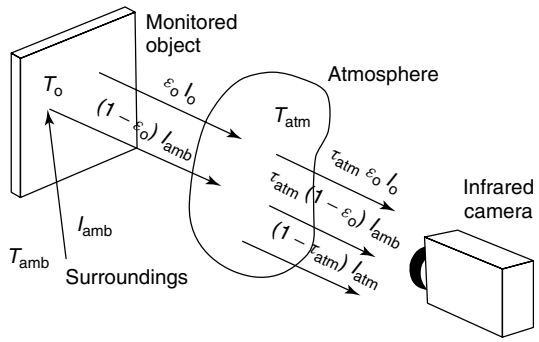


Figure 2. The complexity of the energy balance in a thermographic measurement.

signal depends only on the temperature and the emissivity of the object. These considerations explain why measurements of absolute temperatures by thermography are not very accurate. Generally, the absolute temperature accuracy is near 2 K or 2% of the range [9]. Hopefully, in most of the applications the measurement is relative: the interesting information is the difference of temperature, ΔT, between the monitored point and a reference one, which can be another point of the structure that is supposedly sound, or the same point at a different time, for instance, before stimulation in active thermography.

3.2 Thermographic hardware

Today, two main types of FPA cameras exist: cameras with cooled quantum detectors and bolometric cameras. The first type has the best noise equivalent temperature difference (NETD), the most important characteristic for ΔT measurements. Typically, the NETD is near 20 mK. The second type has a NETD of roughly 100 mK, but does not require a cooling system and its cost is lower. The size of the array is a second discriminating parameter because it influences the space resolution of the thermal images. There are now large size arrays, such as 320×256 or 640×512, and it is possible to make windowing with a correlative frame rate increase (several thousands of frames per second).

It is possible to enhance the NETD by simple data processing [10]: mean value of adjacent pixels ($3 \times 3, 5 \times 5, \ldots$) or mean image of successive images. The NETD is improved in both cases, but with the drawback of respectively decreasing the space or the time resolution. A third possibility that recently appeared is applicable when the monitored thermal phenomena are periodic. In these cases, a lock-in thermographic system permits imaging the modulated part of the thermal field, in amplitude and phase, with an NETD, which can be better than 1 mK, depending on the number of raw images considered by the data processing [11]. Lock-in thermography is particularly useful for experimental mechanics and NDE applications.

3.3 Analysis of the thermographic measurement

The procedure for performing a thermography is presented in Figure 3. There are three "actors" in this process: the thermographer, the thermographic system used, and the monitored structure. The thermographer can be passive, just monitoring the steady state or the time evolution of the structure. This structure state, permanent or evolving, is governed by the interactions with the surroundings. These interactions are the heat transfer, which takes place by conduction, convection, and radiation. A second possibility for the thermographer consists in stimulating the structure with the aim of monitoring its reaction to the stimulation. The result of the stimulation must be thermal (heat sources) to be detected by thermography, but the nature of the stimulation can be extremely varied because heat is the degraded form of all energies: mechanical, electromagnetic, thermal, etc. The information recovered by the thermographic system is the space distribution and/or the time evolution of the structure's surface (assuming the structure is composed of opaque materials). The surface temperature depends

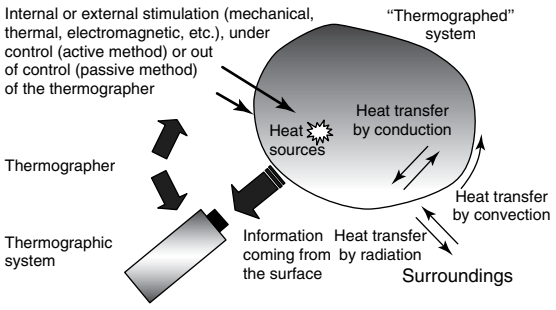


Figure 3. The process of performing thermography.

on (i) the internal structure, (ii) its thermal properties, (iii) its other properties governing the relations existing between the stimulating phenomena and the heat production, and (iv) the possible abnormalities or damages existing inside the structure.

From this analysis we can conclude that (i) to achieve an appropriate and quantitative diagnostic result, the thermographer must have a basic knowledge of heat transfer science; (ii) an active attitude, when possible, is more efficient for establishing a diagnosis; (iii) using mechanical stimulation (pressure, static or dynamic loads, vibrations, waves, etc.) may allow the identification of possible abnormalities of mechanical parameters of the structure, which is useful information to judge the structural health; and (iv) using varied types of stimulations, it is possible to detect thermal abnormalities induced by the presence of defects such as voids, delaminations, debonds, cracks, etc. The last two points are developed next.

4 STRUCTURAL HEALTH MONITORING USING INFRARED THERMOGRAPHY

4.1 Diagnostic based on thermomechanical analysis

The basics of thermomechanical coupling are derived from thermodynamics. We use the generalized standard materials formalism to derive the local heat equation [12]. The equilibrium state of each volume material element is characterized by a set of n state variables: T, the absolute temperature, ε, a strain tensor, and $(\alpha_1, \ldots, \alpha_{n-2})$, the $n-2$ scalar components of the vector $\boldsymbol{\alpha}$ of "internal" variables that characterize the microstructural state of the material. Combining both the first and second principles of thermodynamics, the local heat equation is obtained as follows:

$$\rho C\dot{T} - \mathrm{div}(k\cdot\mathrm{gradT}) = \rho T\frac{\partial^2\psi}{\partial T\,\partial\varepsilon} : \dot{\varepsilon} + \rho T\frac{\partial^2\psi}{\partial T\,\partial\alpha}\cdot\dot{\alpha} + \left(\alpha - \rho\frac{\partial\psi}{\partial\varepsilon}\right) : \dot{\varepsilon} - \rho\frac{\partial\psi}{\partial\alpha}\cdot\dot{\alpha} + r_{\mathrm{e}} \qquad (2)$$

where ψ, the thermodynamic potential, is the Helmholtz free energy, and ρ, C, and k, respectively, the density, heat capacity, and thermal conductivity. On the left side of this equation, two terms describe the changes of temperature versus time and space that can be measured by an IR camera. This left side is in fact the heat diffusion equation without sources. On the right side, the energy production terms (sources) are found. The first two terms represent the thermomechanical coupling sources. Sometimes, these terms are called the reversible coupling between heat sources and mechanical sources. The first term is linked to the time derivative of the strain and represents the heat produced by the change of local strain. The second term, introducing the parameter α, is linked to the internal properties of the material. It may exist and it represents possible interactions between temperature and microstructure. Let us call the sum of these two terms w_{ctm}, the thermomechanical coupling source term. The third and fourth terms are linked to dissipative sources (mainly the plastic energy). The third term, linked to the rate of change of strain, is the product of stress by strain. The fourth term is linked to internal variable α describing internal phenomena. The sum of these two terms, d_1, is the dissipative source term. The last term, r_{e}, represents all external sources.

To solve the heat diffusion equation easily, the following hypotheses are required: the material is thermally linear (thermal properties not depending on temperature) and isotropic (thermal and mechanical properties independent of the direction). Instead of using the absolute temperatures, T, the changes in temperature, θ, can be considered. In this case, the left side of the heat equation can be simplified using the Laplacian of the temperature change:

$$\rho C\dot{\theta} - k\nabla\theta = d_1 + w_{\mathrm{ctm}} + r \qquad (3)$$

In equation (3), temperature and energy are not simply linked, but are related to time and space. To solve this equation, the question of how to link more simply the local value of temperature and the density of energy must be answered. An extra hypothesis is required: the condition of adiabaticity, which assumes that the rate of stress change is such that the conduction inside the material can be considered negligible compared to the rate of heat production. In these conditions, the heat conduction

equation becomes very simple:

$$\rho C\dot{\theta} = d_1 + w_{\text{ctm}} + r \tag{4}$$

The change of temperature is directly linked to the production of energy by the different sources. This result also supposes that the heat losses in the environment are negligible. Under these conditions, variations in temperature are, in fact, equivalent to energy variations. The key question is how can we practically extract energy production from temperature data? There are two ways to solve the problem: (i) standard thermography, which requires recording the full field of the surface temperature, precisely knowing the limit conditions, working on thermally thin specimens, and controlling all heat transfer not connected to the mechanical loading; and (ii) periodic loading and lock-in thermography, which enables one to directly measure energy variations, eliminating all external sources.

Standard thermography gives temperature variations with time for every pixel. With lock-in thermography, the reference parameter is not time but the applied load. Lock-in thermography gives temperature as a function of load. The second solution is more interesting. In periodic regimes (angular frequency ω), adiabaticity is verified if the thermal diffusion length in the specimen, $L_{\text{D}} = (2\kappa/\omega)^{1/2}$, κ being the diffusivity, is smaller than the spatial resolution δ of the thermographic system in the test conditions. This result leads to a condition on the modulation frequency, which depends on instrumentation, test conditions, and the specimen thermal properties: $f \gg \kappa/(\pi\delta^2)$. Critical frequencies calculated using this formula are given in Table 1. Increasing the space resolution requires changing the lens and increasing the modulation frequency of the applied loads at the same time. It is possible with modern cameras to respect these minimum frequencies, in particular, using windowing options for the highest frequencies.

4.1.1 Thermoelasticity and strain measurement

Let us consider the simple case of thermoelasticity. The heat equation is

$$\rho C\dot{\theta} = w_{\text{ctm}} \tag{5}$$

The thermoelastic coupling term of the local heat equation, considering an isotropic standard material, in adiabatic conditions, with harmonic loads and Young modulus independent of temperature, reduces to a simple expression in which α is the thermal expansion coefficient:

$$w_{\text{ctm}} = -T\cdot\alpha\cdot\frac{\mathrm{d}(tr(\sigma))}{\mathrm{d}t} \tag{6}$$

Solving the heat equation leads to the simple following expression for the amplitude, $\Delta\theta$, of the modulated part of the temperature induced by the harmonic loading:

$$\Delta\theta = -K_{\text{M}}\cdot T\cdot\Delta tr(\sigma) \tag{7}$$

which shows that the variation of the temperature is linked to the variation of the applied load or more exactly to the variation of the sum of the principal stresses. This relationship introduces the thermoelastic coefficient of the material, $K_{\text{M}} = \alpha/(\rho C)$. Table 2 presents values of the thermoelastic coefficient for some usual metallic materials.

For composite materials, which are often orthotropic, the relationship is

$$\Delta\theta = -K_M T\left(\Delta\sigma_{11} + \Delta\sigma_{12}\cdot\frac{\alpha_{12}}{\alpha_{11}}\right) \tag{8}$$

Table 1. Critical frequencies to obtain adiabaticity in periodic thermomechanical regime

Critical frequency for adiabaticity (Hz)		Material		
		Carbon/epoxy $\kappa = 1.2\times10^{-6}\text{m}^2\,\text{s}^{-1}$	Stainless steel $\kappa = 7.4\times10^{-6}\text{m}^2\,\text{s}^{-1}$	Aluminum $\kappa = 9.8\times10^{-5}\text{m}^2\,\text{s}^{-1}$
Spatial resolution (mm)	0.50	1.5	9.0	125
	0.15	17	105	1400

Table 2. Thermoelastic coefficient of some usual metallic materials

Material	Cast iron	Iron	Steel	Aluminum
$K_M\ (MPa^{-1})$	2.6×10^{-5}	3.5×10^{-6}	3.1×10^{-6}	2.0×10^{-6}

with $K_M = \alpha_{11}/(\rho C)$, showing that the temperature change is a linear combination of the principal stresses along the material orthotropic axes on the surface.

It is then possible to obtain maps of stresses and, consequently, by comparing the experimental values to references values (theory or experiment on sound structures), to establish a diagnosis of the structural health, and to detect and localize defective areas. In practice, strain maps can be obtained with varied types of mechanical loading: static, periodical, random, etc. Figure 4 presents a stress distribution measured by a thermoelastic stress analyzer CEDIP Infrared System. The system comprises of (i) a JADE IR camera, with a frame rate between 170 Hz and 23 kHz (windowing) and a thermal resolution of 20 mK; (ii) a real-time lock-in module with a frequency range of 0–19 kHz giving a stress resolution of 2 MPa for steel and an adiabatic spatial resolution down to 10 μm; and (iii) a dedicated software, Altair-Li. Stresses are measured during a fatigue test of the structure (a steering knuckle). The load frequency is 5 Hz. Four types of images are produced: (i) a thermal image presenting the mean value of the absolute temperature, in a gray scale; (ii) the peak-to-peak value of the modulated part of temperature, in false colors; (iii) the phase in a color spectrum between blue (areas in tension) and red (areas in compression), black areas corresponding to pixels where amplitudes are null and phases have no sense and are not calculated; and (iv) the local stress distribution. This last image is based on the knowledge of the thermoelastic coefficients and the absolute temperature distribution. It is clear that thanks to the good signal-to-noise ratio of this image the presence of a damage, a crack, for instance, would produce a very local stress contrast, which would be easily detected in the fourth image.

An illustration of the technique at higher frequencies is given in Figure 5. The camera integration time is lower than 100 μs permitting frequency exploration up to 10 kHz. The same system produces stress distributions for the first four modes of vibration of a turbine blade. All images are obtained with an equal processing time of 7 s. A localized damage mechanism would be easily detected and it could be possible to compare the images corresponding to the different modes to deduce information on the size of the defective area.

4.1.2 Dissipative phenomena and fatigue

Structures in service are often subject to fluctuating stresses. Repeated loads, which may or may not be periodic, change and degrade the material properties and may lead, in the long term, to failure (fatigue damage). This phenomenon occurs at levels of stress significantly lower than the resistance limit measured in static conditions, and even lower than the yield strength. Fatigue phenomena produce a continuous degradation of the microstructure. Intracrystalline microcracks first appear, gradually evolving because the stress in their vicinity can be much higher than the mean stress level. Then, they cross the grains when the number of cycles increases (nucleation phase) and eventually coalesce to form a macrocrack oriented in the maximum shear stress direction. Later in the growth phase, the macrocrack propagates perpendicularly to the maximum tensile stress direction, until it reaches a critical size at which the cracked part becomes unstable. The crack then propagates rapidly, leading to catastrophic failure.

Generally, damage occurs without any previous visible modification of shape or aspect. That is why fatigue failure is so threatening. Therefore, there is an interest, in the frame of the SHM approach, for the early detection of damage and also for evaluation of the residual life. This need in SHM is why thermomechanical coupling analysis by means of thermography is a promising technique [13].

These irreversible processes correspond to heat sources described by the term d_l of equation (2). During the mechanical loading of a structure, thermal effects related to these dissipative phenomena can be measured. Let us consider the application of a periodic load. The total temperature change, ΔT, results

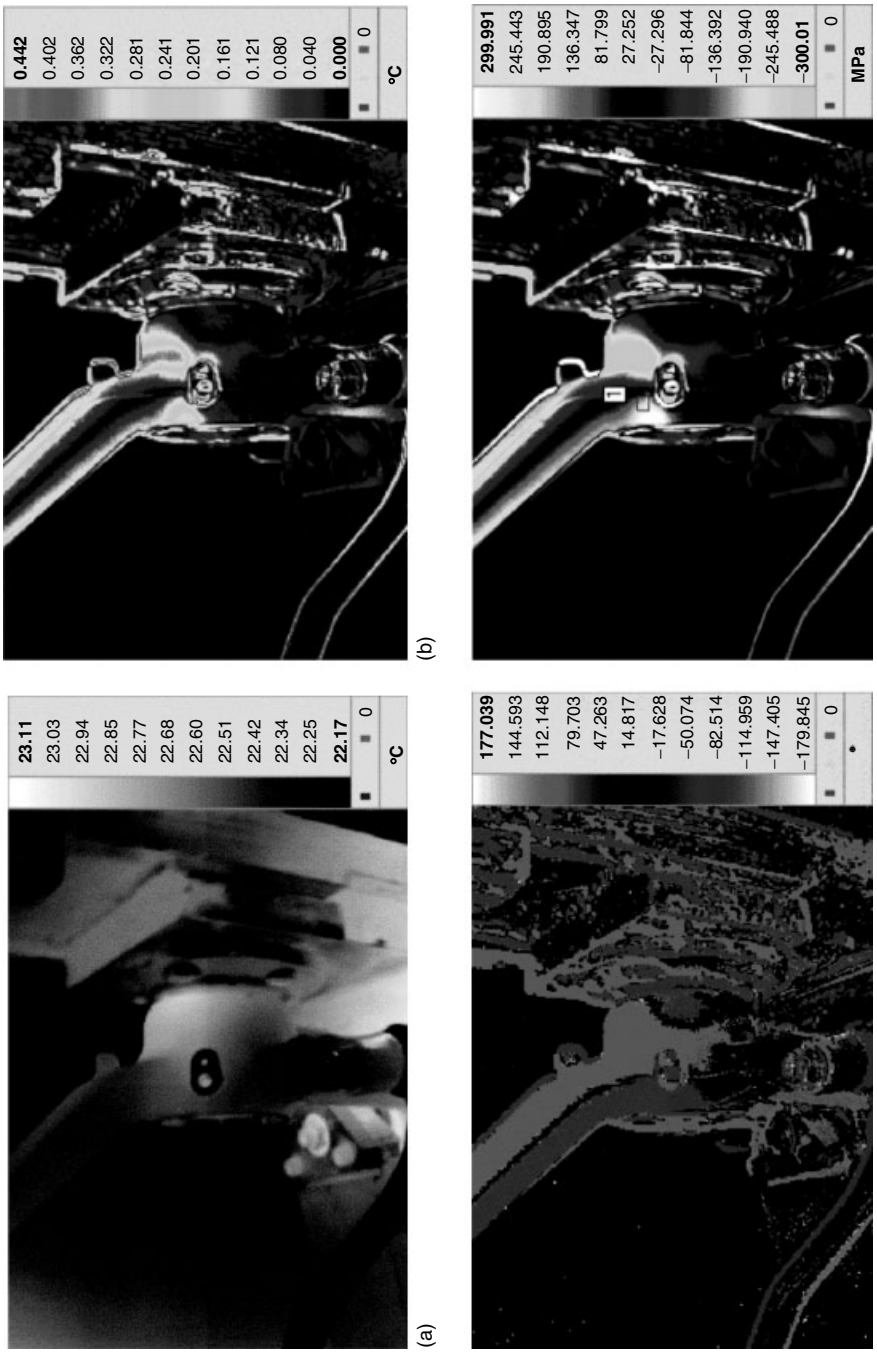


Figure 4. Images of a steering knuckle obtained with a thermographic stress analyzer. (a) Raw thermal image, (b) modulated temperature amplitude, (c) phase image, and (d) stress map. [Reproduced by permission of CEDIP Infrared System.]

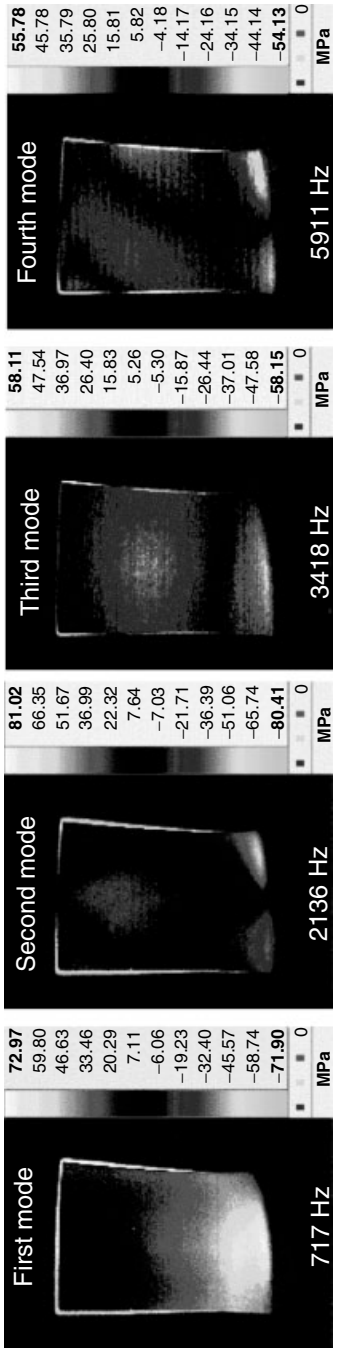


Figure 5. Visualization of the thermoelastic coupling linked to the first modes of a turbine blade. [Reproduced by permission of CEDIP Infrared System.]

from the superposition of temperature variations due to the thermoelastic effect, ΔT_{el}, the dissipated heat, ΔT_{diss}, and the heat energy losses to the environment. Figure 6 schematically presents the variation of the material temperature, ΔT, during a sinusoidal mechanical loading process. Just after the start of the mechanical loading (short-term loading), the time-dependent temperature increase is adiabatic (no influence of the heat losses) and the dissipative contribution, ΔT_{diss}, is linear with time. The mean slope of the curve only depends on dissipation, defining a thermal parameter characteristic of these irreversible phenomena: $\Delta\tau_{diss}$, the temperature change due to heat dissipation per single mechanical loading cycle ($\Delta\tau_{diss} = \Delta T_{diss}/N$, where N is the number of initial cycles). The slope progressively decreases and finally becomes zero when an equilibrium is reached between the internal generation of heat and the heat losses. Measuring this initial slope with standard thermography is a possible technique to monitor the damage process. Rösner *et al.* [14] applied a 30-Hz-cyclic loading ($\sigma_{alt} = 400\,\text{MPa}$, $\sigma_{mean} = 468\,\text{MPa}$) to Ti–6Al–4V specimens and monitored the temperature evolution with an IR camera. The chosen stress amplitude led to failure after about 8×10^4 cycles. Temperature increased continuously until failure occurred. After stopping the cyclic loading and then reloading, the new $\Delta\tau_{diss}$ was shown to be an increasing function of the number of already applied stress cycles.

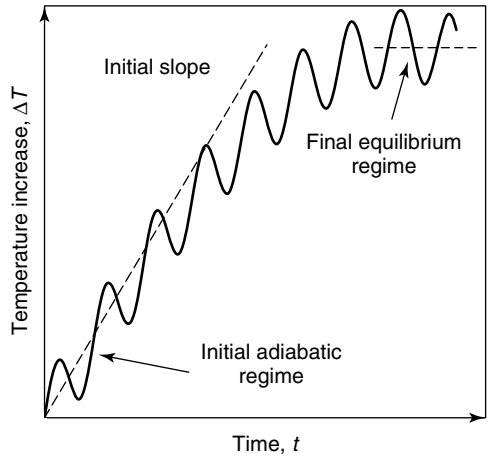


Figure 6. Temperature–time variation during short-term mechanical loading and evaluation of the initial slope.

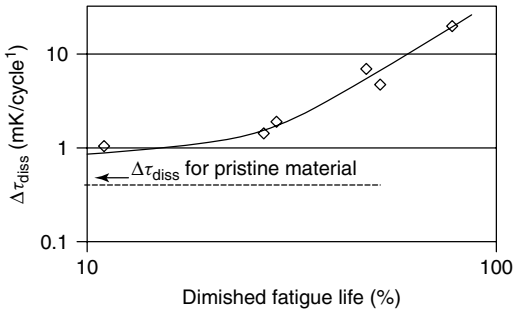


Figure 7. Comparison of the thermal parameter measured on several specimens with varied percentage of diminished fatigue life.

Thus, heat dissipation measurement yields information about the residual life of the structure Figure 7). The same authors later demonstrated that for heat dissipation measurements, short-term loading with a reduced stress could be applied too. For this purpose, they applied both a tensile stressing at 30 Hz and an ultrasonic excitation at 20 kHz [15].

4.2 Damage detection by stimulated thermography

In the previous section of this article, thermographic techniques used in experimental mechanics were presented as candidate techniques for SHM. They have the potential of assessing the distribution of stresses and dissipative heat sources.

Another approach in the use of thermography for SHM consists of detecting the damage itself acting as a heterogeneity of thermal properties or a discontinuity in the internal structure arrangement. This approach comes from classical NDE. The technique involved in this approach is called *stimulated thermography*. The stimulation is generally a flux of photons (*photothermal thermography*). This type of technique is described in the following section.

A new and promising way to monitor defects consists of propagating ultrasound in the structure and detecting the heat sources generated by their interaction with the defects. This approach is very briefly mentioned here.

4.2.1 Photothermal thermography

There are mainly two types of photothermal thermography: the pulsed technique, which can be applied with standard thermographic systems, and the modulated technique, one requiring lock-in thermographic systems.

Pulsed or transient photothermal thermography
Using thermal models [16–18] developed for photothermal radiometry at the end of the 1980s and inversion methods [19–23], the first quantitative thermographic works appeared at the beginning of the 1990s. During this period, this approach was the subject of numerous works and led to robust procedures allowing the production of *depthgrams* (equivalent to an ultrasonic D-scan) and evaluation of the equivalent thermal resistance for defects parallel to the structure surface [24–27].

With such inversion procedures, pulsed photothermal thermography has become a well-established technique in NDE, very attractive for the rapid obtainment of extended images. The comparison of such a technique [25] with a very different NDE technique (Compton backscatter) is presented in Figure 8 to show the ability of thermography to recover the exact shape of delaminations inside a complex composite structure.

Modulated or lock-in photothermal thermography
Lock-in thermography has a similar history, with its basis coming from photothermal radiometry in the 1980s [28]. Very soon, it appeared that the problem of the lack of rapidity of the measurements due to point-by-point scanning could be solved by a lock-in thermography technique [29–31]. In the 1990s several groups actively developed this technique and correlated inversion procedures for NDE purposes [32–35].

An example of application of the technique to an aircraft, the Dornier Do328, is given in Figure 9. The modulation frequency is 0.015 Hz. The area of the aircraft monitored is the tail core. The stiffeners are very visible on the phase image obtained. The image confirms the integrity of the structure. Of course, the analysis of this IR image requires a thorough knowledge of the internal structure.

Both techniques, pulsed and modulated, have their advantages and drawbacks. The lock-in technique, thanks to the production of phase images, does not have the problem of normalization caused by the unavoidable nonuniformity of practical heat depositions, but requires several experiments when the depth

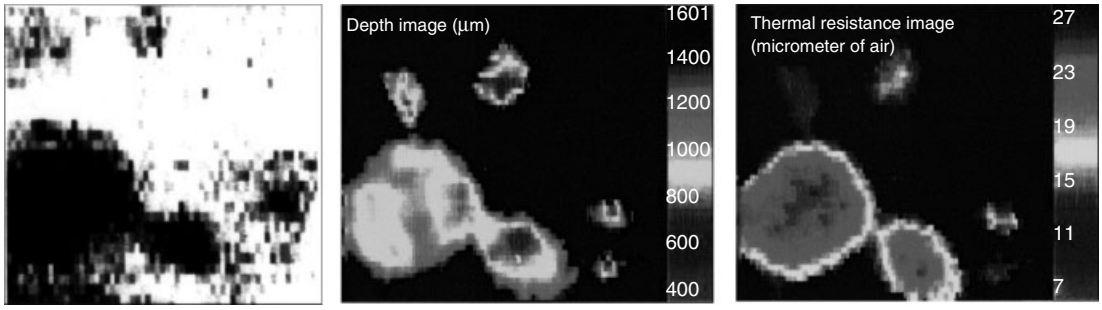


Figure 8. Defects in the external thermal barrier of the Hermes space shuttle (C/C and SiC impregnated C/C composites) as seen by (a) Compton backscatter and (b, c) pulsed photothermal thermography techniques.

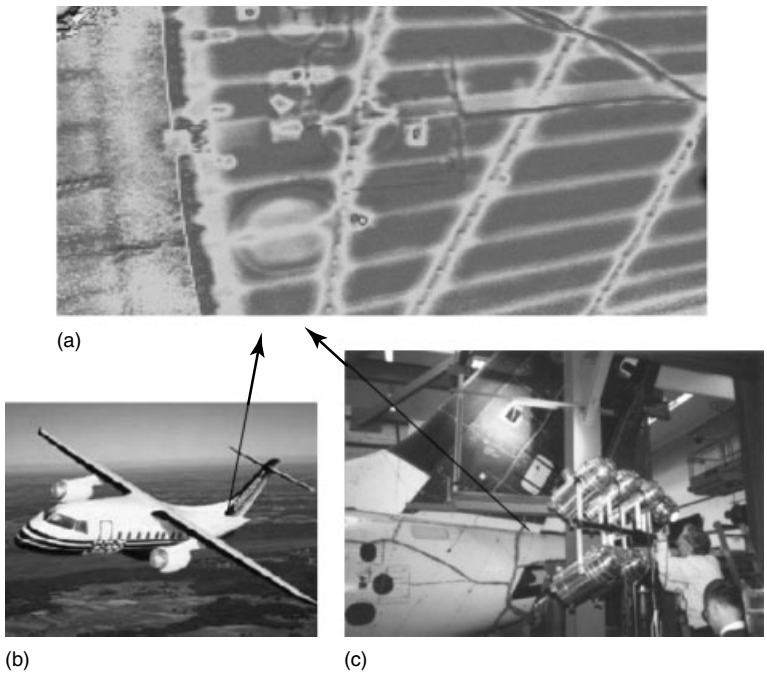


Figure 9. Inspection of the tail of an aircraft using lock-in thermography NDE. (a) Thermographic image showing the in-depth structure of the base of the tail of the Dornier Do328 aircraft, (b) full view of the aircraft, and (c) lock-in thermography setup during ground tests showing, in particular, the bank of high intensity lamps powered by an electric current with a tension modulated at very low frequency (<1 Hz). Images taken from [36]. [Reproduced by permission of IKT-ZfP.]

of the damage is unknown. Furthermore, the inverse problem is not simple. On the contrary, the pulsed technique permits finding with a unique test a damage site lying at an unknown depth; the inverse problem is more simple, allowing the use of analytical algorithms to deduce the depth and equivalent thermal resistance of the damage [19]. In compensation, normalization is required. The most accurate solution consists of using the early detection of the thermal contrast [26] and not the maximum contrast as commonly used by most of the experimenters.

More recently, a hybrid approach was proposed [37]: the pulsed thermogram is analyzed by a Fourier transform, which permits the establishment of phase

images not requiring normalization. This technique, called *pulsed phase thermography*, is frequently used nowadays, but until now no really objective and quantitative comparative studies between this technique and the other two have been done.

4.2.2 Ultrasonic vibrothermography

Ultrasonic vibrothermography is an NDE technique based on the application of modulated mechanical stresses produced by the propagation of ultrasounds in the structure, while an IR camera maps the surface temperature [38]. Thermomechanical coupling is responsible for heat production, particularly in damaged regions, which convert energy into heat through enhanced viscoelastic dissipation, collisions, and/or rubbing of internal free surfaces present in delaminations and cracks. Surface and internal defects appear hotter in surface temperature images. This recent technique is presented in detail elsewhere in the Encyclopedia (*see* **Nondestructive Evaluation of Cooperative Structures (NDECS)**, Section 6).

5 CONCLUSIONS

After a recall on radiometry and thermal imagery, a very general analysis of the thermographic measurement has been made, emphasizing the particular interest of active thermography, which can be used to monitor the structural health state of a structure through the detection of abnormalities in the strain distribution, dissipative heat sources or disturbances of the temperature field after a stimulation of the structure. Then, two types of active thermography applicable to SHM have been detailed: thermography with thermomechanical stimulations and photothermal thermography. Ultrasonic vibrothermography, a third type of active thermography, has been only briefly mentioned since this technique is presented elsewhere in detail (*see* **Nondestructive Evaluation of Cooperative Structures (NDECS)**).

These powerful techniques are more NDE techniques than SHM techniques because the detection is achieved with a camera, without contact, from outside the structure. Nevertheless, an attempt to couple this detection system with embedded stimulations could lead to a hybrid or semi-integrated approach (*see* **Nondestructive Evaluation of Cooperative Structures (NDECS)**).

RELATED ARTICLES

Static Damage Phenomena and Models

Damage Evolution Phenomena and Models

Failure Modes of Aerospace Materials

REFERENCES

[1] Lei JF. High temperature thin film strain gages. *HITEMP Review* 1994 **I**:25.1–25.13, NASA CP-10146.

[2] Stasiek J. Thermochromic liquid crystals and true colour image processing in heat transfer and fluid flow research. *Heat and Mass Transfer* 1997 **33**:27–29.

[3] Ireland PT. The response time of a surface thermometer employing encapsulated thermochromic liquid crystals. *Journal of Physics E: Scientific Instruments* 1987 **20**:1195–1199.

[4] Hobbs RJ, Currall JEP, Gimingham CH. The use of 'thermocolor' pyrometers in the study of heath fire behaviour. *Journal of Ecology* 1984 **72**:241–250.

[5] Allison SW, Gillies GT. Remote thermometry with thermographic phosphors: instrumentation and applications. *Review of Scientific Instruments* 1997 **68**(7):2615–2650.

[6] DeWitt DP, Nutter GD. *Theory and Practice of Radiation Thermometry*. John Wiley & Sons: New York, 1988, pp. 1138.

[7] Gaussorgue G. *La Thermographie Infrarouge: Principle, Technologie, Applications*. Editions Tech & Doc: Paris, 1999, pp. 586.

[8] Maldague X. *Theory and Practice of Infrared Technology for Non Destructive Testing*, *Wiley Series in Microwave and Optical Engineering*, Chang K (ed). Wiley: 2001, pp. 684.

[9] Pron H, Bissieux C. Focal plane array infrared cameras as research tools. *QIRT Journal* 2004 **1**(2):229–240.

[10] Levesque L, Brémond P, Lasserre LL, Paupert A, Balageas DL. Performance of FPA IR cameras and of their improvement by time, space and frequency data processing. Part I: intrinsic characterization of the thermographic system. *QIRT Journal* 2005 **2**(1):97–112.

[11] Breitenstein O. *Lock-in IR Thermography—Basics and Use for Functional Diagnostics of Electronic Components*. Springer: Berlin, 2003.

[12] Germain P, Nguyen QS, Suquet P. Continuum thermodynamics. *Journal of Applied Mechanics* 1983 **50**:1010–1020.

[13] Krapez JC, Pacou D. Thermography detection of damage initiation during fatigue tests. Thermosense XXIV. *Proceedings of SPIE 4710*. SPIE: Bellingham, WA, 2002; pp. 435–449.

[14] Rösner H, Sathish S, Meyendorf N. Nondestructive characterization of fatigue damage with thermography. *Proceedings of SPIE 4336*. SPIE: Bellingham, WA, 2001; pp. 167–175.

[15] Meyendorf N, Rösner H, Kramb V, Sathish S. Thermo-acoustic fatigue characterization. *Ultrasonics* 2002 **40**:427–434.

[16] Vavilov V, Taylor R. Theoretical and practical aspects of the thermal NDT of bonded structures. *Research Techniques in Non Destructive Testing* 1982 **5**:239–280.

[17] Vavilov V, Ivanov A. Pulse thermal testing of multilayer specimens. *Soviet Journal of Nondestructive Testing* 1984 **6**:39–47.

[18] Balageas D, Krapez JC, Cielo P. Pulse photothermal modeling of layered materials. *Journal of Applied Physics* 1986 **59**(2):348–357.

[19] Balageas DL, Déom AA, Boscher DM. Characterization and non destructive testing of Carbon-epoxy composites by a pulsed photothermal method. *Materials Evaluation* 1987 **45**(4):461–465.

[20] Balageas DL, Boscher DM, Déom AA. Temporal moment method in pulsed photothermal radiometry. Application to carbon-epoxy NDT. *Photoacoustic and Photothermal Phenomena*, *Springer Series in Optical Sciences*, Springer: 1987; Vol. 58, pp. 500–502.

[21] Maclachlan Spicer JW, Kerns WD, Aamodt LC, Murphy JC. Time-resolved infrared radiometry of multiplayer organic coatings using surface and subsurface heating. Thermosense XIV. *Proceedings of SPIE 1467*. SPIE: Bellingham, WA, 1991; pp. 311–321.

[22] Krapez JC, Maldague X, Cielo P. Thermographic NDE: Data inversion procedure (Part II: 2D analysis and experimental results). *Research in Nondestructive Evaluation* 1991 **3**(2):101–124.

[23] Favro LD, Crowther DJ, Kuo PK, Thomas RL. Inversion of pulse-echo thermal- wave images. Thermosense XIV. *Proceedings of SPIE 1933*. SPIE: Bellingham, WA, 1993; pp. 138–141.

[24] Krapez JC, Boscher DM, Delpech PH, Déom AA, Balageas DL. Time resolved pulsed stimulated infrared thermography applied to carbon-epoxy non destructive evaluation. *Proceedings of QIRT Conference*. Editions Europ. Therm. et Ind.: Paris, 1992; pp. 195–200.

[25] Delpech PH, Boscher DM, Lepoutre F, Déom AA, Balageas DL. Time resolved pulsed stimulated infrared thermography applied to carbon-carbon non destructive evaluation. Proceedings of QIRT Conference. Editions Europ. Therm. et Ind.: Paris, 1992; pp. 201–206.

[26] Krapez JC, Balageas D, Déom A, Lepoutre F. Early detection by stimulated infrared thermography. Comparison with ultrasonics and holo/shearography. *Advances in Signal Processing for Nondestructive Evaluation of Materials*, *NATO ASI Series E*. Kluwer Academic Publishers, 1994; Vol. 262, pp. 303–321.

[27] Vavilov VP, Kourtenkov DG, Grinzato E, Bison PG, Marinetti S, Bressan C. Inversion of experimental data and thermal tomography using "Thermo.Heat" and "Termidge" software. *Proceedings of QIRT Conference*. Ed. Europ. Therm. et Ind.: Paris, 1994; pp. 272–278.

[28] Nordal PE, Kanstad SO. Photothermal radiometry. *Physica Scripta* 1979 **20**:659–662.

[29] Carlomagno GM, Berardi PG. Unsteady thermography in nondestructive testing. *Proceedings of the 3rd Biannual Information Exchange*. St Louis, MO, 1976; pp. 33–39.

[30] Beaudoin JL, Mérienne E, Danjoux R, Egée M. Numerical system for infrared scanners and application to the subsurface control of materials by photothermal radiometry. Infrared technology and applications. *Proceedings of SPIE 590*. SPIE: Bellingham, WA, 1985; pp. 287.

[31] Kuo PK, Feng ZJ, Ahmed T, Favro LD, Thomas RL, Hartikainen J. Parallel thermal wave imaging using a vector lock-in video technique. *Photoacoustic and Photothermal Phenomena*, *Springer Series in Optical Sciences*, Springer: 1988; Vol. 58, pp. 415–418.

[32] Busse G, Wu D, Karpen W. Thermal wave imaging with phase sensitive modulated thermography. *Journal of Applied Physics* 1992 **71**:3962–3965.

[33] Almond DP, Patel PM. A quantitative thermal wave assessment of the characteristics of sub-surface defects. *Proceedings of QIRT Conference*. Editions Europ. Therm. et Ind.: Paris, 1992; 367–370.

[34] Favro LD, Ouyang Z, Wang Li, Han Xun, Feng Z, Thomas RL. Infrared video lock-in imaging at high frequencies. *Proceedings of SPIE 3056*. SPIE: Bellingham, WA, 1997; pp. 184–187.

[35] Krapez JC. Compared performances of four algorithms used for modulation thermography. *Proceedings of QIRT Conference*. Akademickie Centrum Graficzno-Marketingowe Lodart S.A.: Lodz, Poland Paris, 1998; pp. 148–153.

[36] Wu D, Salerno A, Malter U, Aoki R, Kochendorfer R, Kächele PK, Woithe K, Pfister K, Busse G. Inspection of aircraft structural components using lockin thermography. *Proceedings of QIRT Conference*. Edizioni ETS: Italy, 1996; pp. 251–256.

[37] Maldague X, Marinetti S. Pulse phase infrared thermography. *Journal of Applied Physics* 1996 **79**(5):2694–2698.

[38] Krapez JC, Taillade F, Balageas D. Ultrasound-lock-in thermography NDE of composite plates with low power actuators. Experimental investigation of the influence of the Lamb wave frequency. *QIRT Journal* 2005 **2**(2):191–206.

Chapter 20
Eddy-current Methods

Neil Goldfine, Andrew Washabaugh, Yanko Sheiretov and Mark Windoloski
JENTEK Sensors, Inc., Waltham, MA, USA

1 INTRODUCTION

Eddy-current testing (ET) is a nondestructive testing (NDT) technique that uses time-varying magnetic fields to induce eddy currents in a conducting material. ET is used to assess the material under test conditions or to detect flaws. A time-varying current in a drive coil, at a prescribed frequency or with a prescribed shape in time (i.e., a pulse), creates a time-varying magnetic field that induces an eddy current in the material with a pattern that follows the drive coil geometry. The electrical conductivity, magnetic permeability, and thicknesses of material layers (or process-affected zones), as well as the sensor proximity to the surface (i.e., the liftoff), affect the sensed response at the terminals of the drive coil or at the terminals of one or more sensing elements. Sensing elements can be inductive coils that respond to time-varying fields or alternative field sensors, e.g., magnetoresistive sensors. The sensitivity of different sensing element types is not as important as the actual signal to noise achievable for a given application. The signal to noise is a measure of the sensor response to a change in specific material condition or to a specific flaw type and size of interest compared to all relevant sources of noise/errors. This comparison must be made under actual (or at least representative) inspection conditions. The presence of a flaw (e.g., crack, inclusion, and porosity) or a variation in a material condition (e.g., residual stress effects on magnetic permeability) is sensed by the ET sensor (one sensing element) or ET array (multiple sensing elements) in the region immediately under the sensor. Thus, ET is a local measurement, but ET can provide rapid wide area coverage using manual or automated scanners.

In the context of structural health monitoring (SHM), ET is one of several techniques that the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

maintainers of high-value assets can select as part of an asset life management program. In this article, conventional ET and advanced ET methods are described. The focus is on scanning and point-by-point ET using handheld or automated systems.

In a companion article on eddy-current *in situ* sensors (*see* **Eddy-current *in situ* Sensors for SHM**), the focus is on surface-mounted and embedded sensors for SHM, using onboard instrumentation or portable data-acquisition units. Using a portable data-acquisition system with onboard ET sensors is a direct replacement for conventional ET (NDT), but with the advantage that difficult-to-access locations can be inspected without disassembly to gain access. However, as described in this and the companion article, the achievable signal-to-noise ratio will vary for three types of ET: (i) single inspections with point-by-point measurements or scanning arrays, (ii) embedded or surface-mounted sensors with portable instruments for scheduled or unscheduled inspections of difficult-to-access locations, and (iii) continuous monitoring of onboard sensors with onboard instrumentation.

This article, in combination with the *in situ* sensor article *see* **Eddy-current *in situ* Sensors for SHM**, is designed to help operators and maintainers make intelligent decisions based on the realistic capabilities of today's NDT and SHM solutions, as well as the anticipated new NDT and SHM developments expected within the near future. It is particularly important to remember that NDT capabilities are not stagnant and continue to evolve. Thus, future SHM developments should be compared to anticipated advances in NDT and not only current methods.

This article is organized as follows: (i) Section 2 provides the authors' perspective on the roles of NDT and SHM in life management of legacy and new assets; (ii) Section 3 lists different electromagnetic methods, such as magnetic flux leakage (MFL) and remote field eddy-current methods, and explains how these methods relate to ET; (iii) Section 4 describes the important issues that must be addressed when selecting and applying ET methods; (iv) Section 5 provides a brief introduction to ET as a foundation for understanding the new capabilities offered by advanced methods; (v) Section 6 describes the MWM®-Array ET method, as an advanced ET method that addresses many of the limitations of conventional ET; and (vi) Section 7 describes the importance of real-crack standards when using NDT methods and provides an overview of ET methods for the SHM encyclopedia, because surface-mounted ET sensors offer a new method for generating real-crack standards.

2 BACKGROUND

This encyclopedia of SHM is designed to provide a foundation for maintainers and operators of high-value assets to select the appropriate solution for asset life management. Historically, SHM has been limited to measurement of usage states such as strain, temperature, and vibrations. However, direct monitoring capabilities for fatigue and corrosion damage are now becoming available. In this transition period, for the next decade or so, a balance between conventional NDT, advanced NDT, and SHM solutions will most often be appropriate. Even in the long term, SHM is not expected to replace NDT completely for legacy or even next-generation aircraft, ships, pipelines, bridges, and other high-value assets. Thus, a comprehensive adaptive life management (ALM) approach would incorporate both SHM and NDT data for decision support.

Furthermore, diagnostic methods that temporarily place and/or monitor sensors on structures, e.g., for stress or temperature sensing, offer valuable solutions for some applications, such as bridge assessments and machinery testing, without providing an integrated SHM capability. New magnetic stress gauges and guided-wave ultrasonic methods offer new promise for such diagnostic needs, without requiring significant surface preparation and with limited interference with asset availability.

NDT has clear advantages over SHM for some applications. For example, advanced NDT methods such as locally applied phased array ultrasonics or meandering winding magnetometer (MWM) and MWM-Array ET are highly sensitive and reliable for scanning of large areas and for detecting relatively small cracks. Currently available wide area SHM methods cannot match the sensitivity and reliability of these advanced NDT methods, but new methods such as guided-wave ultrasonic testing (GWUT) may offer solutions for some wide area inspection needs. Other local SHM methods, such as surface-mounted ET, may offer high sensitivity, but still will not offer all of the advantages and reliability of scanning methods for all applications.

This is an important practical distinction. Highly sensitive results have been demonstrated in recent years by SHM tests performed in a laboratory environment, but this sensitivity does not always translate into achievable measurement capabilities for installed systems on aircraft. Controlled laboratory testing over relatively short periods does not capture the realistic conditions encountered when monitoring components over years or decades in service (e.g., cable and sensor degradation and instrumentation drift). For example, one advantage of continuous fatigue monitoring in a laboratory is that permanently mounted sensors (e.g., ET or ultrasonic testing (UT)) can monitor response changes over relatively short periods (such as hours or days) avoiding substantial instrumentation variations, while continuously monitoring the crack growth without interruption.

In practice, continuous monitoring with SHM requires highly reliable, *in situ* (i.e., on aircraft) instrumentation. This is necessary to avoid gaps in health monitoring, during which damage can grow substantially (or an event, e.g., overload or impact, can occur) reducing the reliability of such onboard methods for monitoring a specific defect initiation and growth. The problem with SHM is that onboard sensors can lose their reference frame if cracks or other damage can grow when data is not being acquired. For NDT, neighboring areas with no damage can usually be scanned for comparisons (it is not always sufficient to add channels in areas that do not experience damage, since, for example, channel drift may vary). In the near term (over the next five years) before onboard instrumentation is widely available, SHM using direct damage monitoring sensors, such as ET and UT or even GWUT, will rely on portable data-acquisition units for most implementations; thus, performance will be similar or possibly substantially worse than that achievable with conventional or advanced NDT scanning methods.

Consequently, the main advantage of direct damage monitoring SHM in the near term will be access to difficult-to-access locations without requiring substantial disassembly. This is a big advantage in many applications in which a few (up to 50) critical locations must be inspected to enable life extension/sustainment of legacy fleets. There are, however, many applications that cannot be addressed by NDT. For example, NDT methods cannot protect against rapid damage evolution, e.g., from foreign object damage (FOD) in flight. Thus, onboard vibration sensing or other sensing modes for rapid (and large) damage evolution offer immediate benefits for avoiding catastrophic failures from such unpredictable events.

Thus, before selecting an onboard sensor solution for SHM, it is important to consider the advancements of ET and other NDT methods. Many of these advances have been aimed at increasing the reliability of the measurement. Unfortunately, the range of performance offered by different sensor designs and data analysis methods has broadened, making it increasingly difficult to define ET capabilities in general terms. Although probability of detection (POD) studies on real-crack specimens are recommended, in practice, the costs and convenience of such studies are not always practical. Many studies use electrical discharge machined (EDM) notches or other simulated flaws (flat bottom holes) that are far from representative of actual damage. Care should be taken to assess the actual achievable signal-to-noise ratio for in-service generated flaws, before setting safety margins based on POD performance on simulated flaws. This is often difficult.

This article does not attempt to provide a comprehensive history or catalog of methods; instead, the goal is to provide a perspective on the advances of the last decade and the promise of improved capabilities in the next decade. Conventional ET methods are being increasingly used in the aerospace, automotive, energy, infrastructure, and manufacturing sectors. In many cases, these methods have demonstrated high reliability compared to visual inspection and other NDT methods, resulting in wide acceptance for in-service inspection and manufacturing quality assurance. However, anecdotal stories about conventional NDT performance failures are common and performance in-service is often well below the expectations set by POD studies. It is not unusual to hear comments from world renowned NDT/POD experts such as "sometimes ... we are performing inspection ceremonies", or from high-level military officials, saying "it's as if it was never inspected" after large cracks were found on multiple aircraft in a fleet, even after repeated inspections.

The growing divide between the performance that can be achieved with a modern automated method and that which can be achieved with a conventional operator-driven ET or UT method further

confuses this issue. For many applications, conventional methods are more than adequate. However, for some applications, such as critical engine component and landing gear inspection of shot peened surfaces, conventional methods simply do not work reliably (for detection of relatively small cracks—as now required in many cases, to ensure safety), and more advanced methods are needed.

Furthermore, conventional ET methods require different training than automated ET methods, not only for operation but also for assessment. Understanding how to use a conventional ET method, even for "Level II" and "Level III" practitioners, is not necessarily adequate background for assessing advanced methods. At times, this makes it difficult for senior management to obtain correct assessments from their NDT departments regarding the capabilities of advanced NDT methods, especially when compared to the conventional methods that these practitioners are most comfortable with. Unbiased management oversight is absolutely needed to ensure that the best solutions are used for each application. Entrenched practices that give comfort to management (and continuity with historical practice) and do not evolve with changing mission requirements are common and can result in significant lapses in safety, and sometimes huge increases in life cycle costs (i.e., when large component populations are unnecessarily retired, or when undetected damage progresses to levels requiring far more expensive repairs than necessary).

It is a common perception that there is value in maintaining the consistency of future inspections with historical ET and UT databases even though the performance of such methods may be far inferior to that achievable with newer advanced methods. This can be avoided. Historical ET and UT data can be replaced with modern methods without losing continuity by providing an overlap period during which both methods are employed. This is already happening in isolated cases. Fortunately, we expect this to be the rule moving forward—not the exception.

Still, advanced methods must be proven before they are introduced widely. One often misstated disadvantage of advanced NDT methods is that they are "proprietary" and therefore will result in higher costs. This is simply untrue and the evidence strongly indicates that many implementations of advanced methods offer substantial savings in reduced labor costs, increased reliability, and reduced false indication rates. Furthermore, even conventional implementations of NDT by original equipment manufacturers (OEM)s are often considered proprietary; thus, the apparent advantages of conventional (open access) methods are more perceived than real.

Another common misconception is that very sensitive methods will detect defects so small that too many "good" parts will be retired. Modern methods such as the MWM-Array ET method, described later, offer new size thresholding capabilities that can suppress detection of defects below a prescribed threshold, while more reliably detecting flaws with sizes above the threshold. This discrimination capability will actually result in retirement of far fewer "good" parts, and dramatically reduce the most important safety risk, i.e., missing large defects.

The goal of this article is to provide the reader with an understanding of the capabilities of both conventional and advanced/automated ET methods. This is not intended to replace training in specific methods. Any NDT method must be qualified for each application by an experienced professional. Even fully automated NDT methods require training to enable proper interpretation of results for life management decision support.

3 ELECTROMAGNETIC TESTING METHODS

Electromagnetic fields are commonly used in NDT to detect defects and characterize material conditions, including manufacturing quality and serviceability. Electromagnetic NDT methods can be divided into three groups: those that operate in the dynamic regime (typically above 100 MHz), those that operate in the quasistatic regime (typically between 10 Hz and 50 MHz), and those that use constant fields (generated by permanent magnets or direct current (dc) sources) [1, 2]:

- Electromagnetic methods in the dynamic regime include
  - ground penetrating radar (GPR, frequency range: 30–3000 MHz);
  - visual inspection (frequencies of order 10^{15} Hz);

  - microwave (frequency range: 300 MHz to 300 GHz).
- Electromagnetic methods that operate in the quasistatic regime include
  - low-frequency and high-frequency ET (10 Hz to 40 MHz); and
  - capacitive sensing (i.e., dielectric spectroscopy; 0.0001 Hz to 10 MHz).
- The quasistatic regime also includes electromagnetic methods that use constant fields such as
  - MFL, which uses a permanent magnetic source; and
  - magnetic particle inspection (MPI), which magnetizes ferromagnetic components.

Some methods can operate in both the quasistatic or constant field modes. For example, some methods use inductive coils to excite either time-varying (alternating current, ac source) or constant fields (dc source) and use sensors that measure field intensity (i.e., using Hall sensors or magnetoresistive sensors) instead of coils that sense the rate of change of the field (as accomplished in a conventional ET method).

This article describes methods that operate in the quasistatic mode, called *magneto-quasistatics* [1]. This mode includes both quasistatic, time-varying magnetic fields and constant magnetic fields. Magneto-quasistatics is defined as the regime within which the wavelength of traveling waves is significantly longer than the characteristic dimensions of the sensor, coatings/process-affected zone, liftoff, or other system dimensions.[a] This includes ET, MFL, MPI, and other methods that take advantage of time-varying or constant field sources in the quasistatic regime.

The choices of winding constructs are broad and it is no longer practical to divide these methods into narrowly defined categories such as remote field ET or MFL or "reflectance" probes or even remote field ET. There are clear strengths and limitations to coil designs, however. For example, the use of conventional differential coils was driven, over a decade ago, by the need to reduce noise and increase sensitivity for crack detection. Today, much of this motivation for such differential designs no longer exists. Improvements in cable designs and instrumentation have reduced many electrical noise sources so that material noise from grains, inclusions, and inconsequential defects can produce eddy-current sensor signals that are larger than noise levels of existing instruments. Thus, using differential coils to reduce the impact of instrumentation noise no longer impacts performance for most advanced applications of modern eddy-current systems. Instead, the focus is now on reducing material noise and other noise sources (e.g., from edges, roughness, and other interferences).

Also, continuous wave versus pulsed ET modes are often discussed. Continuous wave methods use one or more frequencies for the input current to the ET drive winding. In general, these single and multiple frequency methods offer substantial advantages over pulsed ET methods that excite time domain signals. Historically, it was less expensive to use pulsed ET methods. One inherent *disadvantage* of pulsed ET methods is the wide bandwidth of frequency content excited in a pulsed ET mode. Because pulsed ET signals include high-frequency and low-frequency content, these methods excite noise sources in a material or construct (e.g., near-surface anomalies or inconsequential structure behind layers) that can otherwise be suppressed in a multiple frequency continuous wave method. Also, efforts to "deconvolve" pulsed ET data have been ongoing for decades. Researchers often state that it is an *advantage* to "take all the multiple frequency data at the same time." This is actually a substantial disadvantage because numerically, when one combines information and then tries to "deconvolve" it, noise will increase. Thus, when possible, continuous wave methods are generally recommended. However, in some limited applications, pulsed ET can actually reduce noise levels (pulsed ET methods are not discussed in detail here). In any case, selection of methods and performance comparisons should always be based on achievable signal to noise (and other reliability metrics) for specific applications, not on anecdotal perceptions.

4 POD STUDIES, NOISE ISSUES, CALIBRATION, AND PERFORMANCE VERIFICATION

Performance assessment for both NDT and SHM solutions should be divided into two key areas including:

- logistics (data format convenience, throughput, portability, required support equipment, size/weight, scanning/coverage, ease of use, training requirements, operator capability/experience requirements, cost, etc.) and
- reliability (reproducibility, signal to noise/sensitivity, selectivity/robustness/classification of defects, etc.).

The use of POD studies can play an integral role in providing a performance assessment of NDT systems. Unfortunately, the common practice of using POD studies, performed under unrealistic conditions and on unrealistic test specimens, has confused the practical assessment of actual NDT capabilities. It is essential that actual performance on service-generated damage (e.g., cracks) is verified whenever practical. Inspection of actual components is practical, at least to evaluate noise levels, for *all* applications. For example, the first 10 or more components inspected should provide a basis for an initial assessment of inherent noise levels under actual inspection conditions. No system should be considered qualified for a specific application before scanning on actual hardware to assess noise levels. It is not relevant to test on a large set of cracks or EDM notch specimens that do not have representative noise levels and the likely variety of noise sources. For example, flat and smooth specimens are often used to assess ET capability, when varied curvatures and surface roughness variations are two of the major sources of ET sensor noise and consequently, performance limitations.

Also, calibration is extremely important. It must be possible to verify that the ET calibration and signal to noise achieved in an actual inspection is practically the same as that in the POD study. Otherwise, the POD study results should not be referenced when making NDT/SHM decisions. For example, all ET methods are sensitive to liftoff (proximity to the surface) variations. Thus, if performance on the test samples is limited to a fixed liftoff range (i.e., the test specimens are smooth and flat), and the actual service-exposed components have higher levels of roughness or curvature, then the resulting higher liftoffs can substantially reduce detection performance.

This introduces a paradox, since most conventional ET methods cannot accurately record liftoff at each inspection location. Fortunately, new advanced systems can provide this capability. This is particularly important for curved component surfaces and rough (e.g., shot peened) or fretted regions. Thus, recording the actual liftoff at each location within an inspection area is one way of ensuring proper performance.

Consistent evaluation criteria are critical to selection of NDT/SHM solutions. Performance for the specific problem being addressed may be very different from performance in a generic POD study sample set. Generic "round robin" POD studies often avoid the most important challenges of specific field and depot inspection applications. Eliminating these challenges makes it impossible to assess the advantages of advanced methods, often making methods look essentially equivalent in performance. Also, advanced methods may require more customization, so low-cost comparisons are generally not possible. This is another reason that, sometimes, inferior conventional methods continue to receive more use than far-superior advanced methods. Fortunately, this is also changing.

Many mistakes are common in anecdotal comparisons of ET methods; these include (i) focusing on signal and not on signal to noise, i.e., using ferrite cores and lots of turns in drive and sensing elements, at the expense of model-based predictability of ET response—*this is one example where performance would be improved on flat smooth specimens but not be realized on actual components*, (ii) assuming that the response to an EDM notch is related in any relevant way to the response on an actual crack—*note for relatively large cracks or buried cracks this may be a good assumption, but for relatively small and tight surface-breaking flaws this is not a good assumption*, and (iii) assuming that noise sources from the field or depot can be represented sufficiently in laboratory studies—*and not including evaluations on actual components early in comparisons of advanced and conventional methods*. These and other mistakes can result in poor selection of NDT/SHM solutions.

One area of critical importance is the inspection of shot peened components in landing gear and engines. It is a mistake to assume that NDT performance for EDM notches or even for real cracks grown in specimens that are not shot peened (even for cracks grown from EDM notches in shot peened

specimens) will be useful for performance evaluations. Shot peening and other cold working methods, such as burnishing, and laser shock processing introduce compressive stresses at the surface; this extends fatigue life. Thus, if a fatigue crack grows in the service of a shot peened part, when the part is at rest, then the crack will be "tight", because it will be under compression. Thus, ET or UT or other NDT methods must be tested with real cracks under the same conditions.

The generation of real-crack standards is addressed near the end of this article. Today, methods are available to generate real-crack specimens in shot peened components without using starter notches, at a comparable cost to EDM notches. Thus, it is no longer necessary to make compromises in performance testing that are so unrealistic.

The focus in future developments of advanced ET should be on reducing the impact of material noise. Two ways to address material noise (the primary limitation for some advanced ET methods) are (i) to use spatial filters, multiple frequencies and two- and three-dimensional defect models, and to enable discrimination between signals from different noise sources and defects of interest and (ii) to use time-based filtering, if possible, to image the material before damage occurs, and then subtract or otherwise filter out the impact of material noise that is not associated with the evolution of the damage/defect of interest. Of course, the latter only works for in-service defect generation, as opposed to manufacturing-induced defects for which precrack imaging may not be possible.

5 CONVENTIONAL ET METHODS

This section briefly reviews some of the capabilities of conventional ET methods, but the focus is on the limitations of these conventional methods to lay the foundation for understanding the advantages of advanced ET methods and ET-based SHM. There are other more comprehensive references on conventional ET methods [3].

NDT methods use sensors to inspect a material for defects or condition/quality. ET uses time-varying magnetic fields to induce eddy currents in a material under test. For crack detection, the induced eddy-current pattern is interrupted by the presence of a crack in the material. For condition/quality assessment (also referred to as *material characterization*), the properties of the material or a coating/process-affected layer are measured and related to a property of interest such as heat-affected zone thickness or cold-work (shot peen[b]) quality. Conventional ET can be used to examine both nonferrous and ferrous metals. However, protective coatings, such as cadmium plating, and surface processing, such as shot peening, traditionally lead to difficulties and poor performance for ferrous alloy inspection.

Historically, ET developed from numerous scientific investigations involving electricity and magnetism that started in the nineteenth century. For example, Han Christian Oersted discovered the presence of a magnetic field in the vicinity of a current-carrying wire. This was followed shortly thereafter by Michael Faraday demonstrating electromagnetic induction, in which current could be induced in a coil by the motion of magnet relative to the coil or by passing a time-varying current through an adjacent coil. Mathematical formulations for electromagnetic induction were developed by Heinrich Friedrich Lenz and Hermann L. von Helmholtz, with James Clerk Maxwell ultimately developing the basic expressions governing electromagnetic phenomenon.

Three simple applications of conventional ET are briefly described: (i) conventional ET-based crack detection, (ii) conventional ET-based conductivity measurement, and (iii) conventional ET-based coating thickness measurement.

5.1 Conventional ET crack detection

For crack detection, conventional ET methods use impedance plane analysis to compensate for liftoff. As shown in Figure 1, the liftoff response is rotated to the horizontal axis and the crack signal is measured in the vertical direction. An operator views the impedance plane, and cracks above a given threshold are determined to be crack indications. To calibrate, first the liftoff direction is determined and then the scale is typically set using an EDM notch (e.g., 0.03 in. long by 0.015 in. deep, and a few thousandths wide). For example, the EDM notch response may be set to 80% of the scale and any ET response above 50% may be recorded as a crack indication.

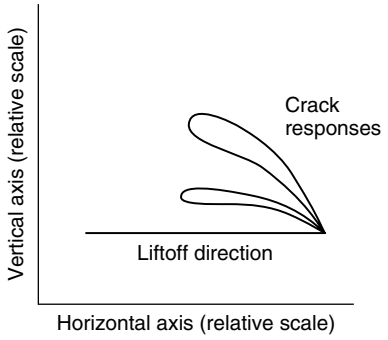


Figure 1. Typical impedance representation of a crack response.

There are many concerns with conventional ET methods for crack detection; some are listed below:

1. Calibration verification
Detection sensitivity of the probe is assumed to be the same as that of a different probe used, perhaps in a POD study. If the sensor is significantly different (as indicated by Auld and Moulder [4], where the fields from conventional ET probes were shown to vary by as much as 35% even when the terminal response was essentially identical), then the signal-to-noise level may be much lower than expected, producing significant false indications even though the "signal" level for a given EDM notch has been set to the same value as for the POD study. Also, if the conditions change from calibration on the test block to measurement on the part (e.g., temperature shifts, roughness differences, curvature variations, probe cable damage, or just cable motion), there is no method to verify that the correct calibration is sustained throughout the actual inspection. For example, a conventional ET probe calibrated in a ship bay at room temperature will respond substantially differently when the same probe is used in more extreme temperatures shipboard.

2. Curvature, roughness, and paint thickness
Variations in the material under test curvature, surface roughness, paint thickness, and other effects such as probe tilt will each reduce sensitivity significantly compared to performance on a flat, smooth, and unpainted calibration or test specimen. Although this is also the case for advanced methods, some advanced methods enable the operator to verify that these variations are maintained to within acceptable levels to insure acceptable performance, and some methods use conformable sensors and other means to substantially limit these response variations.

3. Human error
These can manifest themselves in a variety of forms. Conventional methods do little to avoid such errors. These errors include (i) selecting the wrong probe, (ii) not verifying the noise level on the standard, both away from flaws and by multiple repeated measurements both before and after an inspection, (iii) improper setting of displays and scales, (iv) scanning at speeds above that used in the POD study, (iv) tilting the probe, (v) not scanning the entire surface of interest (for manual scanning) or scanning with less sensor overlap on successive incremental scans, (vi) calibrating on the wrong crack/EDM notch standard, and (vi) not accounting well for temperature variations in the standards and in the component under test. There are also many less obvious human error sources.

4. Tight cracks in shot peened surfaces
This is a severe limitation of conventional methods. Conventional ET probes cannot reliably detect tight cracks in shot peened specimens. As described in Section 6, a linear drive design has advantages over conventional (more circular) coil designs, and use of higher frequencies is often necessary to achieve sufficient sensitivity.

5. Limited frequency range
Most ET probes have a limited frequency range of operation (e.g., 1–5 kHz, 50–500 kHz, or 1–2.5 MHz). This narrow frequency range is a limitation for crack-depth sizing using two frequency methods. Furthermore, conventional ET instruments and probes are typically limited to operation below 5 MHz. Some advanced probes have extended this range, but only a few provide reliable operation at these elevated frequencies. For example, high frequencies above 10 MHz and high reliability are needed to detect small cracks in shot peened superalloy components (e.g., 0.016 in. length by 0.008 in. depth).

6. Throughput
Conventional ET probes are limited in their ability to provide rapid imaging capability compared to array-based solutions that can scan wide areas with 10–50 times the throughput. Fully parallel instrument

channels are needed to provide reliable throughput increases with complete and uniform coverage control. Some instruments that use excessive multiplexing during scanning have most channels turned off at any one time. This introduces concerns about the uniformity of coverage of the inspected area, compared to that in the POD study. It is essential that these are the same or better on a component if the POD performance is to be assumed. Furthermore, even arrays of conventional ET probes are limited in that the crack signal varies dramatically across the sensor and between sensing elements. Some advanced eddy-current arrays have solved these issues, but most are plagued by poor reliability in actual implementations.

7. Poor sensitivity to tight cracks
For both shot peened surfaces and for detection of smaller (shallow) cracks, circular drive winding configurations are typically less sensitive than some more advanced linear drive designs.

8. Sensitivity to mechanical damage
This is a major weakness that dramatically increases false indications. Scratches, nicks, and dings can produce false indications for conventional ET, forcing detection thresholds to be raised to avoid costly false indications. This is particularly difficult in regions with fretting damage, e.g., for engine disk slots and bolt holes.

9. Severe scanner and alignment requirements
The high sensitivity of conventional ET probes to tilt and liftoff variations and the need to scan large areas with a single coil have introduced severe scanner requirements. Huge steel gantry structures are common for costly ET scanners. The high cost and large footprints of these scanners can be avoided by using modern conformable ET array technologies and using self-diagnostics to ensure proper performance. This is true both for depot/production ET scanners and for portable field scanners.

10. No self-diagnostics capability
Conventional ET, using either differential or absolute coils and calibrating on crack standards, is inherently a relative method (i.e., measurement amplitudes are relative to the reference standard properties). Thus, there is no mechanism for verifying that either the liftoff value or the effective conductivity measurements on the material under test are within a range that would indicate proper performance. This issue has been addressed by some advanced ET methods.

11. Other limitations
Many ET sensor arrays consist of a collection of conventional ET coils (with individual drive and sensing elements). These arrays have several major limitations:

(a) Cross talk between sensing elements
Neighboring drives produce signals that are not only captured by the associated sensing element but also by neighboring sensing elements. One common practice is to multiplex the sensors in an array to avoid measuring on neighboring sensing elements, reducing cross-talk effects. This approach introduces concerns regarding coverage and must be performed with care to ensure POD performance is maintained. Other advanced arrays such as the MWM-Array use linear drive wires that eliminate this cross-talk issue entirely.

(b) Variable sensing element responses
Arrays of conventional ET sensors and variability in lead configurations for each sensing element cause variability in the individual sensing element responses to cracks. This makes it difficult to assess POD performance using a conventional ET array. Advanced flexible arrays use microfabrication methods to produce nearly identical sensing elements but many of these designs still suffer from sensing element response variations due to variability in instrumentation and leads. MWM-Array methods address this need by using a patented calibration methodology that corrects for these element-to-element variations, while using microfabrication and a linear drive to limit the severity of these variations. It is anticipated that arrays such as the MWM-Array that address these concerns will become more and more widely used in the near future.

5.2 Conventional ET conductivity measurement

Conventional ET conductivity measurements use one or more conductivity reference standards to calibrate

the instrumentation. The standards have an electrical conductivity comparable to the expected conductivity of the material to be examined and preferably bracket the measured conductivity. Depending upon the probe and instrument configuration, the standards can be used to determine the liftoff direction as with ET crack detection. Many of the limitations described above are also relevant for single sensing element methods or array-based methods for conductivity measurement. The use of conductivity standards to calibrate conventional ET sensors introduces opportunities for human error and eliminates the capability to provide self-diagnostics by verifying performance on a conductivity standard instead of calibrating on such a standard. Also, such methods are typically limited by the available range of conductivity standards.

In most applications, conventional ET examinations use a circular coil. This creates a rotationally symmetric magnetic field so that the probe response is independent of orientation on the material surface. As a consequence, anisotropic (directional) conductivity variations cannot be measured. While this can be an advantage because it removes sensitivity to such variations, in some situations this anisotropy is in itself the material variation of interest for quality or degradation assessments. For example, for the roller burnishing quality inspection of the C-130 and P-3 propellers, the conductivity is measured in two directions (using an MWM® sensor) and the ratio is used to assess the roller burnishing quality.

Advanced methods, which do not require the use of standards for calibration and are calibrated using model-based methods, offer wider range of conductivity measurement reliability and robustness. Furthermore, the use of air as a calibration material, as described in ASTM E2338 [5], offers a self-diagnostic capability that can allow the sensor performance to be verified on a reference standard for each individual sensing element in an ET array. This ASTM standard describes the air calibration method in detail.

5.3 Conventional ET coating thickness measurement

For coating thickness measurements, conventional ET requires a set of coating thickness reference standards [6]. As with the conductivity measurement, this yields a limited performance range and eliminates much of the potential for self-diagnostics. More important, however, is that conventional ET coating property measurements require that the conductivity of the coating standards and the conductivity of the actual coating being tested are essentially the same (i.e., close enough to ensure that the thickness measurement is sufficiently accurate). Advanced methods, performed in accordance with ASTM E2338 [5], offer substantial improvements by enabling calibration without coating standards and by independently measuring the coating conductivity and thickness. This is an enabling feature not only for coating thickness measurement but also for detection of cracks in components with coatings, for identification of coating thermal aging, and for detection of manufacturing quality lapses, e.g., excessive porosity.

6 ADVANCED ET METHOD EXAMPLE: THE MWM®-ARRAY

The MWM-Array is a conformable ET array that was specifically designed for model-based inverse methods. Multivariate inverse methods convert the sensor response into absolute property estimates. Multiple sensor responses, for each individual sensing element at each location in an image, are converted into estimates of several material properties at each location along the inspected surface of the material under test. These methods can also be used to estimate geometric parameters (e.g., liftoff or coating thicknesses) and other dependent properties, such as stress (typically related to either magnetic permeability or electrical conductivity) and temperature (typically related to electrical conductivity).

The MWM-Array addresses most of the limitations listed above for crack detection, conductivity imaging, and coating characterization. For example, the MWM-Array is currently used by the US military for detection of fretting fatigue cracks in engine disk slots. Direct comparison of performance between this advanced method and conventional ET and liquid penetrant inspection (LPI) for actual implementations illustrates the dramatic advances that are possible. Also, this method and the JENTEK Sensors, Inc., Federal Aviation Administration (FAA), Navy, and Air Force team members were selected for the FAA 2007 “Better Way” Award, specifically for engine

component automated disk and blade dovetail inspection solutions.

In one direct comparison, on a US Navy engine population, over 3000 engine disk slots were inspected during an 18-month period at a Navy Depot using a 36-channel MWM-Array with a fully parallel architecture JENTEK impedance instrument. The automated MWM-Array system performed these inspections and (i) there were zero false indications, (ii) the speed achieved was less than 1 min per slot, (iii) over 9 disks in over 130 were detected with cracks, (iv) all detected cracks were verified by destructive testing, i.e., fractography, or other means, and (v) most of these cracks were missed by conventional ET and LPI.

Of course, this unusual level of performance reliability (POD and false indication rate) cannot be expected in all applications of the MWM-Array. However, such advanced automated ET systems are now available at a fraction of the cost of previous automated solutions and with substantial improvements in ease of use, inspection throughput (e.g., 1 min per slot vs 20 min per slot); they offer substantially smaller system footprint, and in some cases are semiportable.

This is an important example because it highlights the difficulty of managing NDT decisions and selecting solutions. Conventional NDT (ET and LPI) are expected to perform reliably; however, instances continue to occur where the original performance expectations (based on POD studies that may not have represented realistic conditions) are not being sustained in service. This highlights the need to improve performance evaluation methods.

Figure 2 illustrates measurement data using an advanced ET method that enables rapid conversion of sensor data into conductivity and liftoff data. The bottom right shows a model-based predicted sensor response as the conductivity and liftoff of a test material are varied. The sensor response, expressed in complex notation in terms of the real (in-phase) and imaginary (out-of-phase) components, depends upon the conductivity and liftoff in a nonlinear fashion. The measurement data for a single channel of a sensor array shows liftoff variation as the sensor array is scanned over the part, as well as the presence of a crack. Using advanced instrumentation, the acquisition rate for the data can be fast compared to the scan speed, which allows multiple data points (in this case 22) to be acquired over the crack. Other references describe how this same method can be used for more unknowns (e.g., for the thickness and conductivity of a coating, as well as the substrate conductivity and liftoff).

As shown in the upper left of Figure 2, a conductivity image is created using a precomputed database of sensor responses called a *measurement grid*. This database enables the rapid conversion of ET data into both conductivity and liftoff values, at each location along the surface being inspected. Multiple frequencies can be used to obtain effective conductivity measurements at each applied frequency. As shown on the bottom left of Figure 2, a typical ET impedance plane response, similar to those obtained for conventional ET methods, is produced. However, here the software automatically interprets the raw sensor data and builds images in near real time.

The key to making this method practical is the design of sensors, such as the MWM-Array, that eliminate most unmodeled dynamics and enable calibration in air without the use of crack, conductivity, or coating standards. The next three subsections describe three brief case studies that illustrate this capability for engine disk slot inspection, blade dovetail inspection, and bolt-hole inspections.

6.1 Case study I: engine disk slot inspection

Figure 3 shows a photograph of the MWM-Array systems at a NAVAIR Depot (Cherry Point) used for engine disk slot inspection, along with conductivity and liftoff image results. The engine disks are placed onto the conical supports and the sensor arrays are scanned through complex-shaped slots in the disks. The performance of these two systems is summarized below:

- *in use at* NAVAIR Depot, Cherry Point, since April 2005;
- *nine disks* with cracks detected, several of these *large and small cracks not detected by conventional ET and LPI;*
- *all crack detections verified by destructive testing or other means;*
- *no false indications* on 131 disks (over 3000 slots inspected with zero false calls resulting in a false call rate per slot of $<0.03\%$);

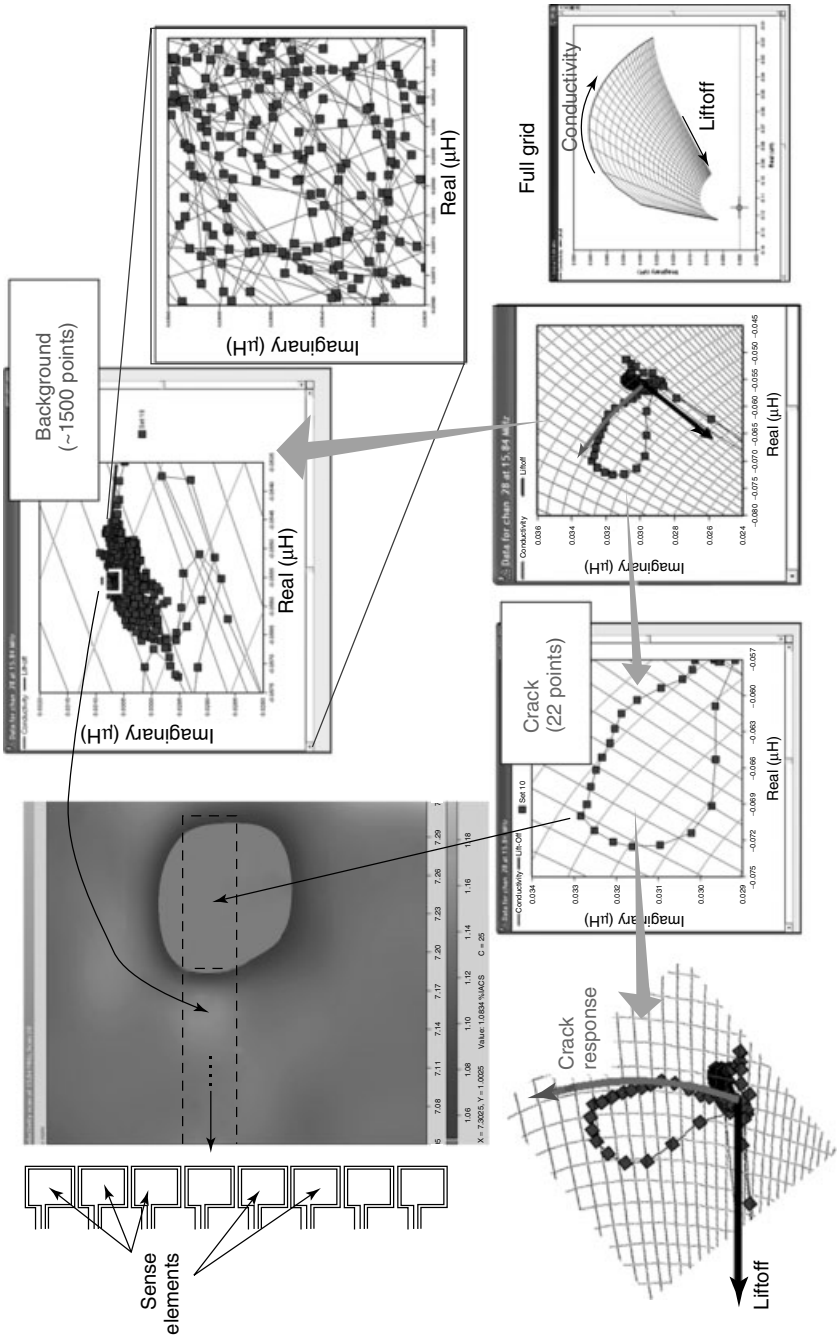


Figure 2. Example of rapid data processing using grid methods (precomputed databases), to convert impedance data for each MWM-Array channel (sense element) into conductivity and liftoff at each point in the image (patents issued and pending).

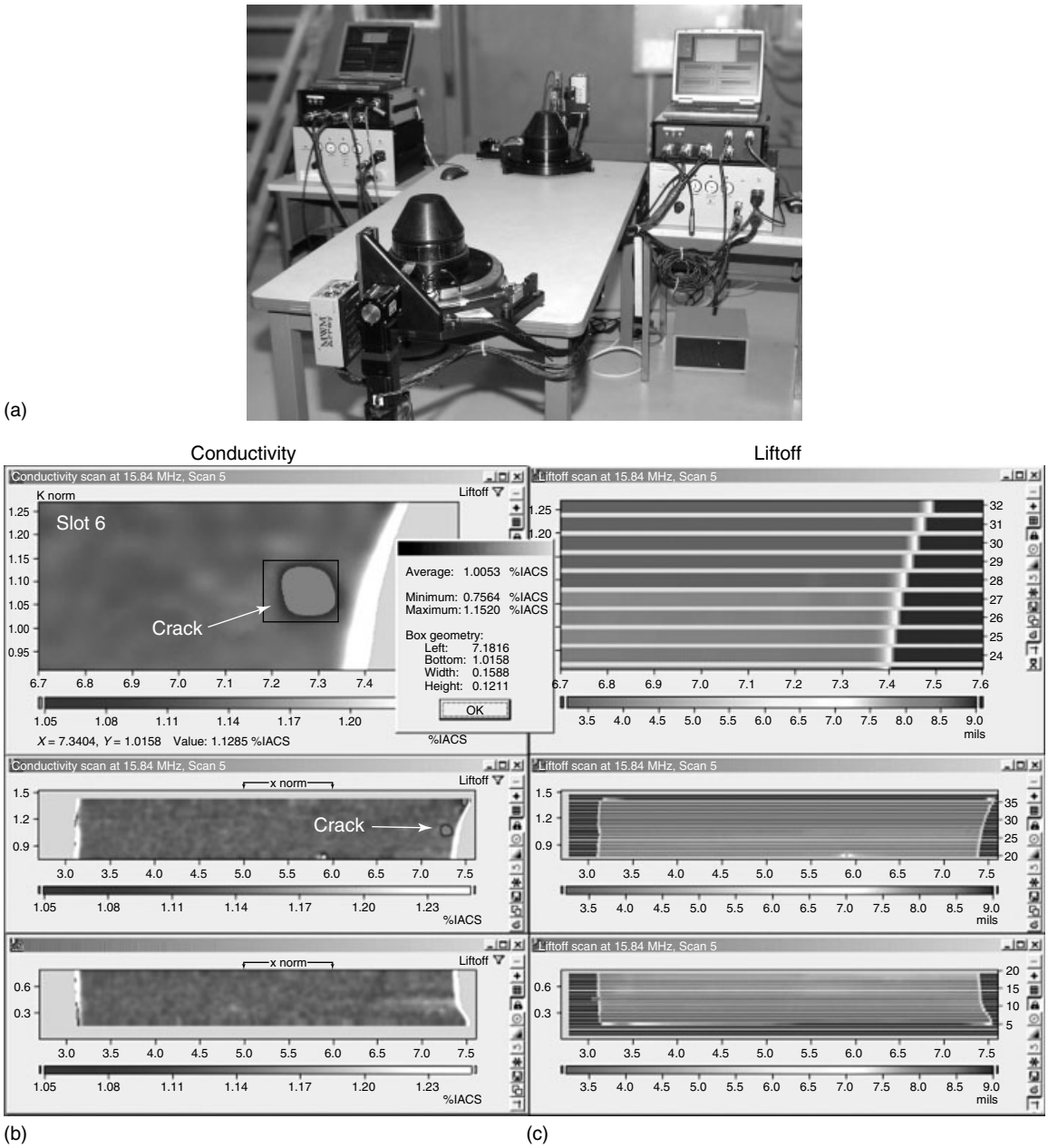


Figure 3. (a) Two automated GridStation Systems at NAVAIR Depot, Cherry Point, for engine disk slot inspection, (b) 37-channel, MWM-Array eddy-current sensor inspecting an engine disk slot, and (c) conductivity and liftoff images of MWM-Array results.

- *all disks now subjected to mandatory inspection* using the MWM-Array system during engine processing at this depot.

This data continues to be accumulated on this particular engine disk slot population. This data is a valuable asset for developing next generation

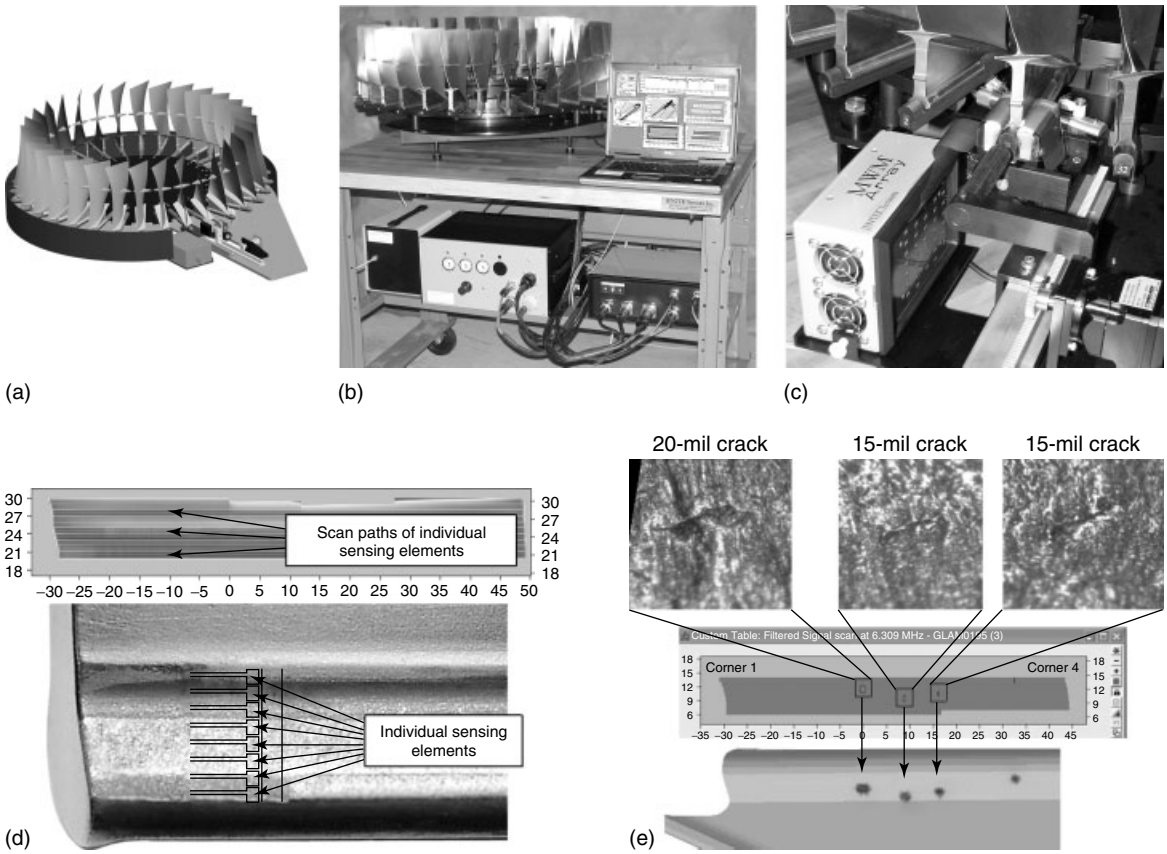


Figure 4. (a) Early design for automated blade dovetail inspection system; (b, c) photograph and detail of system delivered to the US Navy in 2007; (d) close-up of blade dovetail and liftoff image showing scan paths of individual sensing elements and acceptable liftoff values (indicted by green on a green-to-yellow-to-red color scale for liftoff); and (e) MWM-Array results with cracks indicated in the conductivity image (blue), verified by optical microscopy, and mapped onto a 3D image of the blade dovetail.

ALM approaches, i.e., as disks are reinspected after years in service, damage evolution history can be investigated.

The images in Figure 3(b) and (c) are for conductivity and liftoff respectively. These liftoff images enable real-time self-diagnostics by verifying that the liftoff is within an acceptable range. And the blue conductivity images further confirm that the absolute conductivity is within an acceptable range. The combination of these two absolute measurements at each of thousands of locations within an engine disk slot offers a built-in verification of sensor performance for each slot. Furthermore, this inspection is performed using an air/shunt calibration as described in ASTM Standard E2338.

6.2 Case study II: blade dovetail inspection

The inspection of blade dovetails has been funded by NAVAIR under a similar effort. Pilot line testing is ongoing at NAS, Jacksonville. Figure 4(a–c) shows the JENTEK Blade Dovetail Carousel and GridStation® System for automated inspection of an entire blade set in about 1 h. In this system, the sensor array is scanned over the complex-shaped dovetail of the blades. A pneumatic actuator clamps the sensor array to the dovetail so that the surfaces on both sides are examined at the same time. Figure 4(d) and (e) shows where the dovetail is inspected with the conformable MWM-Array sensor (which is shown in

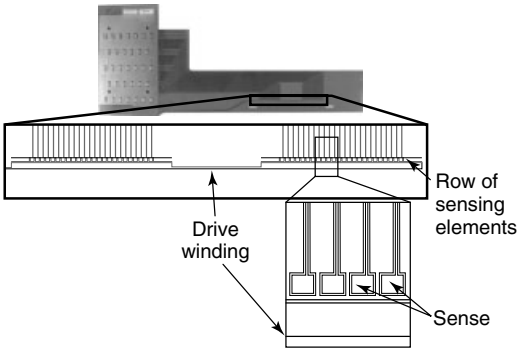


Plate 5. MWM-Array used for disk slot and blade dovetail inspections of regions experiencing fretting damage, for case studies I and II (patents issued and pending). MWM is a registered trademark of JENTEK Sensors, Inc.

Figure 5 and is the same sensor used for the disk slot inspection shown in Figure 3) and resulting MWM-Array images with associated crack indications.

6.3 Case study III: bolt holes

Bolt-hole inspection is also possible with MWM-Arrays. This demonstration illustrates the capability to inspect bolt holes and track crack-growth behavior. Figure 6(a) and (b), respectively, shows the bolt-hole scanning apparatus and MWM-Array sensor used for this demonstration. Figure 6(c) provides an example of conductivity (blue, with red crack indication) and liftoff (green-to-yellow-to-red color scale, where green indicates acceptable liftoff). Note that the conductivity images are cut off when the liftoff is above a specified level, automatically indicating the edge location.

Figure 7 provides the filtered response and the corresponding images of a crack (in a bolt hole) at three different stages of growth during a coupon test. This data might be stored for multiple coupon tests in a database (empirically generated) and then searched and calibrated using results from service components to enable not only estimation of crack sizes but also as a means for estimating damage evolution behavior (e.g., the probability of a failure occurring before the next inspection).

Figure 8 shows the response of a surface-mounted MWM-Array used during the coupon test. These arrays provide real-time, continuous fatigue test monitoring and are used to schedule the C-Scan imaging with the bolt-hole MWM-Arrays. This allows for very efficient generation of time-sequenced NDT (scanning) images at various damage levels, and significantly reduces the costs and labor needed to generate empirical NDT databases.

As illustrated in Figure 8, the surface-mounted MWM-Arrays detected the fatigue crack long before it reached a crack length of 0.0175 in. Also, the high signal to noise for the scanning MWM-Array data in Figure 7 makes it likely that much smaller cracks can be detected and the damage progression tracked.

7 REAL-CRACK STANDARDS

There are several practical methods available for generating real-crack specimens. Figures 7 and 8 illustrate one such method that uses advanced ET arrays either mounted permanently on the specimen or scanned periodically across the specimen during fatigue coupon testing. The configuration in Figures 7 and 8 is for bolt holes. Figure 9 shows a similar capability for generating real cracks in curved and shot peened or coated components. In this example, mechanical damage in the form of a ding is included to illustrate the generation of real-crack specimens with crack initiation at a mechanical damage site.

Thus, generation of real-crack specimens for NDT POD studies are no longer significantly more expensive than that of specimens with EDM notches. Crack specimens with real cracks can be generated without starter notches in bolt holes and curved surfaces. This can be accomplished for shot peened surfaces and even with coatings. Thus, it is no longer necessary to use EDM notches to perform POD studies for ET methods on surface-breaking flaws.

8 SUMMARY AND CONCLUSIONS

The fundamental developments of advanced ET methods that can address the limitations of conventional ET are the following:

1. *Low-cost processing power*—this enables the generation of precomputed databases of sensor responses and/or real-time processing of large volumes of sensor digital sensor data.
2. *Digital signal-processing chips*—these enable the development of highly accurate, absolute,

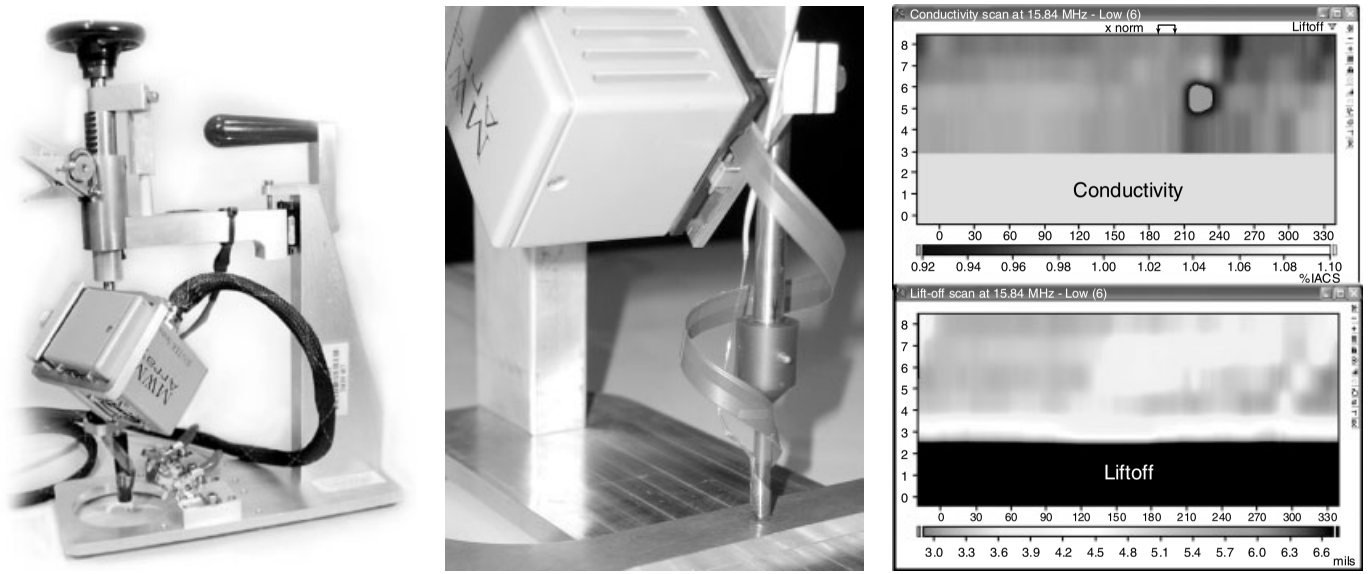


Figure 6. (a) Bolt-hole scanning apparatus, (b) close-up view of MWM-Array sensor scanning a fatigue coupon with a hole (*see* **Eddy-current *in situ* Sensors for SHM**, Figure 3(c) for schematic of 7-channel sensor), and (c) MWM-Array conductivity (blue) and liftoff (green) images.

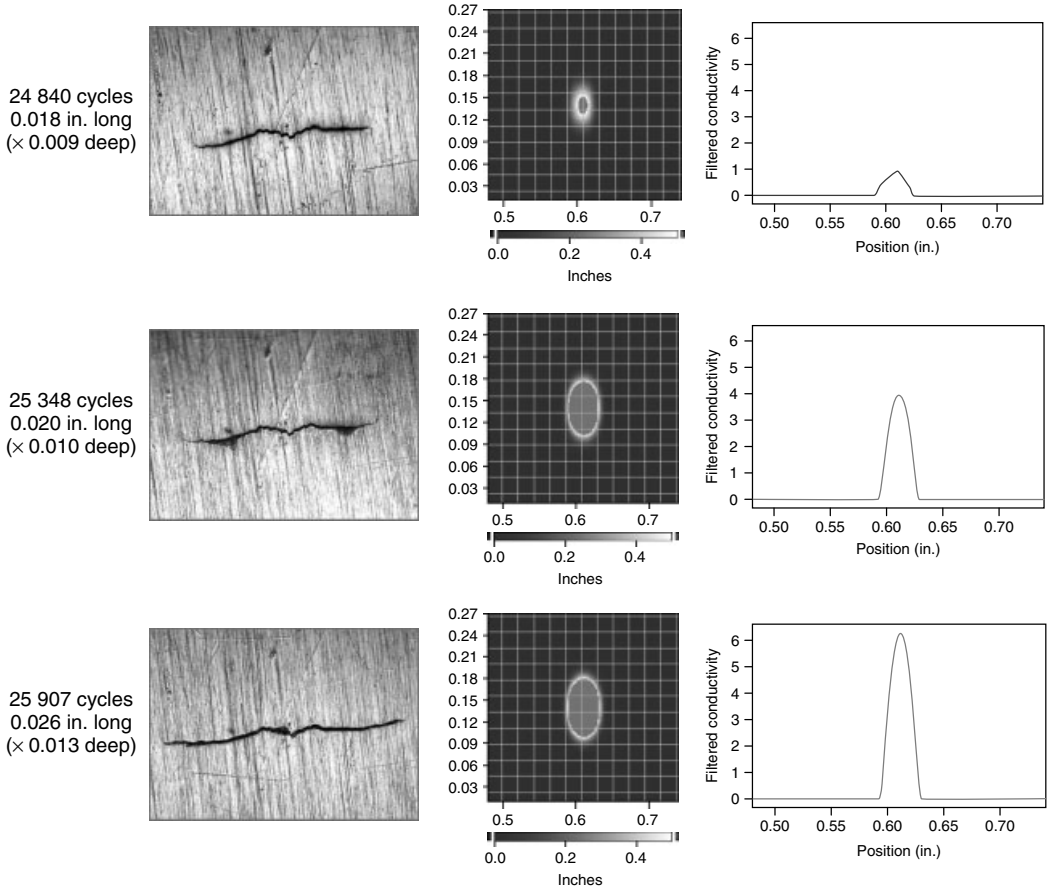


Figure 7. Spatially filtered MWM-Array scan images of a crack produced inside a hole in a Ti-6Al-4V fatigue coupon with two holes. Data are from scans of the hole with the MWM-Array FA43 sensor (*see* **Eddy-current *in situ* Sensors for SHM**, Figure 3(c) for photo) at various stages of the fatigue test. Crack photomicrographs on the left are from acetate replicas taken after each MWM-Array scan.

high-throughput parallel architecture impedance instrumentation that is suitable for use with model-based inverse methods (that require such highly accurate, absolute impedance measurements).

3. Low-cost instrumentation fabrication—this enables the fabrication of relatively low-cost, portable, and powerful high-throughput parallel architecture and high-fidelity impedance instrumentation.
4. Rapid modeling methods—these combined with processing power enable the generation of precomputed databases of sensor responses and/or real-time processing of large volumes of sensor digital sensor data using model-based inverse methods.
5. Rapid multivariate inverse methods—these enable the rapid processing of sensor data without running the model successively. This alone has a huge impact on throughput and reliability.
6. ET sensor designs that can be accurately represented by rapid models >—these combined with the rapid inverse methods enable many of the features needed by advanced methods to overcome the limitations of conventional methods.
7. Flexible ET arrays with no sensing element cross talk—this is required to enable rapid

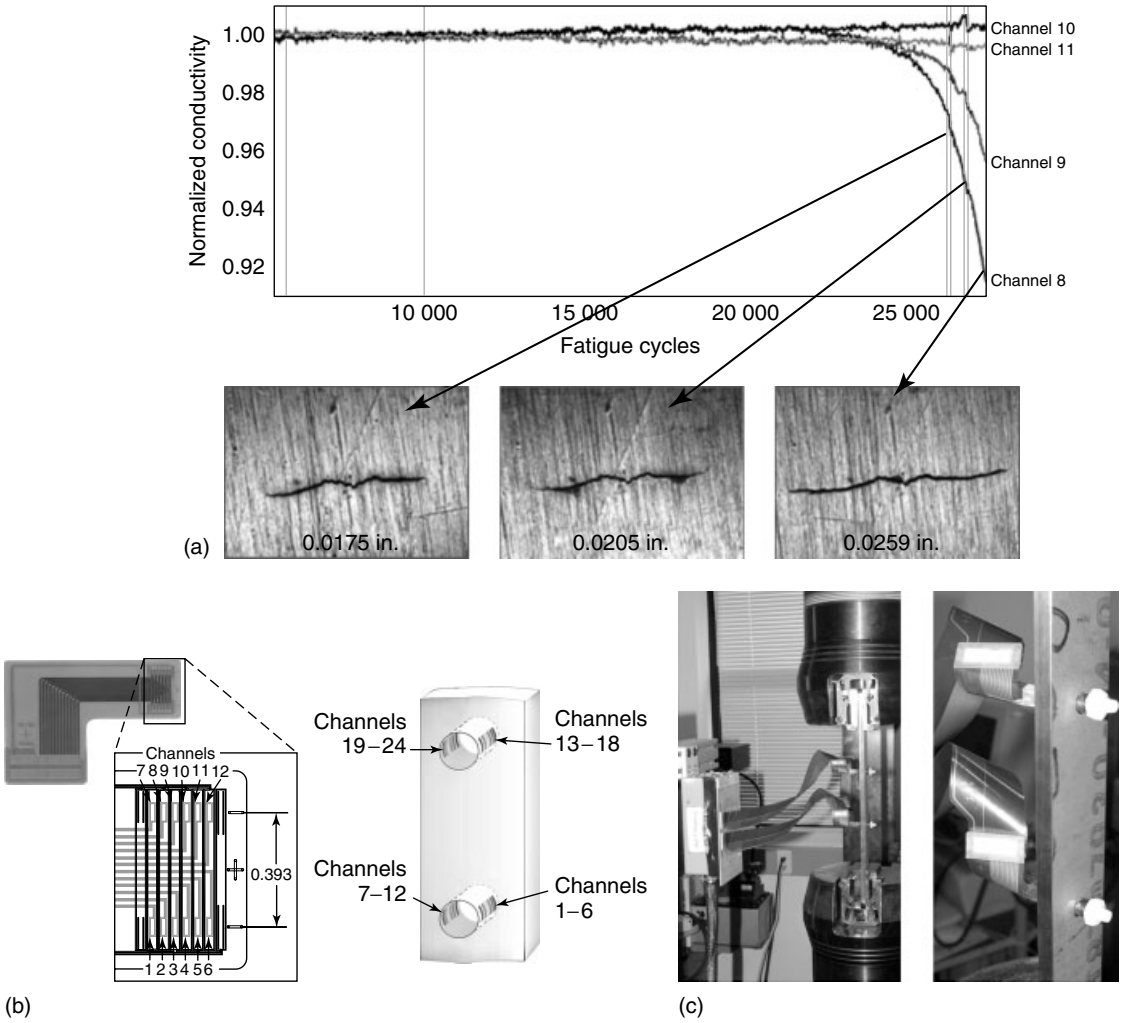


Figure 8. (a) MWM-Array data for the same Ti-6Al-4V coupon as shown in Figure 7, (b) photograph and schematic of the FA75 MWM-Array demonstrating how it was mounted in the two holes, and (c) setup for fatigue coupon testing.

wide area scanning of complex curved surfaces. Many "advanced" ET arrays use more conventional designs for which each sensing element responds differently and do not provide consistent responses for each element in the array, making detection reliability unpredictable.

8. Air calibration methods—as described in ASTM Standard E2338-04 that enable reliable calibration of ET arrays and single sensing element ET sensors and also provide the foundation for inherent self-diagnostics.
9. Self-diagnostics—by reliably measuring conductivity and liftoff at all noncrack locations within and ET image to ensure that performance of at each inspection location remains within the assumed range and that the sensor and instrumentation continue to perform properly.
10. Linear drive configurations—these provide a high-density eddy-current flow in the direction perpendicular to the crack (for improved crack detection sensitivity, particularly for shot peened components) and to provide a reduced

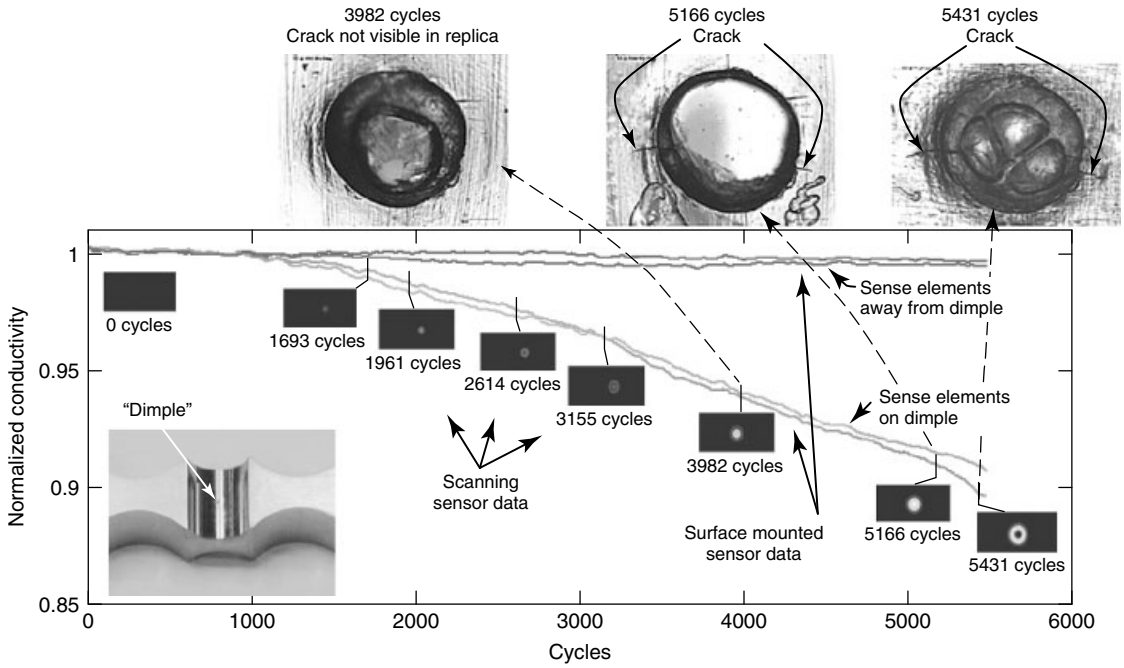


Figure 9. Progression of fatigue damage in Ti–6Al–4V captured by three different methods—(top) optical microscopy of acetate replicas, (solid lines) output of surface-mounted MWM-Arrays, and (purple images) eight C-scan images from periodic inspection with a scanning MWM-Array. At 5166 cycles, the crack length outside the dimple was 0.006 in. (left side) and 0.005 in. (right side). At 5431 both cracks were approximately 0.008 in. in length along the surface. The distance from tip to tip at 5431 cycles was 0.054 in.

sensitivity to crack position within the footprint, as well as a substantial reduction of edge effects associated with drive position relative to the edge.

11. High frequencies and wide dynamic range—this is required to enable improved coating characterization, high sensitivity to tight cracks, and measurement of crack depths (although accurate crack-depth measurement remains a difficult problem when crack morphologies vary significantly or cracks occur in clusters).

Furthermore, repeated imaging to *map and track* early damage growth is a new practice in NDT that is expected to be used more and more as NDT methods are upgraded. This will be complementary to the evolution of SHM and is expected to result in an integrated suite of advanced NDT/SHM solutions over time.

One advantage of using embedded SHM sensors compared to conventional NDT solutions is the capability to monitor damage at a single location before and after damage initiation. This enables removal of component-to-component variations and use of the undamaged material response as a reference for damage detection.

Conventional NDT solutions have, historically, not offered either the reproducibility or the digital archiving necessary to enable such comparisons of component conditions before and after damage initiation. In the next decade, this will change dramatically, since both advanced (scanning) NDT methods (with newly achieved levels of repeatability) and onboard sensors can monitor the damage sites both before and after damage occurs. The relative performance of each method will thus depend on the reproducibility and overall reliability of this repeated measurement data over long periods of time (weeks and months in some cases, to many years in other cases).

There are many factors that will introduce noise (structured and random) in both advanced NDT and

SHM sensor data. Thus, in every application, the relative performance of each should be evaluated. In some cases, reduction of uncertainty to required levels will require a combined solution to accommodate anticipated random and deterministic error/noise sources.

END NOTES

[a.] Wavelength—In dynamics, the wavelength is determined by the velocity of an electromagnetic traveling wave in a medium divided by the oscillation frequency. For example, the wavelength of a traveling wave at 13 MHz in air is 23 m. In quasistatics, this wavelength is long compared to the relevant system dimensions.

[b.] Shot peening is a cold working method that induces residual stresses at the surface of a component to extend fatigue life. Cold working processes include shot peening, burnishing, and laser shock processing.

REFERENCES

[1] Haus HA, Melcher JR. *Electromagnetic Fields and Energy*. Prentice Hall: Englewood Cliffs, NJ, 1989.

[2] Goldfine N. Magnetometers for improved materials characterization in aerospace applications. *ASNT Materials Evaluation* 1993 **51**(3):396–405.

[3] Udpa SS (Technical editor), Moore PO (eds). Electromagnetic testing. In *Nondestructive Testing Handbook, Third Edition*. American Society for Nondestructive Testing: Columbus, OH, May 2004; Vol. 5.

[4] Auld BA, Moulder JC. Review of advances in quantitative eddy current nondestructive evaluation. *Journal of Nondestructive Evaluation* 1999 **18**(1):3–36.

[5] ASTM E2338-06, *Standard Practice for Characterization of Coatings Using Conformable Eddy-Current Sensors without Coating Reference Standards*. ASTM International, Book of Standards, 2006; Vol. 03.

[6] ASTM E376-06, *Standard Practice for Measuring Coating Thickness by Magnetic Field or Eddy-Current (Electromagnetic) Examination Methods*. ASTM International, Book of Standards, 2006; Vol. 03.

Chapter 65

Full-field Sensing: Three-dimensional Computer Vision and Digital Image Correlation for Noncontacting Shape and Deformation Measurements

Michael A. Sutton

Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

1.1 Area of application

The area of application is known as *three-dimensional digital image correlation* (3D-DIC), a noncontacting measurement method capable of accurately quantifying small or large three-dimensional surface displacements and surface strains on specimens ranging from a 10^{-3} to 10 m or larger.

1.2 Motivation

With the increase in computer processing speed, the ability to computationally predict a wide range of complex phenomena has increased dramatically (*see* **Free and Forced Vibration Models**; **Fundamentals of Guided Elastic Waves in Solids**; **Acoustic Emission**; **Electromechanical Impedance Modeling**; **Thermomechanical Models**; **Civil Infrastructure Load Models for Structural Health Monitoring**; **Static Damage Phenomena and Models**; **Damage Evolution Phenomena and**

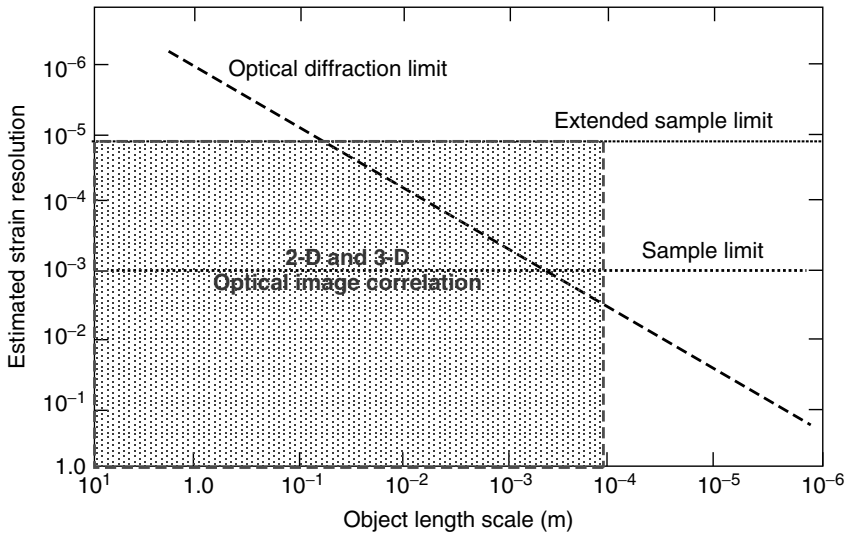


Figure 1. Schematic relating strain resolution to object dimensions. The graph focuses primarily on small-scale 2-D measurements. IC denotes *image correlation*, which appears to be a viable method for measurements over several orders of magnitude in out length scale. Optical IC includes both 2-D and 3-D methods. (The graph shown is a modified version of one developed by Prof. K.S. Kim, Brown University.)

Models; **Failure Modes of Aerospace Materials**; **Modeling Aspects in Finite Elements**; **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators**; **A Simplified Damage Model for SHM Metallic and Composite Structures**; **Modeling for Detection of Degraded Zones in Metallic and Composite Structures**; **Fatigue Life Assessment of Structures**). Phenomena such as stable crack growth, dynamic crack growth, mechanical response of complex structures, impact and blast loading, as well as a wide range of other phenomenon can be simulated with modern computational algorithms and associated software.

Until recent years, the ability to obtain accurate, full-field measurements under general conditions including (i) three-dimensional motions, (ii) large deformations, and (iii) high rate loading has been limited by the methods available. To validate the simulations, one approach is to compare specific predictions to experimental measurements. For example, if a fracture criterion is assumed as part of a simulation (*see* **Static Damage Phenomena and Models**; **Damage Evolution Phenomena and Models**; **Failure Modes of Aerospace Materials**), then measurements such as the following can be directly compared with theoretical predictions to provide quantitative measures of the quality of the predictions:

- load-crack extension
- crack-tip strain field during crack extension
- crack opening displacement.

Figure 1 presents one view[a] of the relationship between strain sensitivity and specimen length scale for digital image correlation (IC) methods. As shown in this figure, for structural measurements on components on the order of 100 μm or larger, optical IC methods offer investigators the ability to quantify strains on the order of 10^{-4} or larger. Though originally developed for 2-D measurement methods (e.g., 2-D IC), stereovision concepts have been used to formulate a general, noncontacting method designated *three-dimensional digital image correlation* (3D-DIC) that has an accuracy similar to that of two-dimensional digital image correlation (2D-DIC), while extending the capability to 3-D motion measurements on 3-D objects. Applications in the past two decades have shown conclusively that 3D-DIC is capable of making full-field deformation

measurements on curved or planar specimens in a wide range of applications, with strain accuracy on the order of 10^{-4}.

1.3 Historical background

A cursory review of the literature indicates that the earliest developments in image-based shape measurements resulted in the formation of the field of photogrammetry. Originally focused on extracting height/shape information through comparative photography, it appears that some of the first work in the area of "image correlation" was performed by Gilbert Hobrough in the 1950s. Hobrough compared analog representations for photographs to register features from various views [1], and later designed and built an instrument to "correlate high-resolution reconnaissance photography with high precision survey photography in order to enable more precise measurement of changeable ground conditions" [2], thereby being one of the first investigators to attempt a form of digital IC to extract height information from the IC/matching process.

As digitized images became available in the 1960s and 1970s, researchers in artificial intelligence and robotics began to develop vision-based algorithms and stereovision methodologies in parallel with photogrammetry applications in aerial photography (see Rosenfeld [3] for an extensive bibliography). As noted by Rosenfeld, engineering applications for shape and deformation measurements using digital images were either nonexistent or rare up to 1980.

While digital image analysis methods were undergoing explosive growth in many areas, much of the field of experimental solid mechanics was focused on applying recently developed laser technology. Holography [4–6], laser speckle [7], laser speckle photography [8, 9], laser speckle interferometry [10], speckle shearing interferometry [11], moire' methods [12, 13], holographic interferometry [14] and fiber optic sensors (*see* **Free and Forced Vibration Models**; **Fundamentals of Guided Elastic Waves in Solids**; **Acoustic Emission**; **Electromechanical Impedance Modeling**; **Thermomechanical Models**; **Civil Infrastructure Load Models for Structural Health Monitoring**; **Static Damage Phenomena and Models**; **Damage Evolution Phenomena and Models**; **Failure Modes of Aerospace Materials**; **Modeling Aspects in Finite Elements**; **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators**; **A Simplified Damage Model for SHM Metallic and Composite Structures**; **Modeling for Detection of Degraded Zones in Metallic and Composite Structures**; **Fatigue Life Assessment of Structures**) are typical examples of the type of measurement techniques developed for use with coherent light sources. In all cases except fiber optic sensing, the measurement data (surface slopes and displacements) was most often embedded in the photographic medium, typically in the form of a fringe pattern. Since the photographic recording process is generally nonlinear, resulting in difficulties in extracting partial fringe positions with high accuracy, the most common process employed by experimental mechanicians was a laborious determination of estimates for fringe center locations at a few points.

Given the difficulties encountered by experimental mechanicians during the postprocessing of photographically recorded measurement data, it was natural for researchers to employ recent progress in digital imaging technology and develop (i) methods for digitally recording images containing measurement data, (ii) algorithms to analyze the digital images and extract the measurement data, and (iii) approaches for automating the entire process.

One of the earliest papers that proposed the use of computer-based image acquisition and deformation measurements in material systems was written by Peters and Ranson in 1982 [15]. Originally envisioned as a method for use with ultrasonic waves, the authors suggested comparing small regions (known as *subsets*) from each of the digitally recorded ultrasonic images before and after deformation. Using fundamental continuum mechanics concepts governing the deformation of small areas, subset matching throughout each image was proposed to obtain a dense set of full-field, two-dimensional displacement measurements. Over the next decade, the basic concepts were extended and applied to optical images of an undeformed and deformed object. The proposed methodology was modified and refined, resulting in a set of numerical algorithms that were validated through a combination of

experimental and computational studies [16–23]. The 2D-DIC method has been used to measure crack-tip strain fields [24–27], creep deformations at elevated temperature [28], and tensile deformations of thin paper sheets [29, 30]; a high contrast random pattern was applied to the paper sheets using Xerox toner power. For these measurements, a random pattern and incoherent illumination were used to obtain high-contrast, white-light speckle images. By selecting subsets from the pattern and comparing deformed and undeformed patterns, the matching process is used to obtain full-field displacements.

More recently, investigators have begun to probe the fundamentals of the image-matching process and quantify the potential accuracy of the method. For example, Schreier *et al.* [31] showed that intensity interpolation must be performed to improve the accuracy of the displacement measurements. In follow-on on studies, Schreier and Sutton [32] have shown that quadratic shape functions provide some advantages when performing the matching process, especially for nonuniform strain fields. Relative to the importance of distortions in pattern matching, Schreier *et al.* [33] developed and applied nonparametric distortion measurement and removal methodologies.

Since 2D-DIC requires predominantly in-plane displacements and strains, relatively small out-of-plane motion of the object will change the magnification and introduces errors in the measured in-plane displacement. To overcome this limitation, stereovision principles developed for robotics, photogrammetry, and other shape and motion measurement applications were modified and used by Chao *et al.* to successfully develop and apply a two-camera stereovision system for the measurement of three-dimensional crack-tip deformations [34–36]. To overcome some of the key limitations of the method (square subsets remained square in both cameras, mismatch in the triangulation of corresponding points, and a calibration process that was laborious and time consuming), the stereovision method was modified to include (i) the effects of perspective on subset shape and (ii) constraints on the analysis to include the presence of epipolar lines [37]. The method has been used successfully in several small- and large-scale applications [38–40].

2 KEY TECHNOLOGIES USED IN THREE-DIMENSIONAL COMPUTER VISION (3D-DIC) FOR STRUCTURAL MEASUREMENTS

Figure 2 shows several actual stereovision systems that have been used successfully to measure 3-D shape and 3-D displacement fields. Key technologies used to construct 3D-DIC measurement systems include the following:

- High-resolution, scientific-grade, charge coupled device (CCD) or complementary metal oxide semiconductor (CMOS) cameras
  - Typical cameras record in monochrome with 8 bits.
  - Spatial resolution is on the order of 1024 × 1024 pixels.
  - Pixels are square in physical dimension.
- Quality optical lenses for imaging during experiments
  - Nikon, Canon, Sigma, and Schneider lenses are typical brands used.
  - Lenses with focal lengths in the range from f19 to f200 are commonly used.
- Stable, durable camera support structures
  - Heavy-duty tripods, or equally strong structures, to support the cameras.
  - Rigid cross members to maintain the relative positions of cameras during an experiment.
- Translation stages to adjust camera position(s)
  - Digitally controlled stages for automation of adjustment process have been used successfully.
  - Rigid cross members with cameras attached can be moved as a unit without affecting the calibration.
- Temporal synchronization unit for simultaneous multicamera image acquisition during calibration and experiment

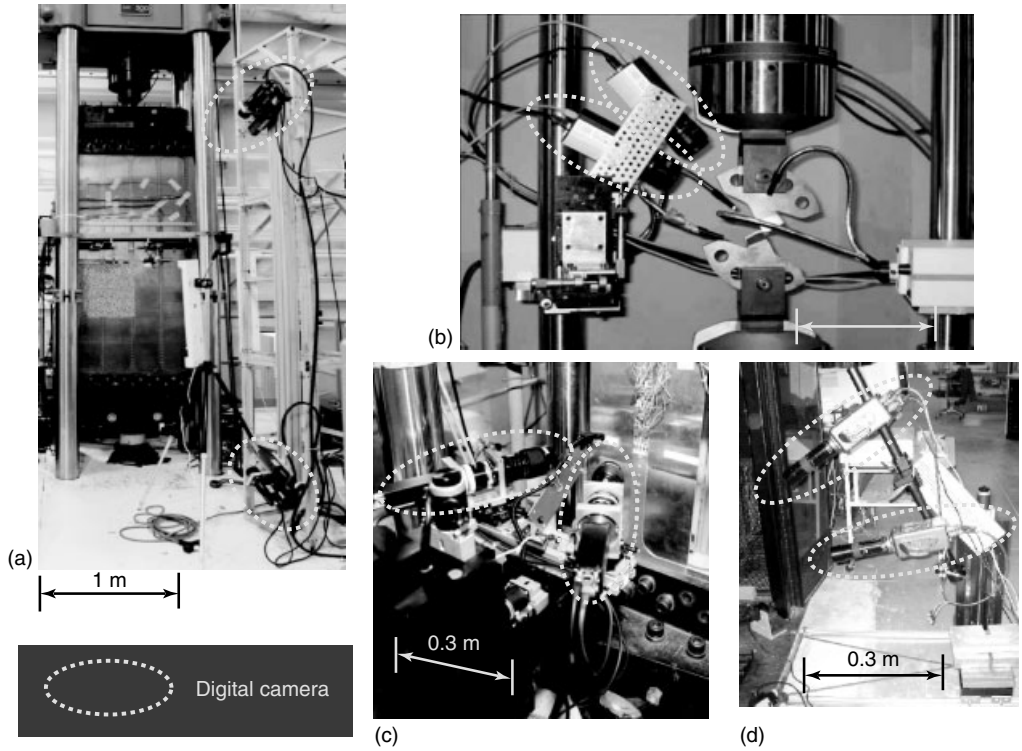


Figure 2. Typical stereovision systems used for laboratory measurements; (a) Pulnix cameras attached to 3-m gantry structure to view 1-m-wide specimen in 300-kips (1334 kN) loading frame; (b) Q-imaging cameras mounted to loading frames via translation stages to view specimen—the small angle between the cameras is due to space constraints for this application; (c) Pulnix cameras viewing rivet holes in 0.3-m-wide specimen undergoing tensile loading; and (d) high-speed Phantom V7.1 cameras viewing specimen mounted inside drop tower.

- Computer
  - Image acquisition and storage.
  - Postprocessing of images.
- Software[b]
  - Image acquisition.
  - Postprocessing of images.
  - Profile, displacement, and strain measurements.
  - Graphical presentation of measurements.
- Lighting to maintain adequate pattern contrast during calibration and experimentation.

Figure 3 presents a flow chart for a typical experiment using a two-camera stereovision system to acquire images of an object subjected to known mechanical loading and environmental conditions. In practice, the computer control system may also serve as (a) the image storage device and (b) the platform for postprocessing and data analysis.

3 BASIC CONCEPTS IN THREE-DIMENSIONAL COMPUTER VISION (3D-DIC) FOR STRUCTURAL MEASUREMENTS OF DEFORMATION AND SHAPE

A stereovision system consists of at least two views of the object from different orientations and positions. Figure 4 presents a schematic containing the key parameters related to the imaging process.

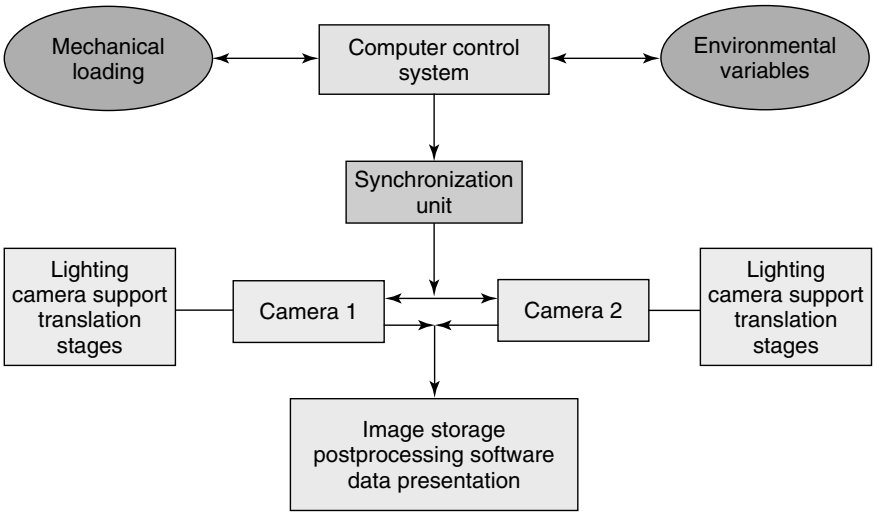


Figure 3. Flow chart for acquisition of synchronized images using a typical stereovision system.

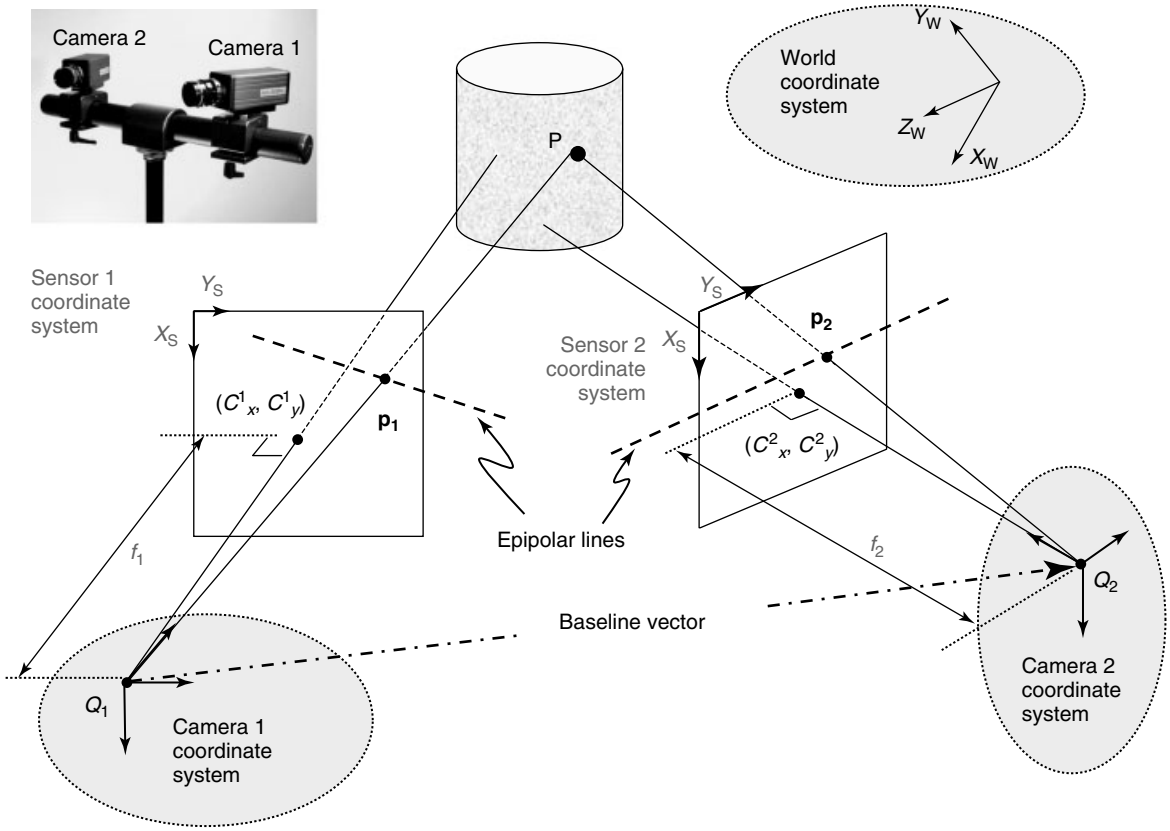


Figure 4. Actual stereovision system and schematic showing pinhole camera models for each camera.; the parameters in the model must be determined to convert the image positions ($\mathbf{p}_1$ and $\mathbf{p}_2$) into an accurate 3-D location of the point **P**.

In this study, each image system is modeled as a pinhole camera. Intrinsic parameters for each camera include (i) the focal length, f, (ii) the image plane center, (C_x, C_y), (iii) lens distortion coefficient, κ, and (iv) scale factors, λ_x and λ_y relating metric distance on the object to pixel position in the image plane. Extrinsic parameters for each camera include (i) three independent components of the rotation matrix [**R**] and (ii) three components of the translation vector to orient the world coordinate system (WCS), with axes (X_W, Y_W, Z_W), relative to a camera coordinate system (CCS) located at a camera pinhole, either Q_1 and/or Q_2.

In practice, the WCS axes oftentimes are defined to be at a specific object position during the calibration process. For example, if a two-dimensional planar grid is used for calibration then (i) the plane of the grid lines can be assumed to be the plane $(X_W, Y_W, 0)$, (ii) the grid intersection at the lower left corner can be defined to be the origin (0, 0, 0), and (iii) the Z_W axis is perpendicular to the planar grid.

In a similar manner, each CCS is generally located at the pinhole, Q. In many cases, the Z_C axis is aligned with the optical axis through the points Q–C, while X_C and Y_C are oriented to align with the camera sensor axes.

The final coordinate system defines the sensor coordinate system (SCS) with (X_S, Y_S) aligned with the row and column directions of the sensor plane, and Z_S perpendicular to the sensor plane. Located in the retinal or image plane of the pinhole camera, the SCS has units of pixels for (X_S, Y_S) to define sensor locations.

Assuming the rows and columns in the sensor plane are orthogonal, transformations among the WCS, CCS, and SCS are performed to develop the relationship between sensor plane coordinates (X_S, Y_S) of point **p** and a 3-D position of a point **P** at (X_W, Y_W, Z_W). The resulting scalar form can be written as

$$\left\{ \begin{array}{l} X_S = C_x + f\lambda_x \dfrac{R_{11}\ X_W + R_{12}\ Y_W + R_{13}\ Z_W + t_x}{R_{31}\ X_W + R_{32}\ Y_W + R_{33}\ Z_W + t_z} \\ Y_S = C_y + f\lambda_y \dfrac{R_{21}\ X_W + R_{22}\ Y_W + R_{23}\ Z_W + t_y}{R_{31}\ X_W + R_{32}\ Y_W + R_{33}\ Z_W + t_z} \end{array} \right\} \tag{1}$$

In equation (1), [**R**] is the rotation matrix and has three independent angles to define all components in [**R**].

Since equation (1) assumes an ideal, undistorted imaging system, improved accuracy in the measured positions can be obtained by including the effect of radial lens distortion at any point in the sensor plane. Such distortions are typically defined as the difference between undistorted and distorted sensor plane position using a cubic function of the radial distance from the image center. Defining the distortion vector, $\boldsymbol{d}$, and the distorted image position by (X^d_S, Y^d_S), then the corrected position, (X_S, Y_S), can be written as follows:

$$\begin{aligned} (X_S, Y_S) &= (X^d_S - d_x, Y^d_S - d_y) \\ d_x &= \kappa[(X^d_S - C_x)^2 + (Y^d_S - C_y)^2]^{3/2} \bullet \cos(\zeta) \\ d_y &= \kappa[(X^d_S - C_x)^2 + (Y^d_S - C_y)^2]^{3/2} \bullet \sin(\zeta) \\ \boldsymbol{r}(\mathbf{p}) &= [|\boldsymbol{r}| \bullet \cos(\zeta)]\boldsymbol{e}_x + [|\boldsymbol{r}| \bullet \sin(\zeta)]\boldsymbol{e}_y \\ &= \text{vector location of point } \mathbf{p} \text{ relative} \\ &\quad \text{to image center } (C_x, C_y) \\ \zeta &= \text{counterclockwise polar angle from} \\ &\quad X_S \text{ axis, with origin at } (C_x, C_y) \\ &\quad \text{with } (-\pi < \zeta \leq \pi) \\ \kappa &= \text{ratial distortion coefficient} \\ (\boldsymbol{e}_x, \boldsymbol{e}_y) &= \text{unit vectors in } X_S \text{ and } Y_S \\ &\quad \text{directions, respectively} \end{aligned} \tag{2}$$

It is worth noting that all digitized intensity patterns are recorded at integer pixel locations. Since the distortion-corrected positions typically form a nonuniform grid at noninteger pixel positions, interpolation of the pixel values is required to make measurements with optimal, subpixel accuracy.

3.1 Camera calibration

As shown in equations (1 and 2), there are 11 independent parameters for each camera to be determined during the calibration process; six extrinsic parameters ($t_x, t_y, t_z, \theta_x, \theta_y$, and θ_z) and five intrinsic parameters ($C_x, C_y, f\lambda_x, f\lambda_y$, and k). A typical camera-calibration process uses a calibration target with known grid spacing. During the calibration process, the target is translated and/or rotated in three

dimensions, with images acquired by both cameras at each position during the motion sequence. By locating at least three noncollinear points in each deformed image, the extrinsic parameters for each view (orientation and translation) are estimated and used to give initial locations for corresponding grid points throughout the view.

Defining an image-based objective function in the form

$$E = \sum_{i=1}^{M} \sum_{j=1}^{N} \{(X_{\mathrm{S\ measured}}^{ij} - X_{\mathrm{S\ model}}^{ij})^2 + (Y_{\mathrm{S\ measured}}^{ij} - Y_{\mathrm{S\ model}}^{ij})^2\} \quad (3)$$

where $X_{\mathrm{S\ model}}$ and $Y_{\mathrm{S\ model}}$ are given by equations (1) and (2), and the measured locations of features in the calibration standard ($X^{\mathrm{d}}{}_{\mathrm{S}}{}^{ij}$, $Y^{\mathrm{d}}{}_{\mathrm{S}}{}^{ij}$) are extracted from $j = 1, 2, \ldots, N$ pixel locations by analyzing all $i = 1, 2, \ldots, M$ images of the calibration standard. Though several approaches have been used to perform the nonlinear optimization of equation (3) for each camera, approaches such as steepest descent, Newton–Raphson, or a combined method such as Levenberg–Marquardt, have been used effectively.

3.2 Three-dimensional measurements

If one considers a specific point, **P**, with position (X_{W}, Y_{W}, Z_{W}), then equation (1) relates this position to the corresponding image location (X_{S}, Y_{S}) in the sensor plane for each camera; the image points are denoted by $\mathbf{p_1}$ and $\mathbf{p_2}$ for cameras 1 and 2, respectively, in Figure 4. When equation (1) is applied to both calibrated cameras, there are four equations with three unknowns, specifically the 3-D position of the point **P**. An optimal solution can be obtained through a least-square process to locate the best estimate for the position (X_{W}, Y_{W}, Z_{W}). By repeating this process for each position of interest, a full field of 3-D points can be determined at each load level. The difference between the initial position, $\mathbf{P_0}$, and the deformed position at time t, $\mathbf{P_0(t)}$, is the displacement vector with components $\{(U(\mathbf{P};t), V(\mathbf{P};t), W(\mathbf{P};t))\}$.

3.3 Image-based correlation for subset matching

To determine the corresponding image location (X_{S}, Y_{S}) in the sensor plane of each camera with utmost accuracy (this is essential for accurate 3-D measurement), the procedure used in this study is to perform optimal matching of image plane subsets; this is shown schematically in Figure 5. Defining an image of the object in one of the cameras (e.g., camera 1 in Figure 1) as the "reference" image, the matching process is typically completed by selecting a dense set of subregions (subsets) in the reference image and performing digital IC to identify the corresponding intensity pattern for each subset in (i) the other view of the undeformed object and (ii) both views of the deformed object (*see* **Statistical Pattern Recognition** for pattern recognition concepts).

Though a variety of matching metrics can be used to optimally locate the point, a normalized cross-correlation is oftentimes used. Since a nonlinear search process is required to locate the best match, most applications use an initial estimate for the parameters to initiate the search process. Typically, initial estimates for the rotation and translations of a subset are determined visually by locating the pixel positions for at least three, noncollinear, matching points and using these approximate matches to initiate the search process. As with the camera-calibration problem, a wide range of optimization methods have been used successfully.

3.4 Overall procedure for 3D vision-based measurements

The following procedure is typical for most practical situations:

1. Set up the stereovision system in preparation for the actual experiment. Cameras and lenses are selected to obtain high spatial resolution over the required visual field of view. For a stereovision system, the cameras and lenses typically are matched to simplify the setup and calibration procedures.
2. Arrange the cameras to meet the physical constraints in the application. In many cases, the

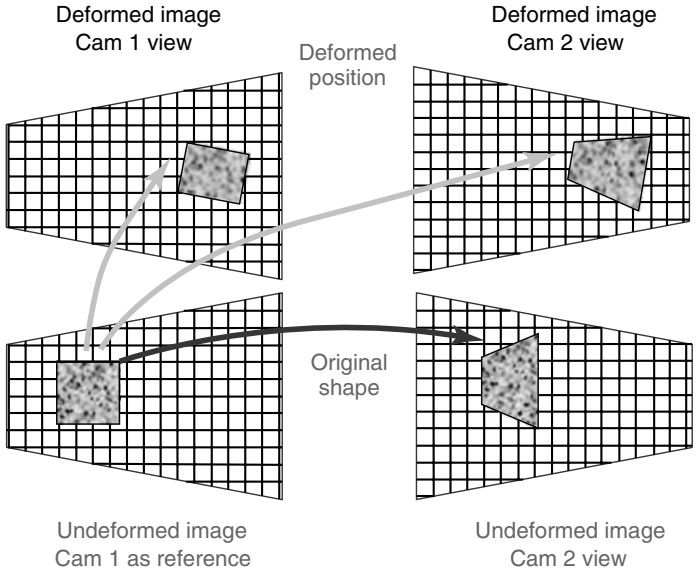


Figure 5. Schematic of image-matching process required to obtain the 3-D position of the object given corresponding images in different views. Camera 1 and camera 2 undeformed views are used to obtain initial 3-D shape of the object. Camera 1 and camera 2 deformed views are combined with camera 1 reference view to obtain the deformed 3-D position of the object.

position of the cameras is dictated by the layout (e.g., laboratory configuration).

3. Carry out a complete calibration of the stereo-vision system, using several motions of a calibration target. During this process, it is recommended that image acquisition be synchronized between both cameras.
4. Remove the calibration grid, install the specimen in the loading system, and verify that the loading is synchronized with the image acquisition process so that both cameras acquire images simultaneously during the experiment.
5. Postprocess all images to determine shape and deformations. After acquiring N pairs of images during the loading process, the analysis process requires that the images from one camera be designated as the *reference* images (e.g., camera 1).
6. To determine the initial shape (see Figure 5), subsets are selected in camera 1 and IC is performed to locate the matching position in the initial image for camera 2. Using the dense set of center-point locations obtained by the correlation process, the procedure described in Section 3.2 is used to obtain a dense set of 3-D points to define the initial object shape. For each loading state, cross-camera image matching and 3-D point generation also obtains the object shape in any configuration.
7. To determine the true 3-D position of a deformed body (see Figure 5, all four images are used as shown), subsets are selected in camera 1 and IC is performed (i) to locate the matching position in the initial image for camera 2, (ii) to locate the matching position in the deformed state as viewed by camera 1, and (iii) to locate the matching position in the same deformed image as viewed by camera 2. Using the dense set of center-point locations obtained by the correlation process, the procedure described in Section 3.2 is used to obtain a dense set of 3-D points for points in both the initial and deformed configurations. These are used to define the 3-D displacement vector for each point $\mathbf{P}$ after undergoing deformation. For each loading state, the process is repeated to obtain the displacement field for each deformed state, generating $\{(U(\mathbf{P};t), V(\mathbf{P};t), W(\mathbf{P};t))\}$.

4 APPLICATION OF THE TECHNIQUE FOR LARGE COMPONENT MEASUREMENTS

Figure 6 shows (i) a 305-mm-wide, 2.3-mm-thick, center-notched Al2024-T3 sheet with notch length/specimen width, $a/w = 1/3$, in the L–T orientation, as well as the tensile grip components, (ii) speckle pattern on the bottom half of the specimen as viewed by the left camera, and (iii) the stereocamera arrangement for acquiring 3-D shape and deformation data on the specimen as it is subjected to far-field tensile loading in a 446-kN tensile loading frame. References [38–40] provide details for similar experiments on a 0.61-m-wide panel; the configuration of cameras and specimen in these previous experiments is shown in Figure 2(a).

Canon lenses with a focal length of 28 mm were attached to Pulnix 9701 cameras with 8-bit intensity resolution and 768 × 484 spatial resolution, and used to image the specimen. As shown in Figure 6, the cameras were separated horizontally with a viewing angle difference of ≈53°. To obtain similar spatial resolution in the horizontal and vertical directions, the camera bodies were rotated 90° so that 768 pixels in the CCD sensor are aligned with the longer vertical direction of the sheet. Magnification is ≈1.5 pixels/mm. Owing to specimen size, the global speckle patterns were applied using a self-adhesive vinyl sheet with a random pattern of appropriate size for the magnification printed on the visible surface.

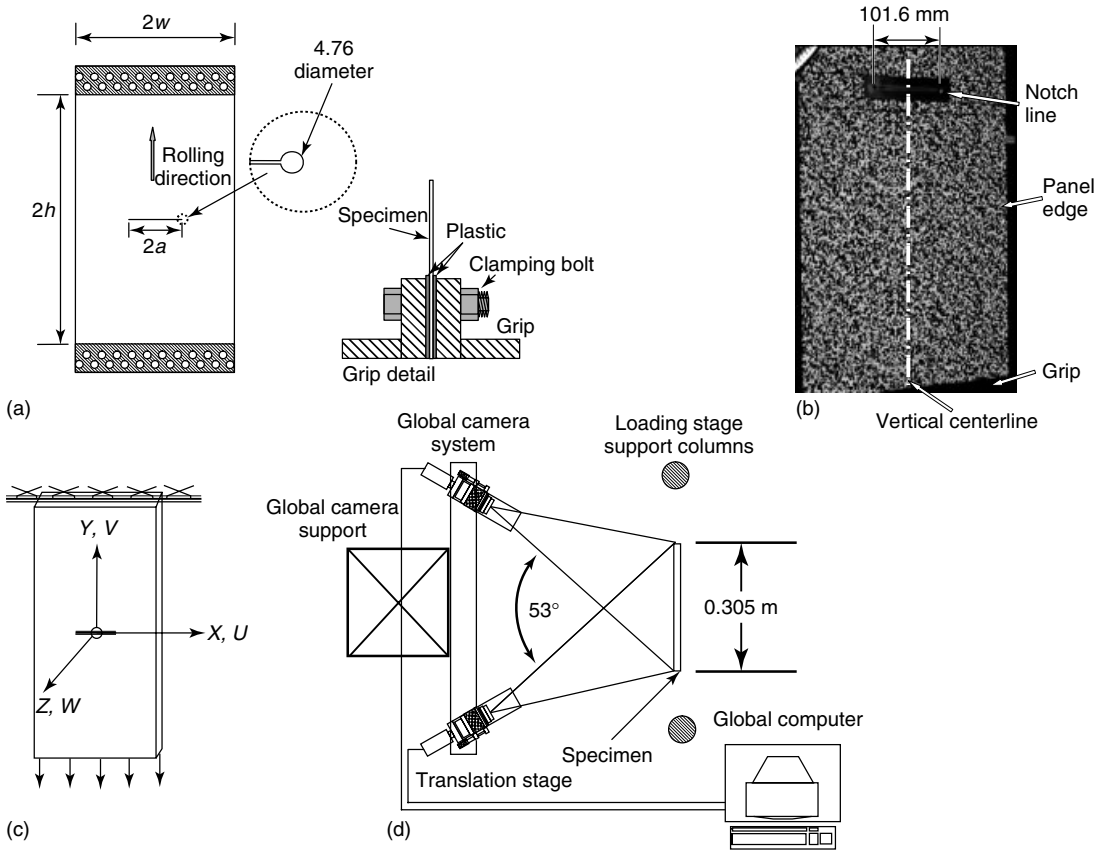


Figure 6. (a) Uniaxial tensile specimen with grips; in this work, $w = 0.1525$ m, $h = 0.330$ m, and $a/w = 1/3$; (b) speckle pattern on the bottom half of the specimen as viewed by the left camera; (c) specimen-based coordinate system to display output results; and (d) schematic of camera arrangement for the experiment. The cameras are mounted horizontal to the gantry system, at a distance of ≈1.2 m from specimen.

Camera calibration is performed using images of a calibration grid with known grid spacing. Owing to the physical size of the specimen field of view, a specially designed glass grid with grid spacing of 101.6 mm is used. To simplify the calibration process, the large glass grid is held stationary and the camera is moved approximately perpendicular to its sensor plane to acquire additional image(s). Using (i) the known spacing of the grid, (ii) the known movement(s) of the camera, and (iii) the location of the grid intersections, as extracted from the calibration images, nonlinear optimization is used to find the camera parameters that best describe the position and operating characteristics of the camera. The process is then repeated, without moving the grid, for the second camera. With this reduced motion process, measurement error had a standard deviation of ± 0.01 mm (± 0.015 pixels) with peak-to-peak values of ± 0.05 mm (± 0.075 pixels)[c]. After completing the calibration process, both CCS were transferred to the specimen shown in Figure 6(d).

The experiment was conducted using grip displacement control with a ramp rate of $2.83\,\mu\text{m}\,\text{s}^{-1}$. The experiment was paused each time the axial load increased by an increment of 2.22 kN, to acquire image data. This process was continued until the panel failed, at a load just beyond 116 kN. Load and ram displacement data were recorded every 5 s throughout the experiment.

4.1 Results

Figure 7 presents the applied load–vertical displacement data for the specimen, where the axial displacement is obtained by averaging the vision-based vertical displacement along the horizontal grip line located near the bottom of speckled image in Figure 6.

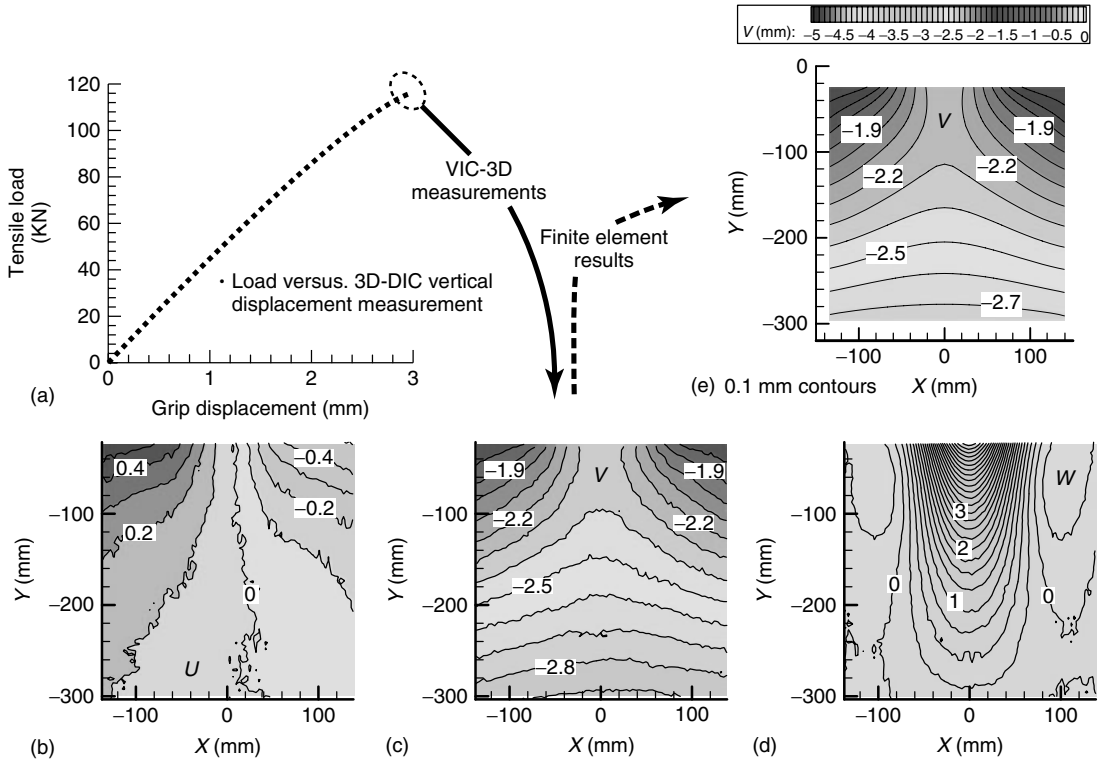


Figure 7. (a) Measured load–displacement data for specimen; (b–d) measured full-field horizontal (U), vertical (V), and out-of-plane (W) displacement fields at maximum tensile load (116 kN); (e) finite element prediction for vertical displacement field ($V(x, y)$) at maximum tensile load, assuming uniform displacement along lower boundary. All displacements are in millimeters. Measurement region is the lower half of specimen shown in Figure 6.

Inspection shows that the load–displacement data is slightly nonlinear for $P > 50$ kN, coinciding with the onset of increased out-of-plane displacements along the notch line.

Also shown in Figure 7 are the full-field plots of the displacement components in the lower half of the sheet, with (i) U, parallel to the notch line; (ii) V, perpendicular to the notch line; and (iii) W, perpendicular to the specimen surface.

As expected, the U-displacement field is antisymmetric relative to the Y axis, with both sides of the sheet moving approximately 0.40 mm toward the centerline. The V-displacement field is symmetric relative to the Y axis and nominally negative, since the bottom edge of the specimen undergoes applied downward displacement (see Figure 6(c)). As expected, the bottom edge of the specimen undergoes nearly constant displacement, confirming that the "rigid" grip was performing as expected. The W-displacement field is also symmetric relative to the Y axis, with maximum out-of-plane motion of 6 mm occurring along the centerline.

To show that the measured displacement fields are consistent with model predictions, Figure 7(e) shows a direct comparison between the V-displacement fields obtained by finite element analysis and stereovision measurements; similar comparisons are obtained for all displacement components.

5 CONCLUSIONS AND REMARKS

At the time of this article, both software and hardware are available commercially; Hardware costs include (i) scientific grade, 8–12 bit digital camera with 1000×1000 pixel array ($4000–6000); (ii) heavy-duty support structures for camera ($1000); (iii) commercial-grade, two-dimensional computer vision software and computer with data-acquisition, data-analysis, and data-presentation capability ($10 000–20 000); and (iv) commercial-grade, three-dimensional computer vision software and computer with data-acquisition, data-analysis, and data-presentation capability ($40 000–65 000).

The combination of modern digital image processing with stereovision imaging offers a unique opportunity to obtain quantitative deformation data on specimens that are undergoing a combination of in-plane and out-of-plane deformations. As long as the specimen remains within the depth of field for both cameras, the method is capable of measuring full-field 3-D shape, 3-D displacements, and surface strains on planar or curved specimen surfaces. The method has been used to make measurements in a wide range of applications including the following:

1. surface strains in excess of 100% on a highly ductile elastic–plastic metallic materials;
2. thermal strains on metallic specimens heated to 700 °C (*see* **Thermal Imaging Methods** for thermal imaging methods, **Gas Turbine Engines**);
3. thermal and mechanical strains on specimens being imaged in a scanning electron microscope;
4. deformations on curved fuselage structure during both internal pressurization and external dynamic impact (*see* **Civil Infrastructure Load Models for Structural Health Monitoring**; **Static Damage Phenomena and Models**; **Damage Evolution Phenomena and Models**; **Failure Modes of Aerospace Materials**; **A Simplified Damage Model for SHM Metallic and Composite Structures**; **Fatigue Life Assessment of Structures**; **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**; **Damage Presence/Growth Monitoring Sensors**; **Principles of Structural Degradation Monitoring**; **Design Principles for Aerospace Structures**; **Design Principles for Civil Structures**; **Risk Monitoring of Aircraft Fatigue Damage Evolution at Critical Locations**; **Risk Monitoring of Civil Structures**; **Environmental Monitoring of Aircraft**; **Maintenance Principles for Civil Structures**; **Use of Leave-in-place Sensors and SHM Methods to Improve Assessments of Aging Structures**; **Usage Management of Military Aircraft Structures**; **Usage Management of Civil Structures**; **Value Assessment Approaches for Structural Life Management**; **Commercial Fixed-wing Aircraft**; **History of SHM for Commercial Transport Aircraft**; **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Unmanned Aerial Vehicles**; **Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance**; **Validation of SHM Sensors in Airbus A380 Full-scale**

Fatigue Test; **Design, Analysis, and SHM of Bonded Composite Repair and Substructure**; **Monitoring of Solid Rocket Motors**.)

5. concrete, asphalt, polymers, and fiber-reinforced composites during mechanical loading (*see* **Lamb Wave-based SHM for Laminated Composite Structures**; **Modeling Aspects in Finite Elements**; **Design, Analysis, and SHM of Bonded Composite Repair and Substructure**; **Fatigue Monitoring in Nuclear Power Plants**).
6. civil engineering structures including bridges and joints (*see* **Applications of Acoustic Emission for SHM: A Review**; **Design Principles for Civil Structures**; **Risk Monitoring of Civil Structures**; **Maintenance Principles for Civil Structures**; **Long-term Monitoring of Dynamic Loads on the Brandenburg Gate**; **Development of a Monitoring System for a Long-span Cantilever Truss Bridge**; **Modular Architecture of SHM System for Cable-supported Bridges**; **Monitoring of Bridges in Korea**; **Bridge Monitoring in Japan**; **Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge**; **Continuous Monitoring of the Øresund Bridge: Data Acquisition and Operational Modal Analysis**; **Condition Compensation in Frequency Analyses—a Basis for Damage Detection**; **Modal Testing of the Vasco da Gama Bridge, Portugal**; **Multiple-model Structural Identification**; **Construction Process Monitoring at the New Berlin Main Station**; **SHM Actions on the Holy Shroud Chapel in Torino**; **SHM of a Tall Building**; **Dynamic Response of Buildings of the Cultural Heritage**; **Suspended Roof of Braga Sports Stadium, Portugal**; **Dams**; **Condamine Floating Dock, Monaco**; **Soil–Structure Interaction and Seismic Effects**; **System Identification for Soil–structure Interaction**).

ACKNOWLEDGMENTS

The author wishes to thank Dr Jeffrey D. Helm for his tireless efforts in completing the bulk of the wide panel experimental work at NASA Langley Research Center. In addition, the technical and editorial support of Dr Hubert Schreier, Dr Stephen R. McNeill, and Dr Junhui Yan in completing this manuscript is deeply appreciated. The financial support of (i) Dr Charles E. Harris, Dr Robert S. Piascik, and Dr James C. Newman, Jr. at NASA Langley Research Center, (ii) Dr Oscar Dillon, Dr Clifford Astill, and Dr Albert S. Kobayashi, former NSF Solid Mechanics and Materials Program Directors, (iii) Dr Bruce LaMattina at the Army Research Office through several grants including ARO 50408-EG-DPS, (iv) Dr Kenneth Chong through NSF CMS-0201345, and (v) the University of South Carolina, Office of Research, are gratefully acknowledged. Finally, the support provided by Correlated Solutions, Inc., through the granting of access to their commercial software for our internal use is deeply appreciated. Through the unwavering technical and financial assistance of all these individuals and organizations, the true potential of image correlation methods is now being realized.

END NOTES

[a.] The original version was developed by Prof. K.S. Kim, Brown University, with emphasis on methods available for measurements on reduced-length scale specimens.

[b.] All image analysis in this work was performed using VIC-2D and VIC-3D software developed by Correlated Solutions Inc.; 120 Kaminer Way, Columbia, SC 29205, www.correlatedsolutions.com.

[c.] Modifications to this procedure have been developed and converted into commercial code. In the version developed by Correlated Solutions, Inc., the grid can be moved and rotated freely, acquiring images simultaneously by all stereo cameras and using equation (3) to efficiently convert grid images into calibration parameter sets for both cameras.

REFERENCES

[1] Doyle FJ. The historical development of analytical photogrammetry. *Photogrammetric Engineering* 1964 **XXX**:259–265.

[2] *The Photogrammetric Record* Gilbert Louis Hobrough. 2003 **18**(104):337–340.

[3] Rosenfeld A. From image analysis to computer vision: an annotated bibliography, 1955–1979. *Computer Vision and Image Understanding* 2001 **84**:298–324.

[4] Gabor D. Microscopy by reconstructed wavefronts. *Proceedings of the Royal Society* 1949 **A197**:454–487.

[5] Haines K, Hildebrand BP. Contour generation by wavefront construction. *Physics Letters* 1965 **21**: 422–423.

[6] Leith EN, Upatnieks J. Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America* 1962 **25**:1123–1130.

[7] Dainty JC (ed). *Laser Speckle and Related Phenomena*, Springer-Verlag: Berlin, 1975.

[8] Archbold E, Burch JM, Ennos AE. Recording of in-plane surface displacements by double exposure speckle photography. *Optica Acta* 1970 **17**:883–898.

[9] Luxmoore AR, Amin FAA, Evans WT. In-plane strain measurement by speckle photography: a practical assessment of the use of Young's fringes. *Journal of Strain Analysis* 1974 **9**:26–34.

[10] Mallik S, Roblin ML. Speckle pattern interferometry applied to the study of phase objects. *Optics Communications* 1972 **6**:45–49.

[11] Leendertz JA, Butters JN. An image shearing speckle pattern interferometer for measuring bending moments. *Journal of Physics E: Scientific Instruments* 1973 **7**:1107–1110.

[12] Post D. White light Moiré interferometry. *Applied Optics* 1979 **24**:4163–4167.

[13] Post D, Ifju P, Han BT. *High Sensitivity Moiré*. Springer-Verlag, 1994.

[14] Vest CM. *Holographic Interferometry*. John Wiley & Sons, 1979.

[15] Peters WH, Ranson WF. Digital imaging techniques in experimental stress analysis. *Optical Engineering* 1982 **21**(3):427–431.

[16] Sutton MA, Wolters WJ, Peters WH, Ranson WF, McNeill SR. Determination of displacements using an improved digital correlation method. *Image and Vision Computing* 1983 **1**(3):133–139.

[17] Chu TC, Ranson WF, Sutton MA, Peters WH. Applications of digital-image-correlation techniques to experimental mechanics. *Experimental Mechanics* 1985 **25**(3):232–244.

[18] Peters WH, Zheng-Hui HE, Sutton MA, Ranson WF. Two-dimensional fluid velocity measurements by use of digital speckle correlation techniques. *Experimental Mechanics* 1984 **24**(2):117–121.

[19] Sutton MA, Chae TL, Turner JL, Bruck HA. Development of a computer vision methodology for the analysis of surface deformations in magnified images, ASTM STP 1094. In *MiCon 90: Advances in Video Technology for Microstructural Control*, Vander Voort GF (ed). ASTM: Conshohocken, PA, 1991, pp. 109–132.

[20] Sutton MA, Cheng MQ, Peters WH, Chao YJ, McNeill SR. Application of an optimized digital correlation method to planar deformation analysis. *Image and Vision Computing* 1986 **4**(3): 143–150.

[21] Sutton MA, McNeill SR, Helm JD, Chao YJ. Advances in two-dimensional and three-dimensional computer vision. In *Photomechanics, Topics in Applied Physics*, Rastogi PK (ed). Springer Verlag: Berlin, 2000.

[22] Bruck HA, McNeill SR, Sutton MA, Peters WH. Digital image correlation using Newton-Raphson method of partial differential correction. *Experimental Mechanics* 1989 **29**(3):261–267.

[23] Sutton MA, Turner JL, Chae TL, Bruck HA. Full field representation of discretely sampled surface deformation for displacement and strain analysis. *Experimental Mechanics* 1991 **31**(2): 168–177.

[24] Amstutz BE, Sutton MA, Dawicke DS. Experimental study of mixed mode I/II stable crack growth in thin 2024-T3 aluminum. *ASTM STP 1256 on Fatigue and Fracture*, ASTM: Conshohocken, PA, 1995; Vol. 26, pp. 256–273.

[25] Han G, Sutton MA, Chao YJ. A study of stationary crack tip deformation fields in thin sheets by computer vision. *Experimental Mechanics* 1994 **34**(2):751–761.

[26] Han G, Sutton MA, Chao YJ. A study of stable crack growth in thin SEC specimens of 304 stainless steel. *Engineering Fracture Mechanics* 1995 **52**(3):525–555.

[27] Liu J, Sutton MA, Lyons JS. Experimental characterization of crack tip deformations in Alloy 718 at High Temperatures. *ASME Journal of Engineering Materials and Technology* 1998 **20**(1): 71–78.

[28] Lyons JS, Liu J, Sutton MA. High-temperature deformation measurements using digital-image correlation. *Experimental Mechanics* 1996 **36**(1): 64–70.

[29] Chao YJ, Sutton MA. Measurement of strains in a paper tensile specimen using computer vision and digital image correlation – Part 1: data acquisition and image analysis system. *Tappi Journal* 1988 **70**(3):173–175.

[30] Chao YJ, Sutton MA. Measurement of strains in a paper tensile specimen using computer vision and digital image correlation – Part 2: tensile specimen test. *Tappi Journal* 1988 **70**(4):153–156.

[31] Schreier HW, Braasch J, Sutton MA. Systematic errors in digital image correlation caused by intensity interpolation. *Optical Engineering* 2000 **39**(11):2915–2921.

[32] Schreier HW, Sutton MA. Systematic errors in digital image correlation due to undermatched subset shape functions. *Experimental Mechanics* 2002 **42**(3):303–310.

[33] Schreier HW, Garcia D, Sutton MA. Advances in light microscope stereo vision. *Experimental Mechanics* 2004 **44**(3):278–288.

[34] Luo PF, Chao YJ, Sutton MA. Computer vision methods for surface deformation measurements in fracture mechanics. *ASME-AMD Novel Experimental Methods in Fracture* 1993 **176**:123–133.

[35] Luo PF, Chao YJ, Sutton MA, Peters WH. Accurate measurement of three-dimensional deformations in deformable and rigid bodies using computer vision. *Experimental Mechanics* 1993 **33**(2):123–132.

[36] Luo PF, Chao YJ, Sutton MA. Application of stereo vision to three-dimensional deformation analyses in fracture experiments. *Optical Engineering* 1994 **33**(3):981–990.

[37] Helm JD, McNeill SR, Sutton MA. Improved 3-D image correlation for surface displacement measurement. *Optical Engineering* 1996 **35**(7):1911–1920.

[38] Helm JD, Sutton MA, Dawicke DS, Hanna G. Three-dimensional computer vision applications for aircraft fuselage materials and structures, *Proceedings of 1st Joint DoD/FAA/NASA Conference on Aging Aircraft*, Ogden, Utah, 1997; Vol. 1, pp. 1327–1341.

[39] Helm JD, Sutton MA, McNeill SR. Deformations in wide, center-notched, thin panels, part I: three-dimensional shape and deformation measurements by computer vision. *Optical Engineering* 2003 **42**(5):1293–1305.

[40] Helm JD, Sutton MA, McNeill SR. Deformations in wide, center-notched, thin panels, part II: Finite element analysis and comparison to experimental measurements. *Optical Engineering* 2003 **42**(5):1306–1320.

Chapter 63

Electric and Electromagnetic Properties Sensing

Michel B. Lemistre

Laboratoire SATIE/CNRS, Ecole Normale Supérieure de Cachan, Cachan, France

1 INTRODUCTION

In the domain of aeronautics, particularly in the area concerning composite structures, electromagnetic techniques are little used for structural health monitoring (SHM). This could be explained by the fact that the main researchers working in the field of SHM have mechanical engineering training. One possible exception concerns eddy currents techniques, but these are mainly used for metallic structures [1]. In fact, in the case of composite structures, eddy currents techniques are impossible or very difficult to use and generally give poor results, these kinds of structures being either purely dielectric or poor conductors (e.g., carbon epoxy). However, another method also based on eddy currents, but using a low frequency holographic technique [2–4], gives good results on carbon–epoxy structures; nevertheless, this method is not easily transposable into the SHM domain, the external equipment required being too complicated. However, there is a possible application for SHM by measurement of electrical resistance or electrical potential [5, 6]. One other possibility is a dynamic capacitive method [7, 8] uniquely used in the domain of civil engineering.

A new family of electromagnetic techniques, which makes it possible to obtain good information on the health of structures made of composite (i.e., carbon fiber reinforced plastic (CFRP) and glass fiber reinforced plastic (GFRP)) has been developed. These techniques consist of determining the state of the health of a structure by measurement of its two electrical parameters, the electric conductivity σ and/or the dielectric permittivity ε, since damages induces locally significant variations of these parameters. The goal of this article is to explain these techniques and to give their field of application.

First, we shall provide a recall of the electromagnetic theory necessary for gaining a full understanding of the electromagnetic techniques developed.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Next, the various techniques will be explained with examples of application. The transposition of these techniques in the domain of SHM are fully explained in **The HELP-Layer® System**.

2 THEORETICAL CONSIDERATIONS

2.1 Surface impedance

Let us consider a plane structure made of a material having the following electrical properties:

- magnetic permeability $\mu = \mu_0$;
- dielectric permittivity $\varepsilon = \varepsilon_0 \varepsilon_\mathrm{r}$;
- electric conductivity σ;
- thickness d.

The structure is illuminated by a plane wave with an inclined incidence (see Figure 1).

Electric and magnetic components of the incident wave $\vec{p}_\mathrm{i}$ are $\vec{E}_\mathrm{i}$ and $\vec{H}_\mathrm{i}$ respectively, $\vec{E}_\mathrm{r}$ and $\vec{H}_\mathrm{r}$ being the components of the refracted wave $\vec{p}_\mathrm{r}$; θ_1 and θ_2 are the angles of incidence and refraction, respectively; N_1 is the refractive index of the external medium and N_2 is the refractive index of the structure. θ_1 and θ_2 are linked by the following relation:

$$N_1 \sin\theta_1 = N_2 \sin\theta_2 \tag{1}$$

If one considers that the external medium is in free space, $N_1 = 1$. Taking into account the complex relative permittivity of the material $\varepsilon_\mathrm{r}^* = \varepsilon'_\mathrm{r} - j\varepsilon''_r$, N_2 the index of the material is given by

$$N_2 = \sqrt{\varepsilon_\mathrm{r}^*} = \sqrt{\varepsilon'_\mathrm{r} - j\frac{\sigma}{\omega\varepsilon_0}} \tag{2}$$

If one considers the material as a good conductive material (i.e., $\sigma \gg j\omega\varepsilon$), N_2 can be reduced to the following:

$$N_2 = \sqrt{-j\frac{\sigma}{\omega\varepsilon_0}} \tag{3}$$

then the relation (1) becomes

$$\sin\theta_2 = \frac{1}{N_2}\sin\theta_1 \tag{4}$$

or

$$\cos\theta_2 = \left(1 - \left(\frac{1}{N_2}\right)^2 \sin^2\theta_1\right)^{\frac{1}{2}} = \left(1 + j\frac{\omega\varepsilon_0}{\sigma}\sin^2\theta_1\right)^{\frac{1}{2}} \tag{5}$$

For good conductive materials, neglecting the second term of the parenthesis, the relation (5) yields

$$\cos\theta_2 \cong 1 \tag{6}$$

Therefore, the wave penetrates through the material perpendicular to the surface of the structure ($\theta_2 = 0$). Figure 2 shows the field's configuration inside the structure.

The two fields $\vec{E}(z)$ and $\vec{H}(z)$ can be written as follows:

$$\vec{E}(z) = \vec{E}_a \exp(-\gamma z) + \vec{E}_b \exp(+\gamma z) \tag{7a}$$

$$\vec{H}(z) = \left(\vec{n} \times \vec{E}_a\right)\frac{\exp(-\gamma z)}{Z} + \left(\vec{n} \times \vec{E}_b\right)\frac{\exp(+\gamma z)}{Z} \tag{7b}$$

with Z being the wave impedance inside the material defined by the relation

$$Z = \sqrt{\frac{\mu_0}{\varepsilon^*}} = (j+1)\sqrt{\frac{\mu f \pi}{\sigma}} = \frac{j+1}{\sigma\delta} \tag{8}$$

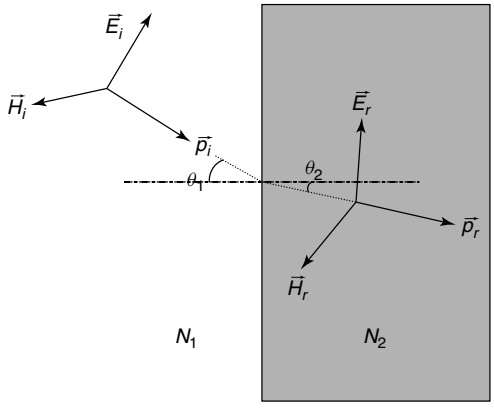


Figure 1. Illumination of a conductive material.

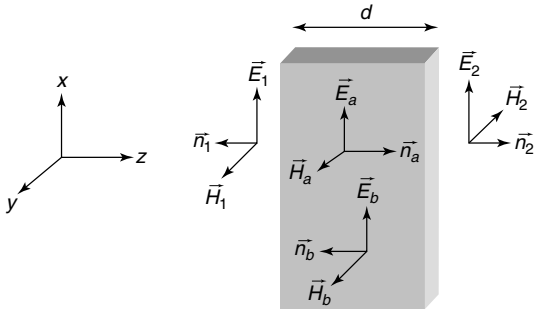


Figure 2. Field configuration inside a structure of thickness d.

and γ the propagation constant given by

$$\gamma = j\omega\sqrt{\varepsilon^*\mu_0} = (j+1)\sqrt{\mu f \pi \sigma} = \frac{j+1}{\delta} \quad (9)$$

where $\vec{n}$ is the unit vector in the z direction (i.e., the thickness of the material), f the frequency of the incident wave, and δ the skin depth defined by the following relation:

$$\delta = \frac{1}{\sqrt{\mu f \pi \sigma}} \quad (10)$$

Electric and magnetic fields tangential to the structure, on each one of its faces, $\vec{E}_1$, $\vec{H}_1$ and $\vec{E}_2$, $\vec{H}_2$ (Figure 2), are given by the following equations:

$$\vec{E}_1 = \frac{Z}{th(\gamma d)}\vec{n}_1 \times \vec{H}_1 + \frac{Z}{sh(\gamma d)}\vec{n}_2 \times \vec{H}_2 \quad (11a)$$

$$\vec{E}_2 = \frac{Z}{sh(\gamma d)}\vec{n}_1 \times \vec{H}_1 + \frac{Z}{th(\gamma d)}\vec{n}_2 \times \vec{H}_2 \quad (11b)$$

One can define a surface current density $\vec{J}_s$, which is the integral of the volume current density $\vec{J}$ in the thickness of the structure. One can write the boundary equations:

$$\vec{J}_s = \vec{n}_1 \times \left(\vec{H}_1 - \vec{H}_2\right) = \vec{n}_2 \times \left(\vec{H}_2 - \vec{H}_1\right) \quad (12)$$

For low frequencies such as $d \ll \delta$, that is, $|\gamma d| \ll 1$, one can write the following approximations:

$$sh(\gamma d) \cong \gamma d \quad (13a)$$

and

$$th(\gamma d) \cong \gamma d \quad (13b)$$

Relations (11a) and (11b) yield

$$\vec{E}_1 = \vec{E}_2 = \vec{E}_{tg} = \frac{1}{\sigma d}\vec{J}_s \quad (14)$$

E_{tg} being the electric field tangential to the surface of the structure, it is now possible to define the surface impedance Z_s as

$$\vec{E}_{tg} = Z_s \vec{J}_s \quad (15)$$

with

$$Z_s = \frac{1}{\sigma d} \quad (16)$$

One can see that at low frequencies, the current density is distributed uniformly in the material. One can then neglect the skin effect and resonances inside the material. The condition $|\gamma d| \ll 1$ (or $\delta > d$) defines the domain of application of the concept of surface impedance; the material is then called *electrically thin*. In the case of $\delta < d$, the material is called *electrically thick*; losses by attenuation inside the material and by reflection are then the main phenomena.

The surface impedance is given in "square ohm" (symbol Ω_c); it is the impedance of an electrically thin square sample, having a conductivity σ. One can represent the equivalent electric diagram to the surface impedance by Figure 3.

2.2 Diffraction by a circular aperture

For an incident plane wave, Bethe [9] has given an approximate analytical representation of diffracted

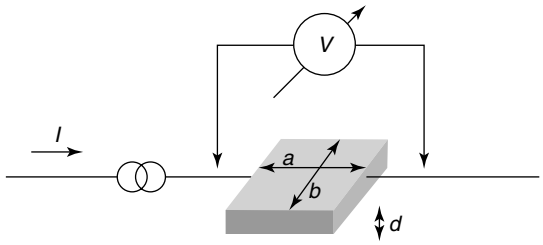


Figure 3. Equivalent electric diagram of the surface impedance.

fields by a small circular aperture in a structure that is considered as infinitely conductive (a good conductive metal such as aluminum can be considered as infinitely conductive), having an infinite surface (i.e., large compared with the diameter of the aperture), with the following assumptions:

- The size of the aperture is small compared with the wavelength of the incident field.
- The fields are calculated at a large distance compared with the size of the aperture.

Bethe has given the concept of "short-circuit fields" $\vec{E}_{cc}$ and $\vec{H}_{cc}$, which represent the fields on the aperture loaded by a perfectly conductive material. These fields are defined in the following manner:

$$\vec{E}_{cc} = 2\vec{E}_0 \tag{17a}$$

and

$$\vec{H}_{cc} = 2\vec{H}_0 \tag{17b}$$

where $\vec{E}_0$ and $\vec{H}_0$ are the orthogonal electric component and the tangential magnetic component of the incident field, respectively. The diffracted fields by the aperture are the sum of the radiated fields by an electric dipole having a moment $\vec{P}_e$ and a magnetic dipole having a moment $\vec{P}_m$. These two dipoles model the orthogonal and tangential fields, respectively as shown in Figure 4.

Let us introduce the concept of "polarizability"; the dipolar moments are related to short-circuit fields by the following relations:

$$\vec{P}_e = \varepsilon\alpha_e\vec{E}_{cc} \tag{18a}$$

and

$$\vec{P}_m = -\overline{\overline{\alpha}}_m\vec{H}_{cc} \tag{18b}$$

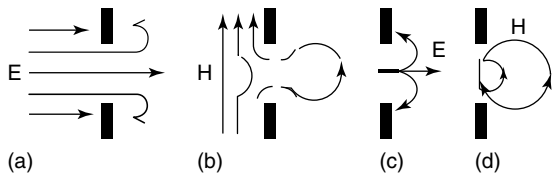


Figure 4. Incident fields and equivalent dipoles: (a) orthogonal electric field, (b) tangential magnetic field, (c) equivalent electric dipole, and (d) equivalent magnetic dipole.

where α_e and $\overline{\overline{\alpha}}_m$ are the electric polarizability and the magnetic polarizability, respectively, given by an aperture of radius a.

The polarizability depends on the geometry of the structure and the aperture. For a plane aperture of any geometry, the electric polarizability is a scalar number and the magnetic polarizability is a tensor of rank 1. For a plane circular aperture of radius a, the magnetic polarizability is a diagonal tensor with the two diagonal terms being equal.

$$\alpha_e = \frac{2}{3}a^3 \tag{19a}$$

$$\alpha_{mxy} = \alpha_{myx} = 0 \tag{19b}$$

$$\alpha_{mxx} = \alpha_{myy} = \alpha_m = \frac{4}{3}a^3 \tag{19c}$$

With Bethe's assumptions, Casey [10] has calculated the moment of the magnetic dipole for a plane circular aperture having a radius a, loaded by a nonperfect conductive material such as carbon epoxy. Casey defines a resistance of "electrical gasket" R_g as the contact resistance between the conductive structure and the material loading the aperture. R_g is given by the product ohm × meter (length of contact between the material and the structure). Casey's method involves the solution of an integral equation. Two solutions are proposed—the first one being exact leads to a semianalytical expression of the magnetic dipole, including a coefficient that must be calculated numerically; the second one, obtained by an approximate method, leads to the dipolar moment $\vec{P}_m$ under the form of the transfer function of a first order low pass filter.

$$\frac{\vec{P}_m}{\vec{P}_{m0}} = \frac{1}{1 + j\dfrac{f}{f_c}} \tag{20}$$

In this expression, $\vec{P}_{m0}$ is the magnetic dipolar moment given by a free aperture, f represents the frequency of the incident wave, and f_c is the cutoff frequency given by the following relation:

$$f_c = \frac{3}{8\mu_0}\frac{Z_s}{a}\left(1 + \frac{R_g}{aZ_s}\right) \tag{21}$$

It should be noted here that the cut-off frequency is a function of the surface impedance Z_s and thus

the conductivity σ of the material (relation 16). This cutoff frequency is called *Casey's frequency*.

2.3 Polarization of dielectrics

All materials that have a conductivity $\sigma \leq 10^{-20}$ $S{\cdot}m^{-1}$ at room temperature ($\approx$300 K) are called *dielectrics*. On a large scale, dielectrics seem to be electrically neutral. However, on a microscopic scale, dielectrics show an "assembly" of elementary electric dipoles having a random space orientation. For a large number of dipoles, one can consider dielectrics as statistically neutral. If a dielectric material is subjected to an electric field, elementary dipoles tend to orient in the direction of the incident electric field; one can define a polarization vector per unit volume $\vec{P}$ as

$$\vec{P} = Nq\vec{\delta} \tag{22}$$

with q elementary charges (per atom or molecule), separated by a distance $\vec{\delta}$ and N the number of atoms (or molecules) per unit volume. The product $q\vec{\delta}$ represents the elementary dipolar moment $\vec{p}$ for each atom (or molecule); this dipolar moment is related to the local electric field $\vec{E}_0$ by $\vec{p} = \alpha\vec{E}_0$. In this relation, the proportionality factor α is called *polarizability*, and it is a function of the electrical characteristics of the medium (i.e., the dielectric) and more precisely of its dielectric relative permittivity ε_r, defined by the Clausius–Mossoti's relation [11]:

$$\alpha = \frac{3}{4\pi N}\left(\frac{\varepsilon_r - 1}{\varepsilon_r + 2}\right) \tag{23}$$

The polarization phenomenon in dielectric medium results from three different sources: electronic polarization, ionic polarization, and orientation polarization. These three sources admit polarizability coefficients α_e, α_i, and α_o, respectively; the real part of the coefficient α is the sum of the three real parts of elementary polarizability.

Electronic polarization arises because the center of local electronic charge cloud around the nucleus is displaced under the action of the electric field $\vec{P}_e = N\alpha_e\vec{E}_0$.

Ionic polarization occurs in ionic materials because the electric field displaces positives and negatives ions in opposite directions $\vec{P}_i = N\alpha_i\vec{E}_0$.

Orientation polarization can occur in materials composed of molecules that have permanent electric dipoles. The alignment of these dipoles depends on temperature and leads to an "orientational polarizability" per molecule $\alpha_o = \frac{p^2}{3KT}$, where p is the permanent dipolar moment per molecule, K is the Boltzmann constant, and T is the temperature.

Because of the different nature of these three polarization processes, the response of a dielectric solid to an applied electric field will strongly depend on the frequency of the field. The resonance of the electronic excitation takes place in the ultraviolet part of the electromagnetic spectrum; the characteristic frequency of the ions' vibration is located in the infrared, while the orientation of dipoles requires fields of much lower frequencies (below 10^9 Hz). This response to electric field of different frequencies is shown in Figure 5.

For a low-frequency excitation (i.e., 1 kHz to 10 MHz), one can consider the response of the dielectric medium to be quasi-static. So, it is possible to describe the electric field inside the dielectric by the following equation:

$$\nabla{\cdot}\vec{E} = \frac{\rho_l + \rho_p}{\varepsilon_0} \tag{24}$$

where ρ_l is the free charge density and ρ_p an apparent density of charges due to the polarization phenomenon. The field $\vec{E}$ can be considered as the resultant of two components: the field resulting from the free charges $\vec{E}_f$ and the field resulting from the polarization phenomenon $\vec{E}_p$, that is, $\vec{E} = \vec{E}_f + \vec{E}_p$.

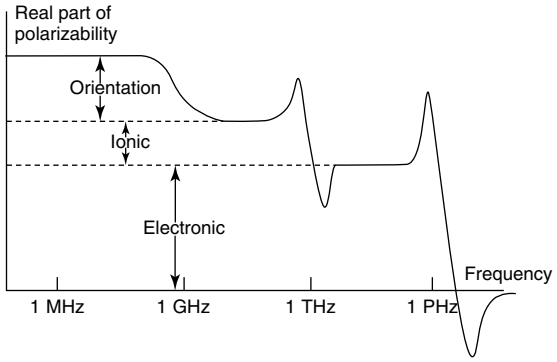


Figure 5. Frequency dependence of the different contribution to polarizability. From Handbook of Chemistry and Physics.

Then

$$\nabla\cdot\vec{E} = \nabla\cdot\vec{E}_{\mathrm{f}} + \frac{\rho_{\mathrm{p}}}{\varepsilon_0} \tag{25}$$

One can define a polarization vector $\vec{P}$ by $\rho_{\mathrm{p}} = -\nabla\cdot\vec{P}$ (see [12]). Equation (25) can be rewritten as

$$\vec{E} = \vec{E}_{\mathrm{f}} + \frac{\vec{P}}{\varepsilon_0} \tag{26}$$

The vector $\vec{P}$ is linked with the incident field $\vec{E}_{\mathrm{i}}$ by the following relation:

$$\vec{P} = \chi_{\mathrm{e}}\varepsilon_0\vec{E}_{\mathrm{i}} \tag{27}$$

where χ_{e} is the electric susceptibility and is linked with the relative electric permittivity ε_{r} by the expression

$$\varepsilon_{\mathrm{r}} = 1 + \chi_{\mathrm{e}} \tag{28}$$

Taking into account relations (26–28), one has

$$\vec{E} = \vec{E}_{\mathrm{f}} + \left(\varepsilon_{\mathrm{r}} - 1\right)\vec{E}_{\mathrm{i}} \tag{29}$$

However, in most of dielectrics, the term due to the free charges can be neglected and the value of the electric field $\vec{E}$ is reduced to the polarization term $(\varepsilon_{\mathrm{r}} - 1)\vec{E}_{\mathrm{i}}$.

3 ELECTRICAL CAPACITANCE SENSORS

Let us consider a dielectric material having an electric relative permittivity ε_{r}, in the presence of an electric field $\vec{E}_{\mathrm{i}}$, the material being considered in a macroscopic manner (i.e., quasi-isotropic). The electric field $\vec{E}_{\mathrm{i}}$ induces a phenomenon of polarization inside the material (see Section 2.3), characterized by the vector $\vec{P}$. The total electric field $\vec{E}_{\mathrm{T}}$ measured at point A (Figure 6) can be written as

$$\vec{E}_{\mathrm{T}} = \vec{E}_{\mathrm{i}} + \frac{\vec{P}}{\varepsilon_0} = \vec{E}_{\mathrm{i}} + \left(\varepsilon_{\mathrm{r}} - 1\right)\vec{E}_{\mathrm{i}} \tag{30}$$

After subtraction of the incident electric field $\vec{E}_{\mathrm{i}}$, one can obtain directly the value of ε_{r}. Note that, the frequency of excitation must be lower than the cutoff frequency of the orientation polarization phenomenon (i.e., 10 MHz).

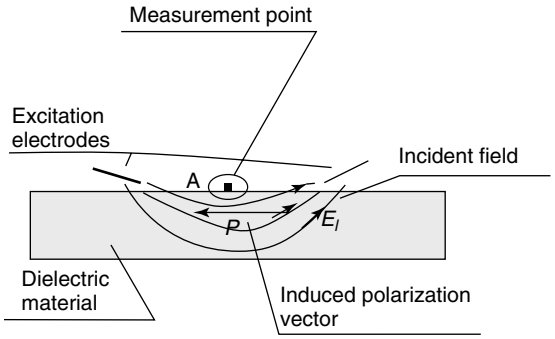


Figure 6. Configuration of measurement applied to dielectric materials.

This method can be applied to all dielectric composite materials such as GFRP, sandwich, etc. An electric sensor has been designed [13] allowing to detect some defects in dielectric composites. This sensor performs a differential measurement between two adjacent zones (Figure 7).

Figure 8 shows an example of detection performed on a sample of glass epoxy sandwich with a lack of foam in the middle of the lower part. Figure 8(b) shows the electric image obtained by scanning the material.

This sensor allows designing a system of SHM dedicated to glass epoxy or sandwich structures and more generally all dielectric composite structures (*see* **The HELP-Layer® System**).

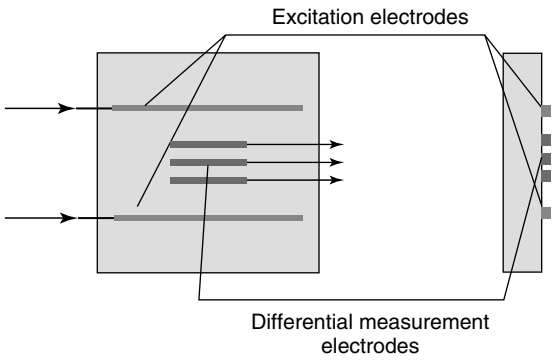


Figure 7. Differential electric capacitive sensor for detection of damages.

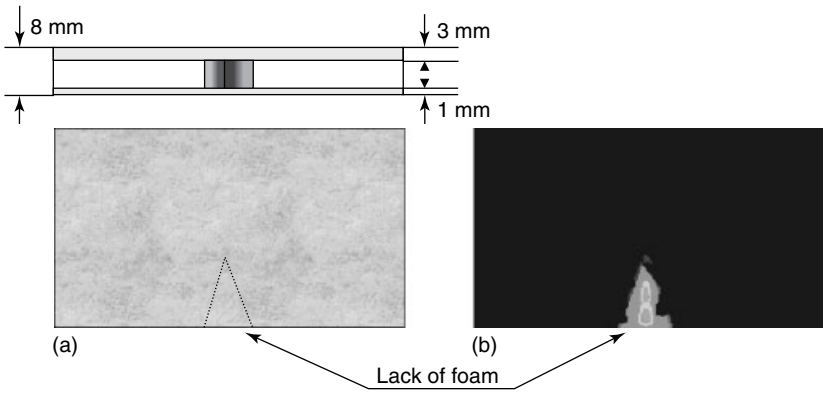


Figure 8. Example of damage detection in a sandwich structure.

4 ELECTROMAGNETIC SENSORS FOR COMPOSITE MATERIALS

4.1 Magnetic sensor

The conductive materials such as metallic structures having a conductivity σ between 10^7 and $10^8\ \mathrm{S{\cdot}m^{-1}}$ will not be considered here. For such materials, the effectiveness of classical methods using eddy currents is well established and is not necessary to demonstrate here. Here we consider composite materials having a mean conductivity σ about $10^4\ \mathrm{S{\cdot}m^{-1}}$ such as CFRP, so-called nonperfectly conductive materials.

One has seen, in Section 2.2 that the magnetic dipolar moment $\vec{P}_{\mathrm{m}}$ of an aperture loaded by a nonperfectly conductive material, can be represented in the form of a transfer function of a first-order low pass filter. The dipolar moment being directly proportional to the magnetic field, the terms $\vec{P}_{\mathrm{m}}$ and $\vec{P}_{\mathrm{m0}}$ can be replaced by $\vec{H}$ and $\vec{H}_0$, respectively in relation (20):

$$\frac{\vec{H}}{\vec{H}_0} = \frac{1}{1 + j\dfrac{f}{f_{\mathrm{c}}}} \tag{31}$$

where $\vec{H}_0$ is the magnetic field measured through a free aperture and $\vec{H}$ is the magnetic field measured through a loaded aperture by a conductive material. The cut-off frequency f_{c} being a function of the surface impedance Z_{s}, one has direct access to the value of the conductivity σ of the considered material by using the relation (21).

Let us now consider a local excitation by a near magnetic field (e.g., with a Hertz loop), and a local measurement of the resulting magnetic field, as shown in Figure 9; one can omit the infinite conductive plane and the contact resistance R_{g}.

An analytical calculus gives a new transfer function between $\vec{H}$ and $\vec{H}_0$:

$$\frac{\vec{H}(f,r)}{\vec{H}_0(f,r)} = \left(1 + \left(\frac{r}{a}\right)^2\right)^{\frac{3}{2}} \int_0^{\infty} \frac{u^2}{u + j\dfrac{f}{f_{\mathrm{c}}}} \times J_1(u)\, \mathrm{e}^{-u\frac{r}{a}}\, \mathrm{d}u \tag{32}$$

with r being the distance between the two loops, a the radius of the emission loop, and J_1 the Bessel function of first order, and the cutoff frequency f_{c}

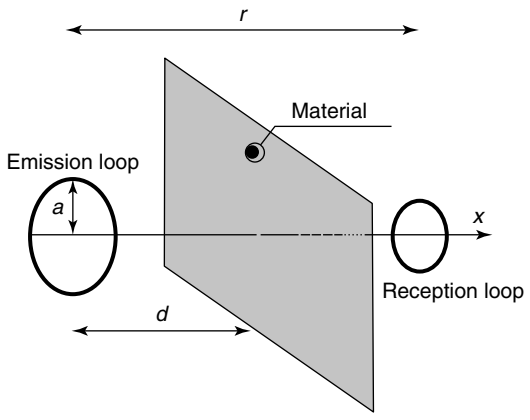


Figure 9. Excitation with near magnetic field.

having the following expression:

$$f_c = \frac{1.4}{\pi \mu_0 \sigma a e} \tag{33}$$

where a represents the radius of the emission loop and e the thickness of the material. The distance between the emission loop and the material d has no influence on the transfer function; it is possible to put the material immediately "after" the reception loop as shown in Figure 10—that is, allowing to test the material only on a single face.

However, in this case, the transfer function $\vec{H}/\vec{H}_0$ is given by the sum of the two following terms T_1 and T_2:

$$T_1 = 1 - \frac{\left(1 + \left(\frac{r}{a}\right)^2\right)^{\frac{3}{2}}}{\left(1 + \left(\frac{2d - r}{a}\right)^2\right)^{\frac{3}{2}}} \tag{34a}$$

$$T_2 = \left(1 + \left(\frac{r}{a}\right)^2\right)^{\frac{3}{2}} \int_0^{\infty} \frac{u^2}{u + j\frac{f}{f_c}} \times J_1(u) \mathrm{e}^{-u\frac{2d-r}{a}} \mathrm{d}u \tag{34b}$$

One can state that when f tends toward infinity, the function $T_1 + T_2$ tends toward the constant T_1. So, if the distance d increases, the value of the constant T_1 becomes close to the value of the transfer function before attenuation (i.e., $f < f_c$); this is the problem of the "lift-off" well known in the eddy current techniques.

The goal of this kind of analysis is not necessarily to measure the exact value of the conductivity of a material under test, but to detect damages inducing a local variation of this conductivity. A differential sensor [13, 14] measuring the "contrast" of the conductivity between two adjacent zones has been designed (Figure 11). This sensor compares the magnitude of the magnetic field between the two zones, 1 and 2 (Figure 11); when the two magnetic fields are different, the voltage V_{out} is nonzero.

The excitation frequency f is set between two characteristic frequencies: the skin frequency f_s and the Casey's frequency f_c (relation 33). The excitation frequency must be smaller than the skin frequency f_s but greater than the Casey's frequency f_c to have a significant variation of the measured magnetic field due to the variation of the local conductivity σ (relation 34). Figure 12 shows the evolution of these two frequencies as a function of the thickness of a quasi-isotropic carbon–epoxy multilayer structure.

By scanning a structure, one can build an image in which damaged areas appear clearly; an example is given in Figure 13. This figure shows the magnetic

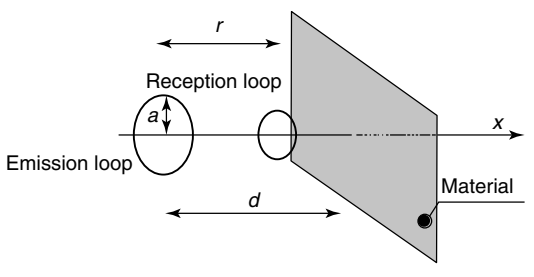


Figure 10. Excitation and measurement on a single face.

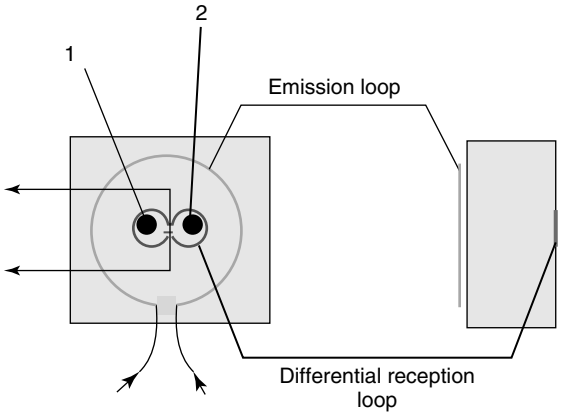


Figure 11. Differential magnetic sensor.

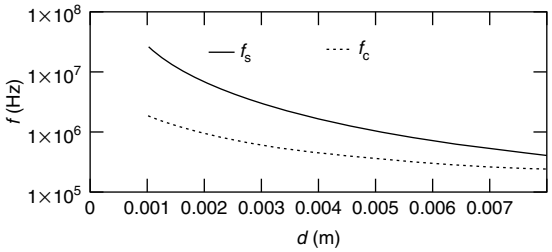


Figure 12. Evolution of f_s and f_c as a function of the thickness d.

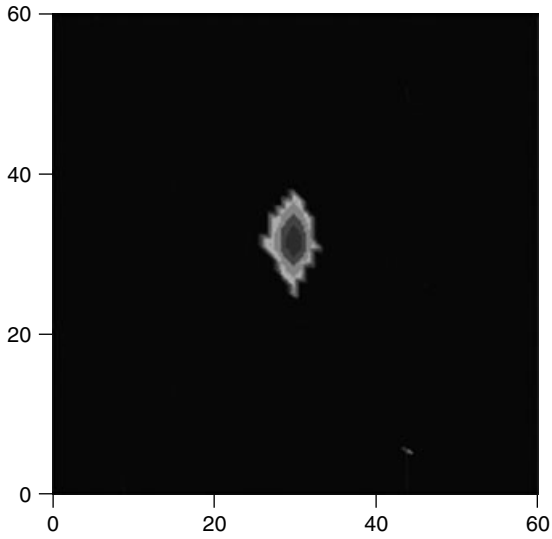


Figure 13. Magnetic image performed on a carbon–epoxy sample including a delamination.

image (i.e., σ contrast) performed on a quasi-isotropic carbon–epoxy multilayer sample of 60 mm × 60 mm ×2 mm dimension, including a delamination and fiber break. The delamination has been performed by a calibrated impact with energy 3 J.

4.2 Hybrid electromagnetic sensor

One can consider that a carbon–epoxy structure is made up of two different media: a conductive medium, the carbon fibers (conductivity $\sigma \approx 10^4\,\mathrm{S{\cdot}m^{-1}}$), and a dielectric medium, the resin (relative dielectric permittivity $\varepsilon_r \approx 4$).

The local electric field $\vec{E}_l$ induced inside a conductive structure by a magnetic induction $\vec{B}$, can be represented by the following Maxwell's equation:

$$\nabla \times \vec{E}_l = -\frac{\partial \vec{B}}{\partial t} \tag{35}$$

This electric field itself induces a current density $\vec{J} = \sigma \vec{E}_l$. However, in a carbon–epoxy media, the current density $\vec{J}$ is the sum of two terms: $\vec{J} = \vec{J}_c + \vec{J}_d$. The first term $\vec{J}_c$ (conductive current density) is due to the conductivity of the carbon fibers; the second one $\vec{J}_d$ (displacement current density) is a transient term due to the polarization phenomenon in the resin epoxy. Let us consider the quasi-static hypothesis, frequency below 10^7 Hz (i.e., below the cutoff frequency of the orientational polarization phenomenon); taking into account equation (26), the measured local electric field $\vec{E}_m$ can be written as follows:

$$\vec{E}_m = \frac{\vec{J}_c}{\sigma} + \frac{\vec{P}}{\varepsilon_0} \tag{36}$$

where $\vec{P}$ is the polarization vector due to the resin epoxy. With the relation (29) one can write

$$\vec{E}_m = \frac{\vec{J}_c}{\sigma} + \vec{E}_l\left(\varepsilon_r - 1\right) \tag{37}$$

So the measured electric field $\vec{E}_m$ is a function of the conductivity σ of the medium and also of its relative dielectric permittivity ε_r. This technique based on the magnetic induction (i.e., eddy currents) and on the analysis of the resulting electric field is called *hybrid electromagnetic method*.

The possibility to have simultaneous access to the main electric properties of the medium presents a great potential for carbon–epoxy structures used in the aeronautical domain, for detecting main damages that are found in these structures. These damages can be classified into three categories:

- The damages having a mechanical origin—they are provoked by impacts, inducing delaminations and generally, fiber breakage. These critical damages are common for carbon (fibers); so in term of electrical properties, they induce uniquely a local variation of the conductivity.
- The thermal damages arise either from proximity to a hot body or from an electrical impact (i.e., spark, lightning). In the first case, one can detect the damage uniquely by conductivity variation. In the second case, if the burn is light, there is not much σ variation, but only significant ε_r variation due to the pyrolysis phenomenon of the resin. Nevertheless, this kind of damage generally affects the two electrical parameters.
- The damages resulting from liquid ingress (water, oil, fuel) can be critical because of the fact that the liquid can start a chemical reaction and weaken the structure. This kind of damage induces uniquely a variation in ε_r due to the

presence of a new medium having a different dielectric permittivity (i.e., the liquid).

From this hybrid concept, a new sensor has been designed [15]. This sensor can be considered as a combination of the electric capacitance sensor and the magnetic sensor, including some improvements. Figure 14(a) presents a photograph of a probe based on this technique, designed with a dielectric parallelepiped (3 cm × 3 cm × 1 cm) including on one face an inductive coil and a differential dipole for the measurement of the electric field on the opposite face. Figure 14(b) shows a schematic view of this probe.

The inductive coil appears as a Moebius loop made with a hard coaxial cable (Figure 14c).

This geometry allows the formation of a double induction loop with preservation of real impedance equal to the characteristic impedance of the coaxial (i.e., 50 Ω), for the entire frequency domain, from 100 kHz to 10 MHz. The measurement unit appears as double crossed dipole (Figure 14d) that allows to perform differential measurements with a sensitivity to the two orthogonal components of the tangential electric field $\vec{E}_x$ and $\vec{E}_y$.

The operating frequency f is not submitted to the same imperative as magnetic measurement. It is only necessary to operate with a frequency below the orientational polarization phenomenon cutoff (i.e., $f < 10^7$ Hz). However, the magnitude of the electric field induced inside the material is directly proportional to the frequency of the inductive magnetic field (equation 35), and it is preferable to use a frequency as high as possible. Nevertheless, too high a frequency does not allow penetration through the total thickness of the material such as carbon–epoxy multilayer, due to the skin effect. For this kind

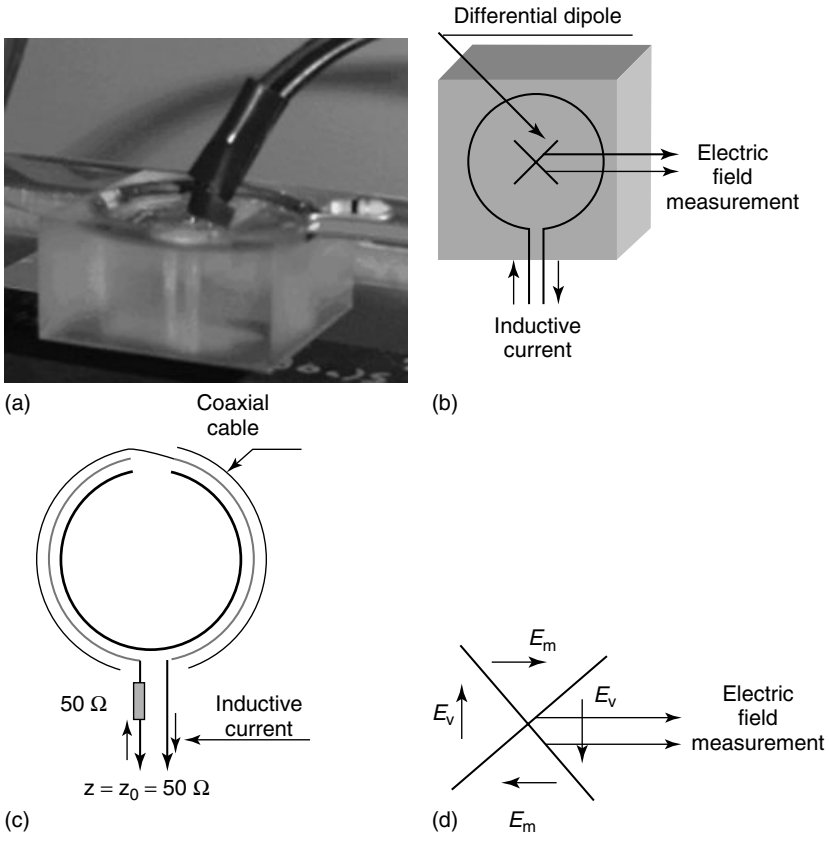
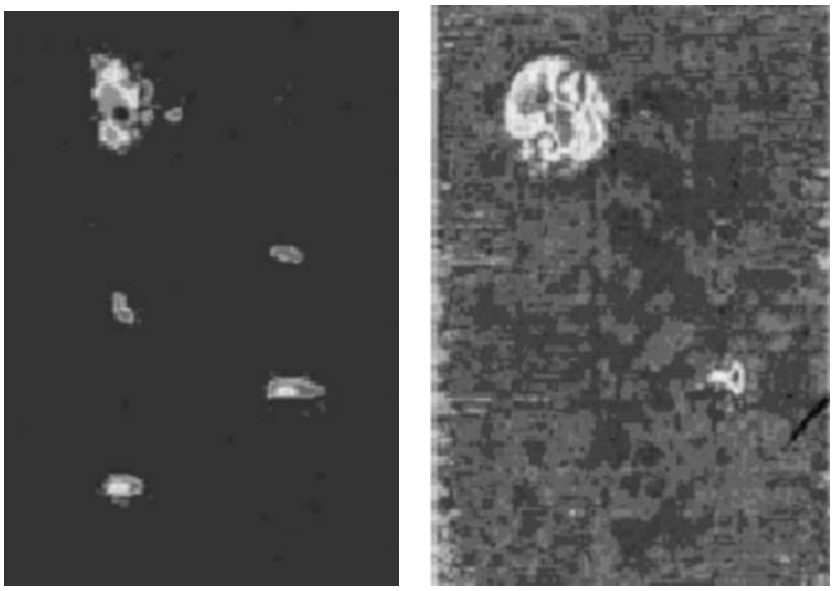


Figure 14. Hybrid electromagnetic probe: (a) photograph of the probe, (b) global view, (c) inductive coil, and (d) electric field sensor.



(a) (b)

Figure 15. Investigation of a composite sample including various burns: (a) electromagnetic image and (b) ultrasonic C-scan image.

of material, the good frequency domain is between 100 kHz and 10 MHz.

Figure 15(a) shows the electromagnetic image of a quasi-isotropic sample of carbon epoxy (200 mm × 150 mm × 4 mm) with various burns. These results are compared with a C-scan ultrasonic image shown in Figure 15(b). Only the electromagnetic hybrid investigation is able to detect light burns.

5 CONCLUSION

The examples presented are results of various methods of signal processing used to extract relevant information, particularly the methods using wavelet transform. It is a fact that the resulting signal is very weak and the signal/noise ratio is close to 1. So, the main process for data reduction consists of performing a noise reduction by discrete wavelet transform (DWT) by using Donoho's method [16]. With regard to electromagnetic techniques, one can say that it is possible to detect practically all kinds of damage and their severity. However, the sensitivity of these methods is lower than the classical ultrasonic C-scan method as regards damages having a mechanical origin, such as light delaminations without fiber breaking, because of the fact that there is no variation of electric conductivity and the very weak variation of dielectric permittivity induced by the air layer of the delamination is not significant. Conversely, these techniques show greatest sensitivity as regards to other kinds of damages (i.e., thermal damages, liquid ingress, etc.). One of the interests of these methods lies in the fact that it is very easy to transpose them in the domain of SHM and it is possible to design a fully integrated SHM system for composite structures.

RELATED ARTICLES

Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection

Fiber-optic Sensor Principles

REFERENCES

[1] Goldfine NJ, Zilberstein VA, Schlicker DE, Sheiretov Y, Walrath K, Washabaugh AP, Van Otterloo S. Surface mounted periodic field eddy currents sensors

for Structural Health Monitoring. *Proceedings of SPIE* 2001 **4335**:20–34.

[2] Madaoui N, Savin A, Premel D, Venard O, Grimberg R. An approach for quantitative nondestructive evaluation of discontinuities in flat conductive materials using eddy currents. 5th International Workshop on Electromagnetic Nondestructive Evaluation. IA, August 1999.

[3] Grimberg R, Savin A, Premel D, Mihalache O. Nondestructive evaluation of the severity of discontinuities in flat conductive materials using eddy currents transducer with orthogonal coils. *IEEE Transaction on Magnetics* 2000 **35**(1):299–331.

[4] Grimberg R, Premel D, Lemistre MB, Balageas DL, Placko D. Compared NDE of damages in graphite epoxy composites by electromagnetic methods. *Proceedings of SPIE* 2001 **4336**:65–72.

[5] Salvia M, Abry J.C. SHM using electrical resistance. In *Structural Health Monitoring*, Balageas D, Fritzen CP, Guemes A (eds). ISTE: London, 2006; Chapter 5, pp. 379–405.

[6] Abry J.C, Choi YK, Chateauminois A, Dalloz B, Giraud G, Salvia M. In situ monitoring of damage in CFRP laminates by means of AC and DC measurements. *Composite Science and Technology* 2001 **61**(6):855–864.

[7] Derobert X, Iaquinta J. Capacitive Methods for Structural Health Monitoring in Civil Engineering. In *Structural Health Monitoring*, Balageas D, Fritzen CP, Guemes A (eds). ISTE: London, 2006; Chapter 7, pp. 463–489.

[8] Iaquinta J. Contribution of capacitance probes for the inspection of external prestressing ducts. *Proceedings of the 16th World Conference on Nondestructive Testing*. Montréal, Canada, 2004.

[9] Bethe HA. Theory of diffraction by small holes. *Physical Review* 1944 **7–8**:163–175.

[10] Casey KF. Low frequency electromagnetic penetration of loaded apertures. *IEEE Transaction on Electromagnetic Compatibility* 1981 **23**(4): 367–377.

[11] Coelho R, Aladenize B. *Les Diélectriques*. Hermès: Paris, 1993.

[12] Feynman RP. *Electromagnétisme*. InterEditions: Paris, 1984; Vol. 2, 217–227.

[13] Lemistre MB. Electromagnetic structural health monitoring for composite materials. In *Structural Health Monitoring, The Demands and Challenges*, Chang FK (ed). CRC Press, 2001, 1281–1290.

[14] Lemistre MB, Gouyon R, Balageas D. Electromagnetic localization of defects in carbon epoxy materials. *Proceedings of SPIE* 1998 **3399**:89–96.

[15] Lemistre MB, Deom A. Détection de brûlures dans les composites à base de carbone. In *Nouvelles Méthodes D'instrumentation*, Lavoisier H (ed). 2004; Vol. 2, 305–312.

[16] Donoho D, Johnstone I. *Ideal Denoising in an Orthonormal Basis Chosen from a library of Bases*. C.R. French Academy of Science: Paris, 1994; Serie I.

Chapter 59
Fiber-optic Sensor Principles

Kara Peters

Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

1 INTRODUCTION

Optical fiber sensors present numerous advantages for the measurement of strain, temperature, humidity, pressure, and other parameters. For a review of their application to structural health monitoring, *see* Measures [1] and **Fiber-optic Sensors**. Optical fiber sensors are immune to electromagnetic interference, do not present an ignition hazard, are lightweight, relatively unobtrusive, and are durable and resistant to corrosion (with appropriate coatings). Additionally, a large number of sensors can be multiplexed into a single optical fiber, drastically reducing the amount of cabling required to access the sensors. This can be an important weight and fire hazard reduction. This article presents the fundamental principles common to all optical fiber sensors, as well as general issues concerning their application to structural health monitoring systems. More detailed descriptions of specific optical fiber sensors can be found in **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors** and **Fiber Bragg Grating Sensors**.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 LIGHTWAVE PROPAGATION IN OPTICAL FIBERS

Optical fibers generally consist of a fused silica (SiO_2) core and cladding. During drawing of the optical fiber, dopants are added to the silica to provide an index of refraction distribution throughout the cross section of the optical fiber [2]. Typical core diameters are $5-10\,\mu m$ for single-mode optical fibers and greater than $50-200\,\mu m$ for multimode optical fibers. The cladding diameter for almost all optical fibers is standardized at $125\,\mu m$ to permit standardized couplers. Small diameter fibers ($80\,\mu m$) have also been developed for sensing applications [3]. The

small diameter makes the fiber sensor less invasive when the fiber is embedded in a host material system and increases the sensitivity of the sensor to applied loads.

Standard optical fibers are coated with a UV-cured acrylate to prevent moisture from entering the fiber and make the fiber more durable for handling. Humidity can induce microcracking in the fiber and cause premature failure of the fiber. A standard diameter for acrylate coatings is 250 μm. For optical fiber sensors, different coatings may be applied such as polyimide, which is stiffer, to provide better protection and strain transfer and is able to withstand higher working temperatures. Removal of the polyimide coating for connectors is considerably more difficult, however, than removal of acrylate coatings. Sensing applications also often require bare optical fibers or coatings that are specially fabricated to increase sensitivity of the optical fiber to outside parameters such as humidity (*see* **Fiber Bragg Grating Sensors**) [4–7].

2.1 Step-index fibers

The cross section of a step-index optical fiber is shown in Figure 1. The index of refraction distribution is only a function of the radius with

$$n(r,\theta) = n_1 \quad r \le a$$
$$n(r,\theta) = n_2 \quad r > a \tag{1}$$

In order for propagation of guided lightwaves to occur $n_1 > n_2$.

The index of refraction difference between the core and cladding is typically less than 1%, therefore we can use the weakly guiding assumption to describe lightwave propagation through the fiber [8]. This assumption states that lightwave can be divided into transverse electric (TE) and transverse magnetic (TM) fields that are orthogonal to one another as they propagate. The two fields also have the same propagation constant, i.e., phase shift per unit distance. Therefore, we can study either the TE or TM mode individually and apply the same results to the other component. Applying the weakly guiding assumption, each component of the lightwave propagating along the fiber in the z direction satisfies the scalar wave equation,

$$\psi(r,\theta,z,t) = \bar{\psi}(r,\theta)\, e^{i(\omega t - \beta z)} \tag{2}$$

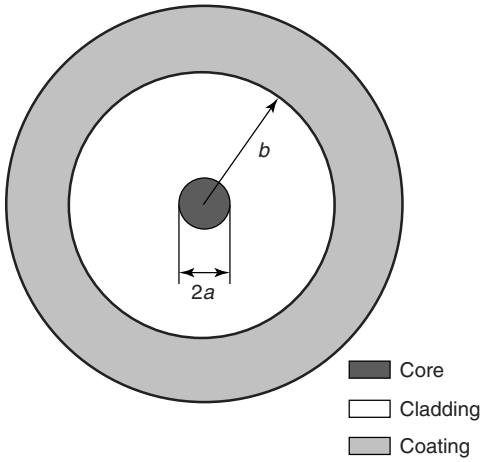


Figure 1. Cross section of step-index optical fiber. Figure not to scale.

where $\bar{\psi}$ is the energy distribution in the plane perpendicular to the propagation direction, ω is the angular frequency ($\omega = 2\pi c/\lambda$ where c is the speed of light in a vacuum and λ is the free space wavelength of the lightwave), and β is the propagation constant of the mode (phase shift per unit length). We assume an infinite diameter cladding and solve over equation (2) in each of the index of refraction regions defined by equation (1). Next, we apply continuity conditions at $r = a$ and find a number of distinct guided modes,[a] which are commonly referred to as the *linearly polarized* (LP) modes.

The energy distribution of the $LP_{\ell m}$ mode in the cross section of the fiber can be written as

$$\bar{\psi}(r,\theta) = \begin{cases} A_0[J_\ell(pr)/J_\ell(pa)]\cos(\ell\theta) & r < a \\ A_0[K_\ell(\gamma r)/K_\ell(\gamma a)]\cos(\ell\theta) & r > a \end{cases} \tag{3}$$

where $J_\ell(r)$ is the Bessel function of the first kind of order ℓ, $K_\ell(r)$ is the modified Bessel function of the first kind of order ℓ, $p = \sqrt{4\pi^2 n_1^2/\lambda^2 - \beta^2}$ and $\gamma = \sqrt{\beta^2 - 4\pi^2 n_2^2/\lambda^2}$ [8]. For each choice of ℓ, there exists zero, one, or multiple solutions to β within the guided range $n_2 < (\beta\lambda)/(2\pi) < n_1$. We order these solutions beginning with the highest value, and thus the $LP_{\ell m}$ corresponds to the mth solution. The energy distribution of equation (3) is different for each mode, due to the different values of β and therefore p and γ. The intensity distribution, $I(r,\theta) = |\bar{\psi}(r,\theta)|^2$, for several of the lowest propagating

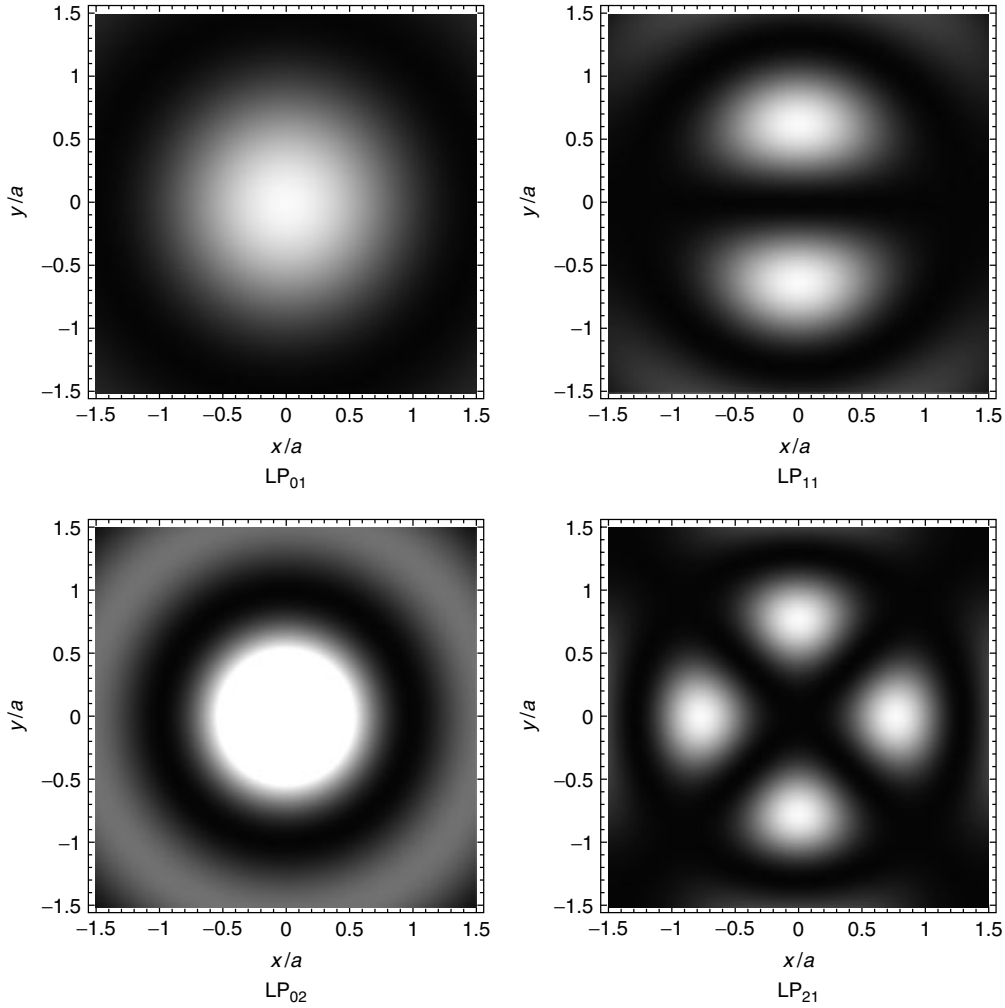


Figure 2. Intensity distribution plots for first four modes (LP_{01}, LP_{11}, LP_{02}, and LP_{21}) for step-index fiber. Radius of core is a. White area represents maximum intensity. [Reproduced from Ref. 2. © Cambridge University Press, 1998.]

modes are plotted in Figure 2. For each propagating mode, we can also define an effective index of refraction, $n_{\text{eff}} = \beta\lambda/(2\pi)$, where n_{eff} corresponds to the index of refraction of an equivalent homogeneous material for which the planar wave would propagate with the same propagation constant β as through the step-index fiber above. For guided modes, $n_2 < n_{\text{eff}} < n_1$. The effective index of refraction, or mode propagation constant β, will play a critical role in the response of many optical fiber sensors (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**; **Fiber Bragg Grating Sensors**).

The number of modes that can propagate at a particular wavelength depends upon the normalized frequency at that wavelength, V,

$$V = \frac{2\pi a}{\lambda}\sqrt{n_1^2 - n_2^2} \tag{4}$$

Figure 3 plots a normalized propagation constant b,

$$b = \frac{n_{\text{eff}}^2 - n_2^2}{n_1^2 - n_2^2} \tag{5}$$

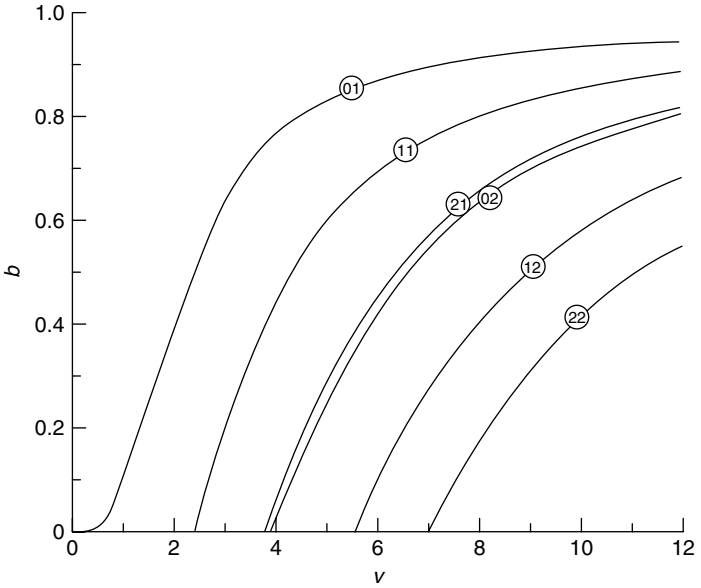


Figure 3. Propagation constant versus normalized frequency for modes LP_{01}, LP_{11}, LP_{21}, LP_{02}, LP_{12}, and LP_{22} of a step-index fiber [2].

for the first six modes of a step-index fiber as a function of the normalized frequency. The normalization allows one to apply Figure 3 to any step-index fiber parameters. As can be seen in Figure 3, each mode has a cutoff value of V such that it will not propagate at any value of V lower than the cutoff value. The cutoff value for the LP_{11} mode is at $V = 2.4048$. Below this value only one mode, the LP_{01} mode, can propagate through the fiber. Under these conditions, the fiber is referred to as a *single-mode fiber*. The LP_{01} is referred to as the *fundamental mode* as it has the highest power density in the core of the optical fiber. For many sensing applications, it is important to utilize fibers that are single mode at the wavelength to be interrogated so that the mode of propagation and coupling is well defined. Additionally, the fundamental mode is the least affected by bending losses, as the energy is concentrated in the center of the fiber. The input light is then coupled to individual modes.

For optical fibers operating at $V > 2.4048$, input lightwaves are coupled into multiple modes. The specific power distribution between the multiple modes is a function of the coupling. Multimode sensors do exist and are generally easier to practically couple to instrumentation as the fiber core is larger and the modes are more spread out throughout the fiber cross section (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**) [9–14]. As the normalized natural frequency V increases, the number of modes rapidly increases, with a good approximation of $N \simeq V^2/2$ for $V \gg 1$. Therefore, when a large number of modes are present, one can assume a uniform distribution of energy throughout the cross section [8].

Finally, as the cladding diameter is not actually infinite, additional mode solutions are possible due to the interface condition between the outer diameter of the fiber and the surrounding medium. The form of these modes depends upon the index of refraction of the surrounding medium, n_3. First, consider the case of $n_3 < n_2$ such as when the fiber has been stripped of any coating and is exposed to air. Solving equation (2), taking into account the cladding diameter yields additional discrete guided modes, called *cladding modes*, with β values below those of the guided modes. The power distributions of the cladding modes are spread over a much larger area of the cross section of the optical fiber, therefore they are far more sensitive to imperfections and microbends in the drawn silica. The cladding modes can be excited locally in long period grating sensors

and exploited for the measurement of cure monitoring, chemical or environmental sensing (*see* **Fiber Bragg Grating Sensors**) [15].

2.2 Other fiber types

Although the step-index fiber is the easiest for which to analyze mode propagation, most commercially available optical fibers have different index distributions as shown in Figure 4 [2]. Such index distributions provide better dispersion properties, important for long-haul telecommunication networks. While most optical fiber manufacturers do not provide detailed information on the index distributions, they do provide data such as the core radius, design operating wavelength, and n_{eff} at common wavelengths (see, for example, [16]). From this information, the user can calculate the response of the optical fiber sensor using the same principles as for the step-index fiber (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**).

A second important class of optical fibers applied for sensing is high-birefringence fibers, also referred to as *polarization maintaining* (PM) fibers (see examples in Figure 5) [17]. PM fibers propagate two separate, LP fundamental modes (LP_{01}) at two separate propagation constants, β_1 and β_2. The modes are polarized about orthogonal axes, referred to as the *fast and slow axes*. As the polarization axes are orthogonal, the two modes do not interfere with one another as they propagate along the optical fiber. These orthogonal modes can be used to independently measure multiple parameters (*see* **Fiber Bragg Grating Sensors**). The fiber types shown in Figure 5 can be divided into two categories. The first group shown in Figure 5(a) and (b), the elliptical core fiber and D-fiber, are PM fibers owing to the fact that the fibers are not geometrically axisymmetric. These fibers are considered mechanically homogeneous. The second group of fibers shown in Figure 5(c–e) achieve birefringence due to the lack of geometric symmetry as well as residual stresses induced by the stress-applying portion (SAP) surrounding the core. This SAP has different mechanical and thermal expansion properties than the fused silica, inducing large residual stresses on the fiber core during cooling of the fiber after it is drawn. The residual stresses induce birefringence at the core through the photoelastic effect. The magnitude of birefringence due to the residual stresses is much higher than the geometric effect, and therefore the fiber types of Figure 5(c–e) are more suitable for sensor applications as the measurands will be easier to distinguish. Owing to the presence of the SAP, these fibers are considered to be mechanically inhomogeneous.

3 LIGHTWAVE SOURCES

Lightwave sources for optical fiber sensors generally fall into three categories: light emitting diodes (LEDs), laser diodes, and superluminescent diodes

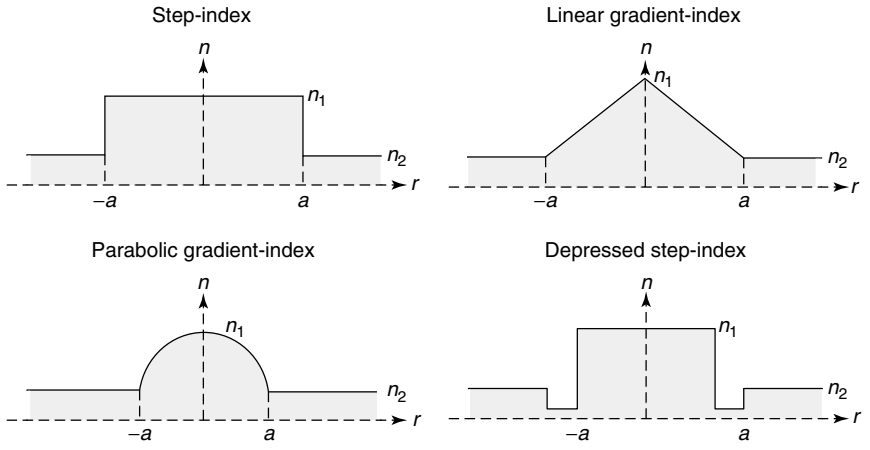


Figure 4. Common fiber index of refraction profiles.

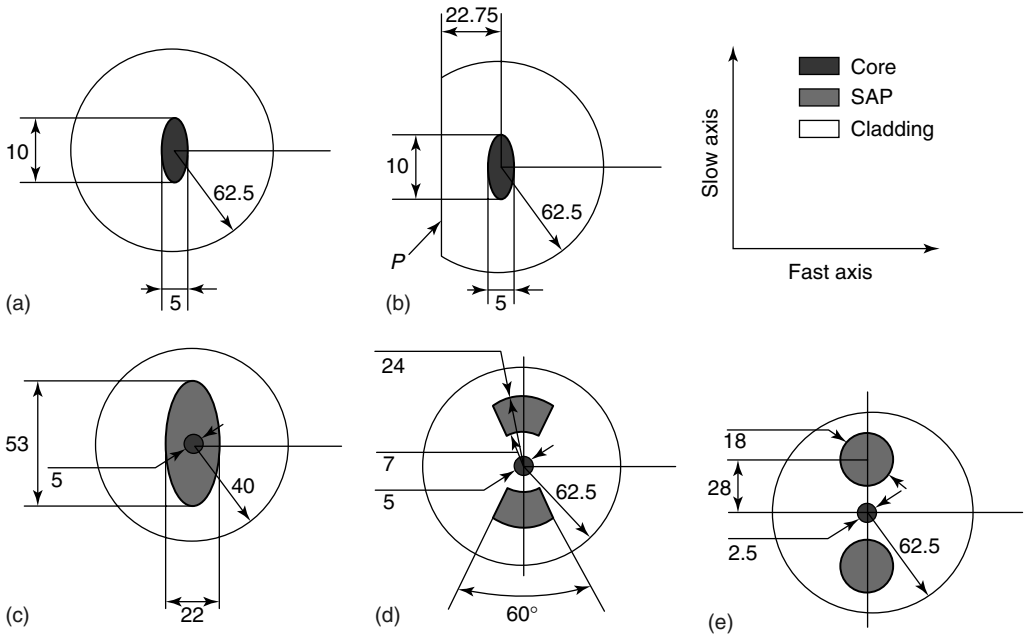


Figure 5. Geometry of common PM fiber types: (a) elliptical core fiber; (b) D-fiber; (c) elliptical core SAP fiber; (d) bow-tie fiber; and (e) panda fiber. The slow and fast axes are also indicated. All dimensions are shown in micrometers.

(SLDs). A summary of each of these is provided in this section; for further information, see [1, 2, 18]. Each of these light sources has advantages and disadvantages based on the output bandwidth, power, coherence length, and cost. The coherence length is defined as the distance over which the amplitude modulation falls off to a certain percentage of the original lightwave amplitude and can be estimated by

$$L_c = \frac{c}{dv} \tag{6}$$

where dv is the spectral frequency width of the light source. The coherence length determines the optical path length difference that can be allowed when interfering with multiple lightwaves from the same original source (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). Additional light sources exist such as gas lasers; however, it is not practical to couple them into optical fibers for field applications.

LEDs provide a broad spectral output at a relatively low cost and low sensitivity to back reflections. LEDs are most suitable for intensity-based sensors. However, their low-coherence length makes them not suitable for interferometric applications other than low-coherence interferometry. LEDs exist in two forms: surface emitting diodes and edge emitting diodes. The surface emitting diodes provide an output spectrum that is difficult to focus in a single direction. Therefore, they are more suitable for coupling to large core size multimode optical fibers than to single-mode fibers. Edge emitting diodes provide a more easily focused output for coupling into optical fibers; however, the total amount of optical power input to the optical fiber tends to be low.

SLDs operate on the principle of amplified spontaneous emission and also provide a broadband spectrum, however, at much higher output power than LEDs. The angular narrowing of the output of the SLD also allows improved coupling into single-mode optical fibers, with output powers on the order of 8 mW.

Laser diodes emit a very narrow bandwidth spectrum with a large coherence length at high output powers. Laser diodes sources are highly susceptible to back reflections; however, these can be reduced or eliminated by applying antireflection coatings to or angling the end of the coupled optical fiber or

through optical isolators. Laser diodes also require accurate temperature control to control the spectral output. One of the advantages of laser diodes is that they can be coupled with single-mode optical fibers through microlenses to produce fiber laser diodes that operate with high output powers, on the order of 10 mW. By combining the fiber laser diode with a Fabry–Perot cavity, one can produce an extremely strong, narrow bandwidth output. Additionally, the cavity length can be scanned to provide a laser with a narrow bandwidth, wavelength tunable output.

4 SENSING MECHANISMS

External sensing parameters are generally encoded into information within lightwaves propagating through an optical fiber in one of four manners, summarized in Figure 6.

4.1 Intensity modulation

The intensity of the lightwave propagating through an optical fiber can be modified through microbending of the optical fiber, a change in coupling from the fundamental mode to other nonguided modes, fracture of the optical fiber or a change in power coupled into the fiber (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). In general, the input power entering the optical fiber is held constant, while the intensity of the light transmitted through the fiber or reflected back to the input of the fiber is measured with a photodetector. Measuring the intensity of the propagating lightwave is relatively simple; however, light sources themselves often fluctuate in intensity. Feedback control loops can be applied to reduce these fluctuations, or the illuminating lightwave divided to create a reference intensity, as shown in Figure 7(a). Intensity-based sensors provide absolute measurements, meaning that they are insensitive to power interruptions between data collections. One drawback to intensity-based sensors, however, is that they cannot be multiplexed for sensor networks.

4.2 Phase modulation

The application of strain or temperature to an optical fiber changes the optical path length traversed by the lightwave propagating through the fiber, $n_{\text{eff}}L$

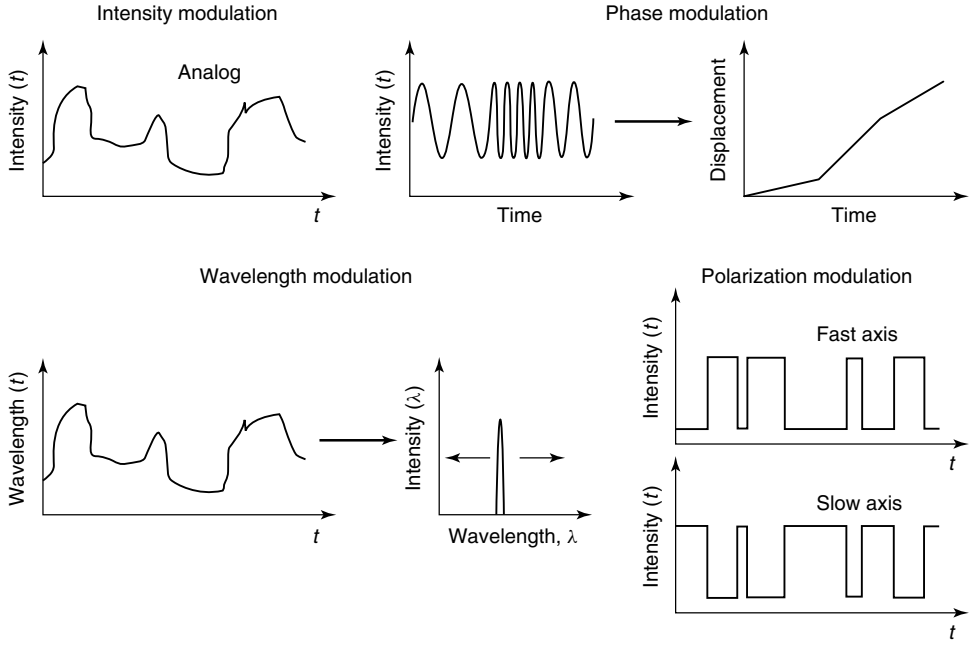


Figure 6. Output signals of optical fiber sensors using four primary sensing mechanisms.

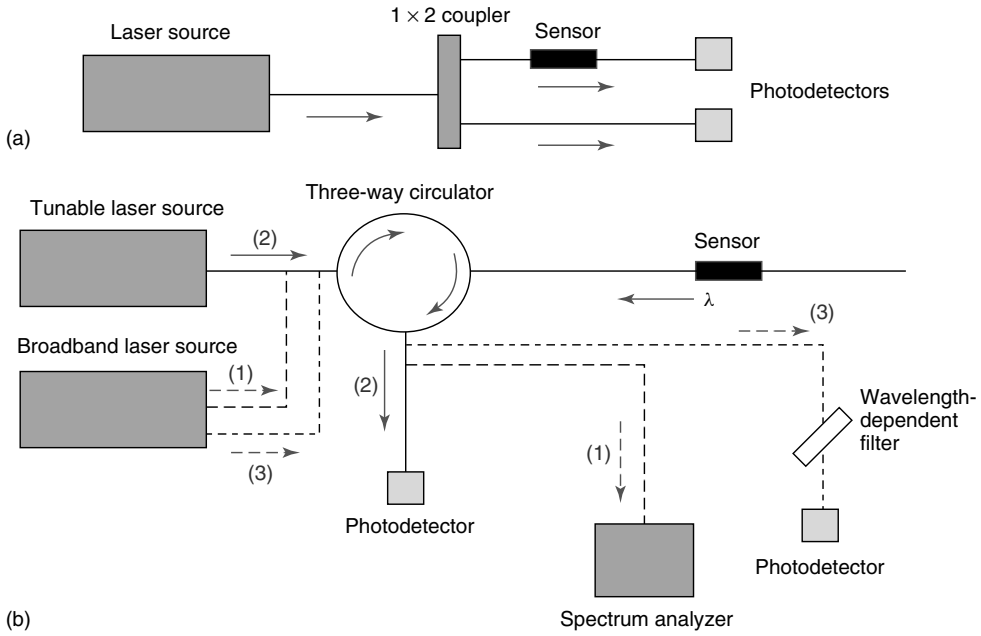


Figure 7. Sensor interrogation schemes for (a) intensity-based sensors and (b) wavelength (frequency) based sensors. For (b), three different possible schemes are shown. See text for details.

(*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). As the phase of the lightwave cannot be measured directly, the lightwave is generally recombined with a reference lightwave from the same laser source (so that the two are coherent). When the reference lightwave is not exposed to the external parameters, their relative phase shift between the signals can be related to the applied strain or temperature. The phase modulation is extremely sensitive to strain and therefore can provide very accurate measurements [1]; however, the measurement of phase shift is not absolute owing to the signal periodicity and is therefore affected by power interruptions between data collections. Additionally, fluctuations in the laser source intensity can be misinterpreted as phase shifts. Several systems are now commercially available to alleviate these problems, each with relative advantages and disadvantages (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). Interferometric measurements also cannot generally be multiplexed. The reference signal can be completely sheltered from external parameters or used for compensation of unwanted measurands such as temperature.

4.3 Spectral modulation

External sensing parameters can also be converted into spectral information of the lightwave, for example, using fiber Bragg gratings (*see* **Fiber Bragg Grating Sensors**) or Fabry–Perot interferometers (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). Such sensors typically act as filters, transmitting certain wavelengths and radiating or reflecting others. Changes in external parameters are therefore converted into a wavelength shift of the transmitted or reflected spectrum of the sensor. A typical signal is shown in the example of Figure 6, for which the reflected wavelength is changing with time. The wavelength encoded signal must then be interpreted, which is commonly performed in using one of three methods: (i) by launching a broadband of light into the fiber and applying a spectrum analyzer to select the reflected wavelength; (ii) by applying a tunable laser for which the output wavelength can be scanned while the reflected intensity is measured by a photodetector; or (iii) by launching a broadband of light and using a wavelength-dependent filter to identify the reflected wavelength (Figure 7b). Other,

more complex methods based on three-way couplers or fast Fourier transforms are also available for data acquisition at higher scan rates or multiplexing sensors (*see* **Fiber Bragg Grating Sensors**).

4.4 Polarization modulation

A final method to encode sensing information for transmission through an optical fiber is through the polarization state of the propagating lightwave [1, 19, 20]. For the example shown in Figure 6, the power of a lightwave propagating through a PM fiber is transferred between the mode polarized about the fast axis and the mode polarized about the slow axis. The power in each mode can be determined by applying a polarizing filter to the output signal. As mentioned above, the two modes propagating through a PM fiber do not normally transfer power as they propagate since they are orthogonally polarized. However, external stimuli such as pressure or twisting of the optical fiber will induce transfer between the two modes.

5 MECHANICAL PROPERTIES OF OPTICAL FIBERS

It is important to know the strength and attenuation characteristics of optical fibers to apply them for sensing applications where severe loading conditions may occur. This is particularly true for many structural health monitoring applications. Additionally, attenuation properties often determine the maximum number of sensors that can be applied in a single network or the physical distance over which the network may span.

5.1 Stiffness and strength

Standard fused silica fibers are linear elastic until failure with the following material properties: the elastic modulus, $E = 72\,\text{GPa}$, Poisson's ratio, $\nu = 0.20$, and thermal expansion coefficient, $\alpha = 5.5 \times 10^{-7}/^{\circ}\text{C}$ [21]. Fused silica optical fibers are extremely brittle and their tensile strength follows a Weibull distribution. The cumulative survival probability, P, of a fiber of length L at an applied axial stress level S is therefore

$$P(S, L) = \exp\left[-\left(\frac{S}{S_0}\right)^m \left(\frac{L}{L_0}\right)\right] \tag{7}$$

where m is the material Weibull parameter and S_0 is the inert strength of the fiber (under pristine conditions), measured for a length of fiber L_0 and a survival probability of 36.8% [22]. The medial strength, σ_{med} corresponds to a failure probability $P = 0.5$. Another important material property is the corrosion susceptibility, n, which determines the growth rate of a microcrack in the fiber,

$$\frac{\mathrm{d}w}{\mathrm{d}t} = AK_I^n \tag{8}$$

where w is the crack length and A is a constant. Typical measured values for standard silica optical fibers are $m = 112, n = 14, \sigma_{\text{med}} = 5.13\,\text{GPa}$ [23]. Kapron and Yuce [22] tested a large set of standard optical fibers in static and dynamic fatigue. The bending strength of optical fibers can be described similarly by a Weibull distribution; however, a general rule is that standard silica sensor fibers can be safely bent to a radius of 5 mm [24].

A second common material class used for optical fibers is polymers. Polymer optical fibers (POFs) have considerably lower stiffness than silica fibers, with $E = 2.4\text{–}3.0\,\text{GPa}$, $\nu = 0.34$ [25]. Their stress–strain response is also extremely sensitive to strain rate, temperature, humidity, and hysteresis effects. POFs are more flexible in bending those silica fibers; therefore, they can be handled and embedded without a protective coating.

5.2 Attenuation and dispersion

The two major factors that determine the maximum length of signal transmission in an optical fiber are the attenuation and dispersion properties of the fiber. For a fixed length of optical fiber, the attenuation, $\bar{\alpha}$, is defined as the ratio of power input, P_{i}, to the power output, P_{o}, of a given length of optical fiber and is measured in decibels [2],

$$\bar{\alpha} = 10 \log_{10} \frac{P_{\text{i}}}{P_{\text{o}}} \tag{9}$$

On the other hand, dispersion (or broadening of propagating pulses) is generally due to the wavelength-dependent properties of silica and, therefore, the wavelength-dependent propagation constants. For telecommunication applications these factors limit the length of an optical fiber that can be used; however, for sensing applications these limits are rarely an issue, except for large structural applications.

For fused silica optical fibers, three main phenomena contribute to attenuation [2, 8]:

- Rayleigh scattering occurs owing to inhomogeneities in the amorphous glasslike silica. The resulting attenuation is proportional to λ^{-4}.
- The infrared absorption properties of silica essentially prevent transmission at wavelengths above 1600 nm.
- OH impurities in the fused silica create "water" absorption peaks at wavelengths of 950, 1240, and 1390 nm.

The resulting attenuation spectrum of a silica optical fiber can be seen in Figure 8. Three primary transmission windows are therefore used in telecommunication applications: (i) around 850 nm with an approximate attenuation of $2\,\mathrm{dB\,km^{-1}}$; (ii) around 1300 nm with an approximate attenuation of $0.4\,\mathrm{dB\,km^{-1}}$ (here the material dispersion is practically zero); and (iii) around 1550 nm with an approximate attenuation of $0.25\,\mathrm{dB\,km^{-1}}$. The transmission windows at 1550 and 1300 nm are most commonly used since the attenuation is the lowest of the three windows at 1550 nm, and at 1300 nm the material dispersion is practically zero. The same three transmissions windows are typically used for optical fiber sensors owing to the relative low cost, wide selection and high quality of components including optical fibers, couplers, and laser sources available for telecommunications applications.

POFs considerably demonstrate different attenuation characteristics than fused silica. Most noticeably, the magnitude of attenuation is orders of magnitude larger than that of silica [25]. Therefore, POFs are most often used for short connector applications that require a flexible fiber. Secondly, whereas the attenuation of silica decreases with wavelength in the infrared region, the attenuation of the polymer materials rapidly increases with wavelength. Therefore,

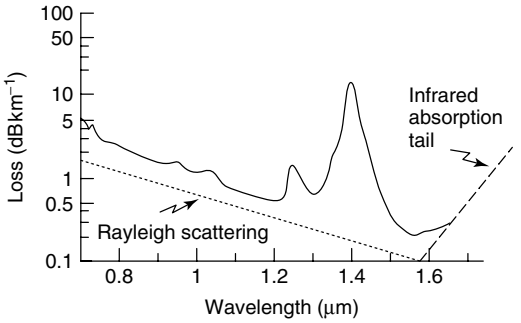


Figure 8. Material attenuation losses as a function of wavelength for fused silica (SiO_2) [2]. [Reproduced with permission from Ref. 2. © Cambridge University Press, 1998.]

POF sensors typically operate in the visible wavelength ranges rather than in the infrared ranges. Most commercially available POFs are multimode fibers in the visible wavelengths; however, recent advances in fabrication techniques have produced single-mode POFs for visible wavelength transmission [26–29] (*see* **Novel Fiber-optic Sensors**).

Other important factors contributing to signal losses in optical fiber sensors are splice losses, component losses, and bending losses. Splice losses depend upon the type of splice applied; fusion splices yield much lower losses than mechanical splices for example. Component losses appear in network components such as couplers or circulators. Finally, one important difference between optical fiber sensor networks and electrical sensor networks is that the optical fiber cannot be bent to an arbitrary radius of curvature. The bend loss in a step-index fiber per unit

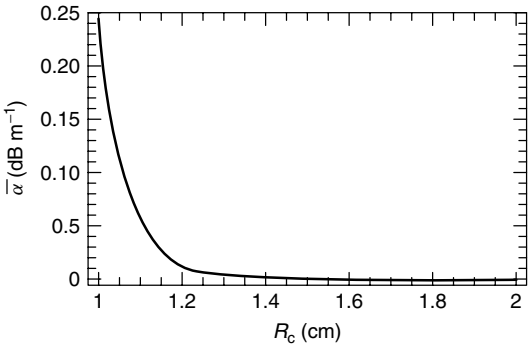


Figure 9. Bending loss as a function of bend radius of curvature ($a = 4\,\mu\mathrm{m}$, $\lambda = 1550\,\mathrm{nm}$, $n_1 = 1.460$, and $n_2 = 1.453$).

fiber length can be approximated by

$$\alpha = 4.343\left(\frac{\pi}{4aR_c}\right)^{1/2}\left[\frac{pa}{K_1(\gamma a)}\right]^2\frac{1}{(\gamma a)^{3/2}} \times \exp\left[-\frac{2(\gamma a)^3}{3k_0^2a^3n_1^2}R_c\right] \quad (10)$$

where R_c is the radius of curvature of the bent fiber and p and γ were defined for equation (3) [2]. Equation (10) is plotted for typical single-mode fiber properties in Figure 9, from which one can see that the bending loss increases exponentially for small radii of curvature.

6 INTEGRATION OF OPTICAL FIBER SENSORS IN STRUCTURAL COMPONENTS

In this section, we summarize the general issues to consider when integrating sensors into structural components, particularly embedded sensors. These include coating options, material compatibilities, strain transfer, and structural integrity. Specific issues related to composite laminates and concrete structures are included.

Various coating materials have been applied for embedded applications including the standard acrylate, which is relatively pliable and allows the fiber to slide within the coating, to reduce crimping of the fiber. However, the acrylate coating does not provide good strain transfer between the sensor and the host material owing to the low stiffness of the coating and the fact that the fiber debonds from the coating well before failure occurs in the host material [30]. As an alternative, polyimide coatings are stiffer and well bonded to the optical fiber. Polyimide coatings provide excellent strain transfer and excellent durability of the sensor during high temperature (up to 600 °C) and high-pressure fabrication conditions [21]. The main disadvantage to polyimide coatings is that removal of the coating for coupling, splicing, etc. must be done chemically. Bare optical fibers have also been embedded for good stress–strain transfer when durability concerns are not as important. Several researchers have calculated optimal coating properties to minimize the obtrusivity of embedded sensors [31, 32] or maximize the measurement capabilities of embedded sensors considering induced strains and their criticality for damage initiation [33]. For composites with other than polymer matrices, surface bonding of the optical sensor is often possible. For example, Figure 10 shows a gold coated silica optical fiber bonded to a titanium matrix composite using a Ni-based plasma spray.

One of the challenges to apply embedded optical fibers for structural health monitoring applications is to calculate the stress–strain state in the host material as a function of the stress–strain state in the optical fiber sensor. A variety of models have been developed for the stress–strain transfer to optical fiber sensors with and without coatings embedded in isotropic materials [35–38]. These models are based on shear-lag analysis, originally derived for stress transfer between reinforcing fibers in fiber reinforced composites. Li *et al.* [39] and LeBlanc [40, 41] extended the shear-lag analysis to include plastic yielding of the coating and debonding at the coating–fiber interface. For the application of optical fiber sensors embedded in orthotropic materials, Van Steenkiste and Springer [42] treated the optical fiber as an inclusion in the composite cross section and applied elasticity to calculate the relationship between the strain in the host material and strain in the optical fiber. Each of these approaches assumes that the strain field is slowly varying; therefore, the average strain at the location of the sensor is sufficient to predict

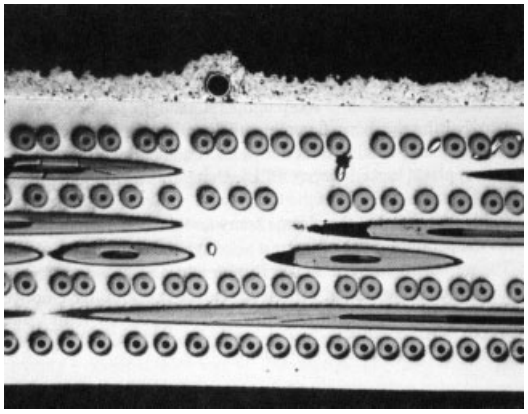


Figure 10. Gold coated silica optical fiber surface bonded to titanium matrix composite with silicon carbide reinforcing fibers using Ni-based plasma spray (image scale 32×) [34]. [Reproduced with permission from Ref. 34. © Cambridge University Press, 1998.]

its response. For more rapidly varying strain fields such as near the location of damage, Prabhugoud and Peters [43] incorporated the shear-lag theory to predict the response of a sensor located near matrix cracking in a unidirectional composite ply.

Finally, one of the major advantages of optical fiber sensors for structural health monitoring of composite structures is that large networks of sensors can be relatively unobtrusively embedded into the laminate for measurement throughout the lifetime of the structure. Owing to their high multiplexing capability (*see* **Fiber Bragg Grating Sensors**), a single optical fiber strand can contain a large number of sensors, requiring only a single ingress and egress location. It is necessary to protect the optical fiber well and eliminate strong stress concentrations that would otherwise appear at the ingress or egress points since the optical fibers are extremely fragile in bending. Commonly, this can be accomplished using polymer cabling; however, for some structural applications more advanced strategies are required [44].

6.1 Laminated composites

Optical fiber sensors are well suited for strain monitoring in polymer matrix–fiber reinforced composites because the failure strain of the optical fiber is significantly larger than that of typical reinforcing fibers such as carbon [45]. A typical carbon–fiber reinforced epoxy composite has carbon reinforcing fibers of 5–10 μm diameter and ply thicknesses

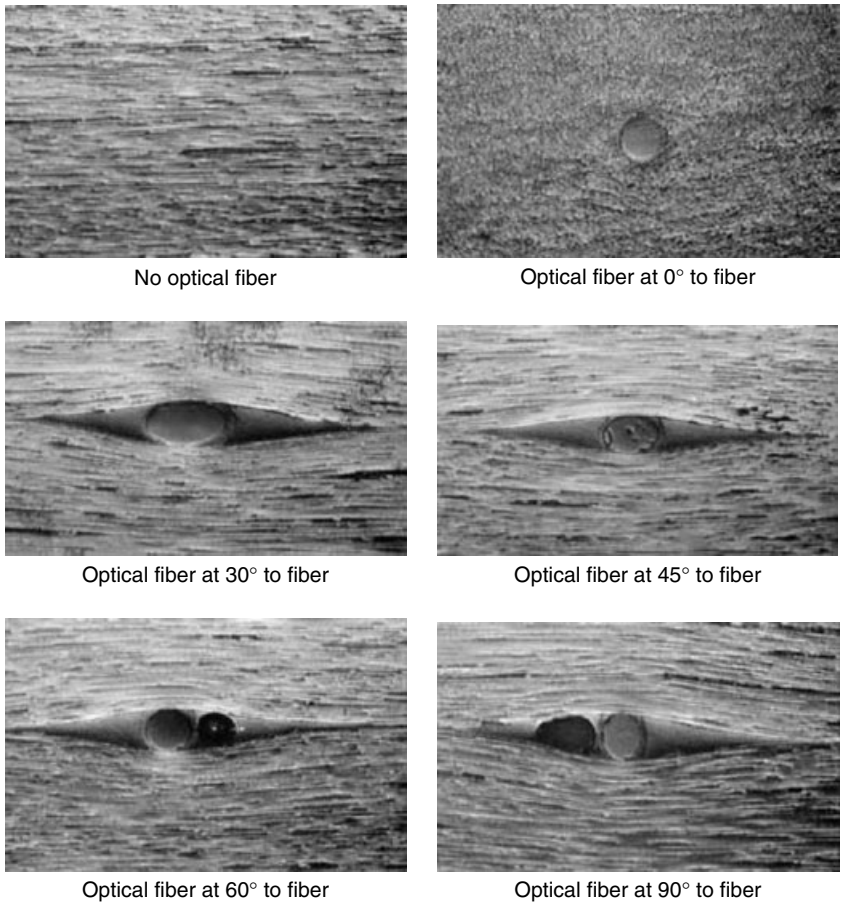


Figure 11. Micrograph of embedded optical fibers at different orientation to reinforcing fibers. Air voids next to optical fibers can be seen for orientations at 60° and 90° [47]. [Reproduced from Ref. 47. © Sage Publications, 2006.]

of 120–140 μm. Therefore, the optical fiber size (without coating) is on the order of one-ply thickness. When embedded between partially cured (prepreg) plies during the lamination fabrication, the presence of the optical fiber therefore introduces a perturbation to the local material system, typically resulting in a resin-rich zone referred to as a *resin eye* which forms around the fiber as seen in Figure 11. This resin eye enhances the local stresses surrounding the optical fiber and potentially induces local matrix cracking in adjacent plies or delamination [31]. For optical fiber sensors embedded within a single ply, the interphase region between the host epoxy and the embedded sensor can also play a significant role in the durability of the composite. Sirkis and Lu [46] used electron backscatter and acoustic emission to measure interphase thicknesses. They also determined that the residual stresses present in the optical fiber sensor after laminate fabrication can be significant.

Sirkis and Lu [46] and Jensen *et al.* [48, 49] demonstrated that optical fibers embedded parallel to the reinforcing fibers do not degrade the axial tensile strength of the laminate; however, the laminate transverse tensile strength and axial compressive strength are reduced by 20–70%. Embedding optical fibers at orientations other than parallel to the reinforcing plies does decrease the axial tensile strength of the composite [50, 51]. Kim *et al.* [51] demonstrated that there was not a significant decrease in axial tensile strength until the relative angle reached 45°, after which point the strength decreases rapidly with relative angle. They attributed this decrease in strength to the aspect ratio of the resin eye, which increased with relative angle. This was later confirmed by Shivakumar [47] (Figure 11) who also observed the presence of air voids next to the optical fiber at high relative angles. Similar observations on the laminate strengths have been made for fatigue loading conditions. The presence of the optical fiber does not noticeably reduce the fatigue life for laminates in tension–tension fatigue [52, 53], however, does create a significant decrease in fatigue life for tension–compression fatigue [53].

Optical fiber sensors embedded between multiple plies of two-dimensional woven composites demonstrate additional challenges. In addition to the resin eye formation, the presence of the optical fiber also introduces additional stresses owing to the undulations in the woven plies surrounding the optical fiber as can be seen in Figure 12. Lebid *et al.* [54] measured varying residual stresses along the optical fiber due to the undulations and observed that this must be accounted for when interpreting the sensor response. Optical fiber sensors do survive when embedded in woven composites when high temperature and pressure are applied during the laminate fabrication, although the transmission losses due to the undulations are more significant than those for nonwoven laminates [55].

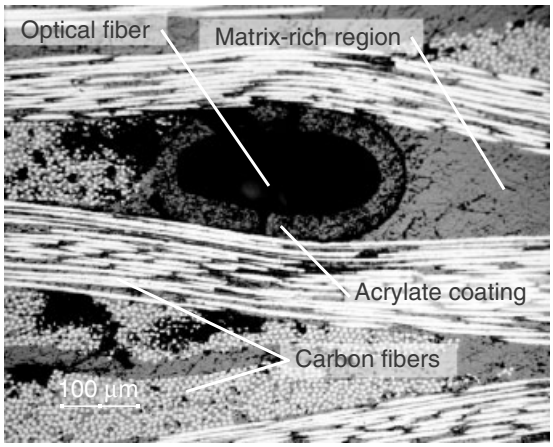


Figure 12. Optical micrograph of embedded optical fiber with acrylate coating in two-dimensional woven graphite–epoxy composite after impact loading. Optical fiber has fractured during polishing, therefore all of surface is not visible.

Finally, researchers have also demonstrated the advantages of embedding optical fiber sensors in composite laminates for impact detection and identification of the resulting internal damage. Similar to the previous loading conditions, optical fibers embedded parallel to the reinforcing fibers showed little or marginal effects on the extent and form of damage induced in the laminates [55–57]. Fibers embedded perpendicular to the reinforcing fibers increased the extent of damage, once again due to the presence of large matrix-rich regions [57].

6.2 Concrete infrastructures

Another promising application for optical fiber sensors is for long-term monitoring of civil infrastructures, specifically concrete structures. Researchers have demonstrated that optical fiber sensors can

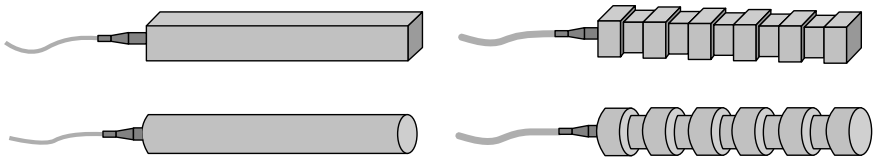


Figure 13. Regular and corrugated preembedded concrete bar sensors [67]. [Reproduced from Ref. 67. © Elsevier, 2000.]

be embedded during casting of concrete structures and provide accurate strain measurements under a variety of conditions [58, 59]. In addition to strain measurements, optical fiber sensors have been used for the measurement of crack opening displacements [60] and chloride penetration [61]. The properties of concrete are significantly different than those of polymer matrix composites; therefore several different issues become important.

For example, the gauge length of the optical fiber sensor must be chosen to be larger than the aggregate size so as to provide useful measurements [62]. Bending of the optical fiber around the individual aggregates can also induce significant signal loss over long sections of the optical fiber [63]. Another challenge to embedding optical fiber sensors in concrete can be the placement and protection of the sensor during the casting process, particularly for large-scale structures. Researchers have bonded optical fibers to rebars prior to casting to prevent movement of the optical fiber during cure of the concrete [64–66]. An alternative idea is to precast the sensor into smaller concrete blocks and then place these precast blocks into the larger structure [67]. Such blocks can be cast with a variety of surface features to improve bonding between the precast block and the concrete structure as shown in Figure 13.

Another significant challenge to embedding optical fiber sensors in concrete structures is the high alkali and moisture content of the concrete, which can degrade the coating and fiber-coating bonding over time [63, 68]. Further, once the coating has been removed, the hydroxide ions diffuse into the optical fiber surface and induce microcracking, leading to eventual failure of the optical fiber. Habel *et al.* [69] and Leung *et al.* [68] studied the long-term performance of potential optical fiber coatings including acrylate, which was rapidly degraded, and polyimide and Tefzel–silicone, which provided excellent long-term protection for the sensor. Alternatively, Kuang *et al.* [70] embedded POFs, which do not degrade as rapidly with moisture and demonstrated that they are compatible with concrete.

7 CONCLUSIONS

Fiber-optic sensors are versatile sensors for structural health monitoring applications owing to the fact that they are durable, immune to electromagnetic interference and corrosion, can be multiplexed into large sensor networks, and can be embedded in a variety of materials including composite laminates and concrete. External parameters for sensing are generally encoded in the lightwave signal through intensity, polarization states, wavelength or phase information. A variety of optical fiber types and coatings have been applied for structural health monitoring applications. The prediction of the response of optical sensors to external parameters requires an understanding of the basic principles of lightwave propagation and its sensitivity to physical changes applied to the optical fiber, which have been introduced in this chapter. For the analysis of specific optical fiber sensors such as fiber-optic interferometers, fiber Bragg gratings, and Brillouin scattering sensors *see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors** and **Fiber Bragg Grating Sensors**. For more recent technologies for fiber-optic sensors, *see* **Novel Fiber-optic Sensors**.

END NOTES

[a.] By guided we mean that the energy remains confined within the optical fiber as it propagates.

RELATED ARTICLES

Lamb Wave-based SHM for Laminated Composite Structures

Reliable Use of Fiber-optic Sensors

REFERENCES

[1] Measures RM. *Structural Monitoring with Fiber Optic Technology*. Academic Press: San Diego, CA, 2001.

[2] Ghatak A, Thyagarajan K. *Introduction to Fiber Optics*. Cambridge University Press: Cambridge, UK, 1998.

[3] Takeda N, Okabe Y. Durability analysis and structural health management of smart composite structures using small-diameter fiber optic sensors. *Science and Engineering of Composite Materials* 2005 **15**:1–12.

[4] Lee ST, Gin J, Nampoori VPN. A sensitive fibre optic pH sensor using multiple sol-gel coatings. *Journal of Optics, A* 2001 **3**:355–359.

[5] Kronenberg P, Rastogi PK, Giaccari P, Limberger HG. Relative humidity sensor with optical fiber Bragg gratings. *Optics Letters* 2002 **27**:1385–1387.

[6] Corres JM, Arregui FJ, Matias IR. Design of humidity sensors based on tapered optical fibers. *Journal of Lightwave Technology* 2006 **24**:4329–4336.

[7] Zilbermann I, Meron E, Maimon E, Soifer L, Elbaz L, Korin E, Bettelheim A. Tautomerism in N-confused porphyrins as the basis of a novel fiber-optic humidity sensor. *Journal of Porphyrins and Phthalocyanines* 2006 **10**:63–66.

[8] Buck JA. *Fundamentals of Optical Fibers*. John Wiley & Sons: Hoboken, NJ, 2004.

[9] Kosaka T, Takeda N, Ichiyama T. Detection of cracks in FRP by using embedded plastic optical fiber. *Materials Science Research International* 1999 **5**:206–209.

[10] Xie GP, Keey SL, Asundi A. Optical time-domain reflectometry for distributed sensing of the structural strain and deformation. *Optics and Lasers in Engineering* 1999 **32**:437–447.

[11] Doyle C, Martin A, Liu T, Wu M, Hayes S, Crosby PA, Powell GR, Brooks D, Fernando G. *In-situ* process and conditioning monitoring of advanced fibre-reinforced composite materials using optical fibre sensors. *Smart Materials and Structures* 1998 **7**:145–158.

[12] Jiang MZ, Gerhard E. A simple strain sensor using a thin film as a low-finesse fiber-optic Fabry-Perot interferometer. *Sensors and Actuators A* 2001 **88**:41–46.

[13] Kuang KSC, Cantwell WJ, Scully PJ. An evaluation of a novel plastic optical fibre sensor for axial strain and bend measurements. *Measurement Science and Technology* 2002 **13**:1523–1534.

[14] Han W, Wang AB. Mode power distribution effect in white-light multimode fiber extrinsic Fabry-Perot interferometric sensor systems. *Optics Letters* 2006 **31**:1202–1204.

[15] James SW, Tatam RP. Optical fibre long-period grating sensors: characteristics and application. *Measurement Science and Technology* 2003 **14**: R49–R61.

[16] Corning SMF-28: Optical Fiber Product Information, Corning Incorporated: New York, 2003.

[17] Emslie C. Polarization maintaining fibers. In *Specialty optical fibers handbook*, Méndez A, Morse TF (eds). Academic Press Amsterdam, NL, 2007.

[18] Udd E. Light sources. In *Fiber Optic Sensors*, Udd E (ed). John Wiley & Sons: Hoboken, NJ, 2006.

[19] Murukeshan VM, Chan PY, Seng OL, Asundi A. On-line health monitoring of smart composite structures using fiber polarimetric sensor. *Smart Materials and Structures* 1999 **8**:544–548.

[20] Spillman WB. Multimode polarization sensors. In *Fiber Optic Sensors*, Udd E (ed). John Wiley & Sons: Hoboken, NJ, 2006.

[21] Carmen GP, Sendeckyj GP. Review of the mechanics of embedded optical sensors. *Journal of Composites Technology & Research* 1995 **17**: 183–193.

[22] Kapron FP, Yuce HH. Theory and measurement for predicting stressed fiber lifetime. *Optical Engineering* 1991 **30**:700–708.

[23] Varelas D, Costantini DM, Limberger HG, Salathé RP. Fabrication of high-mechanical-resistance Bragg gratings in single-mode optical fibers with continuous-wave ultraviolet laser side exposure. *Optics Letters* 1998 **23**:397–399.

[24] Annovazzilodi V, Donati S, Merlo S, Zapelloni G. Statistical analysis of fiber failures under bending-stress fatigue. *Journal of Lightwave Technology* 1997 **15**:288–293.

[25] Zubia J, Arrue J. Plastic optical fibers: an introduction to their technological processes and applications. *Optical Fiber Technology* 2001 **7**:101–140.

[26] Bosc D, Toinen C. Full polymer single-mode optical fiber. *IEEE Photonics Technology Letters* 1992 **4**:749–750.

[27] Kuzyk MG, Garvey DW, Canfield BK, Vigil SR, Welker DJ, Tostenrude J, Breckon C. Characterization of single-mode polymer optical fiber and

electrooptic fiber devices. *Chemical Physics* 1999 **245**:327–340.

[28] Jiang C, Kuzyk MG, Ding JL, Jons WE, Welker D. Fabrication and mechanical behavior of dye-doped polymer optical fiber. *Journal of Applied Physics* 2002 **92**:4–12.

[29] Silva-Lopez M, Fender A, MacPherson WN, Barton JS, Jones JDC. Strain and temperature sensitivity of a single-mode polymer optical fiber. *Optics Letters* 2005 **30**:3129–3131.

[30] Uskokovi PS, Bala I, Rakin M, Puti S, Srekovi M, Aleksi R. Stress field analysis in composites laminates with embedded optical fiber. *Materials Science Forum* 2000 **352**:177–182.

[31] Barton EN, Ogin SL, Thorne AM, Reed GT. Optimisation of the coating of a fibre optical sensor embedded in a cross-ply laminate. *Composites Part A* 2002 **33**:27–34.

[32] Hadjiprocopiou M, Reed GT, Hollaway L, Thorne AM. Optimization of fibre coating properties for fiber optic smart structures. *Smart Materials and Structures* 1996 **5**:441–448.

[33] Jarlås R, Levin K. Location of embedded fiber optic sensors for minimized impact vulnerability. *Journal of Intelligent Material Systems and Structures* 1999 **10**:187–194.

[34] Henkel DP. Microstructure of high temperature smart materials. *Proceedings of Smart Structures and Materials*. SPIE, 1993; Vol. 1916, pp. 97–108.

[35] Pak YE. Longitudinal shear transfer in fiber optic sensors. *Smart Materials and Structures* 1992 **1**:57–62.

[36] Yang HT, Wang ML. Optical fiber sensor system embedded in a member subjected to relatively arbitrary loads. *Smart Material and Structures* 1995 **4**:50–58.

[37] Duck G, LeBlanc M. Arbitrary strain transfer from a host to an embedded fiber-optic sensor. *Smart Materials and Structures* 2000 **9**:492–497.

[38] Ansari F, Libo Y. Mechanics of bond and interface shear transfer in optical fiber sensors. *Journal of Engineering Mechanics* 1998 **124**:385–394.

[39] Li Q, Li G, Wang G, Ansari F, Asce M, Liu Q. Elasto-plastic bonding of embedded optical fiber sensors in concrete. *Journal of Engineering Mechanics* 2002 **128**:471–478.

[40] LeBlanc MJ. Study of interfacial interaction of an optical fibre embedded in a host material by in situ measurement of fibre end displacement—Part 1: theory. *Smart Materials and Structures* 2005 **14**:637–646.

[41] LeBlanc MJ. Study of interfacial interaction of an optical fibre embedded in a host material by in situ measurement of fibre end displacement—Part 2: experiments. *Smart Materials and Structures* 2005 **14**:647–657.

[42] Van Steenkiste RJ, Springer GS. *Strain and Temperature Measurement with Fiber Optic Sensors*. Technomic Publishing: Lancaster, PA, 1997.

[43] Prabhugoud M, Peters K. Efficient simulation of Bragg grating sensors for implementation to damage identification in composites. *Smart Materials and Structures* 2003 **12**:914–924.

[44] Kang HK, Park JW, Ryu CT, Hong CS, Kim CG. Development of fibre optic ingress/egress methods for smart composite structures. *Smart Materials and Structures* 2000 **9**:149–156.

[45] Levin K, Jarlås R. Vulnerability of embedded EFPI-sensors to low-energy impacts. *Smart Materials and Structures* 1997 **6**:369–382.

[46] Sirkis JS, Lu IP. On interphase modeling for optical-fiber sensors embedded in unidirectional composite systems. *Journal of Intelligent Material Systems and Structures* 1995 **6**:199–209.

[47] Shivakumar K, Emmanwori L. Mechanics of failure of composite laminates with an embedded fiber optic sensor. *Journal of Composite Materials* 2004 **38**:669–680.

[48] Jensen DW, Pascual J, August JA. Performance of graphite/bismaleimide laminates with embedded optical fibers. Part II: uniaxial compression. *Smart Materials and Structures* 1992 **1**:31–35.

[49] Jensen DW, Pascual J, August JA. Performance of graphite/bismaleimide laminates with embedded optical fibers. Part I: uniaxial tension. *Smart Materials and Structures* 1992 **1**:24–30.

[50] Kim MS, Lee CS, Hwang W. Effect of the angle between optical fiber and adjacent layer on the mechanical behavior of carbon/epoxy laminates with embedded fiber-optic sensor. *Journal of Materials Science Letters* 2000 **19**:1673–1675.

[51] Lau KT, Chan CC, Zhou LM, Jin W. Strain monitoring in composite-strengthened concrete structures using optical fibre sensors. *Composites, Part B* 2001 **32**:33–45.

[52] Lee DG, Mitrovic M, Friedman A, Carman GP, Richards L. Characterization of fiber optic sensors for structural health monitoring. *Journal of Composite Materials* 2002 **36**:1349–1366.

[53] Badcock RA, Fernando GF. An intensity-based optical fibre sensor for fatigue damage detection in advanced fibre-reinforced composites. *Smart Materials and Structures* 1995 **4**:223–230.

[54] Lebid S, Habel W, Daum W. How to reliably measure composite-embedded fibre Bragg grating sensors influenced by transverse and point-wise deformations? *Measurement Science and Technology* 2004 **15**:1441–1447.

[55] Pearson JD, Zikry MA, Prabhugoud M, Peters K. Global-local assessment of low-velocity impact damage in woven composites. *Journal of Composite Materials* 2007 **41**:2759–2783.

[56] Sirkis JS, Chang CC, Smith BT. Low-velocity impact of optical-fiber embedded laminated graphite-epoxy panels. Part 1: macroscale. *Journal of Composite Materials* 1994 **28**:1347–1370.

[57] Jeon BS, Lee JJ, Kim JK, Huh JS. Low velocity impact and delamination buckling behavior of composite laminates with embedded optical fibers. *Smart Materials and Structures* 1999 **8**:41–48.

[58] Masri SF, Agbabian MS, Abdelghaffar AM, Higazy M, Claus RO, Devries MJ. Experimental study of embedded fiberoptic strain-gauges in concrete structures. *Journal of Engineering Mechanics* 1994 **120**:1696–1717.

[59] Bin LinY, Chern JC, Chang KC, Chan YW, Wang LA. The utilization of fiber Bragg grating sensors to monitor high performance concrete at elevated temperature. *Smart Materials and Structures* 2004 **13**:784–790.

[60] Yuan LB, Ansari F. Embedded white light interferometer fibre optic strain sensor for monitoring crack-tip opening in concrete beams. *Measurement Science and Technology* 1998 **9**:261–266.

[61] Fuhr PL, Huston DR, MacCraith B. Embedded fiber optic sensors for bridge deck chloride penetration measurement. *Optical Engineering* 1998 **37**: 1221–1228.

[62] Bonfiglioli B, Pascale G. Internal strain measurements in concrete elements by fiber optic sensors. *Journal of Materials in Civil Engineering* 2003 **15**:125–133.

[63] Zeng XD, Bao XY, Chhoa CY. Strain measurement in a concrete beam by use of the Brillouin-scattering-based distributed fiber sensor with single-mode fibers embedded in glass fiber reinforced polymer rods and bonded to steel reinforcing bars. *Applied Optics* 2002 **41**:5105–5114.

[64] Spammer SJ, Fuhr PL, Nelson M, Huston D. Rebar-epoxied optical fiber Bragg gratings for civil structures. *Microwave and Optical Technology Letters* 1998 **18**:214–219.

[65] Davis MA, Bellemore DG, Kersey AD. Distributed fiber Bragg grating strain sensing in reinforced concrete structural components. *Cement and Concrete Composites* 1997 **19**:45–57.

[66] Maaskant R, Alavie T, Measures RM, Tadros G, Rizkalla SH, GuhaThakurta A. Fiber-optic Bragg grating sensors for bridge monitoring. *Cement and Concrete Composites* 1997 **19**:21–33.

[67] Yuan LB, Jin W, Zhou LM, Lau KT. The temperature characteristic of fiber-optic pre-embedded concrete bar sensor. *Sensors and Actuators A* 2001 **93**:206–213.

[68] Leung CKY, Darmawangsa D. Interfacial changes of optical fibers in the cementitious environment. *Journal of Materials Science* 2000 **35**:6197–6208.

[69] Habel WR, Hofmann D, Hillemeier B. Deformation measurements of mortars at early ages and of large concrete components on site by means of embedded fiber-optic microstrain sensors. *Cement and Concrete Composites* 1997 **19**:81–102.

[70] Kuang KSC, Akmaluddin CWJ, Thomas C. Crack detection and vertical deflection monitoring in concrete beams, using plastic optical fibre sensors. *Measurement Science and Technology* 2003 **14**: 205–216.

Chapter 61
Fiber Bragg Grating Sensors

Kara Peters
Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

1 INTRODUCTION

Fiber Bragg grating (FBG) sensors have many advantages for strain sensing in structural health monitoring applications including their small size, the potential to multiplex hundreds of sensors with a single ingress/egress fiber, their immunity to electromagnetic interference, and their corrosion resistance [1]. FBG sensors have been applied for a variety of structural health monitoring applications including spacecraft [1, 2], bonded aircraft repairs [3–5], cryogenic composite tanks [6, 7], highway and railway bridges [8, 9], fiber reinforced polymer (FRP) composite bridges [10], FRP strengthened concrete [11], offshore platforms [12], and nuclear reactors [13]. Researchers have also applied FBG sensors for modal-based damage identification [14, 15], the identification of multiple failure modes in composite laminates [16–18], and monitoring of welded and composite joints [19, 20].

The FBG sensor, shown in Figure 1, is a permanent periodical perturbation in the index of refraction of the optical fiber core. This index modulation can be written mathematically as

$$n_{\text{eff}}(z) = n_{\text{eff}} + \overline{\delta n_{\text{eff}}} \left\{ 1 + \nu \cos \left[\frac{2\pi}{\Lambda} + \phi(z) \right] \right\} \quad (1)$$

where ν is the fringe visibility, Λ the grating period, $\phi(z)$ the grating chirp function (which describes any variation in the grating period), n_{eff} the effective index of refraction of the fiber for the fundamental mode, and $\overline{\delta n_{\text{eff}}}$ the "dc" average index change [21]. Hill *et al.* [22] first fabricated permanent Bragg gratings in an optical fiber as a filter for telecommunication applications.

The FBG sensor acts as a "wavelength-dependent filter" as shown in Figure 1. When a broad spectrum of wavelengths is passed through the FBG, a narrow bandwidth of wavelengths is reflected, while all

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

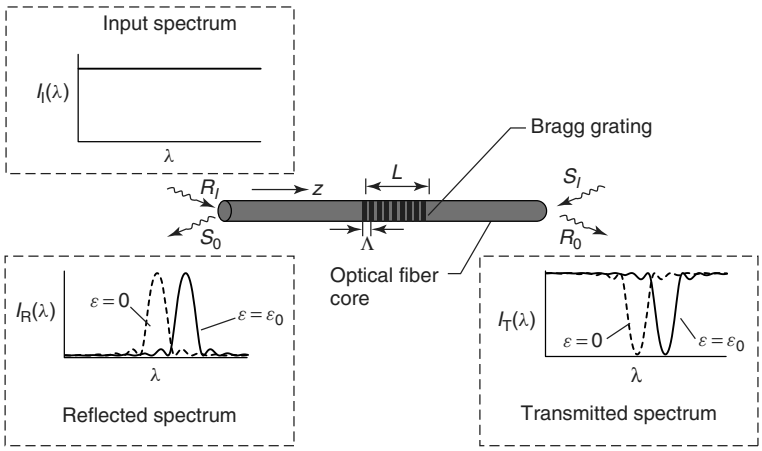


Figure 1. Optical fiber Bragg grating sensor. Reflected and transmitted spectra are shown for a broadband input spectrum. Dashed line is for unstrained FBG and solid line is for strained FBG. [Reproduced with permission. © Springer, 2002.]

others are transmitted. The wavelength at maximum reflectivity is referred to as the Bragg wavelength, λ_B, and is determined by the condition

$$\lambda_B = 2n_{\text{eff}}\Lambda \tag{2}$$

As axial strain, ε, is applied to the FBG, the Bragg wavelength shifts to lower wavelengths (compression) or higher wavelengths (tension). The applied strain is thus encoded in the FBG Bragg wavelength shift. If the applied axial strain is not constant along the length of the grating, the spectrum will distort as well as shift. Section 6 describes how this distortion can be interpreted to measure the applied strain distribution along the FBG.

2 FIBER BRAGG GRATING PARAMETERS

The phenomenon of wavelength selective reflection is due to coupling between various modes propagating through the FBG. For short period FBGs ($\Lambda < 1\,\mu$m), coupling occurs between a forward propagating fundamental (LP_{01}) mode and a backward propagating LP_{01} mode (*see* **Fiber-optic Sensor Principles**). The coupling condition can be described through the coupled mode equations

$$\frac{dR}{dz} = i\hat{\sigma}R(z) + i\kappa S(z)$$

$$\frac{dS}{dz} = -i\hat{\sigma}S(z) - i\kappa R(z) \tag{3}$$

where $R(z)$ and $S(z)$ are the amplitudes of the forward and backward propagating lightwaves, respectively, and z is the coordinate along the axis of the fiber [21]. The coefficients $\hat{\sigma}$ and κ are defined as

$$\hat{\sigma} = \frac{2\pi}{\lambda}\left(n_{\text{eff}} + \overline{\delta n_{\text{eff}}}\right) - \frac{\pi}{\Lambda} - \frac{1}{2}\frac{d\phi}{dz}$$
$$\kappa = \frac{\pi}{\lambda}\nu\overline{\delta n_{\text{eff}}} \tag{4}$$

where λ is the wavelength of the propagating lightwaves.

For a grating with a constant period (i.e., nonchirped grating), we have a constant grating chirp function, $\phi(z) = \phi_0$. Therefore, the parameters $\hat{\sigma}$ and κ are constants and equation (3) can be solved analytically. We find the solution to equation (3) for the reflectivity coefficient,

$$r = \left|\frac{S(-L/2)}{R(-L/2)}\right|^2 = \frac{\sinh^2\left(L\sqrt{\kappa^2 - \hat{\sigma}^2}\right)}{\cosh^2\left(L\sqrt{\kappa^2 - \hat{\sigma}^2}\right) - \dfrac{\hat{\sigma}^2}{\kappa^2}} \tag{5}$$

This solution is plotted for different values of κL in Figure 2. One can see in Figure 2 that for the case of strong coupling ($\kappa L = 6$) the grating is supersaturated. For sensing applications, FBGs with a

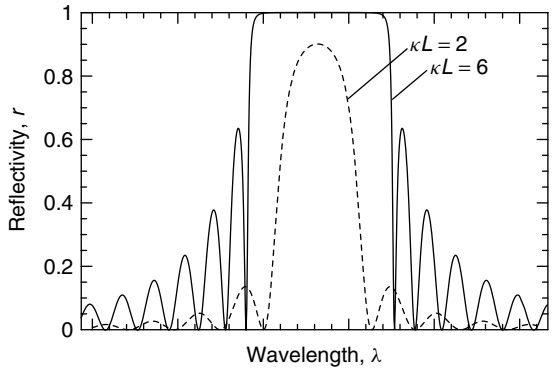


Figure 2. Reflectivity as a function of wavelength for FBG with coupling coefficients $\kappa L = 2$ and $\kappa L = 6$ ($n_{\text{eff}} = 1.46$, $\overline{\delta n_{\text{eff}}} = 3 \times 10^{-6}$, $\nu = 1$, $\lambda_{\text{B}} = 1550\,\text{nm}$).

narrow bandwidth and high reflectivity (i.e., $\kappa L = 2$ in Figure 2) produce the largest signal-to-noise ratio. The maximum reflectivity of the grating, $r_{\max}$, and bandwidth between the first zero crossings, $\Delta\lambda_0$, are found from equation (5),

$$r_{\max} = \tanh^2(\kappa L) \tag{6}$$

$$\frac{\Delta\lambda_0}{\lambda_{\text{B}}} = \sqrt{\left(\frac{\nu\overline{\delta n_{\text{eff}}}}{n_{\text{eff}}}\right)^2 + \left(\frac{2\Lambda}{L}\right)^2} \tag{7}$$

Procedures such apodization are often applied to the FBG sensor during fabrication to reduce the secondary peaks and narrow the bandwidth [23].

As strain is applied to the FBG sensor, the Bragg wavelength shifts according to

$$\Delta\lambda_{\text{B}} = 2\left(\Delta n_{\text{eff}}\Lambda_0 + n_{\text{eff}}\Delta\Lambda_0\right) \tag{8}$$

where Δn_{eff} is due to the strain–optic effect and $\Delta\Lambda_0$ is due to the change in period of the grating. For the specific case of pure axial loading we find (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**),

$$\Delta\lambda_{\text{B}} = \lambda_{\text{B}}\left[1 - \frac{n_{\text{eff}}^2}{2}(p_{12} - \nu p_{12} - \nu p_{11})\right]\varepsilon \tag{9}$$

in terms of the Pockel's constants of silica, p_{11} and p_{12}. Defining the effective photoelastic constant for axial strain, $p_{\text{e}} = n_{\text{eff}}^2(p_{12} - \nu p_{11} - \nu p_{12})/2$ we can write,

$$\frac{\Delta\lambda_{\text{B}}}{\lambda_{\text{B}}} = (1 - p_{\text{e}})\varepsilon \tag{10}$$

Thus, the shift in Bragg wavelength is linearly related to the applied axial strain. Typical values for silica optical fibers are $p_{11} = 0.12$, $p_{12} = 0.27$, and $p_{\text{e}} = 0.22$ [24].

Similar to electrical resistance strain gauges, FBG sensors are also sensitive to temperature through

$$\frac{\Delta\lambda_{\text{B}}}{\lambda_{\text{B}}} = 2\left(\frac{\partial n_{\text{eff}}}{\partial T}\Lambda_0 + \frac{\partial\Lambda}{\partial T}n_{\text{eff}}\right)\Delta T = (\alpha + \zeta)\Delta T \tag{11}$$

where α is the thermal expansion coefficient and ζ is the thermo-optic coefficient of the optical fiber material. A typical value for fused silica is $(\alpha + \zeta) = 6.67 \times 10^{-6}/^{\circ}\text{C}$ [24]. It is important to note that the thermal sensitivity of an FBG sensor is considerably higher than that of its electrical strain gauge counterpart, increasing the need for thermal compensation for strain measurement applications.

One method of temperature compensation is to write two FBGs at different Bragg wavelengths, λ_{B1} and λ_{B2}, and to measure the Bragg wavelength shifts in each grating due to the applied strain and temperature. The strain and temperature can thus be independently determined from the equation

$$\begin{pmatrix}\Delta\lambda_1\\ \Delta\lambda_2\end{pmatrix} = \begin{pmatrix}K_{\varepsilon 1} & K_{T1}\\ K_{\varepsilon 2} & K_{T2}\end{pmatrix}\begin{pmatrix}\varepsilon\\ \Delta T\end{pmatrix} \tag{12}$$

where the K terms are the sensitivities of each grating to strain and temperature. Xu *et al.* [25] fabricated a two FBG sensor with $\lambda_{\text{B1}} = 850\,\text{nm}$ and $\lambda_{\text{B2}} = 1300\,\text{nm}$. Experimental measurements demonstrated a 6.5% difference in $K\varepsilon$, and a 9.8% difference in K_T for the two FBGs. One difficulty with this approach is that λ_{B1} and λ_{B2} must be sufficiently far apart to obtain a significant difference in the FBG sensitivities and therefore an accurate calculation of the strain and temperature when inverting equation (12). This often requires two separate laser sources for the sensor interrogation.

3 FABRICATION AND STRENGTH OF FIBER BRAGG GRATING SENSORS

Several methods are commonly used for the fabrication of FBGs, each based on the exposure of the optical fiber core to ultraviolet (UV) laser light [26]. The UV exposure increases the index of refraction of the silica locally through the process of photosensitivity. Prior to exposure, the optical fiber is typically doped to increase the photosensitivity of the silica, e.g., through high-pressure hydrogen loading of the fiber. Molecular hydrogen diffuses in the central core region due to its small size. This increase in photosensitivity is not a permanent effect since the hydrogen diffuses out over time. However, if the fiber is exposed to UV radiation during this period, the hydrogen molecules react in the silica Si–O–Ge sites forming OH species and UV bleachable germanium oxygen deficiency centers, which are responsible for the enhanced photosensitivity [23].

Several fabrication techniques are commonly applied for FBG sensors. For excellent reviews see [23] and [27]. The interferometric technique is based on a bulk interferometer that splits the incoming UV light into two beams and then recombines them to form an interference pattern that is focused onto the optical fiber core (Figure 3). The interference pattern exposure induces a refractive index modulation in the fiber core. The periodicity of the refractive index modulation is related to the UV source wavelength, λ_{UV}, by

$$\Lambda = \frac{\lambda_{UV}}{2\sin(\theta/2)} \tag{13}$$

By changing the intersecting angle θ between the two writing beams, it is possible to write Bragg gratings for almost any wavelength. However, the stability of the interference pattern is extremely difficult to maintain due to mechanical vibrations that may be present in the bulk interferometer.

Because of its simplicity, the phase mask technique is the most widely used method of FBG fabrication (see Figure 4). The phase masks themselves may be formed holographically or by electron-beam lithography from fused silica. A near-field fringe pattern is produced by the interference of the plus and minus first-order diffracted beams. The period of the fringes produced is one-half of the phase mask period, Λ_{pm}.

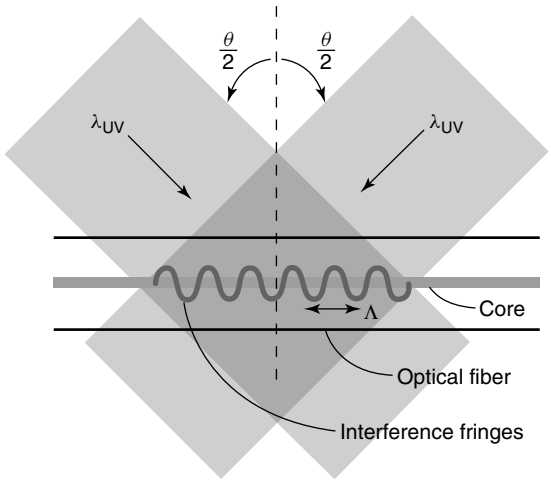


Figure 3. Bulk interferometer arrangement for the fabrication of fiber Bragg gratings.

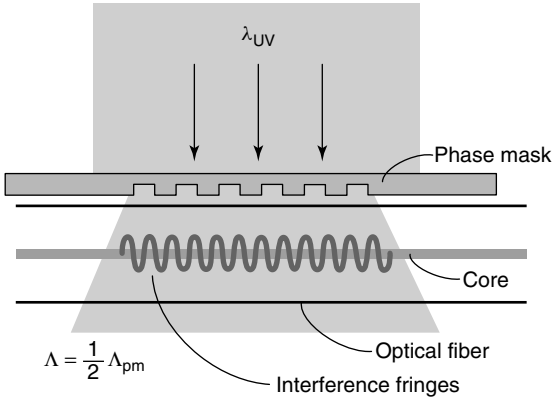


Figure 4. Phase mask arrangement for the fabrication of fiber Bragg gratings.

The simplicity of the phase mask writing technique provides a robust and inherently stable method for producing FBGs. Since the fiber is usually placed close to the mask, the sensitivity to mechanical vibrations is minimized. The main disadvantage of the phase mask technique, however, is the need to produce a phase mask for each desired wavelength. This limitation can be partially overcome by applying tension to the phase mask to vary the period and provide a range of possible Bragg wavelengths (typically $\pm 5\%$ of the original Bragg wavelength).

Stripping the coating from an optical fiber prior to writing the FBG is necessary as the coating will reduce the amount of UV light arriving at the optical

Table 1. Summary of FBG fabrication methods and resulting strengths

Reference	Fiber Bragg grating fabrication method	σ_{med} (Gpa)	m
Wei *et al*. [28]	Pristine fiber	4.95	66.6
	Fiber after chemical stripping	3.38	3.1
	Pulsed laser exposure + chemical stripping	2.52	3.2
Varelas *et al*. [29]	Pristine fiber + chemical stripping	5.13	112
	Continuous wave UV exposure + chemical stripping	> 5	57
Askins *et al*. [30]	Pristine reference fiber (Corning SMF-28)	5.6 – 6.0	90
	Single pulse exposure, writing during draw tower process prior to coating	5.4	45
Gu *et al*. [31]	One-step process of writing through polysiloxane buffer, no stripping required	3.76	13.2
Han *et al*. [32]	LPG fabrication via residual stress relaxation with CO_2 laser (no UV irradiation required)	4.9	10.5

fiber core. However, mechanical stripping of the coating introduces microcracks at the surface of the optical fiber and can significantly reduce the strength of the fiber at the location of the FBG. Researchers have developed several techniques to increase the strength of FBGs, including chemically stripping the coating from the optical fiber and writing the grating in the optical fiber during the drawing of the fiber prior to coating. A summary of the resulting FBG strength properties taken from the literature is listed in Table 1. For definitions of the Weibull modulus and median strength, *see* **Fiber-optic Sensor Principles**. Additionally, description of common coating types for FBG sensors for durability in structural health monitoring applications are provided in **Fiber-optic Sensor Principles**.

4 MULTIPLEXING AND INTERROGATION OF FBG SENSOR NETWORKS

One of the strongest advantages of FBG sensors over other available strain or temperature sensors for structural health monitoring applications (*see* **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**; **Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection**; **Directed Energy Sensors/Actuators**) is the fact that they can be multiplexed into a large sensor network. For structural health monitoring applications, having only a single or limited ingress/egress points and lead cables is a notable advantage because it can significantly reduce the weight of the lead cables and disruptions to the structure itself [1].

Several approaches to multiplexing large FBG sensor networks have been applied in which the sensors are interrogated using wavelength division multiplexing (WDM), time division multiplexing (TDM), and combinations thereof (*see* **Fiber-optic Sensor Principles**). For single FBG sensors, the most common method to measure the peak shifts is to pass the reflected signal from the FBG through a wavelength-dependent filter that outputs a different intensity based on the input wavelength. Examples of wavelength-dependent filters include Fabry–Perot cavities [33], acousto-optic filters [34], chirped Bragg gratings [35, 36], long period gratings (LPGs) [37], and WDM couplers [38, 39]. These techniques have been expanded to WDM FBG sensor networks by scanning the filters to cover the peak wavelengths of multiple sensors. For example, Davis *et al.* [40] demonstrated interrogation of 60 FBGs in 2.5 s with 50 averages per sensor.

The interrogation of a large number of FBGs along a single optical fiber is limited by the power reflected from each grating, since each FBG reflects a portion of the wavelengths in the vicinity of the Bragg wavelength. For WDM applications, reducing the spacing between each FBG reduces the time required to scan the network; however, it also increases the cross talk losses between FBGs. For TDM applications, low reflectivity gratings ($r_{max} \cong 1\%$) should be used to allow for a large number of multiplexed sensors. By combining WDM with TDM, as shown in Figure 5, one can reduce the power loss per FBG,

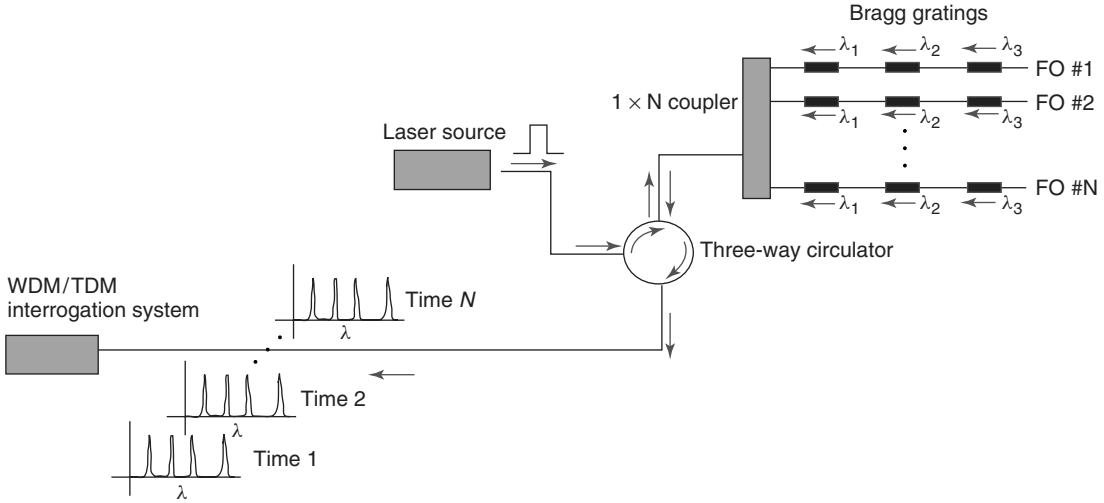


Figure 5. Combined wavelength division multiplexing and time domain multiplexing for FBG sensor network.

as well reduce the total wavelength range that must be scanned to interrogate the entire network [24]. In this case, each parallel network of FBGs has the same wavelength range, while the laser output is pulsed and the particular network being interrogated determined by the time of arrival of the signal.

Alternative approaches to interrogating FBG sensor networks include expanding the output-reflected spectrum via a plane grating or prism to a linear charge-coupled device (CCD) array that would then operate as a CCD spectrometer. In this manner, Askins *et al.* [41] achieved interrogation of 20 sensors at 3.5 kHz. A network of FBG sensors that have the same initial Bragg wavelength can also be interrogated quasi-statically through interpretation of the inverse Fourier transformation of the reflected spectrum [42]. Finally, interferometric methods have also been applied for high-speed interrogation of FBG sensor networks [43–45].

5 MULTIPLE AXIS SENSING

One unique characteristic of the FBG as a strain sensor is its ability to measure and distinguish between multiple strain components. Researchers have fabricated FBG rosettes, following the same strategy as for electrical resistance strain gauges [46, 47]. However, the FBG itself can also be used to monitor axial and transverse strain components through the birefringence that occurs in the optical fiber due to applied transverse loads. Figure 6 shows an example of three independent loads applied to the optical fiber, along with the resulting three principal strain components at the center of the fiber core. The effect of the axial load (P_1) is to shift the reflected peak to higher or lower wavelengths; however, the effect of the transverse loads (P_2 and P_3) is to create peak splitting due to the fast and slow axes that develop in the optical fiber (for a discussion of birefringence, *see* **Fiber-optic Sensor Principles**). A typical example of peak-splitting behavior is also shown in Figure 6.

In general, FBGs written in polarization maintaining (PM) optical fibers exhibit enhanced discrimination between multiple loading components due to geometrical and stress-induced birefringence in the fibers (*see* **Fiber-optic Sensor Principles**). The stress-induced portion of the birefringence can be due to residual stresses and/or the applied transverse loading components. This birefringence effect has been exploited for the independent measurement of multiple strain components and/or temperature changes using single or multiple Bragg gratings [48]. This capability makes the FBG sensor ideal for the monitoring of residual stresses during curing of FRP laminated composites as well as further monitoring of internal stresses during loading of the laminates [17, 49–52]. Additionally, by writing two FBGs with different Bragg wavelengths, λ_{B1} and λ_{B2}, at the same

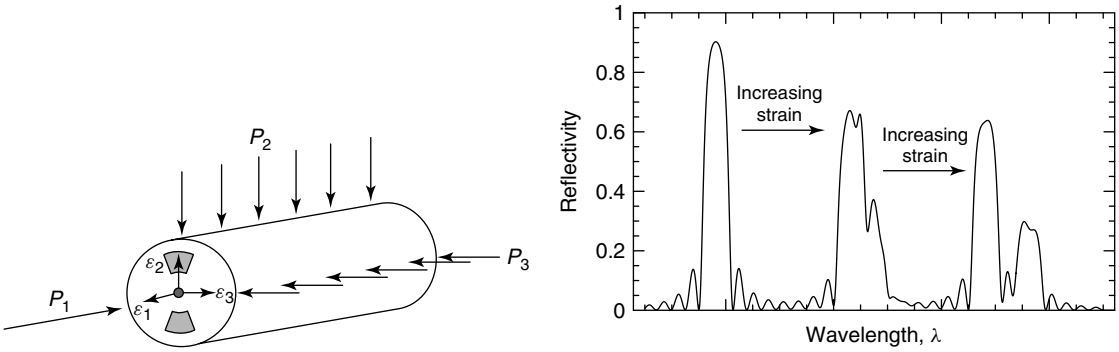


Figure 6. Multiaxis loading applied to FBG sensor written into bow-tie PM fiber. The graph plots the effects of the multiaxis loading on a single FBG (the peak on the left-hand side is the initial peak before load is applied).

location and monitoring the peak wavelength shifts for the fast and slow axes of each grating, one can independently calculate the three principle strains as well as a uniform temperature change applied to the sensor [48],

$$\begin{pmatrix} \Delta\lambda_{1f} \\ \Delta\lambda_{1s} \\ \Delta\lambda_{2f} \\ \Delta\lambda_{2s} \end{pmatrix} = \begin{pmatrix} K_{11f} & K_{12f} & K_{13f} & K_{1Tf} \\ K_{11s} & K_{12s} & K_{13s} & K_{1Ts} \\ K_{21f} & K_{22f} & K_{23f} & K_{2Tf} \\ K_{21s} & K_{22f} & K_{23f} & K_{2Ts} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \Delta T \end{pmatrix} \tag{14}$$

Generally, the relative sensitivities of the FBG written in a PM fiber are calibrated because of the complexity of predicting these sensitivities and the lack of detailed material information for the particular PM fibers used. Urbanczyk *et al.* [53], Chehura *et al.* [54], Lawrence *et al.* [48], and Bosia *et al.* [55] performed experimental measurements of the wavelength shifts in FBGs due to multiaxis loading in various PM fiber types. Chehura *et al.* [54] also performed a comparative experimental study of various PM fiber types to determine which fiber type was optimal for multiaxis strain sensing and temperature monitoring. The authors determined that amongst the specific, commercially available PM fibers studied, the elliptical core stress-applying portion (SAP) fiber provided the maximum sensitivity to transverse loading, while the Panda fiber provided the maximum sensitivity to temperature loading. However, the particular elliptical core SAP fiber used had a significantly smaller cladding diameter than the other fibers, which the authors note was the cause of the increased transverse load sensitivity. The transverse load and temperature sensitivities of the FBG can also be increased through the simultaneous measurement of the response of both the fundamental LP_{01} and LP_{11} modes; however, propagating the LP_{11} mode in the FBG presents substantial challenges [56].

To predict the response of an FBG written in a PM fiber to transverse loading, Kim *et al.* [57] considered the PM fiber to be optically and mechanically homogeneous with orthotropic material properties. As in later models, the assumption is made that most of the energy of the fundamental mode propagating in the fiber is contained in the core; therefore, the principal strains at the center of the fiber are sufficient to estimate the wavelength shift. Sirkis [58] later related the change in Bragg wavelengths, $\Delta\lambda_{B1}$ and $\Delta\lambda_{B2}$, to the principal strains at the center of the core: ε_1, ε_2, and ε_3 (see Figure 6),

$$\begin{aligned} \frac{\Delta\lambda_{B1}}{\lambda_{B1}} &= \varepsilon_1 - \frac{1}{2}n_{\text{eff}}^2\left(p_{11}\varepsilon_2 + p_{12}\varepsilon_3 + p_{12}\varepsilon_1\right) \\ \frac{\Delta\lambda_{B2}}{\lambda_{B2}} &= \varepsilon_1 - \frac{1}{2}n_{\text{eff}}^2\left(p_{12}\varepsilon_2 + p_{11}\varepsilon_3 + p_{12}\varepsilon_1\right) \end{aligned} \tag{15}$$

Lawrence *et al.* [48] and later Bosia *et al.* [55] modeled the mechanical inhomogeneities in the optical fiber due to the SAP and applied the finite element method to calculate the strains at the center of the fiber due to transverse loading. Both the experimental and numerical studies of Lawrence *et al.* [48] and Bosia *et al.* [55] demonstrate that for a PM fiber, the shift in Bragg wavelength is nonlinear with transverse load for certain loading angles. Prabhugoud and Peters [59, 60] performed finite element

analyses of FBGs written in PM fibers analysis incorporating the photoelastic-induced index distribution throughout the cross section of the optical fiber and calculating its contribution to the wavelength shift of the FBG. Comparison of this method with the previous assumption of principal strains at the center of the core in equation (15) for a variety of PM fiber types and loading cases reveals that the approximation of equation (15) is sufficient for most loading cases; however, its accuracy varies between the different fiber types.

6 NONUNIFORM SENSING

Distortion of the grating spectrum due to strain gradients has been observed in several applications of embedded FBG sensors [61–74]. An example of experimentally measured spectral distortion due to the highly nonuniform strain field near a notch tip is shown in Figure 7 [61]. This sensitivity of the sensor response to the form of the strain profile is unique to the optical FBG since other strain gauges (e.g., the classical electrical strain gauge) average the applied strain over the gauge length. This section presents both the forward prediction of the FBG spectral response due to the applied nonuniform strain field and the reverse calculation of the applied strain field from the spectral response.

6.1 Prediction of FBG response

The T-matrix approximation, first introduced by Yamada and Sakuda [75], is widely used to model Bragg gratings with nonconstant properties as explained in Figure 8. The primary advantage of the T-matrix method is that it is computationally efficient as compared to direct numerical integration of equation (3) [21]. This approach divides the grating into M smaller sections each with uniform coupling properties. Defining R_i and S_i to be the amplitudes $R(z)$ and $S(z)$ after each lightwave traverses the ith section, the propagation through this uniform section can be described in the form of a transfer matrix

$$\begin{bmatrix} R_i \\ S_i \end{bmatrix} = F_i \begin{bmatrix} R_{i-1} \\ S_{i-1} \end{bmatrix} \tag{16}$$

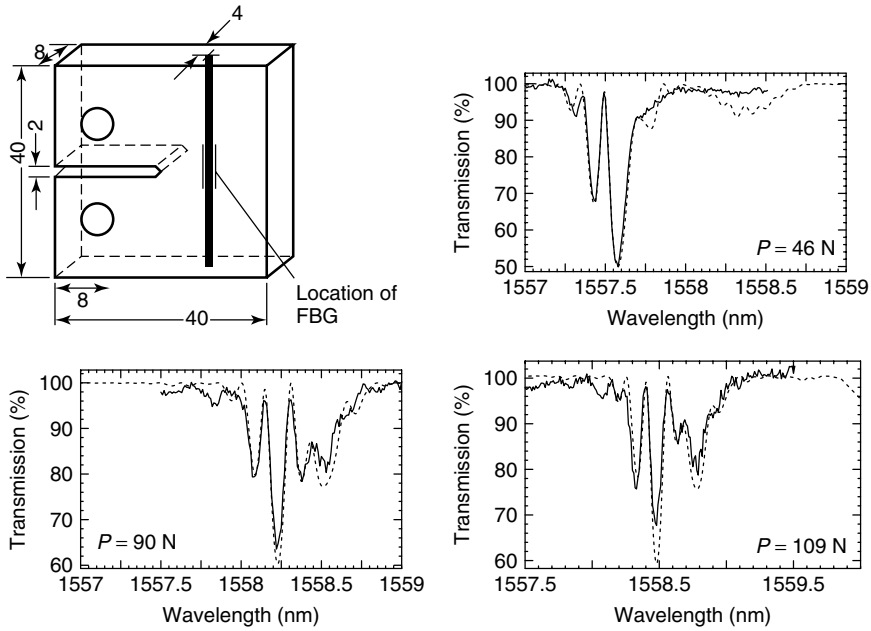


Figure 7. Response of FBG embedded in a compact tension specimen. Location of 10-mm grating is shown in the sketch of the specimen. Experimental grating spectral data in transmission (solid lines) are shown for three load levels (P). The simulated spectra calculated using the transfer matrix method are also plotted (dashed lines). [Reproduced with permission. © Springer, 2001.]

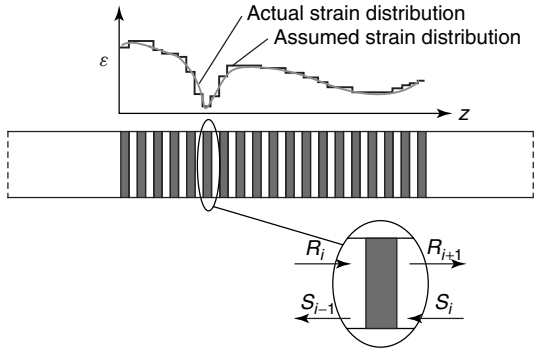


Figure 8. Schematic of transfer matrix approximation for FBG subjected to nonuniform strain field. [Reproduced with permission. © Springer, 2001.].

The optical transfer matrix can be derived as

$$F_i = \begin{bmatrix} \cosh(\gamma_B \Delta z) - i\dfrac{\hat{\sigma}}{\gamma_B}\sinh(\gamma_B \Delta z) & -i\dfrac{\kappa}{\gamma_B}\sinh\left(\gamma_B \Delta z\right) \\ i\dfrac{\kappa}{\gamma_B}\sinh(\gamma_B \Delta z) & \cosh(\gamma_B \Delta z) + i\dfrac{\hat{\sigma}}{\gamma_B}\sinh\left(\gamma_B \Delta z\right) \end{bmatrix} \quad (17)$$

where $\gamma_B = \sqrt{\kappa^2 - \hat{\sigma}^2}$ and Δz is the length of the section. For the entire grating, the combined optical transfer matrix can then be written as

$$\begin{bmatrix} R(-L/2) \\ S(-L/2) \end{bmatrix} = F \begin{bmatrix} R(L/2) \\ S(L/2) \end{bmatrix} \quad (18)$$

where $F = F_M \cdot F_{M-1} \ldots \cdot F_1$. The reflectivity as a function of wavelength can be calculated from equation (18). A limitation of the T-matrix approximation, however, is that the number of sections M cannot be arbitrarily large since several grating periods are required for complete coupling. However, $M \cong 100$ is typically more than sufficient to accurately model chirped gratings.

For axial strain-sensing applications, the T-matrix method was first applied to a Bragg grating subjected to a nonuniform strain distribution by Huang *et al.* [76]. Huang *et al.* [76] approximated the applied strain as a piecewise continuous function, calculating the average period in each grating segment due to the applied strain and substituting this local period into the coupling coefficient for the T-matrix method. The validity of this approximation was later demonstrated by Peters *et al.* [77] for various nonuniform strain fields. Prabhugoud and Peters [78] demonstrated that this prediction works well for small magnitudes of strain gradients. For an arbitrary nonuniform strain field one can enter the strain field into the coupled mode equations or transfer matrix through an "equivalent" period that accounts for both geometrical and photoelastic effects,

$$\Lambda(z) = \Lambda_0\left[1 + (1 - p_e)\varepsilon(z) + (1 - p_e)z\varepsilon'(z)\right] \quad (19)$$

where $(\,)'$ refers the derivative with respect to z [78]. A similar expression for the equivalent period for cases where the grating is originally chirped can be found in Prabhugoud and Peters [78].

6.2 Inversion of measured spectrum

When the axial strain applied to the sensor is nonconstant along its gauge length, one can no longer simply measure the shift in wavelength at maximum reflectivity to obtain useful strain data. Rather, the calculation of the applied strain profile from the spectral response becomes an inverse problem. The most commonly used inversion techniques for optical FBG strain sensor spectra are reviewed by Huang *et al.* [76]. The simplest technique is the intensity-spectrum-based approach, which requires only the amplitude information from the complex reflected spectrum. Although suitable for some applications, this technique is only valid for monotonically varying strain fields and averages out rapidly changing strain fields. A second approach based on the Fourier transform is more suitable to invert nonmonotonic strain profiles. However, this technique requires both intensity and phase information from the reflected spectrum, which significantly increases the cost and complexity of sensor data collection. Several other techniques to invert the spectral information based on heuristic approaches have been developed, including iterative solutions to the coupled mode equations (3) [79], inverse scattering algorithms (*see* **Time–frequency Analysis**) [80–82], time–frequency transform methods [83], neural networks (*see* **Artificial Neural Networks**) [84], and genetic algorithms [85, 86]. Such techniques do not

identify the direction of the strain profile; however, Kitcher *et al.* [87] have demonstrated the presence of directional dependent losses in FBGs, which could be used to identify the orientation of the strain field.

7 LONG PERIOD GRATINGS

While short period FBGs reflect or transmit lightwaves at certain wavelengths based on coupling between counterpropagating core modes, LPG sensors are based on coupling between the forward propagating core mode and forward propagating cladding modes (*see* **Fiber-optic Sensor Principles** for a description of cladding modes). While both of these and other coupling phenomena occur for both types of grating sensors, the large period of the LPG ($100\,\mu\text{m} < \Lambda < 1\,\text{mm}$) results in cladding mode coupling at wavelengths in the near-IR range. Because of the fact that cladding modes attenuate rapidly in optical fibers, the LPG coupling is not measurable from the reflected spectrum, but rather appears as multiple loss peaks in the transmission spectrum (see Figure 9). The phase matching condition for each of these loss peaks is given by

$$\lambda_i = \left[n_{\text{eff}} - n_{\text{clad}}^{(i)}\right]\Lambda \tag{20}$$

where $n_{\text{clad}}^{(i)}$ is the effective index of refraction of the ith cladding mode [88]. The maximum transmission loss for the ith cladding mode is given by

$$T_i = 1 - \sin^2(\kappa_i L) \tag{21}$$

where κ_i is the coupling coefficient for the mode (similar to κ for the short period FBG).

Because of the larger scale of the LPG features as compared to that of short period FBGs, a variety of fabrication procedures have been demonstrated to write LPGs in silica fibers including photosensitivity [89], electric arc discharge [90], etching of the fiber using hydrofluoric acid [91], and CO_2 lasers [92]. Additionally, nonpermanent LPGs have been fabricated in optical fibers by applying mechanical pressure to the fiber with an undulating surface to create a photoelastic effect–induced change in index within the fiber core [93, 94].

While LPG sensors themselves have primarily been applied to the measurement of environmental or chemical parameters because of their strong sensitivity to the index of refraction of the surrounding material system, researchers have also combined them with other optical fiber sensors for multiple parameter sensing. LPGs react to applied strain and temperature fields in much the same manner as short period FBGs, including peak splitting due to transverse loading. A significant difference between the two sensors is that the temperature sensitivity of LPGs typically ranges from 3 nm/100 °C to 10 nm/100 °C, an order of magnitude greater than that of FBGs [88]. Additionally, the strain and temperature sensitivity can be tuned by writing LPGs with different properties. Bhatia *et al.* [95] designed a "strain insensitive" LPG for which the photoelastic effects, Poisson contraction, and period extension as a function of strain canceled each other out, leading to an independent temperature sensor. Patrick *et al.* [96] fabricated a hybrid FBG/LPG sensor for independent strain and temperature measurements. Kim *et al.* [97] fabricated a hybrid LPG/Fabry–Perot interferometer (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**) for the simultaneous measurement of refractive index and temperature.

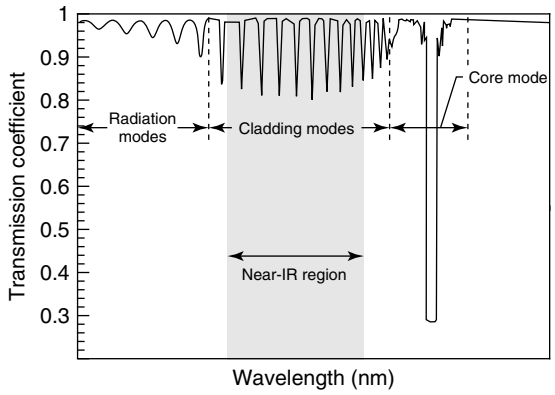


Figure 9. Theoretical transmission spectrum for long period FBG. Portion of spectrum in near-IR region is indicated.

8 CONCLUSIONS

FBGs are versatile and lightweight sensors that can be multiplexed into sensor networks and embedded into various material systems for the measurement of strain and temperature. For structural

health monitoring applications, FBG sensors offer the advantages of immunity to electromagnetic interference and the ability to measure multiaxis strain fields and local strain distributions. As compared to electrical resistance or piezoelectric strain gauges, these advantages come at the cost of expense and data acquisition rate limits for large networks. At the same time, advances in telecommunication networking and devices have led to recent advances in these areas (for example, high-speed tunable microelectromechanical systems (MEMS) filters). Recent developments including the topics of FBG sensors for high-temperature applications and FBGs in polymer and photonic crystal optical fibers can be found in **Novel Fiber-optic Sensors**.

RELATED ARTICLES

Lamb Wave-based SHM for Laminated Composite Structures

Fiber-optic Sensors

Reliable Use of Fiber-optic Sensors

REFERENCES

[1] Friebele EJ, *et al.* Optical fiber sensors for spacecraft applications. *Smart Materials and Structures* 1999 **8**:813–838.

[2] Ecke W, Latka I, Willsch R, Reutlinger A, Graue R. Fibre optic sensor network for spacecraft health monitoring. *Measurement Science and Technology* 2001 **12**:974–980.

[3] McKenzie I, Jones R, Marshall IH, Galea S. Optical fibre sensors for health monitoring of bonded repair systems. *Composite Structures* 2000 **50**:405–416.

[4] Sekine H, Fujimoto SE, Okabe T, Takeda N, Yokobori T. Structural health monitoring of cracked aircraft panels repaired with bonded patches using fiber Bragg grating sensors. *Applied Composite Materials* 2006 **13**:87–98.

[5] Li HCH, Beck F, Dupouy O, Herszberg I, Stoddart PR, Davis CE, Mouritz AP. Strain-based health assessment of bonded composite repairs. *Composite Structures* 2006 **76**:234–242.

[6] Mizutani T, Takeda N, Takeya H. On-board strain measurement of a cryogenic composite tank mounted on a reusable rocket using FBG sensors. *Structural Health Monitoring—An International Journal* 2006 **5**:205–214.

[7] Kang DH, Kim CU, Kim CG. The embedment of fiber Bragg grating sensors into filament wound pressure tanks considering multiplexing. *NDT&E International* 2006 **39**:109–116.

[8] Chan THT, Yu L, Tam HY, Ni YQ, Chung WH, Cheng LK. Fiber Bragg grating sensors for structural health monitoring of Tsing Ma bridge: background and experimental observation. *Engineering Structures* 2006 **28**:648–659.

[9] Zhang W, Gao JQ, Shi B, Cui HL, Zhu H. Health monitoring of rehabilitated concrete bridges using distributed optical fiber sensing. *Computer-Aided Civil and Infrastructure Engineering* 2006 **21**: 411–424.

[10] Kister G, Winter D, Badcock RA, Gebremichael YM, Boyle WJO, Meggitt BT, Grattan KTV, Fernando GF. Structural health monitoring of a composite bridge using Bragg grating sensors. part 1: evaluation of adhesives and protection systems for the optical sensors. *Engineering Structures* 2007 **29**:440–448.

[11] Lau KT, Yuan LB, Zhou LM, Wu JS, Woo CH. Strain monitoring in FRP laminates and concrete beams using FBG sensors. *Composite Structures* 2001 **51**:9–20.

[12] Ren L, Li HN, Zhou J, Li DS, Sun L. Health monitoring system for offshore platform with fiber Bragg grating sensors. *Optical Engineering* 2006 **48**:084401-1–084401-9, Art. No. 084401.

[13] Bin Lin Y, Lin TK, Chen CC, Chiu JC, Chang KC. Online health monitoring and safety evaluation of the relocation of a research reactor using fiber Bragg grating sensors. *Smart Materials and Structures* 2006 **15**:1421–1428.

[14] Todd MD, Johnson GA, Vohra ST. Deployment of a fiber Bragg grating-based measurement system in a structural health monitoring application. *Smart Materials and Structures* 2001 **10**:534–539.

[15] Cusano A, Capoluongo P, Campopiano S, Cutolo A, Giordano M, Felli F, Paolozzi A, Caponero M. Experimental modal analysis of an aircraft model wing by embedded fiber Bragg grating sensors. *IEEE Sensors Journal* 2006 **6**:67–77.

[16] Ling HY, Lau KT, Cheng L. Determination of dynamic strain profile and delamination detection of composite structures using embedded multiplexed fibre-optic sensors. *Composite Structures* 2005 **67**:317–326.

[17] Takeda N, Okabe Y. Durability analysis and structural health management of smart composite structures using small-diameter fiber optic sensors. *Science and Engineering of Composite Materials* 2005 **15**:1–12.

[18] Pearson JD, Zikry MA, Prabhugoud M, Peters K. Global-local assessment of low-velocity impact damage in woven composites. *Journal of Composite Materials* 2007 **41**:2759–2783.

[19] Herszberg I, Li HCH, Dharmawan F, Mouritz AP, Nguyen M, Bayandor J. Damage assessment and monitoring of composite ship joints. *Composite Structures* 2005 **67**:205–216.

[20] Kim MH. A smart health monitoring system with application to welded structures using piezoceramic and fiber optic transducers. *Journal of Intelligent Material Systems and Structures* 2006 **17**:35–44.

[21] Erodgan T. Fibre grating spectra. *Journal of Lightwave Technology* 1997 **15**:1277–1294.

[22] Hill KO, Fujii Y, Johnson DC, Kawasaki BS. Photosensitivty in optical fiber waveguides—application to reflection filter fabrication. *Applied Physics Letters* 1978 **32**:647–649.

[23] Kashyap R. *Fiber Bragg Gratings*. Academic Press San Diego, CA, 1999.

[24] Kersey AD, Davis MA, Patrick HJ, LeBlanc M, Koo KP, Askins CG, Putnam MA, Friebele EJ. Fiber grating sensors. *Journal of Lightwave Technology* 1997 **15**:1442–1463.

[25] Xu MG, Archambault JL, Reekie L, Dakin P. Discrimination between strain and temperature effects using dual-wavelength fibre grating sensors. *Electronics Letters* 1994 **30**:1085–1087.

[26] Meltz G, Morey WW, Glenn WH. Formation of Bragg gratings in optical fibers by a transverse holographic method. *Optics Letters* 1989 **14**:823–825.

[27] Othonos A, Kalli K. *Fiber Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing*. Artech House Norwood, MA, 1999.

[28] Wei CY, Ye CC, James SW, Tatam RP, Irving PE. The influence of hydrogen loading and the fabrication process on the mechanical strength of optical fiber Bragg gratings. *Optical Materials* 2002 **20**:241–251.

[29] Varelas D, Costantini DM, Limberger HG, Salathé RP. Fabrication of high-mechanical-resistance Bragg gratings in single-mode optical fibers with continuous-wave ultraviolet laser side exposure. *Optics Letters* 1998 **23**:397–399.

[30] Askins CG, Putnam MA, Patrick HJ, Friebele EJ. Fibre strength unaffected by on-line writing of single-pulse Bragg gratings. *Electronics Letters* 1997 **33**:1333–1334.

[31] Gu XJ, Guan L, He YF, Zhang HBB, Herman PR. High-strength fiber Bragg gratings for a temperature-sensing array. *IEEE Sensors Journal* 2006 **6**: 668–671.

[32] Han YG, Han WT, Lee BG, Paek UC, Chung YJ. Temperature sensitivity control and mechanical stress effect of boron-doped long-period fiber gratings. *Fiber and Integrated Optics* 2001 **20**:591–600.

[33] Kersey AD, Berkoff TA, Morey WW. Multiplexed fiber Bragg grating strain-sensor system with a fiber Fabry–Perot wavelength filter. *Optics Letters* 1993 **18**:1370–1372.

[34] Xu MG, Geiger H, Archambault JL, Reekie L, Dakin JP. Novel interrogating system for fiber Bragg grating sensors using an acoustooptic tunable filter. *Electronics Letters* 1993 **29**:1510–1511.

[35] Davis MA, Kersey AD. Matched-filter interrogation technique for fiber Bragg grating arrays. *Electronics Letters* 1995 **31**:822–823.

[36] Betz DC, Thursby G, Culshaw B, Staszewski WJ. Acousto-ultrasonic sensing using fiber Bragg gratings. *Smart Materials and Structures* 2003 **12**: 122–128.

[37] Fallon RW, Zhang L, Everall LA, Williams JAR, Bennion I. All-fibre optical sensing system: Bragg grating sensor interrogated by a long-period grating. *Measurement Science and Technology* 1998 **9**:1969–1973.

[38] Davis MA, Kersey AD. All fiber Bragg grating strain sensor demodulation technique using a wavelength division coupler. *Electronics Letters* 1994 **30**:75–77.

[39] Zhang Q, Brown DA, Kung H, Townsend JE, Chen M, Reinhart LJ, Morse TF. Use of highly overcoupled couplers to detect shifts in Bragg wavelength. *Electronics Letters* 1995 **31**:480–482.

[40] Davis MA, Bellemore DG, Putnam MA, Kersey AD. Interrogation of 60 fiber Bragg grating sensor with μstrain resolution capability. *Electronics Letters* 1996 **32**:1393–1394.

[41] Askins CG, Putnam MA, Williams GM, Friebele EJ. Stepped-wavelength optical fiber Bragg grating arrays fabricated in line on a draw tower. *Optics Letters* 1994 **19**:147–149.

[42] Froggatt M, Moore J. Distributed measurement of static strain in an optical fiber with multiple Bragg gratings at nominally equal wavelengths. *Applied Optics* 1998 **37**:1741–1746.

[43] Koo KP, Kersey AD. Bragg grating-based laser sensors systems with interferometric interrogation and wavelength-division multiplexing. *Journal of Lightwave Technology* 1995 **13**:1243–1249.

[44] Todd MD, Johnson GA, Chang CC. Passive, light intensity-independent interferometric method for fibre Bragg grating interrogation. *Electronics Letters* 1999 **35**:1970–1971.

[45] Todd MD, Johnson GA, Althouse BL. A novel Bragg grating sensor interrogation system utilizing a scanning filter, a Mach-Zehnder interferometer and a 3×3 coupler. *Measurement Science and Technology* 2001 **12**:771–777.

[46] Magne S, Rougeault S, Vilela M, Ferdinand P. State-of-strain evaluation with fiber Bragg grating rosettes: application to discrimination between strain and temperature effects in fiber sensors. *Applied Optics* 1997 **36**:9437–9447.

[47] Betz DC, Thursby G, Culshaw B, Staszewski WJ. Advanced layout of a fiber Bragg grating strain gauge rosette. *Journal of Lightwave Technology* 2006 **24**:1019–1026.

[48] Lawrence CM, Nelson DV, Udd E, Bennett T. A fiber optic sensor for transverse strain measurements. *Experimental Mechanics* 1999 **39**:202–209.

[49] Kang HK, Kang DH, Hong CS, Kim CG. Simultaneous monitoring of strain and temperature during and after cure of unsymmetric composite laminate using fibre-optic sensors. *Smart Materials and Structures* 2003 **12**:29–35.

[50] Leng JS, Asundi A. Real-time cure monitoring of smart composite materials using extrinsic Fabry–Perot interferometer and fiber Bragg grating sensors. *Smart Materials and Structures* 2002 **11**:249–255.

[51] Murukeshan VM, Chan PY, Ong LS, Seah LK. Cure monitoring of smart composites using fiber Bragg grating based embedded sensors. *Sensors and Actuators, A* 2000 **79**:153–161.

[52] O'Dwyer MJ, Maistros SM, James SW, Tatam RP, Partridge IK. Relating the state of cure to the real-time internal strain development in a curing composite using in-fibre Bragg gratings and dielectric sensors. *Measurement Science and Technology* 1998 **9**:1153–1158.

[53] Uranczyk W, Chmielewska E, Bock WJ. Measurements of temperature and strain sensitivities of two-mode Bragg gratings imprinted in a bow-tie fibre. *Measurement Science and Technology* 2001 **12**:800–804.

[54] Chehura E, Ye CC, Staines SE, James SW, Tatam RP. Characterization of the response of fibre Bragg gratings fabricated in stress and geometrically induced high birefringence fibres to temperature and transverse load. *Smart Materials and Structures* 2004 **13**:888–895.

[55] Bosia F, Giaccari P, Botsis J, Facchini M, Limberger HG, Salathé R. Characterization of the response of fibre Bragg grating sensors subjected to a two-dimensional strain field. *Smart Materials and Structures* 2003 **12**:925–934.

[56] Doyle C, Martin A, Liu T, Wu M, Hayes S, Crosby PA, Powell GR, Brooks D, Fernando GF. *In-situ* process and conditioning monitoring of advanced fibre-reinforced composite materials using optical fibre sensors. *Smart Materials and Structures* 1998 **7**:145–158.

[57] Kim KS, Kollàr L, Springer GS. A model of embedded fiber optic Fabry–Perot temperature and strain sensors. *Journal of Composite Materials* 1993 **27**:1618–1662.

[58] Sirkis JS. Unified approach to phase-strain-temper ature models for smart structure interferometric optical fiber sensors: part 1, development. *Optical Engineering* 1993 **32**:752–761.

[59] Prabhugoud M, Peters K. Finite element model for embedded fiber Bragg grating sensor. *Smart Materials and Structures* 2006 **15**:550–562.

[60] Prabhugoud M, Peters K. Finite element analysis of multi-axis strain sensitivities of Bragg gratings in PM fibers. *Journal of Intelligent Material Systems and Structures* 2007 **18**:861–873.

[61] Peters K, Studer M, Botsis J, Iocco A, Limberger H, Salathé R. Embedded optical fiber Bragg grating sensor in a nonuniform strain field: measurements and simulations. *Experimental Mechanics* 2001 **41**:19–28.

[62] Studer M, Peters K, Botsis J. Method for determination of crack bridging parameters using long optical fiber Bragg grating sensors. *Composites Part B* 2003 **34**:347–359.

[63] Ling HY, Lau KT, Chen L, Chow KW. Embedded fibre Bragg grating sensors for non-uniform strain sensing in composite structures. *Measurement Science and Technology* 2005 **16**:2415–2424.

[64] Ling HY, Lau KT, Su Z, Wong ETT. Monitoring mode II fracture behavior of composite laminates using embedded fiber-optic sensors. *Composites, Part B* 2007 **38**:488–497.

[65] Kang DH, Park SO, Hong CS, Kim CG. The signal characteristics of reflected spectra of fiber Bragg

grating sensors with strain gradients and grating lengths. *NDT&E International* 2005 **38**:712–718.

[66] Kuang KSC, Kenny R, Whelan MP, Cantwell WJ, Chalker PR. Embedded fibre Bragg grating sensors in advanced composite materials. *Composites Science and Technology* 2001 **61**:1379–1387.

[67] Guemes JA, Menéndez JM. Response of Bragg grating fiber-optic sensors when embedded in composite laminates. *Composites Science and Technology* 2002 **62**:959–966.

[68] Prabhugoud M, Peters K. Efficient simulation of Bragg grating sensors for implementation to damage identification in composites. *Smart Materials and Structures* 2003 **12**:914–924.

[69] Rao YJ, *et al.* Simultaneous strain and temperature measurement of advanced 3-D braided composite materials using an improved EFPI/FBG system. *Optics and Lasers in Engineering* 2002 **38**:557–566.

[70] Yashiro S, Takeda N, Okabe T, Sekine H. A New approach to predicting multiple damage states in composite laminates with embedded FBG sensors. *Composite Science and Technology* 2005 **65**:659–667.

[71] Yashiro S, Okabe T, Toyama N, Takeda N. Monitoring damage in holed CFRP laminates using embedded chirped FBG sensors. *International Journal of Solids and Structures* 2007 **44**:603–613.

[72] Takeda N, Yashiro S, Okabe T. Estimation of the damage patterns in notched laminates with embedded FBG sensors. *Composites Science and Technology* 2006 **66**:684–693.

[73] Yashiro S, Okabe T, Takeda N. Damage identification in a holed CFRP laminate using a chirped fiber Bragg grating sensor. *Composites Science and Technology* 2007 **67**:286–295.

[74] Minakuchi S. Real-time detection of debonding between honeycomb core and facesheet using a small-diameter FBG sensor embedded in adhesive layer. *Journal of Sandwich Structures and Materials* 2007 **9**:9–33.

[75] Yamada M, Sakuda K. Analysis of almost-periodic distributed feedback slab waveguides via a fundamental matrix approach. *Applied Optics* 1987 **26**:3474–3478.

[76] Huang S, Ohn M, LeBlanc M, Measures RM. Continuous arbitrary strain profile measurements with fiber Bragg gratings. *Smart Materials and Structures* 1998 **7**:248–256.

[77] Peters K, Pattis P, Botsis J, Giaccari P. Experimental verification of response of embedded optical fiber Bragg grating sensors in non-homogeneous strain fields. *Optics and Lasers in Engineering* 2000 **33**:107–119.

[78] Prabhugoud M, Peters K. Modified transfer matrix formulation for Bragg grating strain sensors. *Journal of Lightwave Technology* 2004 **22**:2302–2309.

[79] Peral E, Capmany J, Marti J. Iterative solution to the Gel'fand-Levitan-Marchenko coupled equations and application to synthesis of fiber gratings. *IEEE Journal of Quantum Electronics* 1996 **32**:2078–2084.

[80] Feced R, Zervas M, Muriel M. An efficient inverse scattering algorithm for the design of nonuniform fiber Bragg gratings. *IEEE Journal of Quantum Electronics* 1999 **35**:1105–1115.

[81] Giaccari P, Limberger HG, Salathé RP. Local coupling-coefficient characterization in fiber Bragg gratings. *Optics Letters* 2003 **28**:598–600.

[82] Chapeleau X, Leduc D, Lupi C, Le NyR. Experimental synthesis of fiber Bragg gratings using optical low coherence reflectometry. *Applied Physics Letters* 2003 **82**:4227–4229.

[83] Azana J, Muriel M. Fiber Bragg grating period reconstruction using time-frequency signal analysis and application to distributed sensing. *Journal of Lightwave Technology* 2001 **19**:646–654.

[84] Paterno AS, Silva JCC, Milczewski MS, Arruda LVR, Kalinowski HJ. Radial-basis function network for the approximation of FBG sensor spectra with distorted peaks. *Measurement Science and Technology* 2006 **17**:1039–1045.

[85] Casagrande F, Crespi P, Grassi AM, Luilli A, Kenny R, Whelan MP. From the reflected spectrum to the properties of a fiber Bragg grating: a genetic algorithm approach with application to distributed strain sensing. *Applied Optics* 2002 **41**:5238–5244.

[86] Gill A, Peters K, Studer M. Genetic algorithm for the reconstruction of Bragg grating sensor strain profiles. *Measurement Science and Technology* 2004 **15**:1877–1884.

[87] Kitcher DJ, Nand A, Wade SA, Jones R, Baxter GW, Collins SF. Directional dependence of spectra of fiber Bragg gratings due to excess loss. *Journal of Optical Society of America A* 2006 **23**:2906–2911.

[88] James SW, Tatam RP. Optical fibre long-period grating sensors: characteristics and application. *Measurement Science and Technology* 2003 **14**: R49–R61.

[89] DeVries M, Bhatia V, D'Alberto T, Arya V, Claus RO. Photoinduced grating-based optical fiber sensors

for structural analysis and control. *Engineering Structures* 1998 **20**:205–210.

[90] Frazao O, Romero R, Rego G, Marques RVS, Salgado HM, Santos JL. Sampled fibre Bragg grating sensors for simultaneous strain and temperature measurement. *Electronics Letters* 2002 **38**:693–695.

[91] Lin CY, Wang LA, Chern GW. Corrugated long-period fiber gratings as strain, torsion, and bending sensors. *Journal of Lightwave Technology* 2001 **19**:1159–1168.

[92] Wang YP, Wang DN, Jin W. CO2 laser-grooved long period fiber grating temperature sensor system based on intensity modulation. *Applied Optics* 2006 **45**:7966–7970.

[93] Mishra V, Singh N, Jain SC, Kaur P, Luthra R, Singla H, Jindal VK, Bajpai RP. Refractive index and concentration sensing of solutions using mechanically induced long period grating pair. *Optical Engineering* 2005 **44**:094402-1–094402-4, Art. No. 094402.

[94] Cho JY, Lim JH, Lee KS. Optical fiber twist sensor with two orthogonally oriented mechanically induced long-period grating sections. *IEEE Photonics Technology Letters* 2005 **17**:453–455.

[95] Bhatia V, Campbell DK, Sherr D, Ten Eyck GA, Murphy KA, Claus RO. Temperature-insensitive and strain-insensitive long-period grating sensors for smart structures. *Optical Engineering* 1997 **36**: 1872–1876.

[96] Patrick HJ, Williams GM, Kersey AD, Pedrazzani JR, Vengsarkar AM. Hybrid fiber Bragg grating/ long period fiber grating sensor for strain/temperature discrimination. *IEEE Photonics Technology Letters* 1996 **8**:1223–1225.

[97] Kim DW, Shen F, Chen XP, Wang AB. Simultaneous measurement of refractive index and temperature based on a reflection-mode long-period grating and an intrinsic Fabry–Perot interferometer sensor. *Optics Letters* 2005 **30**: 3000–3002.

Chapter 62
Novel Fiber-optic Sensors

Kara Peters

Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

1 INTRODUCTION

Fiber-optic sensors have been applied for a variety of structural health monitoring applications because of their immunity to electromagnetic interference, low weight, small size, and multiplexing capabilities (*see* **Fiber-optic Sensors**). Some of the articles in this encyclopedia review optical fiber concepts useful for understanding fiber-optic sensors (*see* **Fiber-optic Sensor Principles**), as well as the basic operating principles for intensity-based, interferometric, and fiber Bragg grating (FBG) sensors (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**; **Fiber Bragg Grating Sensors**). The goal of this article is to present recent, novel concepts for fiber-optic sensors. Most of these sensors are based on recent advances in optical fiber technology, including single-crystal sapphire, single-mode polymer, and microstructured and multicore optical fibers. Such advances allow the use of optical fiber sensors for high-temperature and large-strain applications.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 EXTREME TEMPERATURE SENSORS

A variety of factors limit the usable temperature range for fiber-optic sensors including the material limits of the optical fibers themselves and temperature limits for any adhesives used in packaging. Another important limit for FBG sensors (*see* **Fiber Bragg Grating Sensors**) is that the index modulation in the FBG written through photosensitivity can be erased at approximately 200 °C. This limitation could be important for many composite applications, as processing temperatures of polymer matrix–based composites may exceed this limit. Other matrix materials have even higher processing temperatures. Coating the optical fiber with polyimide does increase the temperature survival range for the FBG to 500 °C [1]; however, this may not be suitable for many high-temperature applications. The temperature limit for

FBG erasing is well below the material limit for fused silica, typically 1000–1100 °C. As one solution, Luo *et al.* [1] developed a "Bragg stack" by depositing quarter-wavelength-thickness layers on the end of a polyimide-coated silica optical fiber. The alternating layers of silicon nitrite and silicon-rich silicon nitrite acted as a Bragg diffraction grating, which could be interrogated in reflection through the optical fiber. The Bragg stack was applied as a temperature sensor and was demonstrated to survive to at least 800 °C with a 2 °C temperature resolution.

An alternative fabrication method was applied by Lowder *et al.* [2], who fabricated FBGs through first etching a significant portion of the cladding on the flat surface of a D-fiber. Then a surface relief FBG close to the core of the optical fiber was etched. Similar methods have been applied to write long period gratings (*see* **Fiber Bragg Grating Sensors**). As the FBG is created mechanically, rather than chemically as in the case of photosensitive FBGs, the FBG can withstand significantly higher temperatures. Lowder *et al.* [2] demonstrated the linearity of the sensor up to 1100 °C.

To overcome the temperature limit of the silica itself, researchers have also developed optical fiber sensors based on single-crystal sapphire fibers, which have a melting point of 2050 °C [3–5]. Tong *et al.* [4] demonstrated a single-crystal sapphire fiber temperature sensor based on blackbody radiation for high-temperature applications; however, they discovered that surface contamination can significantly degrade the measurements at high temperatures. Wang *et al.* [3] and Xiao *et al.* [5] designed extrinsic Fabry–Perot white-light interferometer sensors, coupling the light-wave from a single mode to a multimode sapphire fiber and achieving a strain resolution of 0.2 microstrain up to 1004 °C. Zhu and Wang [6] later extended the sapphire fiber sensor by bonding the single-crystal sapphire fiber to the surface of a sapphire wafer whose optical path length through the thickness changes with temperature. By further splicing the sapphire fiber to the standard silica fiber lead, removing the need for optical adhesives, the authors demonstrated a temperature resolution of 0.4 °C up to 1170 °C.

At the other end of the usable temperature range for fiber-optic sensors, Zeisberger *et al.* [7] demonstrated that FBGs can be applied for accurate strain measurements at extremely low temperatures, specifically in the temperature range of 30–300 K. The measured wavelength shift of an FBG due to only the thermo-optic effect reported by the authors is plotted in Figure 1. For these measurements, the FBG was constrained so as to not permit thermal expansion.

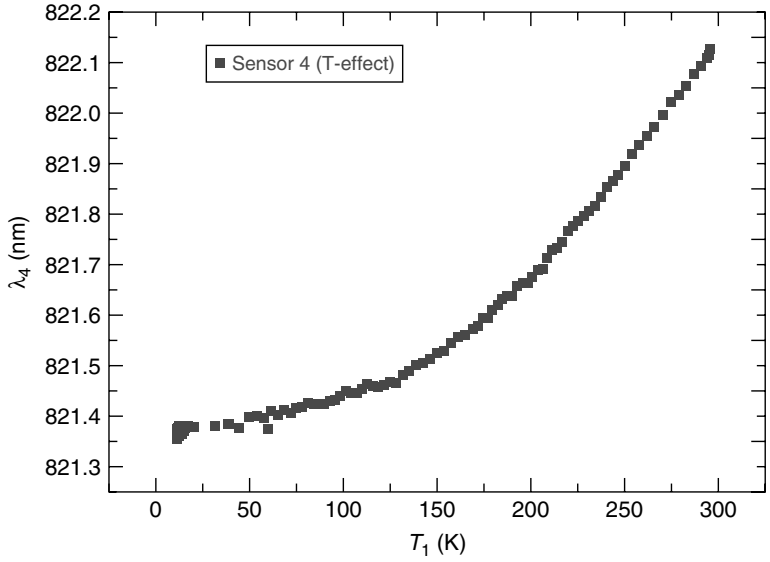


Figure 1. Measured Bragg wavelength of FBG as a function of temperature (change due to thermo-optic effect only). [Reproduced with permission from Ref. 7. © 2005, IEEE.]

The response of the FBG to temperature is linear to approximately 250 K, and then strongly nonlinear at lower temperatures.

3 POLYMER LARGE-DEFORMATION SENSORS

As demonstrated by the large number of optical fiber sensors and their applications (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**; **Fiber Bragg Grating Sensors**; **Fiber-optic Sensors**), optical fiber sensors are extremely versatile. However, the brittle nature of silica limits their maximum strain range to approximately 1–2%, particularly for FBGs [8]. This measurement range is comparable with electrical resistance strain gauges. Polymer optical fibers (POFs) on the other hand offer a much larger potential strain measurement range. In general, the fabrication of POFs is not as well controlled as for silica fibers; therefore, their performance has not been comparable. At the same time, intrinsic material losses in polymers are orders of magnitude higher than for silica. Therefore, POFs have primarily been used for short, flexible optical connections and as sensors operating through bending or environmental intensity losses. However, recent advances in the fabrication of single-mode POFs with minimal attenuation levels have changed the potential for POF sensors [9–13].

Figure 2 plots the inherent material losses for several optical polymers and silica. The material losses in polymethylmethacrylate (PMMA) generally increases with wavelength, and therefore POFs are often designed to operate at wavelengths below the near-IR wavelengths used with silica fibers. Fabricating single-mode fibers at these lower wavelengths requires smaller core diameters in order to keep the normalized frequency, V, in the single-mode range (*see* **Fiber-optic Sensor Principles**). Kuzyk *et al.* [9] first fabricated PMMA optical fibers with a dye-doped core that were single mode at an operating wavelength, $\lambda = 1300$ nm. These fibers

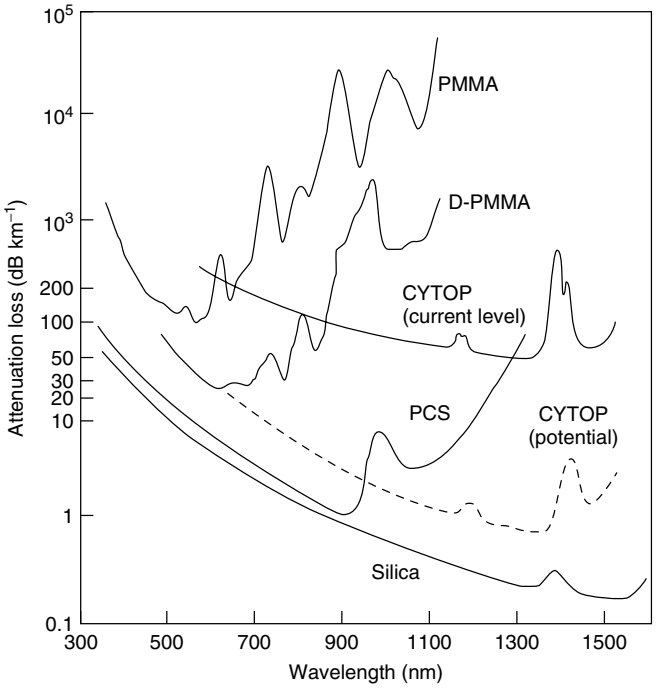


Figure 2. Transmission loss spectra for various polymer optical fiber materials: PMMA, D-PMMA (deuterated), CYTOP, and PCS. Loss spectra for silica is also plotted for comparison. [Reproduced with permission from Ref. 14. © 2001, Elsevier.]

demonstrated attenuation levels of approximately $0.3\,\mathrm{dB\,cm^{-1}}$, close to the intrinsic material attenuation for PMMA at that wavelength. Bosc and Toinen [10] later fabricated single-mode PMMA optical fibers with attenuation levels of $0.25\,\mathrm{dB\,cm^{-1}}$ at $\lambda = 1550\,\mathrm{nm}$ and $0.05\,\mathrm{dB\,cm^{-1}}$ at $\lambda = 850\,\mathrm{nm}$. Finally, Garvey *et al.* [11] improved the process to create single-mode POFs at $\lambda = 1060\,\mathrm{nm}$ with an attenuation level of $0.18\,\mathrm{dB\,cm^{-1}}$.

3.1 Sensing properties of POFs

Silva-López *et al.* [15] first measured the sensitivity of dye-doped single-mode POFs to strain and temperature. The POFs were designed to be single mode at 850 nm with an acrylic cladding ($n = 1.4905$) and doped PMMA core ($n = 1.4923$). The authors operated the fibers with a visible light source at $\lambda = 632\,\mathrm{nm}$, outside of the single-mode region; however, they only observed the fundamental, LP_{01}, mode propagating through the fiber. Using a Mach–Zender interferometer arrangement and loading the optical fiber on a translation stage, they measured a phase sensitivity to displacement of $131 \times 105\,\mathrm{rad\,m^{-1}}$ and temperature sensitivity of $-212\,\mathrm{rad\,m^{-1}\,K^{-1}}$. The phase sensitivity to displacement measured is in good agreement with the properties of bulk PMMA. When compared with silica optical fibers, the sensitivity to displacement of the PMMA fiber is 14% higher than that of the silica, which is an advantage for strain sensing. The difference is primarily due to the large difference in Poisson's ratio and the photoelastic constants (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). On the other hand, the temperature sensitivity of the PMMA fiber is negative, whereas the temperature sensitivity of silica optical fibers is positive. The temperature sensitivity was also about 25% more than predicted. The maximum operating temperature for the fiber was predicted to be in the 80–120 °C range.

The measurements of Silva-López *et al.* [15] provide useful information on the behavior of POF sensors and can be used to predict the response of a variety of sensors based on them. However, the measurements were made in the strain range of 0–0.04% strain, which is a limited portion of the strain range over which they have potential to be applied. For this reason, Kiesel *et al.* [16] derived a formulation to predict the phase sensitivity to strain of the dye-doped PMMA POF for the full usable range of the sensor (≅0–6%) and measured the mechanical properties of the fiber in this range. The formulation includes both the finite deformation of the optical fiber and nonlinear photoelastic effects, both potentially significantly important over the usable strain range. The measured true stress–strain response curve of the PMMA POF is plotted in Figure 3 for multiple strain rates in the range of $0.01–3.05\,\mathrm{min^{-1}}$. One can see from Figure 3 that the response of the sensor is very sensitive to strain rate and that the yield strain increases with strain, typical of polymer fibers [13]. The failure strain of the POF was around 30%, while transmission of the coherent lightwave for sensing was demonstrated to be 15.8%, well beyond the yield strain of the POF [17]. Using the mechanical properties measured from the single-mode POF samples, Kiesel *et al.* [16] predicted the phase sensitivity of the POF over a strain range of 0–6%. This sensitivity is plotted in Figure 4 with the response of a silica fiber for comparison. The important differences between POF sensor and silica sensor from Figure 4 are as follows: (i) the PMMA has a high strain sensitivity throughout the strain range, (ii) the nonlinearity in the PMMA phase response is of the opposite sign and an order of magnitude greater than for silica, and (iii) the importance of including the nonlinearity in the phase response is evident at a much lower strain value (≅1%) for PMMA than for silica (≅3%). Additional effects are expected to be important when applying

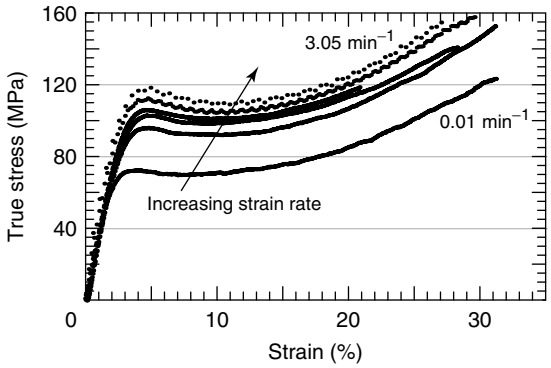


Figure 3. Measured true stress–strain curves for PMMA POF at various strain rates. Strain rates plotted are 0.01, 0.30, 0.60, 0.90, 1.22, and $3.05\,\mathrm{min^{-1}}$. [Reproduced with permission from Ref. 16. © IOP Publishing, Ltd., 2007.]

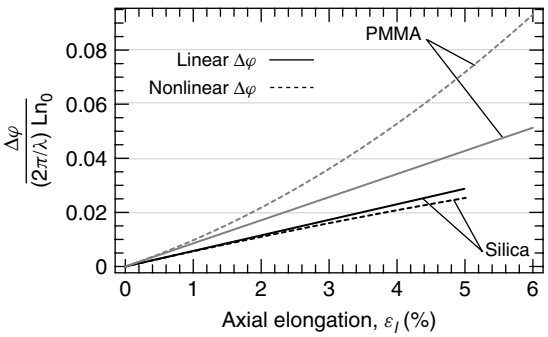


Figure 4. Predicted normalized phase shift of silica optical fiber and PMMA POF, applying linear and nonlinear deformation [16]. For specific material parameters used in calculation, see. [Reproduced with permission from Ref. 16. © IOP Publishing, Ltd., 2007.]

POFs as large-strain sensors, including material hysteresis, creep effects (especially at elevated temperatures), and the anisotropy of the polymer.

3.2 Fiber Bragg gratings in POFs

Peng and Chu [12] first demonstrated the writing of FBGs in POFs through the same photosensitivity process used for writing FBGs in silica optical fibers. The FBGs were fabricated by UV exposure of PMMA single-mode optical fibers (fabricated by the authors) with dye-doped cores to increase the photosensitivity of the polymer. Peng and Chu [12] wrote FBGs operating in near-IR wavelengths ($\lambda_B \cong 1570\,\mathrm{nm}$) with reflectivities >80% and a bandwidth of 1 nm. By applying strain to the FBG, the authors shifted the Bragg wavelength by a total of 73 nm (Figure 5), which is an order of magnitude larger than that previously achieved with FBGs in silica optical fibers. The large wavelength shift was partially due to the increased sensitivity of the POF described in the previous section and partially due to the large yield strain of the POF (Peng and Chu report 6.1% yield strain for the fiber used [12]). Applying thermal loading, the authors tuned the FBG over 20 nm; however, erasing of the FBG occurred when the grating was exposed to thermal loads for extended periods of time [18].

Liu *et al.* [19] later wrote FBGs in CYTOP optical fibers, which have a lower intrinsic material loss than PMMA (Figure 2). The sensitivity of the FBGs in CYTOP fibers was not as high as for those in PMMA fibers, for example, the gratings could only be tuned thermally over 10 nm. However, thermal erasing of the gratings did not occur when they were exposed to thermal loads for long durations.

Liu *et al.* [18] combined FBG sensors in silica and PMMA optical fibers in series to independently measure strain and temperature, through the equation

$$\begin{pmatrix} \Delta\lambda_p \\ \Delta\lambda_s \end{pmatrix} = \begin{pmatrix} K_{\varepsilon p} K_{Tp} \\ K_{\varepsilon s} K_{Ts} \end{pmatrix} \begin{pmatrix} \varepsilon \\ \Delta T \end{pmatrix} \qquad (1)$$

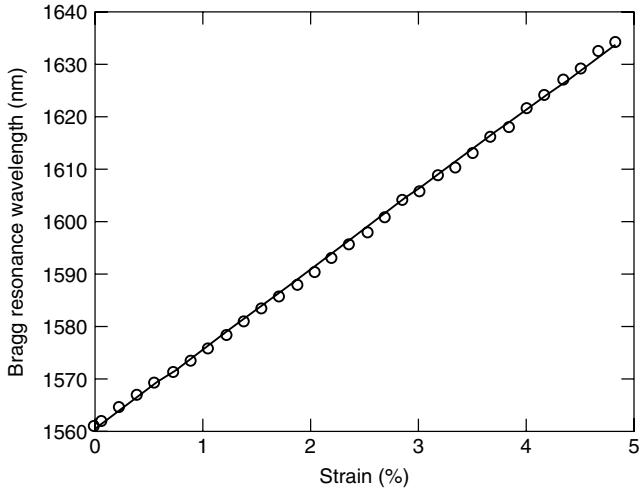


Figure 5. Measured Bragg wavelength shift due to axial strain for FBG written in polymer optical fiber. [Reproduced with permission from Ref. 12. © Taylor & Francis, Ltd., 2000.]

where the K matrix is constructed of the sensitivities of the Bragg wavelength of the FBGs in the polymer ($\Delta\lambda_p$) and silica ($\Delta\lambda_s$) fibers to strain and temperature (*see* **Fiber Bragg Grating Sensors**). These measurements are significantly better than those previously obtained using two FBGs in silica fibers at different wavelengths (*see* **Fiber Bragg Grating Sensors**) due to the fact that the sensitivity to strain and temperature is very different for the polymer and silica. The measured strain sensitivities were $K_{\varepsilon 1} = 1.48\,\text{pm}\ \mu\varepsilon^{-1}$ for the PMMA FBG and $K_{\varepsilon 2} = 1.15\,\text{pm}\,\mu\varepsilon^{-1}$ for the silica FBG. Additionally, the temperature sensitivity of the silica FBG is positive, while the temperature sensitivity of the PMMA FBG is negative.

Challenges to applying polymer FBGs for large-strain sensing include high attenuation in the fibers, difficulties in preparing the fibers for coupling, and thermal erasing of the gratings. This last challenge could be important for many structural health monitoring applications, but is still not fully understood. Similar to the writing process for FBGs in silica fibers, the grating depth increases with exposure time until a threshold is reached, at which point damage to the polymer fiber occurs and transmission losses are introduced in the fiber [20]. On the other hand, Liu *et al.* [21] observed that, when FBGs were written with lower power UV exposures, the FBG peak depth increased, reached a maximum, remained constant, then began to erase during the exposure time. Once the FBG was completely erased, the UV exposure was stopped and the FBG reappeared over a period of 8 h, then was permanent and stable. The authors speculate that the heating of the fiber during the UV exposure temporarily changed the index of refraction, counteracting the change due to photosensitivity.

4 MICROSTRUCTURED FIBER SENSORS

Microstructured optical fibers, also referred to as *photonic crystal fibers*, have been fabricated in a variety of forms including solid core, "holey" fibers, as shown in Figure 6. For a first approximation, the holey fiber acts as an index-guiding waveguide in the same manner as conventional solid optical fibers (*see* **Fiber-optic Sensor Principles**). The air holes

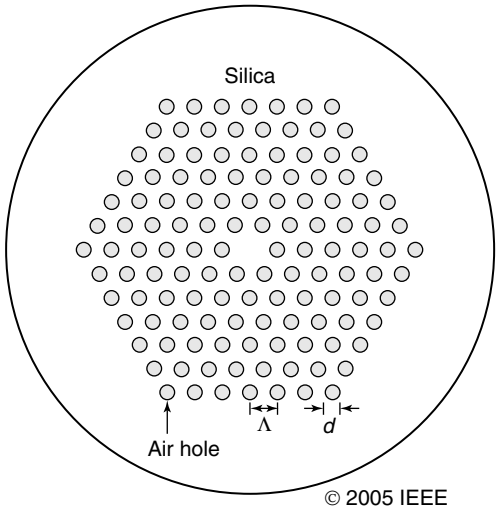


Figure 6. Schematic representation of a typical microstructured holey optical fiber. Λ and d are the hole pitch and diameter respectively. [Reproduced with permission from Ref. 24. © IOP Publishing Ltd., 2005.]

in the cladding region reduce the effective index of the cladding and therefore provide much stronger confinement of the lightwave as it propagates through the fiber [22]. Such microstructured optical fibers are typically fabricated by drawing a preform consisting of capillary tubes with a solid silica rod replacing the center tube to form the solid core. The complex geometry of these holey fibers provides both new sensing possibilities and enhanced response characteristics for conventional fiber-optic sensors applied to these fibers. For example, Ju *et al.* [23] fabricated an interferometric sensor based on a holey fiber and observed a nonmonotonic temperature dependence on wavelength.

Nasilowski *et al.* [25] and Bock *et al.* [26] applied microstructured holey fibers for the measurement of temperature and pressure. Through measurement of the group refractive index change, Bock *et al.* [26] estimated the temperature, axial strain, and pressure sensitivity of such fibers. The primary differences observed as compared to conventional fibers were that the reduced Poisson contraction of the cross section (i) reduced the strain sensitivity slightly and (ii) changed the pressure sensitivity to be negative rather than positive as for solid fibers. Nasilowski *et al.* [25] applied a high-birefringence holey fiber, created by introducing several defects to form a

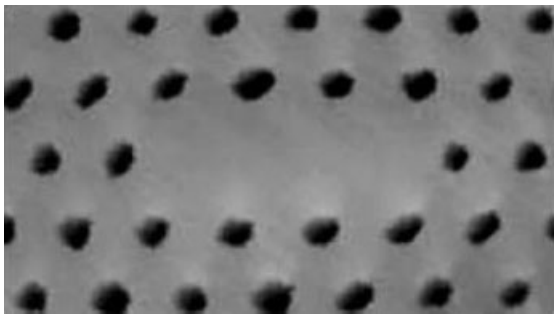


Figure 7. Scanning-electron-microscope image of the core of high-birefringence holey fiber showing the triple defect core. [Reproduced with permission from Ref.23. © Springer Verlag, 2005.]

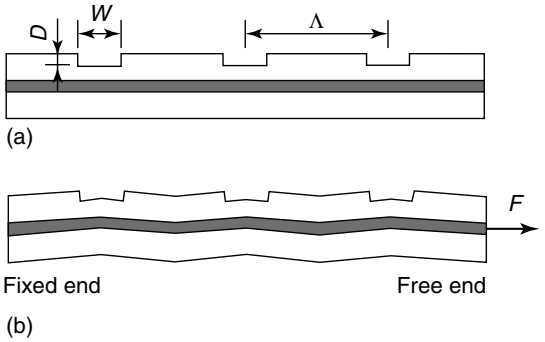


Figure 8. Schematic of CO_2 laser carved LPG (a) before and (b) after axial force is applied to the fiber. Λ, D, and, W are the grating pitch, depth, and, width of the grooves, respectively. [Reproduced with permission from Ref. 27. © OSA Publishing, 2006.]

noncircular core, as a pressure sensor. An image of the fiber core and surrounding region is shown in Figure 7. An important feature of high-birefringence holey fibers is that strong birefringence can be obtained without inducing thermal residual stresses through a different material region in the fiber (*see* **Fiber-optic Sensor Principles**). From a sensing perspective, this birefringence is therefore stable with temperature and does not introduce temperature errors into high-birefringence sensors [25].

Several researchers have written long period fiber Bragg gratings (LPGs) in microstructured fibers through CO_2 laser or electric arc etching [27–29]. These LPGs demonstrate an excellent strain sensitivity, with little or no temperature sensitivity, meaning that temperature compensation is not required [27, 29]. When bending the optical fiber, He *et al.* [28] observed mode splitting and bending directional dependence in the output of the LPG, phenomena not observed with LPGs written in conventional solid silica optical fibers. The directional dependence is due to the fact that the in-plane stiffness of the cross section has been significantly weakened due to the presence of the holes. The response of the LPG does appear to strongly depend on the fabrication method, as Petrovic *et al.* [29] did not observe such phenomena in similar tests. Wang *et al.* [27] also found that, during axial stretching of the optical fiber, the periodic grooves in the LPGs created local microbending around the grooves (Figure 8). This microbending increased the sensitivity of the LPG to axial strain. Demonstrating the fabrication of grating structures through photosensitivity, Dobb *et al.* [30] successfully wrote FBGs in PMMA holey fibers operating at $\lambda_B = 1569\,\text{nm}$ with a 0.5 nm bandwidth. The FBGs were fabricated in both few-mode and single-mode holey fibers.

Zou *et al.* [31] demonstrated independent strain and temperature measurements through stimulated Brillouin scattering in microstructured optical fibers (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). In contrast to scattering in conventional solid fibers, several peaks were observed in the scattered frequency spectrum due to independent interactions between the lightwave and stimulated acoustic waves occurring in the graded-Ge-doped core region, the intermediate silica region, and the microstructured cladding region. The response of the peak in the core region was the same as for a doped solid optical fiber; however, the other peaks had differing sensitivities to temperature and strain.

One final example of a microstructured fiber sensor is the tapered holey fiber sensor demonstrated by Villatoro *et al.* [32]. The authors tapered the middle section of a holey fiber to a diameter of 28 μm (the original diameter was 125 μm), which caused the holes to collapse in the cross section. The output lightwave intensity from the optical fiber was then an interference pattern between the multiple modes of the tapered waist [33]. These interference peaks shifted linearly with strain and were insensitive to temperature up to 180 °C.

5 MULTICORE OPTICAL FIBER SENSORS

Multicore optical fibers have great potential for the measurement of strains and curvatures over large, flexible structures such as aircraft wings or towed hydrophone arrays. When multiplexed, these local strain and curvature measurements can be used to determine the shape of the flexible structure. The concept for curvature measurements based on differential strain measurements between core pairs is shown in Figure 9 for a four-core optical fiber. The presence of multiple cores within a single fiber can provide more stable curvature measurements than those from separate single-mode fibers because more accurate spacing is maintained between the cores [34]. The choice of spacing between the cores can be an important design consideration for multicore sensors. Close spacing of the cores makes the sensor more compact as well as less sensitive to temperature gradients across the fiber cross section. However, reducing the core spacing also increases the coupling between modes propagating through the cores and reduces the accuracy of curvature measurements [35]. Methods to couple the multicore fiber to multiple single-mode fibers for data acquisition include graded index (GRIN) lenses and fan-outs [35]. Multicore optical fibers have also been applied for Doppler differential velocimetry [36] and pitch and roll sensing [35].

Blanchard *et al.* [34] first demonstrated strain and curvature measurements using a multicore optical fiber. They fabricated a three-core holey fiber by drawing silica capillary tubes with silica rods introduced for the cores, shown in Figure 10. The cores were placed sufficiently far apart such that each core acted as a separate single-mode waveguide with no coupling between propagating lightwaves (assuming that no twisting is applied). Light was launched into the optical fiber, creating a far-field interface pattern at the end of the 23-cm sensor length, which was then projected onto a Charge-Coupled Device (CCD) camera. The optical phase difference of each core was extracted from the Fourier transform of the interference pattern. On the basis of the differential phase differences, the sensor demonstrated a strain resolution of 7.4 nε and bend angle resolution of 100 μrad. MacPherson *et al.* [38] later demonstrated that the phase sensitivity to displacement ($\Delta\phi/\Delta L$) of the multicore holey fiber is approximately the same as a solid fiber. Using only a two-core holey fiber, they obtained a bend angle resolution of 170 μrad.

MacPherson *et al.* [37] wrote FBGs into each of the cores of the four-core solid optical fiber of Figure 9

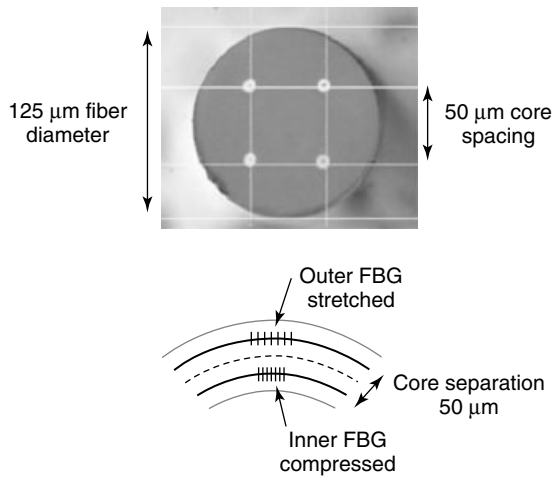


Figure 9. Cross section of multicore fiber illuminated with white light and principle of curvature measurement from FBG pairs in different cores. [Reproduced with permission from Ref. [37]. © IOP Publishing Ltd., 2006.]

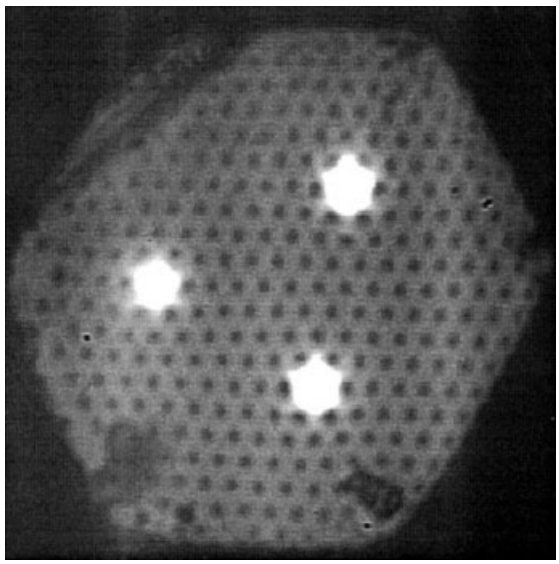


Figure 10. SEM image of the end of a photonic crystal fiber near-field pattern at the output of the three-core fiber with all cores illuminated at 633 nm. [Reproduced with permission from Ref. 34. © IOP Publishing Ltd., 2000.]

for shape monitoring of a tunnel structure. Multiplexing a large number of FBGs along the length of the optical fiber allows for a distributed measurement of the curvature, which can be integrated to obtain the shape of the structure [39]. MacPherson *et al.* obtained a shape resolution of ± 0.1 mm between two sets of FBGs spaced over a length of 5 mm along the fiber. Flockhart *et al.* [40] and Cranch *et al.* [41] used the same fiber configuration, but wrote sets of FBG pairs in each core to act as a low-finesse Fabry–Perot interferometer to enhance the static and dynamic resolution of the shape measurement. Applying differential interferometric demodulation between the grating pairs, the authors obtained a strain resolution of $0.6\,\mathrm{n\varepsilon\,Hz^{-1/2}}$ and curvature resolution of $0.012\,\mathrm{km^{-1}\,Hz^{-1/2}}$ for measurement frequencies above 1 Hz, a 30 times improvement over the previous curvature measurements.

6 MEMS OPTICAL FIBER SENSORS

The ability to multiplex a large number of sensors into a single optical fiber is one of the key advantages to optical fiber sensors for structural health monitoring. Additionally, their relatively small size makes them reasonable to embed or surface mount to a variety of structures (*see* **Fiber-optic Sensor Principles**). However, more and more sensors are being developed for applications where size is critical, thus, necessitating microelectromechanical systems (MEMSs) fabrication techniques. To date, MEMS technology has primarily been applied to optical fiber sensors for Fabry–Perot temperature and pressure sensors (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). Watson *et al.* [42] reviews the state of the art for low-profile external Fabry–Perot diaphragm sensors. However, due to their extrinsic nature, the outside diameters of these sensors are generally larger than the outside diameter of the optical fiber itself. To reduce the outside sensor diameter of end-face Fabry–Perot pressure sensors, Watson *et al.* [42] fabricated a MEMS Fabry–Perot sensor. The authors ablated the end of an optical fiber with an excimer laser, and then bonded an aluminized polycarbonate foil diaphragm to the end of the fiber, which can be seen in Figure 11. The laser ablation

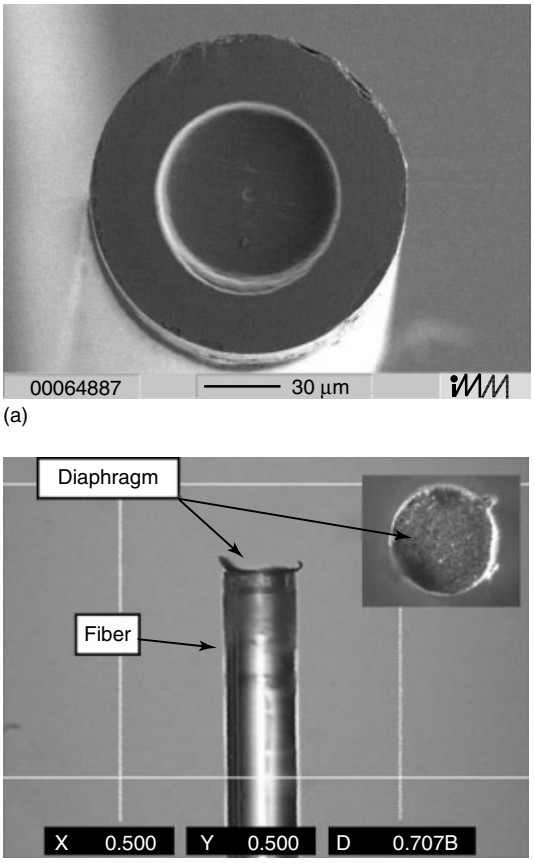


Figure 11. (a) A 70-µm-diameter cavity ablated into the end of a single-core fiber and (b) photograph of the side view of a single-core fiber sensor (the inset shows the top view). [Reproduced with permission from Ref. 42. © OSA Publishing, 2006.]

produced a good quality cavity; however, the authors encountered some difficulties due to tearing of the diaphragm during bonding and poor adhesion of the diaphragm.

The requirement to adhere the diaphragm with an adhesive glue also limits the temperature range over which such MEMS sensors can be applied [43]. Abeysinghe *et al.* [43], therefore, developed a Fabry–Perot temperature and pressure sensor without adhesives, also applying a MEMS fabrication technology. They micromachined the end of the optical fiber through wet etching to create a cavity similar to the last example, but then bonded a crystalline silicone membrane to the end face of the optical

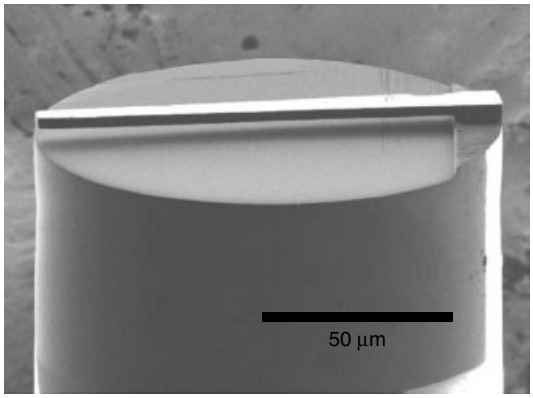


Figure 12. Scanning-electron-microscopy image of the fiber-top cantilever (before evaporation of the silver layer) [44]. Dimensions: length ≅ 112 μm, width ≅ 14 μm, and thickness ≅ 3.7 μm. [Reproduced with permission from Ref.44. © American Institute of Physics, 2006.]

fiber through anodic bonding. The authors experimentally demonstrated the excellent performance of the MEMS sensor in a 25–300 °F thermal environment.

At a smaller scale, Iannuzzi *et al.* [44] and Deladi *et al.* [45] machined a thin rectangular beam out of the cleaved edge of an optical fiber, as shown in Figure 12. The vertical displacement of the cantilever was determined by measuring the amplitude of interference of laser light reflected by the fiber–air interface and reflected by the cantilever. The authors prepared the cantilevered sensor by coating a standard single-mode optical fiber with a thin metallic layer (5-nm Cr, 20-nm Pd) and then machining the cantilever using focused ion beam machining. A 100-nm layer of silver was then deposited on the sensor to increase the reflectivities. For a further description of focused ion beam machining applied to optical fiber sensors, see Nellen and Brönnimann [46]. The authors demonstrated that the subnanometer displacement resolution of the cantilever beam sensor was comparable with results from atomic force microscopy.

7 CONCLUSIONS

This article reviews recent advances in fiber-optic sensor technology based on surface relief gratings and sapphire optical fibers for high-temperature applications; single-mode POFs for large-deformation sensors; and microstructured and multicore optical fibers. Applications of MEMs fabrication technique to fiber-optic sensors are also presented. These advances have extended the range for optical fiber sensors such that they can now be uniquely applied for structural health monitoring applications in extreme environments.

RELATED ARTICLES

Lamb Wave-based SHM for Laminated Composite Structures

Reliable Use of Fiber-optic Sensors

REFERENCES

[1] Luo F, Hernández-Cordero J, Morse TF. Multiplexed fiber-optic Bragg stack sensors (FOBSS) for elevated temperatures. *IEEE Photonics Technology Letters* 2001 **13**:514–516.

[2] Lowder TL, Smith KH, Ipson BL, Hawkins AR, Selfridge RH, Schultz SM. High-temperature sensing using surface relief fiber Bragg gratings. *IEEE Photonics Technology Letters* 2005 **17**:1926–1928.

[3] Wang A, Gollapudi S, May RG, Murphy KA, Claus RO. Sapphire optical fiber-based interferometer for high-temperature environmental applications. *Smart Materials and Structures* 1995 **4**:147–151.

[4] Tong LM, Shen YH, Ye LH. Performance improvement of radiation-based high-temperature fiber-optic sensor by means of curved sapphire fiber. *Sensors and Actuators, A* 1999 **75**:35–40.

[5] Xiao H, Deng J, Pickrell G, May RG, Wang A. Single-crystal sapphire fiber-based strain sensor for high-temperature applications. *Journal of Lightwave Technology* 2003 **21**:2276–2283.

[6] Zhu Y, Wang A. Surface-mount sapphire interferometric temperature sensor. *Applied Optics* 2006 **45**:6071–6076.

[7] Zeisberger M, Latka I, Ecke W, Habisreuther H, Litzkendorf D, Gawalek W. Measurement of the thermal expansion of melt-textured YBCO using optical fibre grating sensors. *Superconductor Science and Technology* 2005 **18**:S202–S205.

[8] Nellen PhM, Mauron P, Frank A, Sennhauser U, Bohnert K, Pequignot P, Bodor P, Brändle H. Reliability of fiber Bragg grating based sensors for downhole applications. *Sensors and Actuators, A* 2003 **103**:364–376.

[9] Kuzyk MG, Paek UC, Dirk CW. Guest-host polymer fibers for nonlinear optics. *Applied Physics Letters* 1991 **59**:902–904.

[10] Bosc D, Toinen C. Full polymer single-mode optical fiber. *IEEE Photonics Technology Letters* 1992 **4**:749–750.

[11] Garvey DW, *et al.* Single-mode nonlinear-optical polymer fibers. *Journal of the Optical Society of America A* 1996 **18**:2017–2023.

[12] Peng GD, Chu PL. Polymer optical fiber photosensitivities and highly tunable fiber gratings. *Fiber and Integrated Optics* 2000 **19**:277–293.

[13] Jiang CH, Kuzyk MG, Ding JL, Johns WE, Welker DJ. Fabrication and mechanical behavior of dye-doped polymer optical fiber. *Journal of Applied Physics* 2002 **92**:4–12.

[14] Zubia J, Arrue J. Plastic optical fibers: an introduction to their technological processes and applications. *Optical Fiber Technology* 2001 **7**:101–140.

[15] Silva-López M, *et al.* Strain and temperature sensitivity of a single-mode polymer optical fiber. *Optics Letters* 2005 **30**:3129–3131.

[16] Kiesel S, Peters K, Hassan T, Kowalsky M. Behaviour of intrinsic polymer optical fibre sensor for large-strain applications. *Measurement Science and Technology* 2007 **18**:3144–3154.

[17] Kiesel S, Peters K, Hassan T, Kowalsky M. Large deformation in-fiber polymer optical fiber sensor. *IEEE Photonics Technology Letters* 2008 **20**: 416–418.

[18] Liu HB, Liu HY, Peng GD, Chu PL. Strain and temperature sensor using a combination of polymer and silica fibre Bragg gratings. *Optics Communications* 2003 **219**:139–142.

[19] Liu HY, Peng GD, Chu PL. Thermal stability of gratings in PMMA and CYTOP polymer fibers. *Optics Communications* 2002 **204**:151–156.

[20] Liu HY, Liu HB, Peng GD, Chu PL. Observation of type I and type II gratings behavior in polymer optical fiber. *Optics Communications* 2003 **220**:337–343.

[21] Liu HB, Liu HY, Peng GD, Chu PL. Novel growth behaviors of fiber Bragg gratings in polymer optical fiber under UV irradiation with low power. *IEEE Photonics Technology Letters* 2004 **16**:159–161.

[22] Zolla F, Renversez G, Nicolet A, Kuhlmey B, Guenneau S, Felbacq D. *Foundations of Photonic Crystal Fibres*. Imperial College Press: London, 2005.

[23] Ju J, Wang Z, Jin W, Demokan MS. Temperature sensitivity of a two-mode photonic crystal fiber interferometer sensor. *IEEE Photonics Technology Letters* 2006 **18**:2168–2170.

[24] Saitoh K, Koshiba M. Numerical modeling of photonic crystal fibers. *Journal of Lightwave Technology* 2005 **23**:3580–3590.

[25] Nasilowski T, *et al.* Temperature and pressure sensitivities of the highly birefringent photonic crystal fiber with core asymmetry. *Applied Physics B* 2005 **81**:325–331.

[26] Bock WJ, Urbanczyk W, Wojcik J. Measurements of sensitivity of the single-mode photonic crystal holey fibre to temperature, elongation and hydrostatic pressure. *Measurement Science and Technology* 2004 **15**:1496–1500.

[27] Wang YP, Xiao L, Wang DN, Jin W. Highly sensitive long-period fiber-grating strain sensor with low temperature sensitivity. *Optics Letters* 2006 **31**:3414–3416.

[28] He Z, Zhu Y, Du H. Effect of macro-bending on resonant wavelength and intensity of long-period gratings in photonic crystal fiber. *Optics Express* 2007 **15**:1804–1810.

[29] Petrovic JS, Dobb H, Mezentsev VK, Kalli K, Webb DJ, Bennion I. Sensitivity of LPGs in PCFs fabricated by an electric arc to temperature, strain and external refractive index. *Journal of Lightwave Technology* 2007 **25**:1306–1312.

[30] Dobb H, Webb DJ, Kalli K, Argyros A, Large MCJ, Van Eijkelenborg MA. Continuous wave ultraviolet light-induced fiber Bragg gratings in few- and single-mode microstructured polymer optical fibers. *Optics Letters* 2005 **30**:3296–3298.

[31] Zou L, Bao X, Chen L. Distributed Brillouin temperature sensing in photonic crystal fiber. *Smart Materials and Structures* 2005 **14**:S8–S11.

[32] Villatoro J, Minkovich VP, Monzon-Hernández D. Temperature-independent strain sensor made from tapered holey fiber. *Optics Letters* 2006 **31**(3): 305–307.

[33] Minkovich VP, Monzon-Hernandez D, Villatoro J, Sotsky AB, Sotskaya LI. Modeling of holey fiber tapers with selective transmission for sensor applications. *Journal of Lightwave Technology* 2006 **24**:4319–4328.

[34] Blanchard PM, *et al.* Two-dimensional bend sensing with a single, multi-core optical fibre. *Smart Materials and Structures* 2000 **9**:132–140.

[35] MacPherson WN, Flockhart GMH, Maier RRJ, Barton JS, Jones JDC, Zhao D, Zhang L, Bennion I. Pitch and roll sensing using fibre Bragg gratings in multicore fibre. *Measurement Science and Technology* 2004 **15**:1642–1646.

[36] MacPherson WN, Jones JDC, Mangan BJ, Knight JC, Russell PStJ. Two-core photonic crystal fibre for Doppler difference velocimetry. *Optics Communications* 2003 **223**:375–380.

[37] MacPherson WN, *et al.* Tunnel monitoring using multicore fibre displacement sensor. *Measurement Science and Technology* 2006 **17**:1180–1185.

[38] MacPherson WN, *et al.* Remotely addressed optical fibre curvature sensor using multicore photonic crystal fibre. *Optics Communications* 2001 **193**: 97–104.

[39] Duncan RG, Froggatt ME, Kreger ST, Seeley RJ, Gifford DK, Sang AK, Wolfe MS. High accuracy fiber-optic shape sensing. In *Proceedings of the SPIE Smart Sensor Systems and Networks: Phenomena, Technology, and Applications for NDE and Health Monitoring (SPIE Vol. 6530)*, Peters K (ed). SPIE: Bellingham, WA, 2007, pp. 65301S-1–65301S-11.

[40] Flockhart GMH, Cranch GA, Kirkendall CK. Differential phase tracking applied to Bragg gratings in multi-core fibre for high accuracy curvature measurement. *Electronics Letters* 2006 **42**: 390–391.

[41] Cranch GA, Flockhart GMH, MacPherson WN, Barton JS, Kirkendall CK. Ultra-high sensitivity two-dimensional bend sensor. *Electronics Letters* 2006 **42**:520–522.

[42] Watson S, Gander MJ, MacPherson WN, Barton JS, Jones JDC, Klotzbuecher T, Braune T, Ott J, Schmitz F. Laser-machined fibers as Fabry-Perot pressure sensors. *Applied Optics* 2006 **45**: 5590–5596.

[43] Abeysinghe DC, Dasgupta S, Jackson HE, Boyd JT. Novel MEMS pressure and temperature sensors fabricated on optical fibers. *Journal of Micromechanics and Microengineering* 2002 **12**:229–235.

[44] Iannuzzi D, Deladi S, Gadgil VJ, Sanders RGP, Schreuders H, Elwenspoek MC. Monolithic fiber-top sensor for critical environments and standard applications. *Applied Physics Letters* 2006 **88**:053501-1–053501-3.

[45] Deladi S, Iannuzzi D, Gadgil VJ, Schreuders H, Elwenspoek MC. Carving fiber-top optomechanical transducers from an optical fiber. *Journal of Micromechanics and Microengineering* 2006 **16**:886–889.

[46] Nellen PhM, Brönnimann R. Milling micro-structures using focused ion beams and its application to photonic components. *Measurement Science and Technology* 2006 **17**:943–948.

Chapter 60

Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors

Kara Peters

Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

1 INTRODUCTION

Optical-fiber sensors present numerous advantages for the measurement of strain, temperature, humidity, pressure, and other parameters. This article focuses on sensors that use the intrinsic properties of optical fibers themselves as a transducer. By evaluating the behavior of a lightwave propagating through an optical fiber, strain and temperature fields can thus be inferred from microbending, microfracture, or scattering losses that occur along the fiber. Alternatively, the phase shift of the propagating lightwave can be used to determine the strain or temperature profile along the fiber through the photoelastic effect. In comparison, fiber Bragg gratings and other optical-fiber sensors are reviewed in **Fiber Bragg Grating Sensors** and **Novel Fiber-optic Sensors**.

A summary of the fundamental principles for optical fibers, useful to understand the sensing phenomena described in this article, can be found in **Fiber-optic Sensor Principles**. General issues concerning the application of all the sensors described in this article to structural health-monitoring systems are also presented in **Fiber-optic Sensor Principles** and **Fiber-optic Sensors**.

2 INTENSITY-BASED OPTICAL-FIBER SENSORS

Intensity-based optical-fiber sensors are potentially the simplest in-fiber sensors to apply. Typically, these are based on the measurement of the loss of intensity transmitted through an optical fiber. Such sensors can be interrogated with low-cost lightwave sources and detection systems such as light emitting diodes (LEDs) and photodetectors. Measures *et al.* [1, 2] presented the first example of an intensity-based optical-fiber sensor for the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

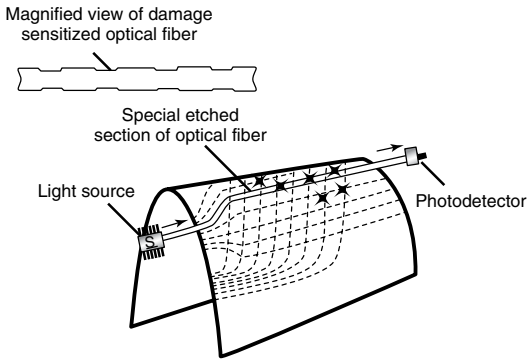


Figure 1. Intensity-based damage-sensitized optical fiber embedded as a two-dimensional sensor network in a composite aircraft wing. [Reproduced with permission from Ref. 3. © Elsevier, 2001.]

measurement of structural damage. As shown in Figure 1, the authors sensitized discrete regions of the optical fiber by etching the cladding to reduce the diameter of the optical fiber. The optical fibers were then embedded to form a two-dimensional grid network in an aircraft wing. The optical fiber in these regions was thus sensitive to damage at approximately the same level as the surrounding aramid fiber reinforced composite. During impact testing of the wing, microfractures in the optical fiber produced the bleeding of light from the fibers. By measuring the total intensity of the lightwave transmitted through each fiber, the researchers could therefore determine the extent of damage within the composite.

Intensity-based sensors that load the optical fiber through microbending have been developed for the measurement of strain, temperature, pressure, or other parameters [4–7]. Significant microbending along the optical fiber creates transmission losses that increase rapidly with the bend radius of curvature (*see* **Fiber-optic Sensor Principles**). Figure 2 presents a schematic of a typical microbend sensor design. As tensile strain is applied (for the example of Figure 2), the microbending applied to the optical fiber is reduced and therefore the transmitted intensity increases. Conversely, if compressive strain is applied, the transmitted intensity is reduced. To prevent errors due to fluctuations in the intensity of the light source, a reference optical fiber and a dual photodetector are also included. While relatively inexpensive, the accuracy of intensity-based sensors can be limited.

3 INTERFEROMETRIC OPTICAL FIBER SENSORS

Butter and Hocker [8] designed and demonstrated the first in-fiber strain sensor. This "fiber-optics strain gauge" was based on the measurement of the change in phase shift of a lightwave propagating through an optical fiber due to axial strain applied to the

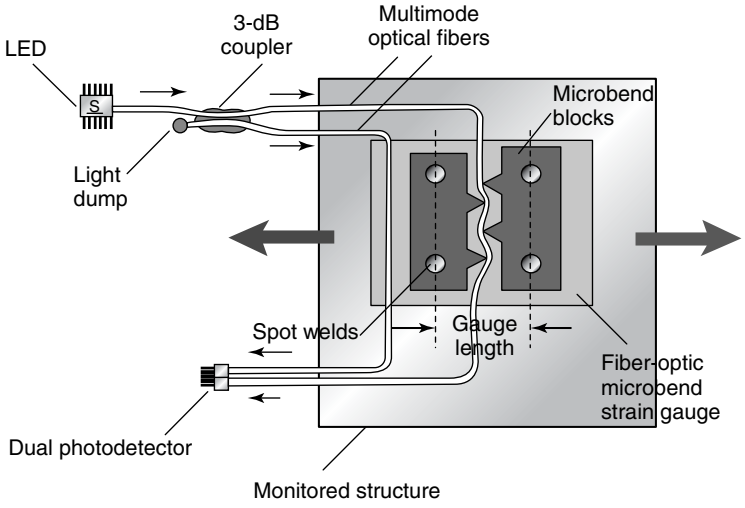


Figure 2. Schematic of a typical intensity-based microbend. [Reproduced with permission from Ref. 3. © Elsevier, 2001.]

optical fiber. This section derives the change in phase shift of the lightwave due to mechanical and thermal loads and presents common interferometer configurations for the measurement of this change in phase shift. Finally, more recent distributed interferometer sensor devices based on low-coherence interferometery and stimulated Brillouin scattering are also described.

3.1 Strain and temperature sensitivity

We assume that the optical-fiber material is an optically isotropic single crystal with an index of refraction in the unstressed state of n_0 (this would be the effective index of refraction of the fundamental mode, *see* **Fiber-optic Sensor Principles**). Since the fiber cross section is a two-dimensional surface, we define the orthogonal coordinate system with the axes p and q shown in Figure 3. These axes are the principal optical axes, which correspond to the propagation axes for the fiber. We write the field vector for an electromagnetic plane wave of wavelength λ, propagating in the axial, 1, direction (Figure 3) as

$$\mathbf{E} = A^p \mathbf{S}^p \sin\left[\omega t - \frac{2\pi n^p}{\lambda} x_1\right] + A^q \mathbf{S}^q \sin\left[\omega t - \frac{2\pi n^q}{\lambda} x_1\right] \qquad (1)$$

where $\mathbf{S}^p$ and $\mathbf{S}^q$ are orthogonal unit vectors in the 2–3 plane in the direction of the principle optical axes, ω is the angular frequency of the lightwave, and A^p and A^q are the amplitudes of the orthogonal components [9]. The propagating lightwave is thus split into two orthogonal lightwaves, both with the same wavelength and frequency as the original wave, but each experiencing a separate index of refraction, n^p and n^q.

The vector field, $\mathbf{E}$, must satisfy the planar wave equation

$$\mathbf{R} \times (\mathbf{R} \times \mathbf{B}\mathbf{E}) + \frac{1}{n^2}\mathbf{E} = 0 \qquad (2)$$

where n is the index of refraction experienced by the field and $\mathbf{R}$ is the unit vector in the direction of propagation; in this case, $\mathbf{R} = (1, 0, 0)$. The tensor $\mathbf{B}$ is the material dielectric impermeability tensor,

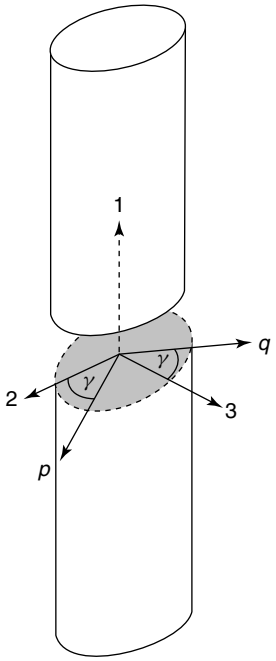


Figure 3. Sketch of optical fiber with coordinate systems highlighted on an arbitrary cross section ($p - q$ are the principal optical axes; 1–2–3 are the applied strain directions).

which is directly related to the index of refraction of the material.

$$\mathbf{B} = \begin{bmatrix} B_1 & B_6 & B_5 \\ B_6 & B_2 & B_4 \\ B_5 & B_4 & B_3 \end{bmatrix} \qquad (3)$$

For an isotropic material, $B_1 = B_2 = B_3 = n_0$, $B_4 = B_5 = B_6 = 0$ [10]. Since the vector field is propagating in the 1 direction, $\mathbf{E}$ has spatial components in the 2 and 3 directions, $\mathbf{E} = (0, E_2, E_3)$. Substituting $\mathbf{E}$ and $\mathbf{R}$ into equation (2) and expanding yields the two equations

$$\begin{bmatrix} B_2 - 1/n^2 & B_4 \\ B_4 & B_3 - 1/n^2 \end{bmatrix} \begin{pmatrix} E_2 \\ E_3 \end{pmatrix} = 0 \qquad (4)$$

Nontrivial solutions for E_2 and E_3 occur only when the determinate of the matrix on the left-hand side is zero. Applying this condition and solving for n yields

$$\frac{1}{n^2} = \frac{(B_2 + B_3) \pm \sqrt{(B_2 - B_3)^2 + 4B_4^2}}{2} \qquad (5)$$

In general, there are therefore two solutions to n, which we previously labeled as n^p and n^q. For the isotropic fiber, we simply have $n^p = n^q = n_0$; therefore, there is only one propagation solution.

The next step is to calculate the change in phase shift of the lightwave propagating through a length L of the optical fiber. As the lightwave, **E**, propagates through the optical fiber, it experiences a total phase shift, φ,

$$\varphi = \beta L = \frac{2\pi}{\lambda} nL \tag{6}$$

β is thus the propagation constant or phase shift per unit length, $\beta = 2\pi n/\lambda$. Once strain or thermal loading is applied to the optical fiber, both the length of the fiber and the effective index of refraction change, altering the phase shift of the propagating lightwave [8],

$$\Delta\varphi = \frac{2\pi}{\lambda}(\Delta nL + n\Delta L) \tag{7}$$

The change in length of the optical fiber is directly related to the axial strain, ε_1, and the coefficient of thermal expansion, α, as $\Delta L = \varepsilon_1 L + \alpha \Delta T$.

We now calculate the term Δn of equation (7). The photoelastic effect describes the change in optical properties of the material due to strain and temperature. Since the components of the tensor **B** are dependent on the material indices of refraction, we can write $B_i = B_i^0 + \Delta B_i$ and apply this to equation (5). Simplifying this expression and using the isotropic values for B_i^0, we find

$$\frac{1}{n^2} = \frac{1}{n_0^2} + \frac{(\Delta B_2 + \Delta B_3) \pm \sqrt{(\Delta B_2 - \Delta B_3)^2 + 4\Delta B_4^2}}{2} \tag{8}$$

Assuming the applied strains are small, we can apply the linear thermoelastic strain-optic effect, which can be written as

$$\Delta B_i = p_{ij}(\varepsilon_j - \alpha_j \Delta T) + W_i \Delta T \tag{9}$$

where the summation convention is applied and the strain components are written in compact notation [9]. Here we consider that all six components of strain can be applied to the optical fiber; more specific loading cases will be considered later. The terms W_i are the changes in B_i due to temperature measured at a constant stress state, $W_i = \partial B_i / \partial \Delta T|_{\sigma=\text{const.}}$. The components of the **p** matrix are commonly referred to as the *Pockel's constants* for the material. For an isotropic material, **p** has the form [10]

$$\mathbf{p} = \begin{bmatrix} p_{11} & p_{12} & p_{12} & 0 & 0 & 0 \\ p_{12} & p_{11} & p_{12} & 0 & 0 & 0 \\ p_{12} & p_{12} & p_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{(p_{11}-p_{12})}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{(p_{11}-p_{12})}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{(p_{11}-p_{12})}{2} \end{bmatrix} \tag{10}$$

Similarly, for an isotropic material,

$$\mathbf{W} = \begin{bmatrix} -\frac{2}{n_0^3}\frac{\mathrm{d}n_0}{\mathrm{d}T} \\ -\frac{2}{n_0^3}\frac{\mathrm{d}n_0}{\mathrm{d}T} \\ -\frac{2}{n_0^3}\frac{\mathrm{d}n_0}{\mathrm{d}T} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{11}$$

Typical values for fused silica fibers are $p_{11} = 0.17$, $p_{12} = 0.36$, $\mathrm{d}n_0/\mathrm{d}T = 1.2 \times 10^{-5}/^{\circ}\mathrm{C}$ [9]. Combining equations (8–11), we find the general expression for the index change due to strain and temperature loading

$$\frac{1}{n^2} = \frac{1}{n_0^2} + p_{12}\varepsilon_1 + \frac{(p_{11}+p_{12})}{2}(\varepsilon_1 + \varepsilon_2) \pm \frac{(p_{11}-p_{12})}{2}\sqrt{(\varepsilon_2 - \varepsilon_3)^2 + 4\varepsilon_4^2} - \frac{2}{n_0^3}\frac{\mathrm{d}n_0}{\mathrm{d}T}\Delta T - \alpha\Delta T(p_{11} + 2p_{12}) \tag{12}$$

where the strain components are the total strain due to both the thermal and mechanical loading.

Considering the specific case of pure axial tension without thermal loading, for which $\varepsilon_1 = \varepsilon$, $\varepsilon_2 = \varepsilon_3 = -\nu\varepsilon$, we find

$$\frac{1}{n^2} = \frac{1}{n_0^2} + \left[p_{12} - \nu(p_{11} + p_{12})\right]\varepsilon \tag{13}$$

Calculating $n - n_0$ and linearizing with respect to strain, we find

$$\Delta n = -\frac{1}{2} n_0^3 \left[p_{12} - \nu(p_{11} + p_{12}) \right] \varepsilon \qquad (14)$$

and substituting into equation (7),

$$\begin{aligned} \Delta\varphi &= \frac{2\pi}{\lambda} n_0 L \varepsilon \left\{ 1 - \frac{1}{2} n_0^2 \left[p_{12} - \nu(p_{11} + p_{12}) \right] \right\} \\ &= \frac{2\pi}{\lambda} n_0 L \varepsilon (1 - p_e) \end{aligned} \qquad (15)$$

where p_e is referred to as the *effective photoelastic constant* for applied axial strain [8]. The phase shift is therefore linearly proportional to the change in length of the optical fiber, or the applied strain.

Adding the effect of a temperature change ΔT to equation (15), where the axial elongation ε is now the total axial elongation, we find

$$\begin{aligned} \Delta\varphi = \frac{2\pi}{\lambda} n_0 L \Bigg\{ &\varepsilon - \frac{1}{2} \varepsilon n_0^2 \left[p_{12} - \nu(p_{11} + p_{12}) \right] \\ &+ \frac{1}{n_0} \frac{\mathrm{d}n_0}{\mathrm{d}T} \Delta T + \frac{n_0^2}{2} (p_{11} + 2p_{12}) \alpha \Delta T \Bigg\} \end{aligned} \qquad (16)$$

[9]. Typical material property values for fused silica are $\nu = 0.16$ and $\alpha = 0.5 \times 10^{-6}/\,^{\circ}\mathrm{C}$.

The above equations consider a constant strain field applied along the gauge length of the optical fiber. In general, the fiber can be mounted in any orientation on a surface; therefore, these strain components can also vary along the length of the gauge. Sirkis and Haslach [11] first calculated the total phase shift, φ, in the optical fiber due to a known strain field. We define the variable s to be the variable along the path length of the fiber. We find the total phase shift by integrating the local phase shift along the optical fiber

$$\begin{aligned} \varphi &= \int_0^L \left(\frac{\mathrm{d}\varphi}{\mathrm{d}s} \right) \mathrm{d}s \\ &= \left(\frac{2\pi}{\lambda} \right) \int_0^L n_0 (1 - p_e)(1 + \varepsilon_n)\, \mathrm{d}s \end{aligned} \qquad (17)$$

where ε_n is the strain component tangent to the fiber path at each location. Sirkis and Haslach also showed that transfer of shear stress and transverse stresses was negligible for surface-mounted sensors; however, these components could be significant for sensors embedded in materials. Details on the inclusion of these strain components for embedded sensors can be found in [9, 12].

3.2 Conventional interferometers

Interferometric sensors present an extremely high sensitivity to external parameters. Additionally, as the optical fiber itself is used as the measurement device, such sensors are beneficial for simplicity and cost as compared to other optical fiber–based sensors. When using an optical fiber as strain sensor, it is not possible to measure the phase shift directly, and therefore we typically measure the interference between the sensor fiber and a second, reference fiber, which is not exposed to the environmental changes and therefore has a constant phase shift. Additionally, this second fiber can be exposed to only some of the loading, for example, temperature, so as to provide compensation during the measurements. The interferometric measurement of phase shifts for in-fiber sensors parallels that for classical free-space interferometers. The two most commonly applied interferometric arrangements are the Mach–Zehnder and Michelson interferometers [13]. Figures 4 and 5 show each of these classical interferometers and their equivalent for in-fiber sensors.

Assuming that the light source has a high coherence length and polarization effects are compensated between the two fibers, the average intensity of the interference pattern varies sinusoidally, as shown in Figure 6. The cyclic form of the intensity measurement presents two challenges when the intensity is near one of the quadrature points, i.e., where $\mathrm{d}I/\mathrm{d}\varphi = 0$. The first challenge is that the direction of fringe movement cannot be determined at the quadrature point (directional ambiguity) and the second is that the measurement sensitivity goes to zero at the same point (signal fading). A variety of signal-processing techniques have been applied to remove these difficulties, generally categorized into passive and active demodulation. Each of these categories can be further divided on the basis of whether the output signal is divided into multiple branches at the same frequency (homodyne) or different frequencies (heterodyne) [14]. One example of a passive homodyne demodulation is shown in Figure 6. For

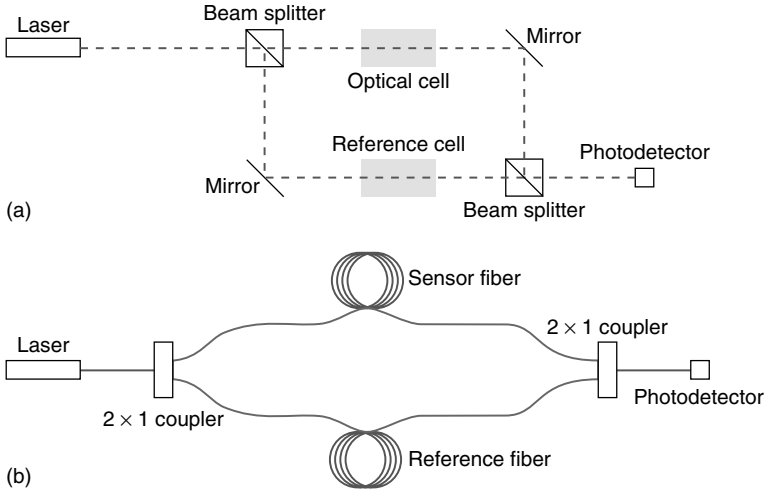


Figure 4. Schematic of Mach–Zehnder interferometer: (a) free-space optics version measures the difference in optical path length through the optical cell and reference cell and (b) in-fiber version measures the difference in optical path length through sensor fiber and reference fiber.

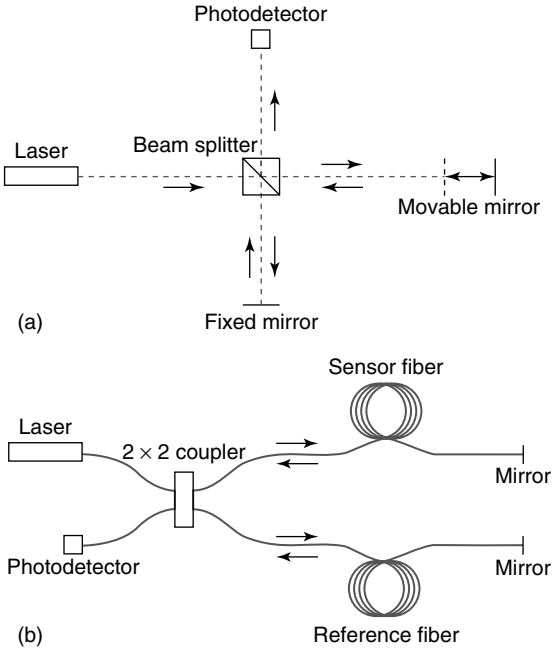


Figure 5. Schematic of Michelson interferometer: (a) free-space optics version and (b) in-fiber version.

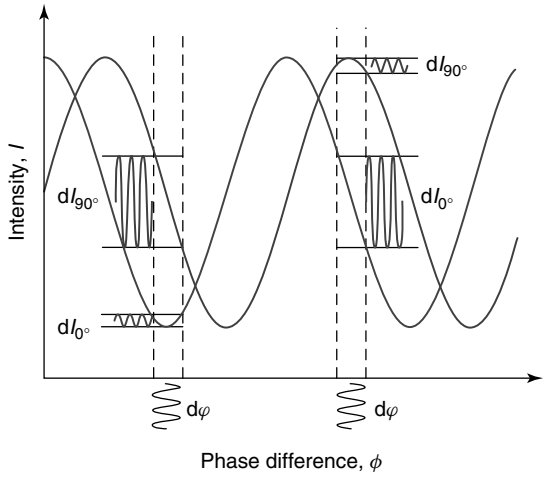


Figure 6. Intensity signals as a function of interferometer phase difference for two sensor signals, phase shifted by 0° and 90°. When one signal is as a quadrature point ($dI/d\varphi = 0$), the other is at the location of maximum sensitivity. [Reproduced from Ref. 14. © John Wiley & Sons Inc., 2006.]

this example, the sensor lightwave is divided into two channels, with one phase shifted by 90°, and interfered with the reference lightwave. As seen in Figure 6, when one signal is near a quadrature point, the other is at the point of maximum sensitivity, removing both the signal fading and directional ambiguity issues. A second popular example of passive demodulation is the use of a 3×3 coupler, exploiting the phase shifts between the various outputs [15]. Active homodyne demodulation involves actively

loading the reference fiber to keep the interferometer signal 90° from the quadrature point (or at the quadrature point for some examples), for example, by wrapping the reference fiber around a piezoelectric cylinder [16]. Kersey [17] reviews multiplexing strategies for interferometric optical-fiber sensors, including time-division multiplexing and frequency-division multiplexing examples.

The measurement of phase shift is not an absolute measurement. Specifically, in order to know the current strain level, the system must be continually operated throughout the lifetime of the structure. This presents significant challenges for monitoring of structures since power interruptions could invalidate all later measurements. One solution to this problem is through low-coherence interferometry, which is described in the following section.

3.3 Low-coherence interferometers

One technique to provide an absolute measurement is through the use of low-coherence interferometry, also referred to as *white-light interferometry*. This measurement technique is particularly useful for large structural applications as the spatial resolution is significantly less than the previous interferometers. However, the system would not be affected by power interruptions that may occur over the lifetime of the structure. Until recently, low-coherence interferometry had also been limited to quasi-static applications; however, it has been demonstrated successfully in many field applications. Bock *et al.* [18] applied low coherence in high-birefringence fibers (*see* **Fiber-optic Sensor Principles**) for the measurement of pressure. Inaudi *et al.* [19] embedded low-coherence interferometric sensors in a concrete structure to measure shrinkage of the concrete during cure and later strain during loading of the structure. Yuan and Ansari [20] applied a similar embedded low-coherence interferometric sensor to measure crack-tip openings in a concrete structure.

The fundamental measurement concept applied to low-coherence interferometric sensors is to interrogate a primary interferometer consisting of a sensor and reference arm (referred to as the *sensing interferometer*) with a second similar interferometer for which the optical path difference between the two arms can be scanned (referred to as the *scanning interferometer*). A typical low-coherence interferometer configuration for optical-fiber sensors is shown in Figure 7. Interference fringes are observed only when the optical path difference in the scanning interferometer is within the coherence length of the laser or when the optical path difference between the two interferometers is within the same coherence length. By selecting a laser source with an extremely low-coherence length such as a superluminescent diode or multimode laser diode, the displacement of the sensing interferometer can be determined within a reasonable precision [21].

The interference between two beams from a low-coherence source can be described by

$$I = I_0\left[1 + V(nx)\cos\left(\frac{2\pi}{\lambda}nx\right)\right] \qquad (18)$$

where nx is the optical path difference in the interferometer, I_0 is the mean intensity, and V is the variation in fringe visibility as a function of optical path difference for the laser source (typically a Gaussian profile) [22]. This equation yields a wave packet whose width is approximately equal to the coherence length and centered around the condition $nx = 0$, as seen in Figure 8. For the interferometer of Figure 7, the optical path difference in the sensing interferometer to be measured is ΔL_1. The movable mirror in the scanning interferometer is displaced while the interference pattern between the two interferometers is measured. When $\Delta L_2 \cong 0$, the strongest interference occurs (as seen in Figure 8), and this condition is used as a reference point. For $\Delta L_2 \cong +\Delta L_1$, the phase imbalance in the sensing interferometer is offset by the phase imbalance of the scanning interferometer for a portion of the lightwave. Thus, interference also occurs at a lower maximum intensity. From these two measurements, ΔL_1 can be determined, which is the optical path difference in the original sensing interferometer. As in the previous section, this optical path difference can then be converted into the physical change in length of the sensing fiber, ΔL, through [19]

$$\Delta L = \frac{\Delta L_1}{n_0(1 - p_e)} \qquad (19)$$

Inaudi *et al.* [19] demonstrated a low-coherence optical-fiber sensor using a LED with a coherence length of 30 μm. By calculating the center of

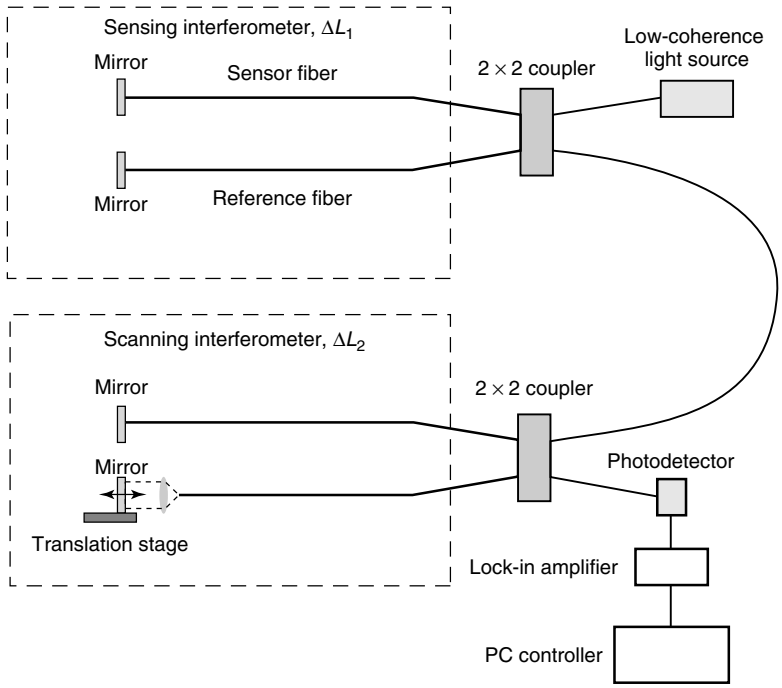


Figure 7. Typical low-coherence in-fiber interferometer configuration. [Reproduced with permission from Ref. 19. © Elsevier, 1994.]

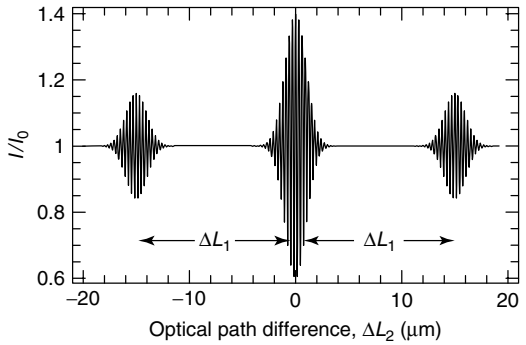


Figure 8. Interference signal from double Michelson interferometer with low-coherence source shown in Figure 7. [Reproduced with permission from Ref. 19. © Elsevier, 1994.]

gravity of the interference pattern, they were able to increase the displacement resolution to 10 μm. Rao and Jackson [23] provide an excellent review of signal processing for low-coherence interferometric sensors, as well as methods to increase the accuracy of the maximum point on the interference envelope. Meggitt *et al.* [22] replaced the translational stage for the scanning mirror by wrapping an optical fiber around a piezoelectric cylinder. Therefore, the optical path difference could be scanned by applying an electrical current to the cylinder, without the need for externally moving components.

To increase the spatial resolution of the basic low-coherence interferometer for structural measurements, Inaudi [24] implemented partial reflectors along the sensing arm to provide multiple interference lengths. By placing the reflectors sufficiently apart and scanning all possible interference distances, it is possible to use the same interrogation system for multiple sensors. A typical output from this system, demonstrating the interference lengths for each sensor, is shown in Figure 9. Later, Yuan and Ansari [25] multiplexed several sensing lengths using partial reflectors and a low-coherence source; however, they applied N balance arms for N sensors and switched between the balance arms to measure each sensor. Most recently, Lloret *et al.* [26] applied amplitude modulation of the low-coherence light source at high frequencies to obtain dynamic strain measurements up to 100 Hz.

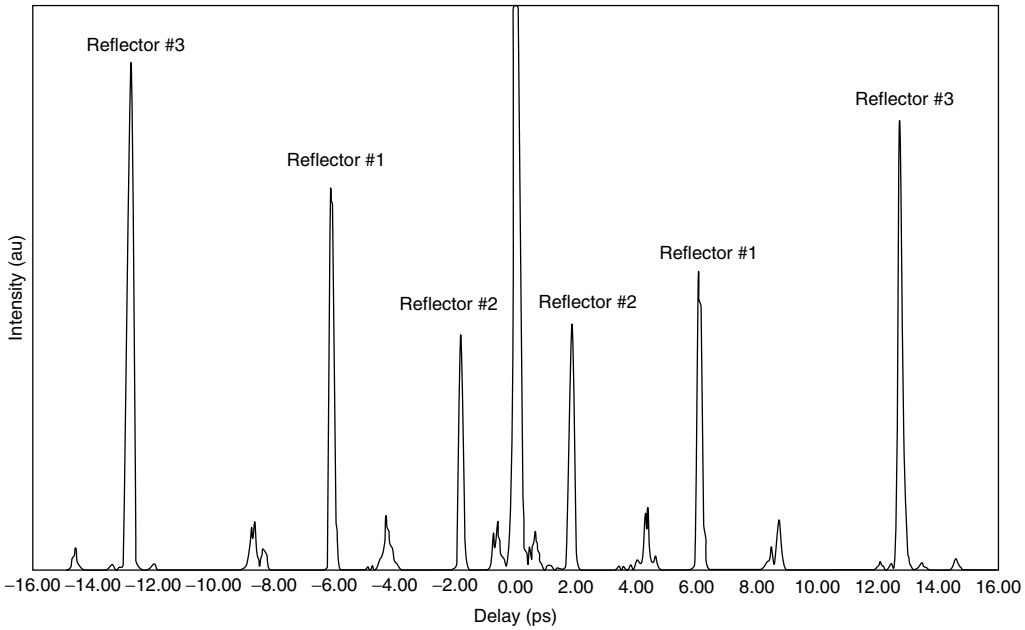


Figure 9. Interference signal from multiplexed low-coherence interferometric sensors (only envelope of demodulation signal is plotted). [Reproduced with permission from Ref. 24. © SPIE, 1995.]

3.4 Fabry–Perot interferometers

A second absolute interferometer configuration is the Fabry–Perot interferometer (FPI), first demonstrated by Murphy *et al.* [27]. FPI sensors provide local strain and temperature measurements. FPI sensors have been surface mounted on composites for the measurement of acoustic emission signals [28] and impact detection [29]. They have also been embedded in composite structures for the detection and measurement of fatigue crack propagation in bonded composite patch repairs [30], delamination and buckling [31], process-induced residual stresses [32], and the development of impact damage [33]. They have also been combined with fiber Bragg grating sensors for the independent measurement of strain and temperature [34] (*see* **Fiber Bragg Grating Sensors**). The miniaturization of FPI sensors is also discussed in **Novel Fiber-optic Sensors**.

A schematic of a Fabry–Perot in an optical fiber is shown in Figure 10(a). The sensor cavity of length L is formed between two partial reflectors with reflectivities r_1 and r_2. These partial reflectors are typically partial mirrors, but fiber Bragg gratings have also been used [35]. For high reflectivity values, the FPI acts as a "multipass" interferometer whose transmission characteristics are wavelength dependent. We can assume that no losses are incurred at the reflectors so that the transmissivities $t_1 = 1 - r_1$ and $t_2 = 1 - r_2$. The cavity itself can be a section of the optical fiber (as shown in Figure 10a) for which the sensor is termed *intrinsic*. Extrinsic FPIs can be formed using external air gaps or other mediums to be analyzed. An example of an extrinsic Fabry–Perot interferometer (EFPI) is shown in Figure 10(b). Extrinsic FPI sensors are larger than intrinsic sensors and therefore more intrusive; however, they are easier to fabricate and have less inherent noise. They also demonstrate low transverse strain sensitivities and low apparent thermal strains [36].

To calculate the response of the FPI, we define the index of refraction of the cavity medium as n. Considering a plane wave (arriving from the left in Figure 10a) that has passed through the first reflector, we can write its time-averaged (or steady-state) amplitude as $E = E_0 \exp(\mathrm{i}\beta z)$ (*see* **Fiber-optic Sensor Principles**). When this wave reaches the second partial reflector $E = E_0 \exp(\mathrm{i}\beta L)$, the transmitted and reflected portions of the waves are

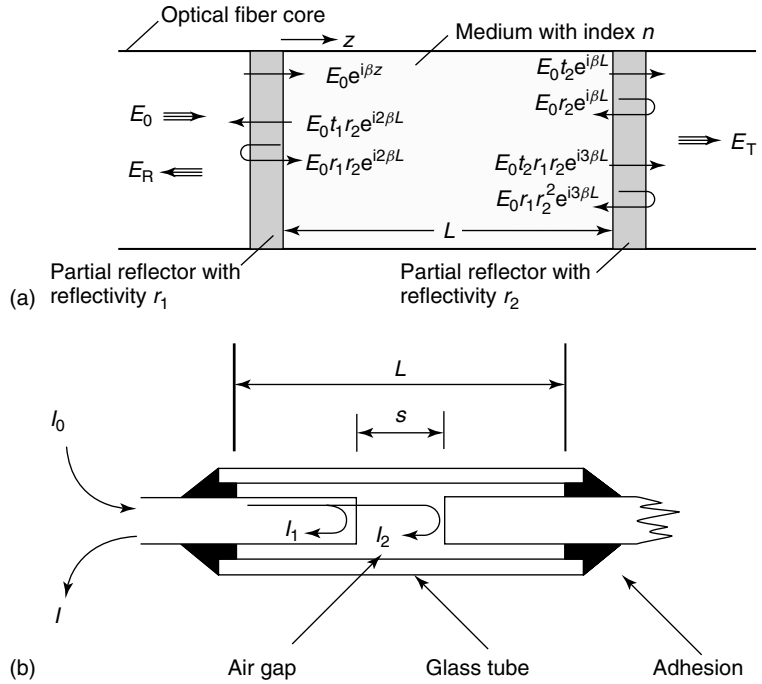


Figure 10. (a) Representation of multipass interference in a Fabry–Perot interferometer and (b) extrinsic Fabry–Perot interferometer. [Reproduced from Ref. 31. © Sage Publications, 2000.]

$E_t = t_2 E_0 \exp(i\beta L)$ and $E_r = r_2 E_0 \exp(i\beta L)$ respectively. The reflected portion then reaches the first partial reflector, at which point $E = r_2 E_0 \exp(i2\beta L)$. The reflected portion of this wave reaches the second reflector, at which point $E_t = t_2 r_1 r_2 E_0 \exp(i2\beta L)$ and $E_r = r_1 r_2^2 E_0 \exp(i2\beta L)$.

Continuing these calculations for an infinite number of passes, we find the total lightwave exiting the FPI to the right,

$$\begin{aligned} E_T &= t_2 E_0 \exp(i\beta L)\left[1 + r_1 r_2 \exp(2i\beta L)\right. \\ &\quad \left. + r_1^2 r_2^2 \exp(4i\beta L) + \cdots\right] \\ &= \frac{E_0 t_2 \exp(i\beta L)}{1 - r_1 r_2 \exp(2i\beta L)} \end{aligned} \tag{20}$$

We could also calculate the total lightwave reflected from the cavity, $E_R = E_0 - E_T$. The intensity of the transmitted lightwave is then

$$I_T = \frac{1}{2}\varepsilon_0 c |E_T|^2 = \frac{I_0 t_2^2}{(1-R)^2[1 + \mathcal{F}\sin^2(\beta L)]} \tag{21}$$

where ε_0 is the permittivity of the vacuum ($\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$), c is the speed of light in a vacuum, $R^2 = r_1 r_2$, and $I_0 = \varepsilon_0 c E_0^2/2$ is the intensity of the input lightwave. The coefficient $\mathcal{F}$ is referred to as the *coefficient of finesse* of the FPI and determines the response characteristics of the FPI. $\mathcal{F}$ is given by the expression

$$\mathcal{F} = \frac{4R}{(1-R)^2} \tag{22}$$

Figure 11 plots the normalized transmitted intensity of the FPI as a function of wavelength. The peaks in the transmission spectrum correspond to the condition $L = m\lambda/2n$, where m is an integer. By measuring the wavelength location of one of these peaks, one can determine the length of the cavity and therefore the applied strain, displacement, pressure, or temperature. For this measurement strategy, it is best to have high finesse values to narrow the peaks as much as possible [37]. Spectral measurements provide absolute strain or temperature measurements without the problems of directional ambiguity and signal fading previously described.

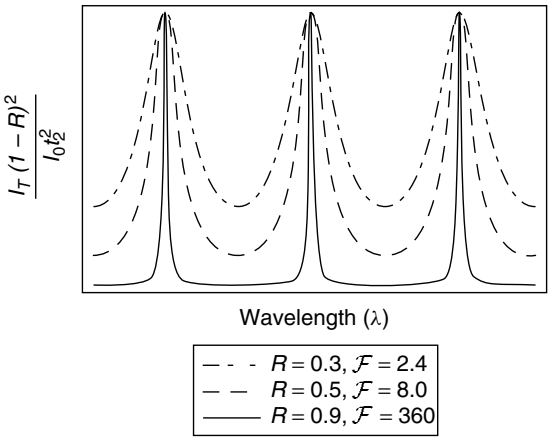


Figure 11. Typical normalized transmitted intensity spectrum for a Fabry–Perot interferometer. Three different finesse values are plotted.

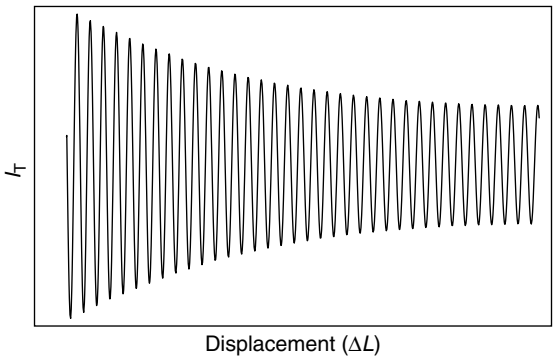


Figure 12. Transmitted intensity for a low-finesse Fabry–Perot interferometer as a function of displacement of one of the reflectors ($R = 0.1, \mathcal{F} = 0.49$).

High-speed interrogation of spectral sensors is discussed in **Fiber Bragg Grating Sensors**. Rao [38] presents a review of multiplexing and signal interrogation of FPI sensors.

Alternatively, low-finesse FPI sensors can also be applied for high-resolution, dynamic strain sensing [27]. For very low values of reflectivity (and finesse), the FPI acts as essentially a two-beam interferometer. A plot of the response of a low-finesse FPI as a function of the displacement of one of the mirrors is shown in Figure 12. For small displacements, this sensor can be interrogated similar to the Mach–Zehnder or Michelson configurations previously described. Various signal interrogation methods have been applied to remove the direction ambiguity including quadrature phase shifted signals [27], path-matched differential interferometry [39], and low-coherence interferometry [40, 41].

4 SCATTERING-BASED OPTICAL-FIBER SENSORS

A final property of optical fibers that has been exploited for the measurement of strain and temperature is the intrinsic scattering property of the fused silica material. As a lightwave propagates through the optical fiber, an extremely small portion is backscattered owing to local inhomogeneities in the material and therefore the index of refraction. Several forms of scattered waves can be detected, including Rayleigh, Brillouin, and Raman components. Figure 13 shows a typical scattering spectrum for an optical fiber. As can be seen, the Rayleigh component is the largest component of the scattering spectrum. Brillouin scattering occurs due to interactions with acoustic waves known as *phonons* and is therefore frequency shifted (through the Doppler effect) as compared to the Rayleigh scattering component. This frequency shift is dependent upon the local density of the fused silica and is therefore linearly related to applied strain or temperature [42]. The Brillouin backscattered light is an extremely weak signal and difficult to measure outside of the laboratory; however, this scattering can be amplified by stimulating acoustic waves through the use of a second pulsed laser connected to the opposite end of the optical fiber. By scanning the frequency of the pulsed laser, the relative frequency shifts at which the stimulated

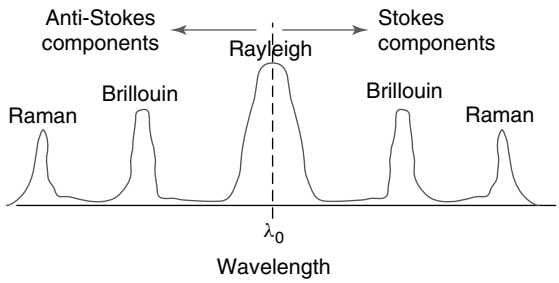


Figure 13. Spectrum of intrinsic material scattering for fused silica. Stokes and anti-Stokes components are indicated.

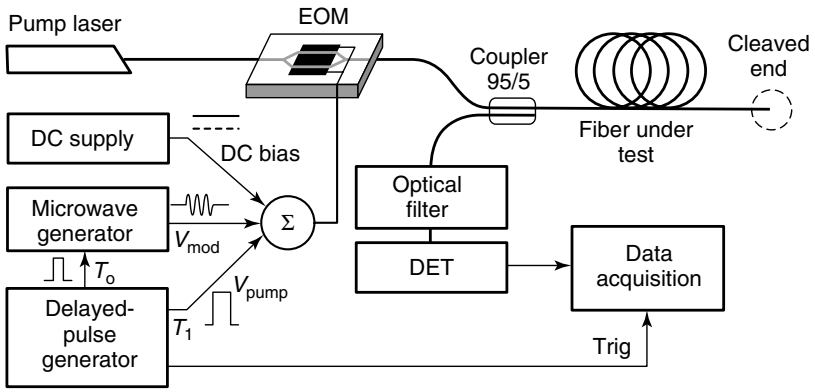


Figure 14. Experimental setup for distributed Brillouin gain spectrum analysis. [Reproduced with permission from Ref. 43. © 1996, OSA Publishing.]

Brillouin scattering occurs can be identified and converted into the applied strain or temperature. To determine the location of each Brillouin scattering event, the time of arrival of the scattered light is interrogated. In this manner, the strain or temperature distribution along an optical fiber can be measured. A schematic of a typical stimulated Brillouin scattering data-acquisition system is shown in Figure 14. Bao *et al.* [42] first demonstrated the use of stimulated Brillouin scattering as a temperature and strain sensor, achieving a strain and temperature resolution of $20\,\mu\varepsilon$ and $2\,^{\circ}\mathrm{C}$ with a spatial resolution of 5 m, over a sensing length of 22 km. Later, Niklès *et al.* [43] improved the strain and temperature resolution to $5\,\mu\varepsilon$ and $0.25\,^{\circ}\mathrm{C}$ using a single laser source, however, with a spatial resolution of 35 m.

Brillouin scattering–based optical-fiber sensors provide a unique distributed sensing capability using only the intrinsic properties of the optical fiber itself. Generally, such sensors have been applied for the monitoring of large-scale structures due to the limits on the spatial resolution. Example applications include composite cure sensing [44], monitoring of steel beam and composite fuselage structures [45, 46], and damage identification in sandwich composites [47]. At the same time, researchers have recently applied signal-processing techniques and hardware configurations to reduce the spatial resolution substantially. For example, Song *et al.* [48] recently demonstrated a spatial resolution of 1.6 mm, applying Brillouin optical correlation domain analysis.

5 CONCLUSIONS

As reviewed in this chapter, optical fibers can be applied as strain and temperature sensors through their intrinsic properties. Intensity-based optical-fiber sensors offer low-cost, simple solutions to complete optical measurements. For more precise measurements, a variety of interferometer configurations and interrogation methods are also available, including variations for high-speed and absolute measurements. Finally, low-coherence interferometry and stimulated Brillouin scattering–based sensors permit distributed measurements for large structural applications. As compared to their electrical counterparts, optical-fiber sensors can require more complicated and expensive instrumentation. On the other hand, they can be readily multiplexed into large sensor networks with a single data-acquisition system and can therefore be less obtrusive to the structure. Additionally, the distributed examples discussed in this article emphasize that simple optical fibers can be used to instrument large structures without the need to monitor a large number of sensors.

RELATED ARTICLES

Lamb Wave-based SHM for Laminated Composite Structures

Reliable Use of Fiber-optic Sensors

REFERENCES

[1] Measures RM, Glossop NDW, Lymer J, LeBlanc M, West J, Dubois S, Tsaw W, Tennyson RC. Structurally integrated fiber optic damage assessment system for composite-materials. *Applied Optics* 1989 **28**:2626–2633.

[2] LeBlanc M, Measures RM. Impact damage assessment in composite-materials with embedded fiber-optic sensors. *Composites Engineering* 1992 **2**:573–596.

[3] Measures RM. *Structural Monitoring with Fiber Optic Technology*. Academic Press: San Diego, CA, 2001.

[4] Lagakos N, Bucaro JA. Fiber optic microbend sensor. *ISA Transactions* 1998 **27**:19–24.

[5] Donlagic D, Culshaw B. Microbend sensor structure for use in distributed and quasi-distributed sensor systems based on selective launching and filtering of the modes in graded index multimode fiber. *Journal of Lightwave Technology* 1999 **17**:1856–1868.

[6] Xie GP, Keey SL, Asundi A. Optical time-domain reflectometry for distributed sensing of the structural strain and deformation. *Optics and Lasers in Engineering* 1999 **32**:437–447.

[7] Pandey NK, Yadav BC. Embedded fibre optic microbend sensor for measurement of high pressure and crack detection. *Sensors and Actuators, A* 2006 **128**:33–36.

[8] Butter CD, Hocker GB. Fiber optics strain-gauge. *Applied Optics* 1978 **17**:2867–2869.

[9] Van Steenkiste RJ, Springer GS. *Strain and Temperature Measurement with Fiber Optic Sensors*. Technomic Publishing: Lancaster, PA, 1997.

[10] Nye JF. *Physical Properties of Crystals*. Oxford Science Publications: Oxford, 1985.

[11] Haslach HW, Sirkis JS. Surface-mounted optical fiber strain sensor design. *Applied Optics* 1991 **30**:4069–4080.

[12] Peters KJ, Washabaugh PD. Balance technique for monitoring *in situ* structural integrity of prismatic structures. *American Institute of Aeronautics and Astronautics Student Journal* 1997 **35**:869–874.

[13] Rogers AJ. Essential optics. In *Optical Fiber Sensors: Principles and Components*, Dakin J, Culshaw B (eds). Artech House: Boston, MA, 1988.

[14] Dandridge A. Fiber optic sensors based on the Mach-Zehnder and Michelson interferometers. In *Fiber Optic Sensors*, Udd E (ed). John Wiley & Sons: Hoboken, NJ, 2006.

[15] Koo KP, Tveten AB, Dandridge A. Passive stabilization scheme for fiber interferometers using (3x3) fiber directional couplers. *Applied Physics Letters* 1982 **41**:616–618.

[16] Jackson DA, Priest R, Dandridge A, Tveten AB. Elimination of drift in a single-mode optical fiber interferometer using a piezoelectrically stretched coiled fiber. *Applied Optics* 1980 **19**:2926–2929.

[17] Kersey AD. Distributed and multiplexed fiber optic sensors. In *Fiber Optic Sensors*, Udd E (ed). John Wiley & Sons: Hoboken, NJ, 2006.

[18] Bock WJ, Urbanczyk W, Wojcik J, Beaulieu M. White-light interferometric fiber-optic pressure sensor. *IEEE Transactions on Instrumentation and Measurement* 1995 **44**:694–697.

[19] Inaudi D, Elamari A, Pflug L, Gisin N, Breguet J, Vurpillot S. Low-coherence deformation sensors for the monitoring of civil-engineering structures. *Sensors and Actuators, A* 1994 **44**:125–130.

[20] Yuan L, Ansari F. Embedded white light interferometer fibre optic strain sensor for monitoring crack-tip opening in concrete beams. *Measurement Science and Technology* 1998 **9**:261–266.

[21] Liu T, Brooks D, Martin A, Badcock R, Ralph B, Fernando GF. A multi-mode extrinsic Fabry-Perot interferometric strain sensor. *Smart Materials and Structures* 1998 **7**:550–556.

[22] Meggitt BT, Hall CJ, Weir K. An all fibre white light interferometric strain measurement system. *Sensors and Actuators, A* 2000 **79**:1–7.

[23] Rao YJ, Jackson DA. Recent progress in fibre optic low-coherence interferometry. *Measurement Science and Technology* 1996 **7**:981–999.

[24] Inaudi D. Coherence multiplexing of in-line displacement and temperature sensors. *Optical Engineering* 1995 **34**:1912–1915.

[25] Yuan L, Ansari F. White-light interferometric fiber-optic distributed strain-sensing system. *Sensors and Actuators, A* 1997 **63**:177–181.

[26] Lloret S, Rastogi P, Thévenaz L, Inaudi D. Measurement of dynamic deformations using a path-unbalance Michelson-interferometer-based optical fiber sensing device. *Optical Engineering* 2003 **42**:662–669.

[27] Murphy KA, Gunther MF, Vengsarkar AM, Claus RO. Quadrature phase-shifted, extrinsic Fabry-Perot optical fiber sensors. *Optics Letters* 1991 **16**:273–275.

[28] Duke JC, Cassino CD, Childers BA, Prosser WH. Characterization of an extrinsic Fabry-Perot interferometric acoustic emission sensor. *Materials Evaluation* 2003 **61**:935–940.

[29] Akhavan F, Watkins SE, Chandrashekhara K. Prediction of impact contact forces of composite plates using fiber optic sensors and neural networks. *Mechanics of Composite Materials and Structures* 2000 **7**:195–205.

[30] Seo DC, Kwon IB, Lee JJ. Fatigue crack growth monitoring by optical fiber sensors in smart composite patch repairs. *Key Engineering Materials* 2006 **321–323**:286–289.

[31] Park JW, Ryu CY, Kang HK, Hong CS. Detection of buckling and crack growth in the delaminated composites using fiber optic sensor. *Journal of Composite Materials* 2000 **34**:1602–1623.

[32] Lawrence CM, Nelson DV, Bennett TE, Spingarn JR. An embedded fiber optic sensor method for determining residual stresses in fiber-reinforced composite materials. *Journal of Intelligent Material Systems and Structures* 1998 **9**:788–799.

[33] Liu TY, Cory J, Jackson DA. Partially multiplexing sensor network exploiting low coherence interferometry. *Applied Optics* 1993 **32**:1100–1103.

[34] Ferreira LA, Ribeiro ABL, Santos JL, Farahi F. Simultaneous measurement of displacement and temperature using a low finesse cavity and a fiber Bragg grating. *Journal of Lightwave Technology* 1996 **11**:1519–1521.

[35] Legoubin S, Douay M, Bernage P, Niay P, Boj S, Delevaque E. Free spectral range variations of gratin-based Fabry-Perot filters photowritten in optical fibers. *Journal of the Optical Society of America A* 1995 **12**:1687–1694.

[36] Sirkis JS, Brennan DD, Putman MA, Berkoff TA, Kersey AD, Friebele EJ. In-line fiber étalon for strain measurement. *Optics Letters* 1993 **18**:1973–1975.

[37] Bhatia V, Murphy KA, Claus RO, Tran TA, Greene JA. Recent developments in optical-fiber-based extrinsic Fabry-Perot interferometric strain sensing technology. *Smart Materials and Structures* 1995 **4**:246–251.

[38] Rao YJ. Recent progress in fiber-optic extrinsic Fabry-Perot interferometric sensors. *Optical Fiber Technology* 2006 **12**:227–237.

[39] Sirkis J, Berkoff TA, Jones RT, Singh H, Kersey AD, Friebele EJ, Putnam MA. In-line fiber etalon (ILFE) fiber-optic strain sensors. *Journal of Lightwave Technology* 1995 **13**: 1256–1263.

[40] Bhatia V, Schmid CA, Murphy KA, Claus RO, Tran TA, Greene JA, Miller MS. Optical fiber sensing technique for edge-induced and internal delamination detection in composites. *Smart Materials and Structures* 1995 **4**:164–169.

[41] Chang CC, Sirkis J. Absolute phase measurement in extrinsic Fabry-Perot optical fiber sensors using multiple path match conditions. *Experimental Mechanics* 1997 **37**:26–32.

[42] Bao X, Webb DJ, Jackson DA. Combined distributed temperature and strain sensor based on Brillouin loss in an optical fiber. *Optics Letters* 1994 **19**:141–143.

[43] Niklès M, Thévenaz L, Robert PA. Simple distributed fiber sensor based on Brillouin gain spectrum analysis. *Optics Letters* 1996 **21**:758–760.

[44] Bao X, Huang C, Zeng X, Arcand A, Sullivan P. Simultaneous strain and temperature monitoring of the composite cure with a Brillouin-scattering-based distributed sensor. *Optical Engineering* 2002 **41**:1496–1501.

[45] Bao X, DeMerchant M, Brown A, Bremner T. Tensile and compressive strain measurement in the lab and field with the distributed Brillouin scattering sensor. *Journal of Lightwave Technology* 2001 **19**:1698–1704.

[46] Yari T, Nagai K, Takeda N. Aircraft structural-health monitoring using optical fiber distributed BOTDR sensors. *Advanced Composite Materials* 2004 **13**:17–26.

[47] Murayama H, Kageyama K, Naruse H, Shimada A. Distributed strain sensing from damaged composite materials based on shape variation of the Brillouin spectrum. *Journal of Intelligent Material Systems and Structures* 2004 **15**:17–25.

[48] Song KY, He Z, Hotate K. Distributed strain measurement with millimeter-order spatial resolution based on Brillouin optical correlation domain analysis. *Optics Letters* 2006 **31**: 2526–2528.

Chapter 55
Piezoelectric Wafer Active Sensors

Lingyu Yu and Victor Giurgiutiu
University of South Carolina, Columbia, SC, USA

1 INTRODUCTION

In recent years, piezoelectric wafer active sensors (PWAS) permanently attached to the host structures have been used for guided-wave generation and detection during the structural health monitoring (SHM) process. PWAS are inexpensive transducers that operate on the piezoelectric principle. PWAS are used for SHM employing the following three methods: (i) modal analysis and transfer function; (ii) electromechanical (E/M) impedance; and (iii) wave propagation. The use of PWAS for high-frequency local modal sensing with the E/M impedance method as well as for damage detection with Lamb wave propagation has been pursued by many researchers [1]. PWAS are no more expensive than conventional high-quality resistance strain gauges. However, PWAS performance by far exceeds that of conventional resistance strain gauges because they can be used as active interrogators. PWAS can be used in high-frequency applications at hundreds of kilohertz and beyond. In summary, PWAS can be used in several ways:

1. Piezoelectric resonator
A PWAS has the property of performing mechanical resonances under direct electrical excitation; thus, very precise frequency standards can be created with a simple setup consisting of the PWAS and a signal generator. The resonant frequencies depend only on the wave speed in the PWAS material and the geometric dimensions. Precise frequency values can be obtained through machining the PWAS geometry.

2. High bandwidth strain exciter and detector
As an exciter, a PWAS directly converts electrical energy into mechanical energy; thus, it can easily induce vibrations and waves in the substrate material. PWAS acts very well as an embedded generator of waves and vibration. High-frequency waves and vibrations are easily excited with input signals as low as 10 V. As a sensor, PWAS directly converts

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

mechanical energy to electrical energy. The conversion effectiveness increases with the signal frequency. In the kilohertz range, signals of the order of hundreds of millivolt are easily obtained. No conditioning amplifiers are needed; the PWAS can be directly connected to a high-impedance measurement instrument, such as a digitizing oscilloscope. These dual-sensing and excitation characteristics of PWAS justify their name of "active sensors". A particularly fruitful application is the PWAS phased-array approach, which uses steered guided waves to effectively scan large areas of thin-wall structures from a single location.

3. High-frequency modal sensor

The PWAS can directly measure the high-frequency modal spectrum of the support structure. This is achieved with the E/M impedance method, which reflects the mechanical impedance of the support structure into the real part of the electrical impedance measured at PWAS terminals. The high-frequency characteristic of this method, which has been proven to operate at hundreds of kilohertz and beyond, cannot be achieved with conventional modal instrumentation techniques. Thus, PWAS are the sensors of choice for high-frequency modal measurement and analysis.

2 PWAS PRINCIPLES

This section addresses the behavior of a bonded (constrained) PWAS when excited by an alternating electric voltage. As shown in Figure 1(a), PWAS are small and unobtrusive and utilize the coupling between in-plane strain and transverse electric field. A 5 mm × 5 mm × 0.2 mm square PWAS weighs about 50 mg and costs about $10.

2.1 1D PWAS analysis

Consider a PWAS of length l_a, width b_a, and thickness t_a that is undergoing longitudinal expansion (u_1) induced by the thickness polarization electric field (E_3), as shown in Figure 1(b). The electric field is produced by the application of a harmonic voltage, $V(t) = \hat{V}e^{j\omega t}$, between the top and bottom surfaces (electrodes). The resulting electric field, $E_3 = V/t$, is assumed to be uniform over the PWAS. Assume that h, b, l have widely separated values ($h \ll b \ll l$) such that the length, width, and thickness motions are practically uncoupled. The motion predominantly in the longitudinal direction (x_1) will be considered (1D assumption). PWAS operate on the piezoelectric principle that couples the electrical and mechanical variables in the material in the form

$$S_1 = s_{11}^E T_1 + d_{31} E_3 \tag{1}$$

$$D_3 = d_{31} T_1 + \varepsilon_{33}^T E_3 \tag{2}$$

where S_1 is the strain, T_1 is the stress, D_3 is the electrical displacement (charge per unit area), s_{11}^E is the mechanical compliance at zero field, ε_{33}^T is the dielectric constant at zero stress, and d_{31} is the induced strain coefficient (mechanical strain per unit electric field). When PWAS is bonded to the structure, the structure constrains the PWAS motion with a structural stiffness (k_{str}); that is to say, PWAS are elastically constrained as shown in Figure 1(b). In this model, the overall structural stiffness applied to the PWAS is split into two equal components, $2k_{\text{str}}$ each, applied to the ends of the PWAS, such that

$$k_{\text{total}} = \left[\left(2k_{\text{str}}\right)^{-1} + \left(2k_{\text{str}}\right)^{-1}\right]^{-1} = k_{\text{str}} \tag{3}$$

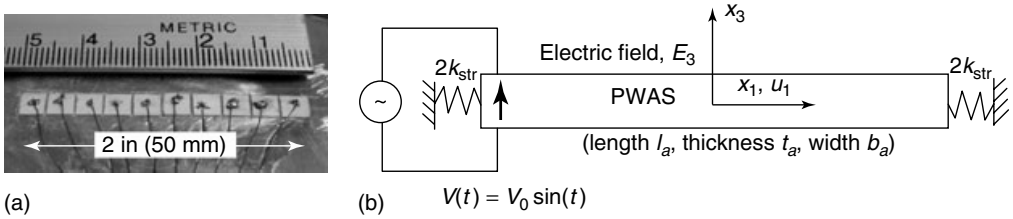


Figure 1. Embedded piezoelectric wafer active sensors: (a) a phased array composed of 10 PWAS, each 5-mm square, permanently attached to the host structure and (b) a one-dimensional bonded PWAS under electric excitation, constrained by structural stiffness k_{str}.

Note that the effective structural stiffness, k_{str}, is a frequency-dependent complex quantity reflecting the structural dynamics. The boundary conditions applied at the ends of the PWAS balance the resulting stress ($T_1 b_a t_a$) with the spring reaction force ($2k_{\text{str}}u_1$)

$$T_1\left(\frac{l_a}{2}\right)b_a t_a = -2k_{\text{str}}u_1\left(\frac{l_a}{2}\right),$$
$$T_1\left(-\frac{l_a}{2}\right)b_a t_a = 2k_{\text{str}}u_1\left(-\frac{l_a}{2}\right) \tag{4}$$

Using the Newton's law of motion ($T_1' = \rho\ddot{u}_1$) and the strain-displacement relation ($S_1 = u_1'$), the axial wave equation can be obtained:

$$\ddot{u}_1 = c_a^2 u_1'' \tag{5}$$

2.1.1 PWAS responses

Introduce the notations: induced strain $S_{\text{ISA}} = d_{31}\hat{E}_3$, displacement $u_{\text{ISA}} = S_{\text{ISA}}l = (d_{31}\hat{E}_3)l$; wave number γ, $\gamma = \omega/c$; wavelength λ, $\lambda = cT = c/f$, $f = \omega/2\pi$, quasi-static PWAS stiffness $k_{\text{PWAS}} = A_a/s_{11}^E l_a$, piezoelectric material wave speed $c_a^2 = 1/\rho s_{11}^E$, and $\frac{\partial}{\partial x_1}(\,) = (\,)'$, $\frac{\partial}{\partial t}(\,) = \dot{(\,)}$, the general solution $u_1(x,t) = \hat{u}(x)e^{j\omega t}$ of equation (5) can be obtained as

$$\hat{u}(x) = \frac{1}{2}u_{\text{ISA}}\frac{\sin\gamma x}{\gamma l\cos(\gamma l/2)/2 + \gamma\sin(\gamma l/2)} \tag{6}$$

if the system determinant is nonzero ($\Delta \neq 0$)

$$\Delta = \begin{vmatrix} \phi\cos\phi + r\sin\phi & -(\phi\sin\phi - r\cos\phi) \\ \phi\cos\phi + r\sin\phi & (\phi\sin\phi - r\cos\phi) \end{vmatrix}$$
$$= 2(\phi\cos\phi + r\sin\phi)(\phi\sin\phi - r\cos\phi) \tag{7}$$

The electrical responses of PWAS under harmonic electric excitation can be represented as the admittance (Y), which is defined as the ratio between the current and the voltage and the admittance (Z), which is the inverse of admittance.

$$Y = \frac{\hat{I}}{\hat{V}} = j\omega C\left[1 - k_{31}^2\left(1 - \frac{1}{r + \phi\cot\phi}\right)\right] \tag{8}$$

$$Z = \frac{\hat{V}}{\hat{I}} = \frac{1}{j\omega C}\left[1 - k_{31}^2\left(1 - \frac{1}{r + \phi\cot\phi}\right)\right]^{-1} \tag{9}$$

with $\phi = \gamma l/2$, $k_{31}^2 = d_{31}^2/(s_{11}^E\varepsilon_{33}^T)$ (E/M coupling coefficient), $C = \varepsilon_{33}^T b_a l_a/t_a$ (stress-free PWAS capacitance), and r as the structural stiffness ratio. For free PWAS, $r = 0$ ($k_{\text{str}} = 0$) and full constrained PWAS, $r \to \infty$($k_{\text{str}} \to \infty$). When the oscillation frequency is so low that the dynamic effects inside the PWAS are negligible, quasi-static conditions are encountered with $\phi = 0$, i.e., $\gamma l = 0$.

2.1.2 PWAS resonances

The determinant in equation (7) is zero when the first or second parenthesis is zero, corresponding to two types of PWAS resonances, mechanical resonances ($\phi\sin\phi - r\cos\phi = 0$) or E/M resonances ($\phi\cos\phi + r\sin\phi = 0$). The E/M resonances are specific to piezoelectric materials. They reflect the coupling between the mechanical and electrical variables. During E/M resonances, there are two possibilities that arise:

- *Resonance*, when $Y \to \infty$; that is, $Z = 0$. This resonance is associated with the situation in which a device draws very large current when excited harmonically with a constant voltage at a given frequency.
- *Antiresonance*, when $Y = 0$; that is, $Z \to \infty$. This antiresonance is associated with the situation in which a device under constant voltage excitation draws almost no current.

2.2 2D circular PWAS

The 2D constrained PWAS behavior can be analyzed using constitutive equations in cylindrical coordinates (r, θ, z),

$$S_{rr} = s_{11}^E T_{rr} + s_{12}^E T_{\theta\theta} + d_{31}E_z \tag{10}$$
$$S_{\theta\theta} = s_{12}^E T_{rr} + s_{11}^E T_{\theta\theta} + d_{31}E_z \tag{11}$$
$$D_z = d_{31}\left(T_{rr} + T_{\theta\theta}\right) + \varepsilon_{33}^T E_z \tag{12}$$

The wave equation in polar coordinates is

$$\frac{\partial^2 u_r}{\partial r^2} + \frac{1}{r}\frac{\partial u_r}{\partial r} - \frac{u_r}{r^2} = \frac{1}{c_p}\frac{\partial^2 u_r}{\partial t^2} \tag{13}$$

where $c_p = \sqrt{1/[\rho s_{11}^E(1-\nu_a^2)]}$, $\nu_a^2 = s_{12}^E/s_{11}^E$. Equation (13) provides a general solution in terms of Bessel functions of the first kind (J_1) in the form

$$u_r(r,t) = AJ_1\left(\frac{wr}{c}\right)e^{j\omega t} \tag{14}$$

where the coefficient A is determined from the boundary conditions. When PWAS is bonded to a structure, at the boundary $r = r_a$, the boundary condition is

$$T_{rr}\left(r_a\right)t_a = k_{\text{str}}(\omega)\,u_r\left(r_a\right) \tag{15}$$

The electric impedance is calculated as the ratio between voltage and current amplitudes

$$Z(\omega) = \left\{ j\omega C\left(1-k_p^2\right)\left[1 + \frac{k_p^2}{1-k_p^2}\frac{(1+\nu_a)J_1(\phi_a)}{\phi_a J_0(\phi_a) - (1-\nu_a)J_1(\phi_a) - \chi(\omega)(1+\nu_a)J_1(\phi_a)}\right]\right\}^{-1} \tag{16}$$

with $\phi_a = \omega r_a/c$, $\chi(\omega) = k_{\text{str}}(\omega)/k_{\text{PWAS}}$ (dynamic stiffness factor), and $k_p^2 = 2d_{31}^2/[s_{11}^E(1-\nu_a)\varepsilon_{33}^T]$ (planar coupling factor). Equation (16) predicts that the E/M impedance spectrum can be measured by the impedance analyzer at the embedded PWAS terminals during an SHM process and it allows for direct comparison between calculated predictions and experimental results. The structural dynamics is reflected in equation (16) through the dynamic stiffness factor $\chi(\omega)$, which is a measure of the dynamic stiffness of the structure.

3 PWAS ULTRASONIC TRANSDUCERS

For embedded nondestructive evaluation (NDE) applications, PWAS can be used as embedded ultrasonic transducers. PWAS act as both Lamb wave exciters and detectors (Figure 2). PWAS couple their in-plane motion with the particle motion of Lamb waves on the material surface. The in-plane PWAS motion is excited by the applied oscillatory voltage thought the d_{31} piezoelectric coupling.

The PWAS ultrasonic transducer operation is fundamentally different from that of conventional ultrasonic probes

1. PWAS achieve Lamb wave excitation and sensing through surface "pinching" (in-plane strains), while conventional ultrasonic probes excite through surface "tapping" (normal stress).
2. PWAS are strongly coupled with the structure and follow the structural dynamics, while conventional ultrasonic probes are relatively free from the structure and follow their own dynamics.
3. PWAS are nonresonant wideband devices, while conventional ultrasonic probes are narrowband resonators.

Optimum Lamb wave excitation and detection happen when the PWAS length is an odd multiple of the half wavelength of particle wave modes. Geometric tuning can be obtained through matching between

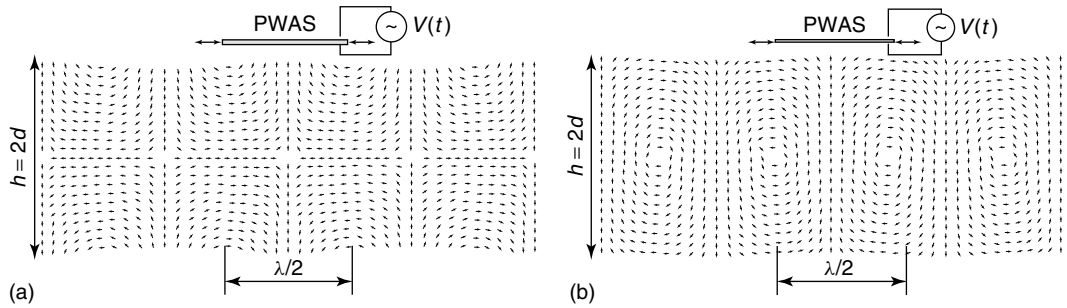


Figure 2. PWAS interaction with Lamb waves in a plate: (a) S_0 mode and (b) A_0 mode.

their characteristic direction and the half wavelength of the excited Lamb wave mode. Circular PWAS excite omnidirectional Lamb waves that propagate in circular wave fronts. Omnidirectional Lamb waves are also generated by square PWAS, though their pattern is somewhat irregular in proximity to the PWAS. At far enough distance ($r \gg a$), the wave front generated by square PWAS is practically identical with that generated by circular PWAS.

3.1 Shear-layer interaction between PWAS and the host structure

The transmission of actuation and sensing between the PWAS and the structure is achieved through the adhesive layer. The adhesive layer acts as a shear layer, in which the mechanical effects are transmitted through shear effects. As shown in Figure 3(a), a thin-wall structure of thickness t and elastic modulus E is attached with a PWAS of thickness t_a and elastic modulus E_a to the upper surface through a bonding layer of thickness t_b and shear modulus G_b. The PWAS length is l_a while the half length is $a = l_a/2$. In addition, $t = 2d$. Upon application of an electric voltage, the PWAS experiences an induced strain of $\varepsilon_{ISA} = d_{31}V/t$. The induced strain is transmitted to the structure through the bonding layer interfacial shear stress (τ). Upon analysis, we obtain the PWAS displacement,

$$u_a(x) = \frac{\alpha}{\alpha+\psi}\varepsilon_{\mathrm{ISA}} a\left(\frac{x}{a} + \frac{\psi}{\alpha}\frac{\sinh\Gamma x}{(\Gamma a)\cosh\Gamma a}\right) \tag{17}$$

the interfacial shear stress in bonding layer,

$$\tau(x) = \frac{t_a}{a}\frac{\psi}{\alpha+\psi}E_a\varepsilon_{\mathrm{ISA}}\left(\Gamma a\frac{\sinh\Gamma x}{\cosh\Gamma a}\right) \tag{18}$$

and the structural displacement at the surface,

$$u(x) = \frac{\alpha}{\alpha+\psi}\varepsilon_{\mathrm{ISA}} a\left(\frac{x}{a} - \frac{\sinh\Gamma x}{(\Gamma a)\cosh\Gamma a}\right) \tag{19}$$

where $\psi = (Et)/(E_a t_a)$, and α is a parameter that depends on the stress and strain distribution across the thickness. Under static and low-frequency dynamic conditions (i.e., uniform stress and strain for the symmetric deformation and linear stress and strain for antisymmetric deformation), elementary analysis yields $\alpha = 4$. For Lamb wave modes, which have complex stress and strain distributions across the thickness, the parameter α varies from mode to mode. These equations apply for $|x| < \alpha$. The shear-lag parameter [$\Gamma^2 = (G_b/E_a)(1/t_a t_b)((\alpha+\psi)/\psi)$] controls the x distribution. The effect of the PWAS is transmitted to the structure through the interfacial shear stress of the bonding layer. A small shear-stress value in the bonding layer produces a gradual transfer of strain from the PWAS to the structure,

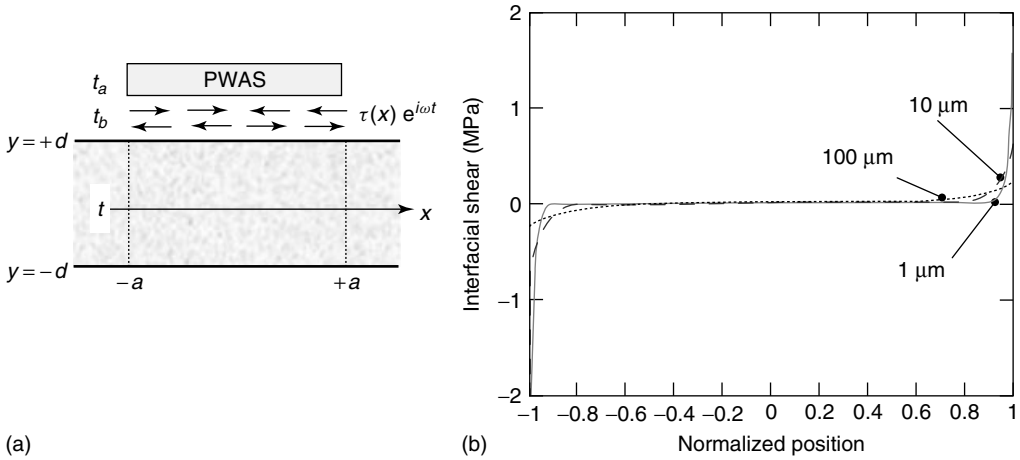


Figure 3. Shear-lag interaction between PWAS and structure: (a) interaction between the PWAS and the structure and (b) variation of shear-lag transfer mechanism with bond thickness.

whereas a large shear stress produces a rapid transfer. Because the PWAS ends are stress free, the build up of strain takes place at the ends, and it is more rapid when the shear stress is more intense. For large value of Γa, the shear transfer process becomes concentrated toward the PWAS ends. Figure 3(b) presents the results of the simulations on an APC-850 PWAS ($E_a = 63\,\text{GPa}$, $t_a = 0.2\,\text{mm}$, $l_a = 7\,\text{mm}$, and $d_{31} = -175\,\text{mm}\,\text{kV}^{-1}$) bonded to a thin-wall aluminum structure ($E = 70\,\text{GPa}$ and $t = 1\,\text{mm}$) using cynoacrylate adhesive ($G_b = 2\,\text{GPa}$) of various thicknesses, $t_b = 1,\ 10,\ 100\,\mu\text{m}$. It reveals that a relatively thick bonding layer produces a low Γa value—a slow transfer over the entire span of the PWAS (the 100-μm curves)—whereas, a very thin bonding layer produces a very rapid transfer (the 1-μm curves) that is confined to the ends, i.e.,

$$\tau(x) = a\tau_0[\delta(x-a) - \delta(x+a)] \tag{20}$$

The shear-lag analysis indicates that in the limit, as $\Gamma a \to \infty$, all the load transfer can be assumed to take place at the PWAS actuator ends. This leads to the concept of ideal bonding, also known as the *pin-force model*, in which all the load transfer takes place over an infinitesimal region at the PWAS ends, and the induced-strain action is assumed to consist of a pair of concentrated forces applied at the ends. The ideal bonding stress is then given by

$$\tau(x) = a\tau_a[\delta(x-a) - \delta(x+a)] \tag{21}$$

Using the shear-lag model, the energy transferred from PWAS to the structure can be found by analyzing either the elastic energy in the structure or work done by the shear stresses at the structural surface.

3.2 PWAS Lamb waves excitation and reception

Lamb waves, also known as *guided plate waves*, are a type of ultrasonic waves and are guided between two parallel free surfaces, such as the upper and lower surfaces of a plate (*see* **Fundamentals of Guided Elastic Waves in Solids**). Lamb waves can exist in two basic types: symmetric (designated as $S_0, S_1, \ldots$.) and antisymmetric (designated as $A_0, A_1, \ldots$.). For each propagation type, a number of modes exist, corresponding to the solutions of the Rayleigh–Lamb equation. The symmetric Lamb waves resemble the axial waves while the antisymmetric Lamb waves resemble the flexural waves. Lamb waves are highly dispersive and their speed depends on the product of frequency and the plate thickness. Details of Lamb wave theory can be found in [1, 2]. Traditional generation of Lamb waves has been through a wedge probe, and modification of the wedge angle and excitation frequency allows the selective tuning of various Lamb wave modes. Another traditional way is to cause excitation through a comb probe, in which the comb pitch is matched with the half wavelength of the targeted Lamb mode. However, both the wedge and comb probes are relatively large and expensive and not appropriate for installation in large numbers in an aerospace structure as part of an SHM system. Being smaller, lighter, more affordable, PWAS could be deployed into the structure being permanently wired and could interrogate at will.

3.2.1 Lamb wave excitation by PWAS

The excitation of Lamb waves with PWAS is studied by considering the excitation applied by the PWAS through a surface stress $\tau = \tau_0(x)e^{j\omega t}$ applied to the upper surface of a plate in the form of shear-lag adhesion stresses over the interval $(-a, +a)$. Applying a space domain Fourier transform analysis of the basic Lamb wave equations to yield the strain wave and displacement wave solutions [3], we have

$$\varepsilon_x(x,t)|_{y=d} = -i\frac{a\tau_0}{\mu}\left[\sum_{\xi^S}\sin\left(\xi^S a\right)\frac{N_S\left(\xi^S\right)}{D_S'\left(\xi^S\right)} \times e^{i\left(\xi^S x-\omega t\right)} + \sum_{\xi^A}\sin\left(\xi^A a\right) \times \frac{N_A\left(\xi^A\right)}{D_A'\left(\xi^A\right)}e^{i\left(\xi^A x-\omega t\right)}\right]$$

$$N_S = \xi\beta\left(\xi^2+\beta^2\right)\cos(\alpha d)\cos(\beta d),$$

$$D_S = \left(\xi^2-\beta^2\right)^2\cos(\alpha d)\sin(\beta d) + 4\xi^2\alpha\beta\sin(\alpha d)\cos(\beta d)$$

$$N_A = \xi\beta\left(\xi^2+\beta^2\right)\sin(\alpha d)\sin(\beta d),$$

$$D_A = \left(\xi^2 - \beta^2\right)^2 \sin(\alpha d)\cos(\beta d) + 4\xi^2\alpha\beta\cos(\alpha d)\sin(\beta d) \qquad (22)$$

where ξ^S and ξ^A are the zeros of D_S and D_A, respectively. We can note that these are the solutions of the Rayleigh–Lamb equation. Raghavan and Cesnik [4] extended these results to the case of a circular transducer coupled with circular-crested Lamb waves and proposed corresponding tuning prediction formulae based on Bessel functions:

$$\varepsilon_r(r,t)|_{z=d} = \pi\frac{\tau_0 a}{\mu}e^{i\omega t}\left[\sum_{\xi^S} J_1\left(\xi^S a\right)\xi^S\frac{N_S\left(\xi^S\right)}{D'_S\left(\xi^S\right)} \times H_1^{(2)}\left(\xi^S r\right) + \sum_{\xi_A} J_1\left(\xi^A a\right)\xi^A \times\frac{N_A\left(\xi^A\right)}{D'_A\left(\xi^A\right)}H_1^{(2)}\left(\xi^A r\right)\right] \qquad (23)$$

3.2.2 PWAS Lamb wave tuning

An important characteristic of PWAS, which distinguishes them from conventional ultrasonic transducers, is their capability of tuning into various guided-wave modes. A comprehensive study of these prediction formulae in comparison with experimental results has recently been performed by Bottai and Giurgiutiu [5]. Simulation plot of the equation (22) is presented in Figure 4(a) using a 7-mm square PWAS installed on 1.07-mm-thick 2024-T3 aluminum alloy plate. Note that the efficient PWAS length for a 7-mm PWAS has been verified as 6.4 mm [5]. Equation (22) contains the $\sin(\xi a)$ behavior that displays maxima when the PWAS length $l_a = 2a$ equals an odd multiple of the half wavelength, and minima when it equals an even multiple of the half wavelength. A complex pattern of such maxima and minima emerges, since several Lamb modes, each with its own different wavelength, coexist at the same time. The plot in Figure 4(a) shows that at 210 kHz, the amplitude of the A_0 mode goes through zero, while that of the S_0 is close to its peak. This represents an excitation "sweet spot" for S_0 Lamb waves. Experimental results confirming this prediction are presented in Figure 4(b).

3.3 PWAS Lamb wave structural health monitoring

As active sensors, PWAS interact directly with the structure and find its state of health and reliability through the use of ultrasonic Lamb waves [6]. Similar to the conventional ultrasonic transducers, PWAS can operate in pitch–catch, pulse–echo, or be wired into phased array to implement structural scanning.

3.3.1 Pitch–catch method

The pitch–catch method detects damage from the changes in the Lamb waves traveling through a

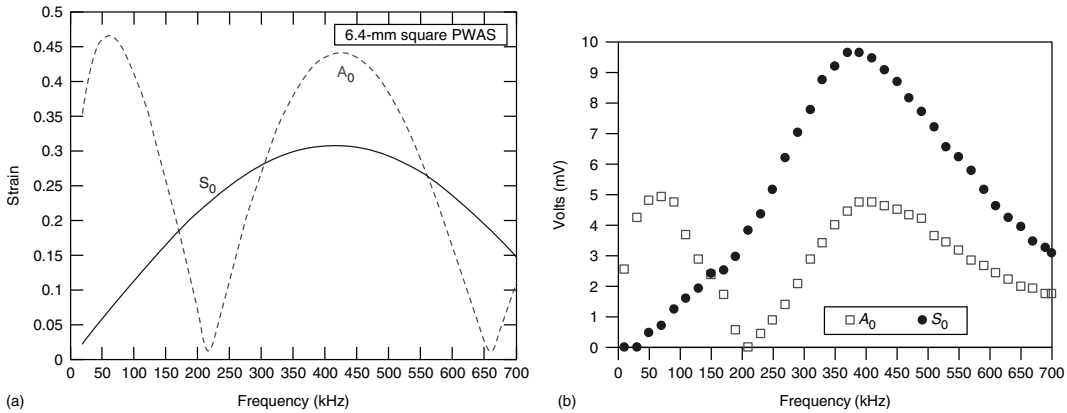


Figure 4. PWAS Lamb wave tuning using a 7-mm square PWAS placed on 1.07-mm 2024-T3 aluminum alloy plate: (a) prediction with equation (22) for 6.4-mm effective sensor length and (b) experimental results.

damaged region. The method uses the transducers in pairs: one as transmitter, and the other as receiver. The pitch–catch SHM method has been used extensively [7–10]. Changes occurring owing to damage in the properties of the transmitted medium induce changes in the signal transmitted through the medium. Of particular interest is the time reversal of the pitch–catch SHM method. In the PWAS time reversal procedure [11], an input signal is reconstructed at an exciter point if an output signal recorded at a receiver point is reversed in time domain and emitted back to the original source point. The reversion is conducted in the frequency domain and depends on the structure transfer function $G(\omega)$. When the reversed signal is reconstructed, it is compared with the original transmission signal to detect the damage occurrence along the transmission–reception line. Figure 5(a) shows the block diagram of PWAS Lamb wave time reversal procedure. With the time reversal method, the original status of the structure is no longer needed for damage detection, so it is known as a *baseline-free* methodology. In PWAS time reversal procedure, it is important to use a single Lamb mode such that the transfer function $G(\omega)$ can remain constant [12]. A simulation with 16-count Hanning window smoothed transmission signal has been conducted in a 1-mm aluminum plate using 7-mm round PWAS. The result of S_0 mode excitation at 290 kHz is shown in Figure 5(b) with noticeable residual waves, while the result of A_0 mode at 30 kHz in Figure 5(c) demonstrating no residual wave packets beside the reconstructed wave.

3.3.2 *Pulse–echo crack detection*

The embedded PWAS pulse–echo method follows the general principles of conventional Lamb wave NDE. A PWAS transducer attached to the structure acts as both transmitter and detector of guided Lamb waves traveling in the structure. The wave sent by the PWAS is partially reflected at the crack. The echo is captured at the same PWAS, which acts as receiver. To ensure the crack detection, an appropriate Lamb wave mode must be selected. It has been verified that the S_0 Lamb waves can give much better reflections from the through-the-thickness cracks than the A_0 Lamb waves [3]. The selection of such a wave is achieved through the Lamb wave tuning. A wave propagation experiment was conducted on an aircraft panel to illustrate crack detection through the pulse–echo method (Figure 6). A baseline signal was first collected, and then after the simulated crack was added, another reading was recorded. By comparing

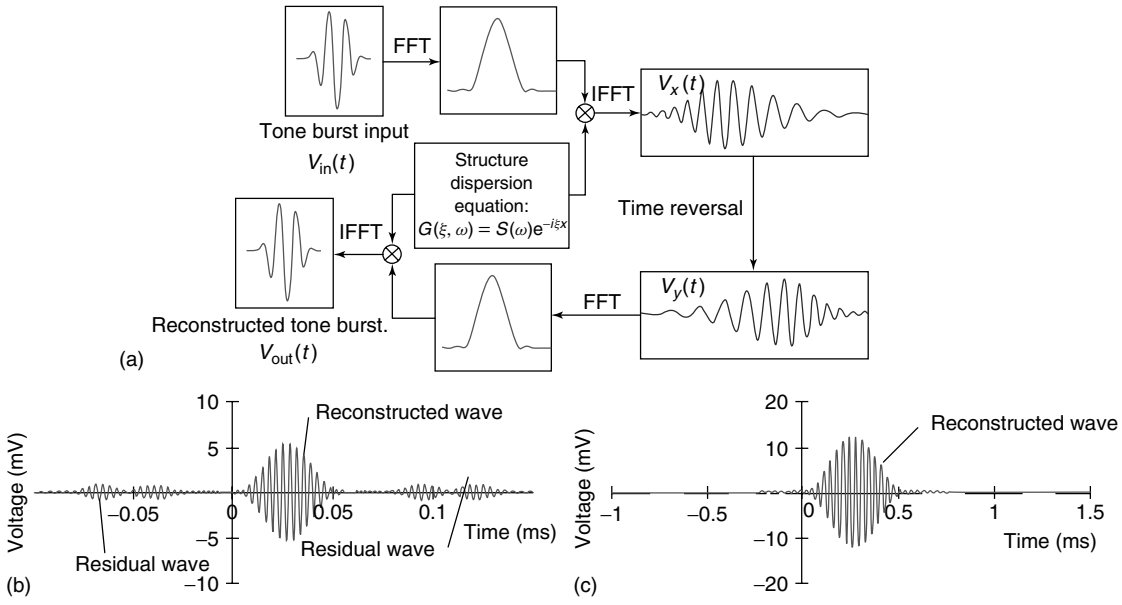


Figure 5. PWAS Lamb wave time reversal procedure. (a) Flowchart block diagram; (b) S_0 mode time reversal at 290 kHz showing residuals; and (c) A_0 mode time reversal at 30 kHz with no residual wave packets present.

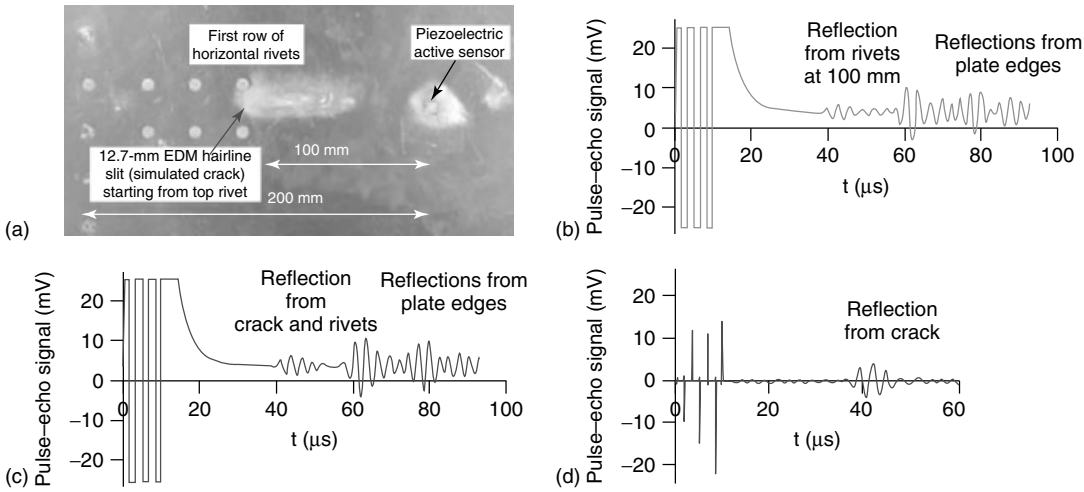


Figure 6. PWAS pulse–echo crack-detection experiment: (a) plate with built-in rivets and a simulated 12.7-mm crack; (b) baseline signal containing reflections from the plate edges and the rivets at 100 mm; (c) reading containing reflections from the crack, plate edges, and rivets; and (d) subtracted signal containing reflection from the crack alone.

the difference between the reading and baseline, the reflection caused by the crack could be extracted and detected.

3.3.3 Phased arrays for large area imaging

PWAS can also be wired as phased arrays to detect damage in thin-wall structures (*see* **Guided-wave Array Methods**). Using guided-wave PWAS phased-array methods, wide coverage could be achieved from a single location. PWAS phased arrays have been developed for thin-wall structures (e.g., aircraft shells, storage tanks, large pipes, etc.) that use Lamb waves to cover a large surface area through beam steering from a central location [13]. Its principle of operation is derived from two general concepts:

1. the generation of tuned guided Lamb wave with PWAS;
2. the principles of conventional phased-array radar.

The embedded ultrasonic structural radar (EUSR) algorithm [13] using PWAS is different from the conventional phased-array approach in two aspects:

1. it uses structurally embedded PWAS transducers;
2. it works in virtual time, not in real time.

The latter observation is very important for SHM, because it allows the phased-array benefits without the drawback of real-time multichannel phased excitation equipment. Whereas real-time phased-array transducers require heavy and complex multichannel phasing equipment, the virtual-time approach adopted by the EUSR method can be carried out with only one channel and very simple equipment. Figure 7(a) shows the typical setup of using a nine 7-mm square PWAS array to detect a broadside crack in a 1-mm-thick aluminum plate. On the investigated structure, the EUSR method captures and stores signals of an M-PWAS array in a round-robin fashion, i.e., one PWAS being activated as transmitter, while all the other PWAS act as receivers. Thus a matrix of $M \times M$ elemental signals is generated. In the post-processing phase, the elemental signals are assembled into the synthetic beamforming response using the synthetic beamforming algorithm and the scanning result is given as a 2D planar image (Figure 7b).

4 PWAS HIGH-FREQUENCY MODAL SENSORS

Modal analysis and dynamic structural identification are an intrinsic part of current engineering practice. Structural frequencies, damping, and modes shapes identified through this process are subsequently used to predict dynamic response, avoid resonance, and even monitor structural change that are indicative of

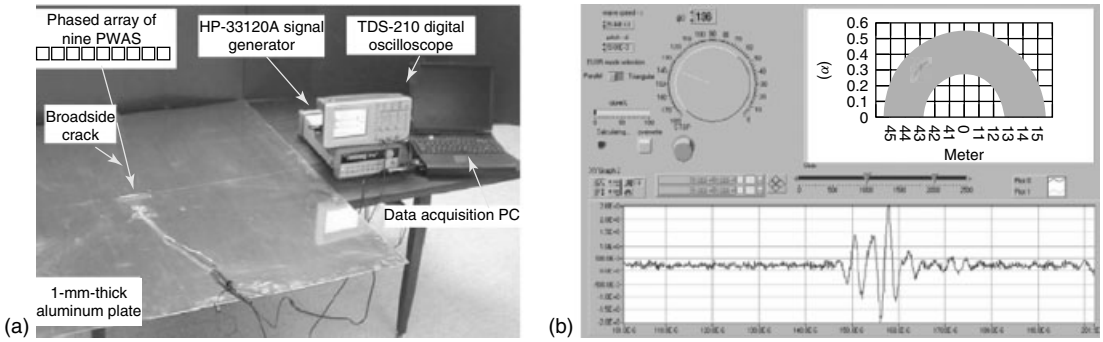


Figure 7. EUSR PWAS phased-array damage detection: (a) equipment setup and (b) EUSR scanning image.

incipient failure. The advantage of using PWAS for damage detection resides in their very high frequency capability, which exceeds by orders of magnitudes the frequency capability of conventional modal analysis sensors. They can be permanently attached to the structural surface and could form sensor and actuator arrays that permit effective modal identification in a wide frequency band. Thus PWAS are able to detect subtle changes in the high-frequency structural dynamics at local scales. Such local changes in the high-frequency structural dynamics are associated with the presence of incipient damage, which would not be detected by conventional modal analysis sensors that operate at low frequencies.

4.1 Analytical model

Consider a 1D structure with a PWAS attached to its surface (Figure 8a). Upon activation, the PWAS expands by $\varepsilon_{\text{PWAS}}$, generating a reaction force F_{PWAS} from the beam onto the PWAS and an equal and opposite force from PWAS onto the beam (Figure 8b).

The force excites the beam. As the PWAS is electrically excited with a high-frequency harmonic signal, it induces elastic waves into the beam structure. The elastic waves travel sideways into the beam structure and set it into oscillation. In a steady-state regime, the structure oscillates at the PWAS excitation frequency. The dynamic stiffness presented by the structure to the PWAS depends on the internal state of the structure, excitation frequency, and the boundary conditions as

$$k_{\text{str}}(\omega) = \frac{\hat{F}_{\text{PWAS}}(\omega)}{\hat{u}_{\text{PWAS}}(\omega)} \tag{24}$$

where $\hat{F}_{\text{PWAS}}(\omega)$ is the reaction force and $\hat{u}_{\text{PWAS}}(\omega)$ is the displacement amplitude at frequency ω. The symbol $\hat{}$ signifies the complex amplitude of a time-varying function. Because the size of the PWAS is very small with respect to the size of the structure, equation (24) represents a pointwise structure

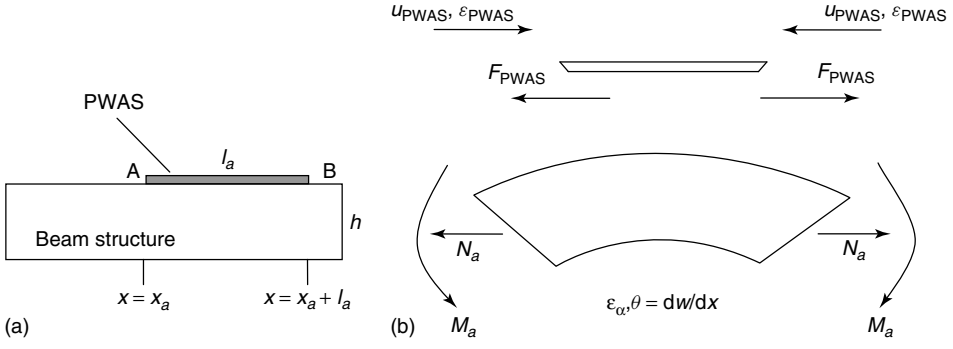


Figure 8. Interaction between a PWAS and a beamlike structural substrate: (a) geometry and (b) forces and moments.

stiffness. The response of the structural substrate to the PWAS excitation can be deduced from the general theory of beam vibrations. However, note that the PWAS excitation departs from the typical textbook formulation since it activates a pair of self-equilibrating axial forces and bending moments that are separated by a small finite distance l_{PWAS}. Kinematic analysis finally gives the dynamic structural stiffness as

$$k_{\text{str}}(\omega) = \rho A \left\{ \sum_{n_u} \frac{\left[U_{n_u}(x_a + l_a) - U_{n_u}(x_a)\right]^2}{\omega_{n_u}^2 + 2j\varsigma_{n_u}\omega_{n_u}\omega - \omega^2} + \left(\frac{h}{2}\right)^2 \sum_{n_w} \frac{\left[W'_{n_w}(x_a + l_a) - W'_{n_w}(x_a)\right]^2}{\omega_{n_w}^2 + 2j\varsigma_{n_w}\omega_{n_w}\omega - \omega^2} \right\}^{-1} \tag{25}$$

where $n_u, \omega_{n_u}, U_{n_u}(x)$ and $n_w, \omega_{n_w}, U_{n_w}(x)$ represent the axial and flexural vibrations frequencies and mode shapes, respectively.

4.2 Simulation and experimental results

The analytical model was used to perform numerical simulations that directly predict the E/M impedance and admittance signature at PWAS terminals during structural identification. Numerical simulations were performed with steel beams assuming damping coefficient at 1% over a modal subspace that incorporated all modal frequencies in the frequency bandwidth of interest. Finite element method (FEM) was also conducted to predict the structural natural frequencies using (i) the purely mechanical response via conventional structural elements and (ii) directly the E/M response via coupled-field elements [14]. The stiffness k_{str} in equation (25) is used to get the structural stiffness and calculate the E/M impedance. Subsequent experiments were performed to verify these predictions using a small steel beam ($100 \times 8 \times 2.6$, in millimeter) ($E = 200\,\text{GPa}$, $\rho = 7750\,\text{kg m}^{-3}$) instrumented with 7-mm square PWAS ($t_a = 0.22\,\text{mm}$) placed at $x_a = 40\,\text{mm}$ from one end (Figure 9(a) and (b)). During the experiments, recording of the E/M impedance real part spectra was performed with the HP4194A impedance analyzer in the $1 \sim 30\,\text{kHz}$ range. Numerical and FEM simulations and experimental results are given in Figure 9(c) and (d). Note that the experimental results are consistent with the predictions regarding the natural frequencies. Comparing the two FEM methods, it is noted that the impedance data obtained from the coupled-field analysis was more close to the experimental data, validating the possibility of using

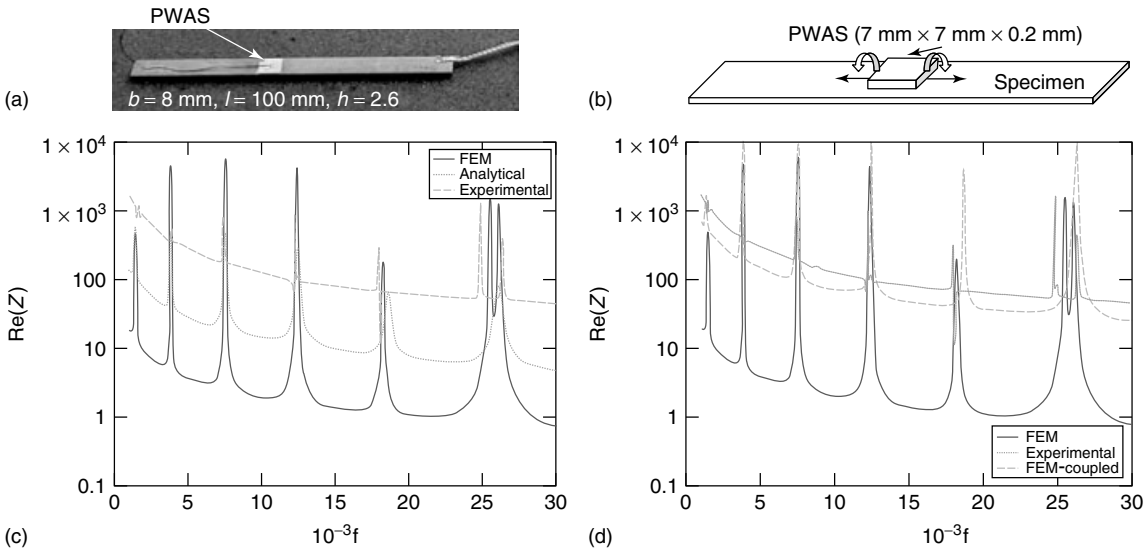


Figure 9. 1D structure simulation and experiments: (a) narrow-beam specimen; (b) beam structure layout for simulation analysis; (c) real part impedance spectra using structure FEM method; and (d) real part impedance spectra using coupled-field FEM method.

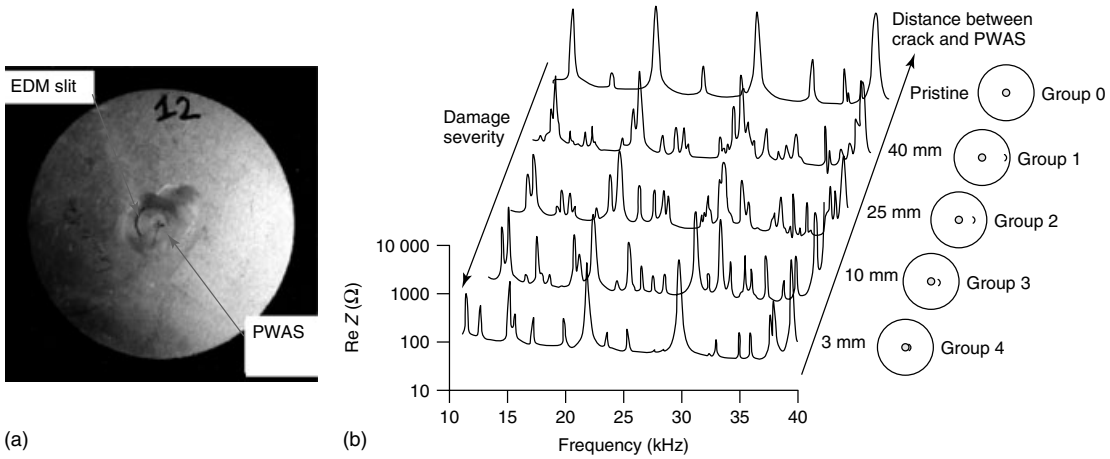


Figure 10. Experiments on dependence of the E/M impedance spectra on the location of damage on metallic plate specimen, E/M impedance in $0.5 \sim 40$ kHz frequency range: (a) the aluminum plate specimen with PWAS installed in the center and (b) E/M impedance spectra at various crack situations.

this method to analyze and simulate the structure with PWAS [14].

4.3 Damage detection with PWAS modal sensors

PWAS and the associated structural dynamics identification methodology based on the E/M impedance response are ideally suited for small machinery parts that have natural frequencies in the kilohertz range [15–17]. PWAS are able to detect subtle changes in the high-frequency structural dynamics at local scales. Such local changes in the high-frequency structural dynamics are associated with the presence of incipient damage, which would not be detected by conventional modal analysis sensors that operate at low frequencies. The E/M impedance of a 7-mm round PWAS has been measured with an HP4194A impedance analyzer at various crack situations on an aluminum metallic plate to assess the crack-detection capabilities, as shown in Figure 10.

5 NOVEL PWAS UNDER DEVELOPMENT

PWAS transducers have been proven to be valuable enablers of SHM systems. As a small embedded sensing device, PWAS have also shown great potential for many other applications such as biomedical applications (bio-PWAS) for *in vivo* monitoring of the bodily reaction to implants [18]. However, the bonded interface between the PWAS and the structure is often the durability weak link for SHM applications. The bonding layer is susceptible to environmental ingression that may lead to loss of contact with the structural substrate. The bonding layer may also induce acoustic impedance mismatch with detrimental effects on damage detection. Additionally, prefabricated piezoelectric sensors such as PWAS do not fit well on surfaces with complex geometry. Better durability may be obtained from a built-in sensor that is incorporated into the structure.

5.1 *In situ* fabricated PWAS

Polymer-based piezoelectric paints have been investigated as a potential substitute for PWAS in certain sensing applications. A new type of smart material, piezoelectric paint that cures at ambient temperature, has been developed by Zhang and Li [19] as sensing material for ultrasonic-based NDE (*see* **Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection**). Piezoelectric paint is a composite piezoelectric material that is composed of tiny piezoelectric particles randomly dispersed within a polymer matrix phase with adjustable material

properties that are not easily attainable in a single-phase material. Through judicious selection of the polymer matrix, the composite properties of the piezoelectric paint can be tailored to meet the specific requirements of an application condition. Because of its ease of application, piezoelectric paint can be easily deposited onto the complex surface of host structures; the paint uniformly cures at ambient temperature. On the basis of the preliminary results from the experimental study and finite element simulation, piezoelectric paint sensor appears to have a great potential for use as distributed sensors for fatigue crack monitoring in metallic structures.

5.2 Thin-film nano-PWAS

An alternative way to incorporate PWAS into the structure is the ferroelectric thin-film PWAS that uses nanofabrication technology to produce the sensors directly on the structural substrate [20]. Ferroelectric thin films have been shown to have piezoelectric properties that are those of single-crystal ferroelectrics, which are an order of magnitude better than common piezoceramics. In addition, the thin films require much smaller poling voltage/power. Thin film can get the same strain as that obtained by the PWAS with a much lower voltage, and through layering, the power of the device can be amplified many times. Figure 11 illustrates the thin-film multilayer PWAS development process [20]. C-axial preferred oriented ferroelectric $BaTiO_3$ thin films have been successfully fabricated on Ni metal tapes with a thin NiO buffered layer by pulsed laser ablation. Ferroelectric polarization measurements have shown the hysteresis loop at room temperature in the film with a large remnant polarization, indicating that the ferroelectric domains have been created in the as-deposited BTO films. These excellent properties in this piezoelectric thin film indicate that the as-fabricated BTO films show promise in the development of SHM systems. Hudak *et al.* [21] successfully pursued the alternative path of magnetostrictive thin-film PWAS and achieved enhanced thin-film performance by optimizing architecture through multilayered films, which demonstrated a 4 times stronger receiving signal in pulse–echo experiments. The guided waves have been used in pulse–echo mode to clearly detect

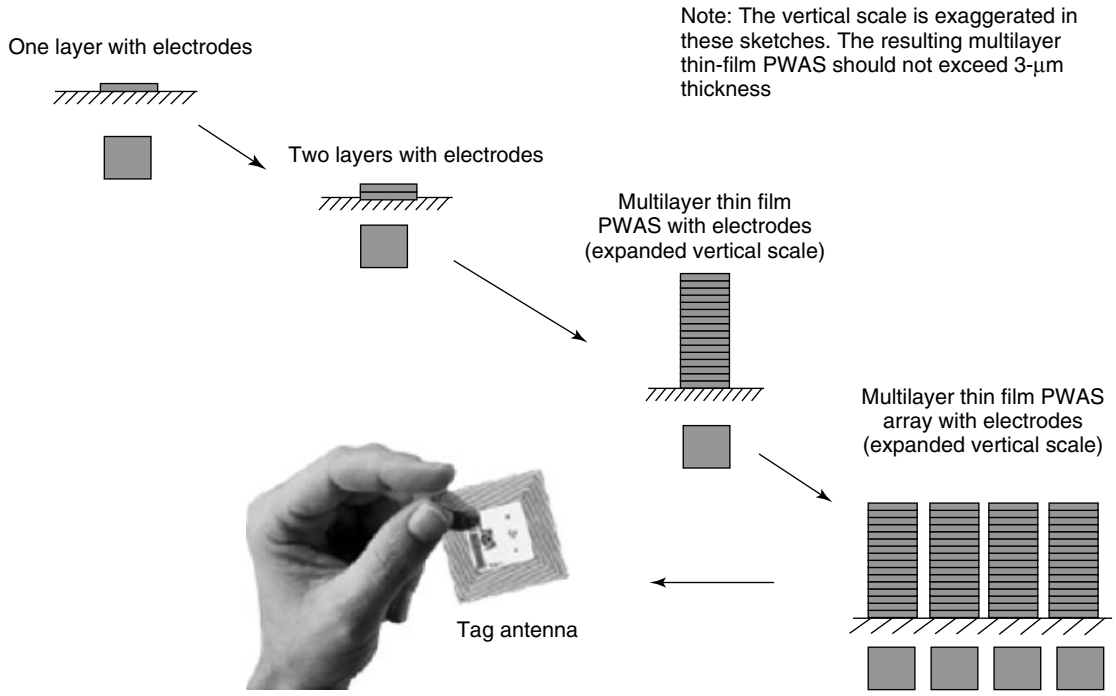


Figure 11. Blueprint of multilayer thin-film PWAS array.

1.52 mm × 7.62 mm defects in laboratory specimen at temperatures up to 554 °F.

6 SUMMARY AND CONCLUSIONS

This article has presented the use of PWAS as resonators, ultrasonic transducers, and high-frequency modal sensors for ultrasonic SHM. The PWAS transducer can be permanently attached to structures for *in situ* health monitoring to determine the structural state of health and predict the remaining life. After reviewing the PWAS operational principles, the PWAS piezoelectric properties (actuation and sensing) were presented in connection with Lamb wave excitation and detection. It was shown analytically and verified experimentally that Lamb wave mode tuning can be achieved by the judicious combination of PWAS dimensions, frequency values, and Lamb wave mode characteristics. For example, the A_0 and S_0 Lamb modes could be separately tuned by using the same PWAS installation but different frequency bands. Next, the use of embedded PWAS transducers as modal sensors was introduced for dynamic structural identification. The PWAS are able to detect subtle changes in the high-frequency structural dynamics at local scales, which are associated with the presence of the incipient damage, which would not be detected by conventional modal analysis sensors that operate at low frequencies. Finally, with all the success of using PWAS transducer for *in situ* SHM, several on-going novel PWAS developments were introduced, including the *in situ* fabricated PWAS and thin-film nano-PWAS.

Although remarkable progress has been made in using PWAS for NDE, considerable work remains to be done. To increase the acceptance of this emerging technology, the refining of the theoretical analysis and calibration against well-planned experiments is needed. However, to deploy the PWAS transducers to *in situ* SHM applications, several hurdles have still to be overcome. In particular, the operational and environmental variations of the monitored structure need to be addressed. In addition, a better understanding of the micromechanical coupling between PWAS and structure for various Lamb wave modes must be achieved. The behavior of the bonding layer between the PWAS and the structure must be clarified as well, such that the predictable and repeatable results are achieved. The durability of this bond under extended environmental exposure must be determined. Last, but not least, the signal analysis methods must be developed to achieve probability of detection values at least comparable with that of conventional NDE methods.

REFERENCES

[1] Giurgiutiu V, Lyshevski SE. *Micro Mechatronics: Modeling, Analysis, and Design with MATLAB*, ISBN 084931593X. CRC Press, 2004.

[2] Rose JL. *Ultrasonic Waves in Solid Media*. Cambridge University Press: Cambridge, 1999.

[3] Giurgiutiu V. Embedded ultrasonics NDE with piezoelectric wafer active sensors. *Journal of Instrumentation, Mesure, Metrologie* 2003 **3**(3–4):149–180.

[4] Raghavan A, Cesnik CES. Modeling of piezoelectric-based Lamb-wave generation and sensing for structural health monitoring. In *Proceedings of SPIE—Smart Structures and Materials 2004: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Liu S-C (ed). SPIE, July 2004; Vol. 5391, pp. 419–430.

[5] Bottai G, Giurgiutiu V. Simulation of the Lamb wave interaction between piezoelectric wafer active sensors and host structure. In *Proceedings of SPIE—Smart Structures and Materials 2005: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Tomizuka M (ed). SPIE, May 2005; Vol. 5765, pp. 259–270.

[6] Staszewski WJ. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. Bobs Books (JRM): Wayne, NJ, 2004.

[7] Chang FK. Built-in damage diagnostics for composite structures. *Proceedings of the 10th International Conference on Composite Structures (ICCM-10)*. Whistler, 14–18 August 1995; Vol. 5, pp. 283–289.

[8] Ihn JB, Chang FK. Multicrack growth monitoring at riveted lap joints using piezoelectric patches. In *Proceedings of SPIE—Smart Nondestructive Evaluation for Health Monitoring of Structural and Biological Systems*, Kundu T (ed). SPIE, June 2002; Vol. 4702, pp. 29–40.

[9] Lin X, Yuan FG. Diagnostic Lamb waves in an integrated piezoelectric sensor/actuator plate: analytical and experimental studies. *Journal of Smart Materials and Structures* 2001 **10**:907–913.

[10] Diamanti K, Soutis C, Hodgkinson JM. Piezoelectric transducer arrangement for the inspection of large

composite structures. *Journal of Composites, Part A* 2007 **38**:1121–1130.

[11] Kim S, Sohn H, Greve D, Oppenheim I. Application of a time reversal process for baseline-free monitoring of a bridge steel girder. *International Workshop on Structural Health Monitoring*. Stanford, CA, 15–17 September 2005.

[12] Xu B, Giurgiutiu V. Lamb waves time-reversal method using frequency tuning technique for structural health monitoring. In *Proceedings of SPIE—Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Tomizuka M (ed). SPIE, 2007; Vol. 6529, pp. 65290R1–65290R12.

[13] Giurgiutiu V, Bao J, Zagrai AN. *Structural Health Monitoring System Utilizing Guided Lamb Waves Embedded Ultrasonic Structural Radar*, US Patent, Patent #US6996480B2, February 2006.

[14] Liu W, Giurgiutiu V. Finite element simulation of piezoelectric wafer active sensors for structural health monitoring with couple-field elements. In *Proceedings of SPIE—Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Tomizuka M (ed). SPIE, 2007; Vol. 6529, pp. 65293R1–65293R13.

[15] Liang C, Sun FP, Rogers CA. Coupled electro-mechanical analysis of adaptive material system-determination of the actuator power consumption and system energy transfer. *Journal of Intelligent Material Systems and Structures* 1994 **5**:12–20.

[16] Sun FP, Liang C, Rogers CA. Experimental modal testing using piezoceramic patches as collocated sensors-actuators. *Proceedings of the 1994 SEM Spring Conference and Exhibits*. Baltimore, MD, 6–8 June 1994.

[17] Chaudhry Z, Joseph T, Sun FP, Rogers CA. Local area health monitoring of aircraft via piezoelectric actuator/sensor patches. *Proceedings of the SPIE North American Conference on Smart Structures and Materials*. Orlando, FL, February 26–March 3 1995.

[18] Bender JW, Friedman HI, Giurgiutiu V, Watson C, Fitzmaurice M. The use of biomedical sensors to monitor capsule formation around soft tissue implants. *Annals of Plastic Surgery* 2006 **56**(1):72–77.

[19] Zhang Y, Li X. Piezoelectric paint sensor for ultrasonic NDE. In *Proceedings of SPIE—Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Tomizuka M (ed). SPIE, 2007; Vol. 6529, pp. 652904-1–652904-12.

[20] Lin B, Giurgiutiu V, Yuan Z, Liu J, Chen C, Bhalla AS, Guo R. Ferroelectric thin-film active sensors for structural health monitoring. In *Proceedings of SPIE—Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Tomizuka M (ed). SPIE, 2007; Vol. 6529, pp. 65290I1–65290I8.

[21] Hudak SJ, Lanning BR, Light GM. A thin-film sensor for monitoring materials damage at elevated temperature. *Proceedings of AeroMat 2005*. Orlando, FL, June 2005.

Chapter 146

Sensor Technologies for Direct Health Monitoring of Tires

Ronald D. Moffitt[1], Scott M. Bland[2], Mohammad R. Sunny[2] and Rakesh K. Kapania[2]

[1] *Institute for Advanced Learning and Research (IALR), Virginia Polytechnic Institute and State University, Danville, VA, USA*

[2] *Aerospace and Ocean Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*

1 INTRODUCTION

An online structural health monitoring system provides information about the integrity of the structures in use in a real-time basis. This is very important to the end user, as it allows an optimal use and reduces the probability of catastrophic failure, and also to the manufacturer as it helps in decision making regarding the improvement of the product. Even in normal operating conditions, both automobile and aircraft tires experience large strain, pressure, and impact force. It makes a tire prone to catastrophic failure in different complex modes. Underinflation causes excessive sidewall flexing, heat generation and consequent tread separation, and sudden blowout. Overinflation and overloading result in excessive pressure and strain and make the tire more vulnerable to damage due to uneven surface. Manufacturing defects result in poor adhesion between tread and belt and may result in tread separation and blowout. These necessitate the development of a real-time structural health monitoring system for tire. Detailed understanding of tire–vehicle system dynamic behavior, development of a multifunctional sensor system, development of a wireless communication, among others are major challenges in the development of a tire structural health monitoring system [1].

The development of tire pressure monitoring systems (TPMSs) has allowed for indirect monitoring of tire health by analyzing both pressure and temperature data. The accuracy of indirect tire health monitoring based on tire pressures and temperatures is low due to the complexity of the tire structure and various operating conditions that a tire commonly encounters. To provide a more accurate means of

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

accessing overall tire health, a variety of new sensing technologies have been developed and employed to measure various characteristics of the tire such as the longitudinal/lateral forces, friction between the tire and the roadway, deformation and displacement of the tire during operation, strain in the tire, acoustic signatures, accelerations, etc. By measuring these quantities, we can get a much more accurate description of the mechanical response of the tire, and therefore a more accurate view of the tire's overall health. The various tire health monitoring sensors can be separated into two groups: noncontacting and embedded. The embedded sensors are either bonded to the inner liner of the tire or are embedded into the tire during manufacture. Noncontacting tire health monitoring sensors commonly utilize optical, electromagnetic, or acoustic phenomena to measure tire deformations, wheel speed, temperature, and friction. Noncontacting systems are generally desirable since they are not mechanically coupled to the tire, which improves reliability and other operational factors, but these systems can be large and may not be appropriate for some service environments. Another issue for some noncontacting measurements is the need for complex signal processing in order to evaluate sensor data, which increases cost and complexity of the system. Embedded sensors allow for direct measurement of the mechanical response such as temperature, strain, displacement, accelerations, etc. of the tire, but must operate in a harsh environment, and transmit data from the tire into the vehicle. There is also some concern about the presence of embedded sensors reducing the structural integrity of the tire due to localized stress phenomena. The following section discusses various sensor technologies that are currently being investigated for use in a tire health monitoring system. There is currently no widely accepted sensor technology that has proven to be the best for tire health monitoring applications, but significant advances in measurement accuracy, reliability, size, power consumption, and cost are continually being made, which will allow adoption of these sensor technologies in the near future.

1.1 Noncontacting sensors

A sidewall torsion sensor developed by Continental Inc. [2] utilizes a specially manufactured tire, which has two magnetized strips incorporated in the tire rubber on the inner sidewall and two sensors mounted on the chassis of the vehicle to measure the magnetic field strength from these strips on the tire. On the basis of phase changes in the upper and lower sensor signals and variations in magnetic field intensity between the two sensors, the deformation of the tire in the longitudinal and lateral directions can be determined and the forces acting on the tire can be estimated. This system was initially designed to be used for antilock brake and electronic stability control systems, but work is continuing to allow for detection of tire defects based on changes in the magnetic sensor signals.

The Apollo project for "Intelligent Tyre for Accident-free Traffic" [3] developed an optical positioning sensor, which used infrared (IR) diodes as a light emitter bonded to various locations on the inner liner of the tire and a position-sensing diode mounted on the wheel. A schematic of the optical positioning sensor is shown in Figure 1.

The displacement of the contact patch relative to the rim can be measured using this sensor. A light beam emitted from the IR diode is focused onto the Position Sensitive Diode (PSD) chip through a lens. On the basis of changes in the current at the corners of the PSD generated from the light from the IR diodes, the movement of the light beams can be determined. This information can then be processed to determine the distance from the rim to the inner liner.

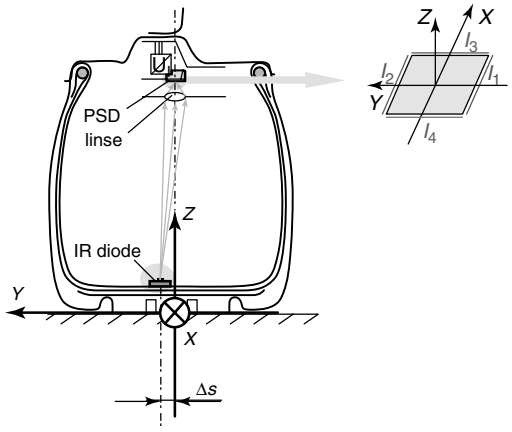


Figure 1. Schematic of optical positioning sensor and location of IR diode and PSD in the tire (image from Apollo final report). [Reproduced with permission from Ref. 3. ©, APOLLO, 2005.]

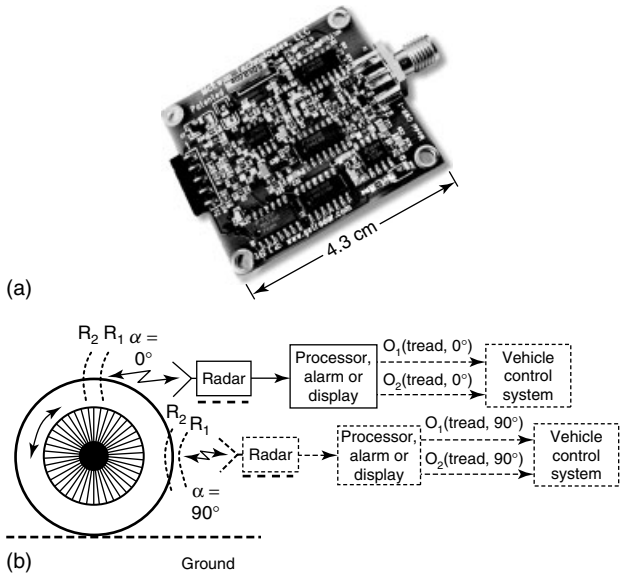


Figure 2. (a) Image of Model R-1 UWB Impulse Radar from McEwan Technologies (picture from: http://www.getradar.com/DemoUnits/Demo8.htm) and (b) MIR-based tire health monitoring system from patent application [5].

Preliminary results showed that this sensor performed well for measuring the displacement of the contact patch, but the diodes consumed a significant amount of power, which could limit its applicability because the sensor is mounted between the tire and the rim.

A German company, Optimess, has developed a laser-sensing system, which is mounted on the chassis of the vehicle to perform noncontact scanning of the tire tread surface during vehicle operation. This system performs scans over the entire tread surface after fixed driving intervals, and can detect tread wear as well as tire expansion during high-speed operation. Another application of the laser sensor could be placed on the rim to measure the contact patch displacement of the tire or measure the road surface conditions during vehicle operation.

Micropower impulse radar (MIR) developed in the early 1990s as a form of ultra wideband (UWB) radar is capable of high-resolution tire deformation measurements by utilizing small, low-power radar transmitter/receivers mounted externally, in close proximity to the tire. A recent patent by Michelin [4] discusses an MIR-based tire health monitoring system, which utilizes a 24-GHz pulse Doppler radar from McEwan Technologies [5] to detect tire abnormalities such as tread delamination, uneven tread wear, foreign object detection, and sidewall bulging. An image of the MIR and schematic of the tire health monitoring system are shown in Figure 2. The system makes use of a signal processor to analyze the signal generated by the MIR to detect the onset of various tire anomalies. This signal processor compares the reflected radio frequency (RF) signal energies over various frequency bands to detect the existence of a tire anomaly. Testing showed that this system was sensitive to tire wear and tread separation, but careful selection of the appropriate frequency bands was required to ensure accurate detection.

UWB radar-based tire health monitoring has not been fully investigated to determine its reliability and accuracy, but initial results show that this is a promising technology. Some complications that may arise during the implementation of this type of system are limitations on the power of the UWB radar by the Federal Communications Commission (FCC) in the United States in order to minimize potential interference. Currently, UWB radar use is not approved for use outside of the United States, which may also hinder potential widespread use of this technology.

Tekgenuity Limited has developed chromasonic sensors [6], which utilize acoustic emissions from the tire/road interface to identify tire anomalies. Various acoustic sensors such as microphones, strain gauges,

and optical microphones can by utilized in an array on the chassis of the vehicle around the tire to measure the acoustic emissions. Signal processing using Gabor transforms is then performed to extract important information from the sensor signals. The chromasonic sensors have shown the ability to detect pressure loss in tires, change in tracking angle, damage to tread area, objects embedded in tire, and changes in tire tread wear.

1.2 Embedded sensors

Embedded sensors for tire health monitoring are required to meet a number of performance requirements. They should have a wide frequency response and dynamic range, high resolution, and acceptable nonlinear characteristics. These sensors also must be compact, light, robust, and consume as little power as possible. Given all these requirements, very few current sensor technologies are acceptable for tire health monitoring, but rapid advances are being made and sensor performance is constantly increasing. The following sections discuss current embedded sensor technologies for tire health monitoring applications.

1.2.1 Acceleration sensing

The use of microelectromechanical systems (MEMSs) technology has drastically reduced the size of numerous sensor packages. MEMS accelerometers, which are both small and inexpensive, have become commercially available in common integrated circuit platforms. Owing to these benefits, MEMS accelerometers have been investigated for use as an embedded tire health monitoring sensor. The Apollo project investigated the use of custom-designed accelerometer circuits [3] to be embedded in the tire tread as well as bonded to the tire inner liner (Figure 3).

Results from testing of these sensors showed that they perform reliably and consume relatively low power, but these sensors were only capable of providing useful data in the radial direction of the tire. This type of sensor information could be used to estimate the wheel loads and possibly the contact patch length, but this information would be difficult to utilize for tire health monitoring applications.

Figure 3. MEMS accelerometer bonded to the inner liner of the tire using silicone rubber (image from Apollo final report [3]).

A significant issue with the use of accelerometers for tire health monitoring is that they detect all accelerations including rotational, gravitational, and vibrations from the mounting point, which requires significant signal processing to extract desired acceleration data. Another complication is the temperature dependence of the accelerometers, which can become a significant issue for sensors embedded in the tread due to frictional heating.

1.2.2 Force/friction sensing

A sensor system developed by Darmstadt University of Technology uses magnets placed in the tread block of the tire and a Hall sensor to monitor movements of the magnet [7]. The position sensor utilizes four monolithically integrated Hall crosses and an additional temperature sensor. This sensor allows movements of the tread block due to friction and forces of the tire/road contact to be measured. Test stand experiments have shown the potential to measure tire pressure, wheel load, and longitudinal/lateral forces. Miniaturization of the sensor makes it possible to integrate the chip and magnet into a single tread lug, so that sensor measurements can be made without interference from steel belts in radial tire applications. Another tire friction sensor is also being developed by Siemens in collaboration with Continental and Darmstadt University of Technology, which utilizes surface acoustic wave-sensing technology [7, 8].

1.2.3 Strain sensing

The direct measurement of strain in the tire would be very useful because tire degradation and tread separation can be directly related to high strains in

various locations on the tire. Owing to the importance of strain measurements in tires, there have been numerous approaches, but few are applicable for tire health monitoring applications. The Apollo project [3] evaluated the use of piezoelectric strain sensors made of Polyvinylidene fluoride (PVDF), which were bonded to the inner liner of the tire. These strain sensors were capable of monitoring the longitudinal and lateral strains in the tire during operation. These strain measurements could be used to estimate the contact patch length, possibly road conditions by analyzing the variations in strain signals. The interpretation of the strain signals was also generally simpler to interpret than the accelerometer data since only local deformations around the strain sensor were sensed. The PVDF material was highly sensitive to temperature variations in the tire, and this required temperature compensation to ensure reasonable accuracy of the measured responses. There were also questions regarding the reliability of this strain sensor for use in commercial scale production.

An alternative strain-sensing method was developed by Luna Innovations [9], which utilized fiber-optic strain sensors embedded in the tire or bonded to the inner liner of the tire after manufacture. This sensing technology allows for numerous strain sensors on a single optical fiber using extrinsic Fabry–Perot interferometry. Owing to the small size of the optical fiber, the effects of embedding the sensor on the structural integrity of the tire were small and strain sensor is less sensitive to temperature variations than other technologies such as PVDF strain sensors. Significant issues exist for using this technology on a commercial basis, such as the cost and size of signal conditioning and processing hardware and efficient methods for condensing and transmitting sensed data. The durability of the optical fibers in the harsh environments in which tires frequently operate also need further study.

Another feasible strain-sensing technology, which has not been fully tested for tire health monitoring applications, is the vibrating string sensor. This technology, developed by Vibstring [7], uses a thin wire set in a plastic tube and a standard microwave radio transceiver. Vibrations of the wire were found to modulate the RF. When strain is applied to the wire, the natural frequency of the wire changes. These changes in natural frequency, and therefore strain, can be determined by demodulating the modulated radio signals. The modulation due to the wire overrides all other vibration effects since the wire acts as an antenna, with its length selected in accordance to the high-frequency radio signal. No research currently exists, which discusses the performance of this sensor for tire health monitoring applications. It does not seem likely that this sensor could be embedded in the tire rubber without compromising the tire, and there are questions regarding the durability over long time periods.

Other more exotic strain sensors have also been proposed such as MetalRubber™ by NanoSonic, Inc. (NanoSonic), a novel and innovative new product that is a freestanding, conductive, elastomeric material based on electrostatic self-assembly comprising multiple alternating layers of electrostatically deposited monolayer metal and latex layers. The material can be cyclically stretched in uniaxial strain to more than 200% with good elastic recovery, exhibiting electrical resistivity as low as 10^{-5} Ω·cm, and remains viable over the temperature range bracketing the tire vulcanization process. MetalRubber™ may be capable of measuring very large strains (greater than 200%), but this material is very new, developmental, and it remains as a part of this research to prove the reliability and survivability of this material in the hostile environment of a radial tire. Another potential challenge is the creation of a strain sensor that minimizes the mismatch in material moduli between the rubber and the sensor material. Because the rubber is significantly less stiff than most sensor materials, there may be difficulty in determining the strain state in rubber with conventional sensors. These technical challenges could be significantly mitigated if the material from which the structure has been fabricated could be used as a matrix to incorporate conductive particles, such as high-aspect ratio single-wall carbon nanotubes (SWCNTs) and/or multiwall carbon nanotubes (MWCNTs) to facilitate the fabrication of a matched-compliance, conductive, elastomeric strain sensor that could be cured *in situ* as part of the tire tread, in the vicinity of the steel belt edges of a tire, where tread separation initiates.

1.2.4 Capacitance sensing

Recently, another form of embedded sensing for tire health monitoring has been developed, which is based

on capacitance changes to measure global strain in the tire. Virtually all these techniques employ the wires in the steel belts used in radial tire construction as electrodes in a circuit to monitor belt separation through changes in electrical circuit capacitance. An important feature and benefit of these systems is that the tire construction serves as the sensor. Therefore, no additional sensing elements are necessary that could induce a mechanical flaw and compromise tire structural integrity. Some examples of applied circuitry vary as follows:

1. Connecting the steel wires as a condenser within a Hartley-type oscillating circuit whereby the changes in the frequency of the circuit correspond to changes in the wire-to-wire capacitance [10]. This technique has been demonstrated successfully in both wired and wireless signal transmission modes but requires an external power source such as a battery to generate the oscillating circuit.
2. Connecting the steel wires as a condenser in an inductance–capacitance *LC* resonator circuit, whereby the changes in the frequency of the passive filter circuit correspond to changes in the wire-to-wire capacitance [11]. The change in the capacitance causes a change in the filtering frequency of a radio wave passed through the filter circuit, which provides a wireless measure of strain in the tire tread.
3. Connecting the steel wires as a condenser in a temperature-compensated capacitance–resistance *CR* resonator circuit, whereby the changes in the frequency of the circuit correspond to changes in the wire-to-wire capacitance [12]. Wireless strain measurements are possible for cyclic loading rates between 1 and 10 Hz. Static compression and dynamic rotation tests confirm the methodology to be applicable to deployment in commercial radial tires.

2 DISCUSSION OF DIRECT TIRE HEALTH MONITORING SENSORS

Numerous sensing technologies exist for tire health monitoring applications. In general, it would be desirable to isolate the sensors from the tire itself for three major reasons: avoiding the harsh operational environment inside and on the tire, avoid reducing the structural integrity or changing performance of the tire by introducing a sensor, and limited means for powering sensors inside the tire. Noncontact sensors offer the ability to measure the deformation of the tire during operation using various electromagnetic, optical, and acoustic methods. These methods currently appear to be the most commercially viable since they avoid the aforementioned issues, but do suffer some drawbacks themselves. These noncontact sensors can be susceptible to poor performance in driving conditions such as rain, dirt, and mud. Noncontact sensor systems also rely heavily on signal-processing systems to extract meaningful data from the raw sensor signals that may be expensive or bulky. Noncontact sensors also do not have the capability to measure the same quantities such as strain, tire forces, and friction, which may reduce their ability to perform tire damage identification and provides less information for tire life prediction methods.

REFERENCES

[1] Makinen T, Wunderlich H. Intelligent tyre promoting accident-free traffic. *The IEE 5th International Conference on Intelligent Transportation Systems*. Singapore, 3–6 September 2002.

[2] Gill J. *Continental Teves and Team of Developers Working Hard to Ready 'Intelligent Tire' for Mass Production*. PRNewswire, 1999.

[3] Technical Research Centre of Finland, Project IST-2001-34372, *Final Report Including Technical Implementation Plan (Annex) Deliverable 22/23*. APOLLO: Intelligent Tyre for Accident-free Traffic, July 25, 2005.

[4] Radar Monitoring System for Tires and Wheels, US Patent Application Publication # 2002/0189336 A1, 2001.

[5] Thiesen J, O'brien GP. *Doppler radar for detecting tire abnormalities*, US Patent No. 7,082,819 B2.

[6] Todd R. Chromasonic technology: a new methodology for the analysis of acoustic signatures for condition monitoring and early stage fault detection. White Paper on Chromasonic Technology, Cheshire, 2003.

[7] Technical Research Centre of Finland, Project IST-2001-34372, *Intelligent Tyre Systems—State of the*

Art and Potential Technologies, Deliverable D7. APOLLO: Intelligent Tyre for Accident-free Traffic, 2003.

[8] Pohl A, Steindl R, Reindl L. The 'intelligent tire' utilizing passive saw sensors—measurement of tire friction. *IEEE Transactions on Instrumentation and Measurement* 1999 **48**:1041–1046.

[9] Palmer M, Boyd C, McManus J, Meller S. Wireless smart-tires for road friction measurement and self state determination 43rd AIAA structures. *Structural Dynamics, and Materials Conference*. AIAA 2002-1548, Denver, CO, 22–25 April 2002.

[10] Todoroki A, Myatani S, Shimamura Y. Wireless strain monitoring using electrical capacitance change of tire: part I—with oscillating circuit. *Smart Materials and Structures* 2003 **12**:403–409.

[11] Todoroki A, Myatani S, Shimamura Y. Wireless strain monitoring using electrical capacitance change of tire: part II—passive. *Smart Materials and Structures* 2003 **12**:410–416.

[12] Matasuzaki R, Todorok A. Wireless strain monitoring of tires using electrical capacitance changes with an oscillating circuit. *Sensors and Actuators A* 2005 **119**:323–331.

Chapter 144
Ship and Offshore Structures

Myung Hyun Kim[1] and Do Hyung Kim[2]

[1] *Department of Naval Architecture and Ocean Engineering, Pusan National University, Busan, Korea*
[2] *Research and Development, Lloyd's Register Asia, Busan, Korea*

1 INTRODUCTION

In the 1980s and early 1990s, a number of losses of tankers and bulk carriers resulted in the environmental pollution owing to oil spills and the loss of lives. For example, from 1990 to mid-May 1997, a total of 99 bulk carriers were lost with the death of 654 people. Even the large tanker accidents, including *Erika* and *Prestige*, occurred relatively in recent years. Figure 1 shows the accident of the tanker *Prestige* (Single hull oil tanker, 81 000 ton) in the coast of Spain.

In this regard, International Maritime Organization (IMO), International Association of Classification Societies (IACS), and individual Classification Societies (Lloyd's Register of Shipping, LR; Det Norske Veritas, DNV; and American Bureau of Shipping, ABS) have recommended the use of hull stress monitoring systems for bulk carriers of 20 000 dwt and above to reduce the risks of structural failure since 1994. More recently, there is an increasing demand for documentation of safety and hull conditions for ships carrying oil and gas owing to the oil and gas sector's movement toward harsher environment. For example, as the offshore oil production moves from the North Sea toward the Barents/Kara Sea in Russian Arctic, LNG (liquefied natural gas) ships, FPSO (floating production storage and offload vessels), and shuttle tankers have to withstand longer, harsher winters and fatigue loads.

Hull monitoring system enables the operator of the vessel to monitor all relevant responses (motions, accelerations, loads, bending moments, stresses,

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Figure 1. The accident of tanker "Prestige" (November 2002).

etc.) and provides rational guidance as preventive measures in heavy weather conditions. The hull stress monitoring system (HSMS) is a system that provides real-time information such as motions and global stress experienced by the ship to crew of the ship while navigating as well as during loading and unloading operations.

Hull monitoring system consists of an affordable number of strain, temperature, or acceleration sensors placed in carefully selected positions on the structure. The sensors are connected to a central processing unit that converts the sensor signals into digital data, which are analyzed by the computer on the ship's navigation deck, allowing an operator to monitor hull stresses even under altering sea states. The system alarms when there is a risk of damage to the hull structures or cargo caused by improper loading or high stress in heavy weather. The lifetime history of the ship recorded by the monitoring system assists the operator in determining hull condition evaluation and fatigue life assessment, thus providing the possibility to prevent cracking and severe casualties. In addition, recorded data is useful when planning hull inspections and maintenance.

Recently, new ideas have been proposed, and advanced techniques have been developed for various measurements and applications. Some hull monitoring systems (HMS) have been already commercialized and user interface computation modules run over a distributed network system.

This article presents examples of current integrated HMS, as well as examples of innovative sensor systems.

2 HULL MONITORING SYSTEM RECOMMENDED BY CLASSIFICATION SOCIETIES

The basic rule requirements of HSMS are in accordance with IMO recommendations for "fitting of HSMS (MSC/Circ.646 6th June 1994, Maritime Safety Committee)". The major classification societies such as LR, DNV, and ABS not only provide similar monitoring guidance for minimum parameters to be monitored but also recognize the fitting of enhanced and more comprehensive monitoring systems through notations. The notations and requirements of classification societies concerning the provision of the typical monitoring for bulk carrier and tankers are summarized in Table 1.

Bulk carrier and tanker hulls could be instrumented with four sensors on the main deck to monitor the dominating sagging and hogging motion. The system monitors the longitudinal bending moments within the structure and compares permissible values that are predetermined by the class rules in conjunction with ship designer. Additional sensors are available to meet higher specifications required by some classification societies' notations, such as DNV HMON-2 and ABS HM3 [1]. The number of sensors that are required depends on the complexity of the structure and the level of ambition of the owner. More number of sensors enables more advanced analyses and the monitoring of more bending modes. The level of analysis can vary from simply monitoring the stress levels and issuing alarms if they approach dangerous levels. Additionally, the level of analysis can be

Table 1. Notations and requirement of the classification societies

Class	Notation	Gauges
LR	SEA (HSS-4)	4 LBSG, 1 bow vertical accelerometer
DNV	HMON-1	4 LBSG, 1 bow vertical accelerometer, 1 midship accelerometer
ABS	HM2 + R	4 LBSG, 1 bow vertical accelerometer

LBSG, long-based strain gauge.

extended to determining overall loads on the structure by converting strains and stresses to global moments, as well as to closely monitoring the fatigue cycles on selected structural details and thereby gather information about the fatigue status.

3 TYPICAL HULL MONITORING SYSTEM

In practice, a typical installation for a large bulk carrier or tanker consists of four long-based strain gauges (LBSGs) and an accelerometer linked to a display monitor, which is illustrated in Figure 2.

Hull monitoring system gives global hull stress response and provides alarm warnings to ship staffs if stress measurement values approach a prespecified level at which corrective action is advisable. The system supports the graphical confirmation that the corrective action has produced the desired reduction in stress levels. Stress signals from the sensors are analyzed in several minutes. Each channel passes through the routine for statistical data analysis, which calculates mean stress, maximum peak stress, standard deviation, peak-to-peak, and mean zero upcrossing periods. Fatigue strength is represented by the number of cumulative stress cycles up to measuring time. The counting is carried out using the rainflow method, and stress range are divided into a number of blocks with a constant stress range. Fatigue monitors indicate the amount of usage of the initial fatigue strength over time relative to the approved scantlings by use of Miner's rule in conjunction with rainflow counting. It is recommended that the measured data can also be used later for crack growth calculations using a method submitted for review. The data is also required to be recorded for later analysis and for new shipbuilding.

3.1 Long-based strain gauge (LBSG)

The LBSG measures the hull girder response of the vertical still water bending moment (SWBM) and wave-induced bending moment (WIBM) at selected points along the vessel. Figure 3 shows an exaggeration of how a ship bends owing to these loads. The permissible bending moment for both sea and harbor condition is provided by classification rules.

By means of continuous measurements and displays in real time using LBSG, the longitudinal stresses at strategic locations on the hull girder are simultaneously monitored in real time. A minimum of four LBSGs are fitted: two at midship and two at approximately 25% from the bow and stern. The precise location may vary per different ships. The LBSG consists of about 2-m stainless steel rod, fixed at one end and free to move against a displacement transducer at the other end. The electrical transducer is a liner variable differential transformer (LVDT).

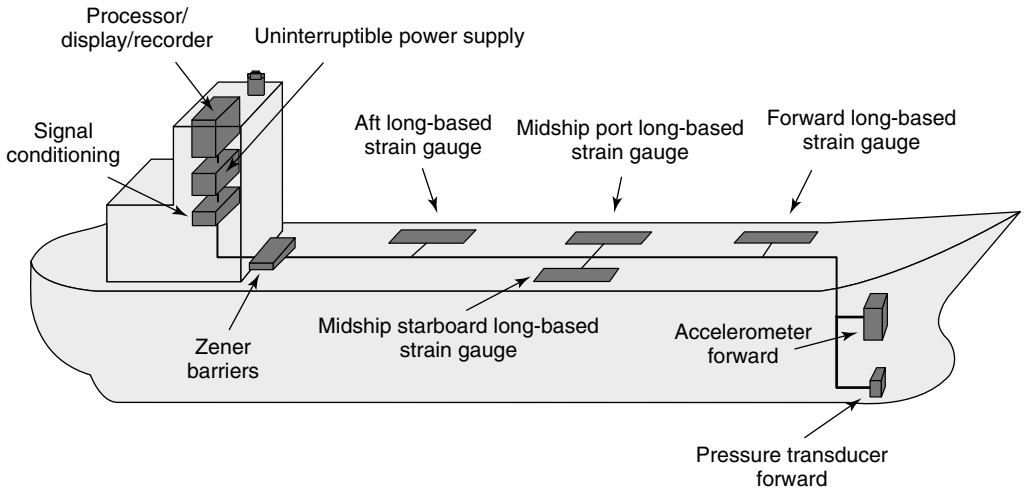


Figure 2. Typical hull monitoring system for a bulk carrier.

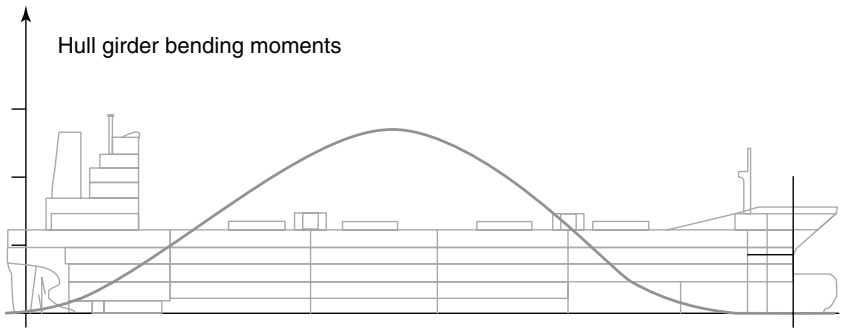


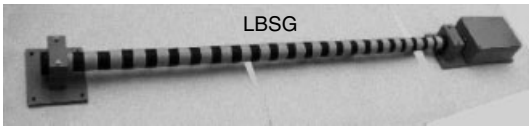
Figure 3. The bending moment display for a bulk carrier.

The end mountings are bolted to plates welded to the deck with centers about 2 m apart and located along a longitudinal deck. The unit is waterproof and fitted with a protective cover. Figure 4 shows the mounting plate with a strain gauge rod and a protect cover.

3.2 Accelerometer

The accelerometer with the range of $\pm 2\,g$ is used to measure the vertical acceleration at the bow. It is used to give an indication of the probability of slamming in heavy weather. It operates via a line amplifier that provides output signal to the display logger unit. The accelerometer unit shown in Figure 5 should be attached to the bracket that is welded to a secure structural member of the ship. In addition, the system should calculate and store the impact energy of a slamming by calculating area under the acceleration versus displacement curve.

(a)



(b)

Figure 4. Long-based strain gauges (LBSG). (a) LBSG installed at an actual ship (KR 2004) and (b) LBSG mounted on a SUS304 rod.

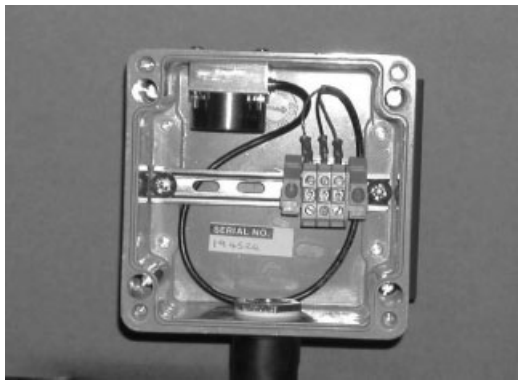


Figure 5. The accelerometer unit.

3.3 Additional sensors

In addition to the typical specification of four LBSG and one bow accelerometer, additional sensors could be added to extend the functionality to meet client-specific requirements, or the higher rules of the certifying authorities. The following lists some examples of sensors that may be installed:

- strain gauge: measures localized stress, e.g., for ice loadings on the bow and stern structures. Strain gauges are typically required to have accuracy better than ± 5 microstrain and be capable of measuring in the 0–5-Hz frequency range;

- six degree of motion sensor: roll, sway, heave, pitch, yaw, and surge accelerations;
- pressure transducer: measures pressures on the bow or stern and the depth of water at the bow and gives indication of draft as well as the detection of the bow leaving the water at the onset of slamming.

4 EMERGING TECHNOLOGIES FOR SHIP HULL MONITORING

In recent years, HMS has attracted much attention in new technology. Fiber-optic sensor (FOS), acoustic emission (AE) sensor, and crack monitoring sensor are promising sensing alternatives in HMS. New HMS was developed to describe and predict how damage evolves in structures before their reliability can be accurately forecasted in real time. This intelligent system offers services in onshore data management to help the maintenance planners and other onshore personnel manage the assets in an optimal way. Reports, which summarize the hull condition, are available based on the data from the HMS. Data can be transferred for onshore analysis by automatic e-mails or by backup media. In addition, the transferred data is useful when planning hull inspections and maintenance.

Monitoring contributes to increased safety and may facilitate an increase in fatigue lifetime for the structure. In the case of offshore structures, the problem is also compounded by harsh marine environment. Offshore structures must withstand cyclic wave loading, severe storms, sea quakes, and the corrosive effects of sea water, particularly for long period of time. Furthermore, the process of visually inspecting marine structures, especially those in deep water, is much more difficult than that for land-based structures.

4.1 Fiber-optic sensor (FOS)

One of the new methods for measuring the hull stress and the health of ship hulls is to apply FOS. Optical fibers, which usually consist of three layers: fiber core, cladding, and jacket, are dielectric devices used to confine and guide light. The majority of optical fibers used in sensing applications have silica glass cores and claddings, and the refractive index of the cladding is lower than that of the core to satisfy the condition of Snell's law for total internal reflection and thus confine the propagation of the light along the fiber core only. The outer layer of a FOS, called jacket, is usually made of plastic to provide the fiber with appropriate mechanical strength and protect it from damage or moisture absorption. In some sensing applications, a specialized jacket is required to enhance the fiber's measurement sensitivity and to accommodate the host structure.

There are a number of different concepts for turning optical fiber into a sensor. The new structure monitoring systems use fiber-optic sensors based on Bragg gratings (FBG), which is shown in Figure 6.

FBG is such a sensor technology, creating an optical strain gauge within the core of an optical fiber through the use of a wavelength-specific filter. As the FBG experiences induced strain along its major axis, its light signal indicates the amount of strain with great accuracy and sensitivity. In FBG applications, where loading may occur in all directions, complex changes take place in the FBG signal.

The generalized strain–optic relationship between the optical path length (Δ_l) and the axial strain, ε_{xx}, induced over the gauge length of a single segment of the optical fiber is given below:

$$\Delta_l = n_{\text{eff}} L \varepsilon_{xx} \tag{1}$$

and

$$n_{\text{eff}} = n - \frac{1}{2}\left[P_{12} - \nu_f (P_{11} + P_{12})\right] \tag{2}$$

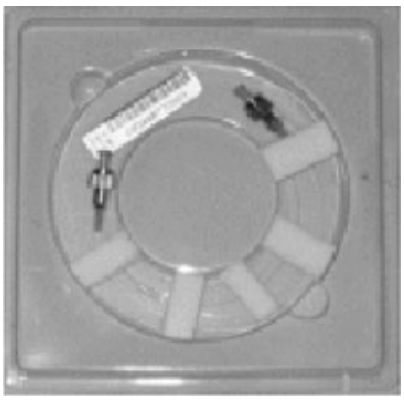


Figure 6. Typical fiber-optic sensors based on Bragg gratings (FBG).

where P_{11} and P_{12} are Pockels constant, ν_f is the Poisson's ratio, n is the refractive index of the fiber, and L is the length (gauge length) of the optical fiber to which the strain is applied.

The FOS have a number of advantages over the electrical alternatives, especially in harsh environments: small size, light weight, immunity to electromagnetic interference (EMI), passive composition, high-temperature performance, large bandwidth, higher sensitivity as compared to existing monitoring techniques, and multiplexing capabilities. Multiplexing of Bragg grating sensor would allow for a number of strain measurements along one fiber, which is shown in Figure 7. This arrangement maximizes the multiplexing potential of Bragg gratings and keeps system cost to a minimum.

Application of fiber Bragg grating sensors for the purpose of ship hull monitoring is reported by Norwegian Defense Research Establishment. FOS are embedded in composite hull, and real-time strains as well as global loads are measured in actual sea-keeping tests [2].

4.2 Acoustic emission (AE) sensor

Crack growth due to fatigue and stress corrosion is normally a slow degradation process up to a point, beyond which failure may be sudden and catastrophic. Detection as early as possible during this initial period of crack growth is essential if the consequences of an unexpected failure are to be avoided.

AE monitoring can overcome many of the difficulties with crack detection in service and is potentially a most powerful method of inspection. It is sensitive to the propagating cracks, that is, the structurally significant defects, and provides information on growth rate under service loading conditions, guiding inspection and repair work for maximum cost-effective maintenance.

Examination of the microstructure of fatigue and stress corrosion cracks, see Figure 8(a), reveals that fracture on a microscale involves crack steps of magnitude comparable to the grain size of the material. In the case of stable crack growth due to fatigue, the crack will extend only if the stress

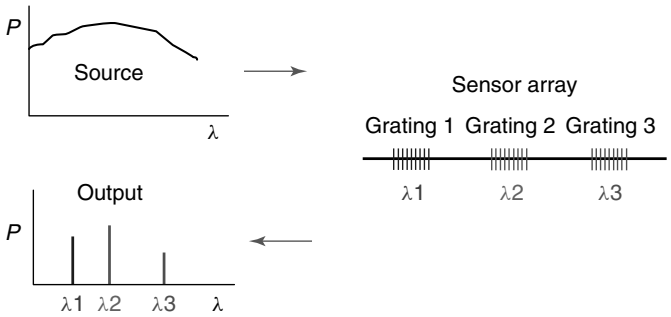


Figure 7. Schematic for multiplexing of Bragg grating fiber-optic sensor.

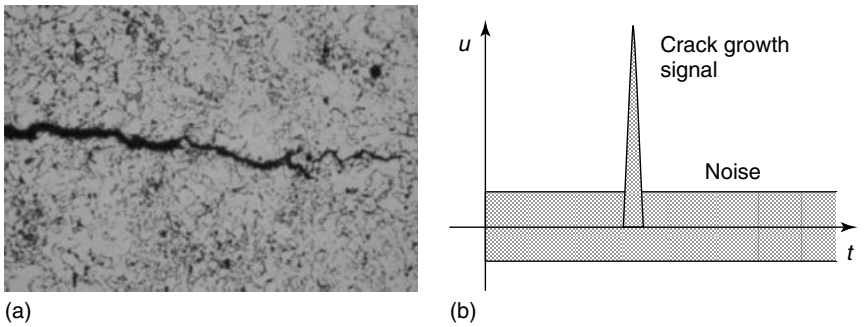


Figure 8. Schematic representation of extremes of fracture behavior. (a) Microstructure of a fatigue crack and (b) fast brittle crack advance in poor toughness material.

concentration is sufficient for the material to yield over a short distance ahead of the crack tip. This distance is referred to as the threshold plastic zone size.

This monitoring technology can provide improved assurance of overall integrity, justifying the additional work and cost involved. By continuously monitoring a vessel or offshore structure for a period of time, usually several weeks or months, depending on the minimum acceptable defect size, enhanced assurance of structural integrity can be obtained. The underwater sensor shown in Figure 9 is attached to a marine structure (ballast tank in vessel or jack-up platform), which is the most effective sensor for monitoring.

The period of monitoring must be predetermined as adequate for a measurable amount of crack growth to occur. It is typically specified by the minimum size of structurally significant size defects (e.g. propagating fatigue cracks) that can be identified. By relating crack growth to the loading and environmental forces driving the crack and the stress concentration due to geometry, it is possible to model the crack growth accurately. In addition, the effectiveness of different methods of inhibiting or arresting crack growth can be evaluated by continuing the monitoring for a period of time.

Sources of AE are defect-related processes such as crack extension and plasticization of material in the highly stressed zone adjacent to the crack tip. In general, the form of the primitive wave changes during propagation through the medium, and the amplified signal from a resonant piezoelectric sensor bears little resemblance to the original pulse. The transducer produces an electrical pulse that can be analyzed to provide information about the original AE source and, hence, facts about the structure.

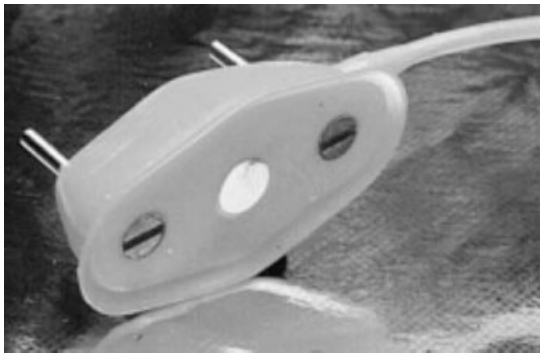


Figure 9. A new AE sensor for use on underwater marine structure.

4.3 Fatigue crack monitoring sensor

Fatigue crack monitoring sensors are developed for detecting fatigue cracks occurring in welded structures. These sensors are cheap, small in size, and made of thin metal pieces as shown in Figure 10. This type of fatigue crack sensors can be placed in front of stress-concentrated area for detecting fatigue cracks, particularly in welded joints.

When fatigue crack propagates through the surface of the sensor due to repeated fatigue loadings, the sensor can be used for measuring the fatigue crack length. The crack length information then can be used for calculating the remaining fatigue life of the welded joint using the following equation:

$$Ds = \frac{\Delta a}{a_0} \tag{3}$$

Here, a_0 and Δa denote the initial crack length and the crack propagation, respectively. $Ds = 0$ indicates intact state, while $Ds = 1$ indicates fatigue fracture. This type of fatigue crack sensors can be placed in many fatigue-prone locations and can be used for monitoring fatigue cracks in ships and offshore structures.

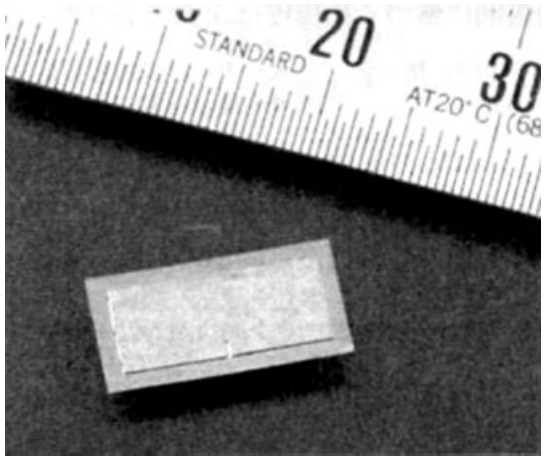


Figure 10. Fatigue crack monitoring sensor.

5 FUTURE TRENDS

The consumption of LNG is increasing and the market for oil transport for cold climate (ice/arctic) is expanding at an unprecedented pace. As there is increase in the size of vessels and cargo containment systems, navigating in new region and the demands on the classification societies are also very challenging. At the same time there is increasing demand for documentation of safety and, in particular, of hull conditions for ships carrying oil and gas.

Sloshing-induced impact load can cause a critical damage on tank structure. Sloshing is especially troublesome in LNG cargo tank. A sloshing monitoring system should measure impact pressure and strain in large LNG carriers. A new line of FOS has been recently developed for monitoring the sloshing pressure in low-temperature insulation tanks. Traditional pressure sensor cannot be used for measuring pressure inside the containment system owing to the cryogenic environment. The newly developed sensors are applicable to LNG cargo containment systems at $-163\,^{\circ}C$ and offer the possibility to measure sloshing pressures in nonintrusive manner without penetrating the primary membrane. The pressure sensors may be fitted with integrated temperature sensors and may be combined with fiber-optic stress sensors on the tank walls.

As the oil and gas sector moves to cold climate environments, the infrastructure must take an increasing load. Ice load as a phenomenon is largely statistical in nature. The arctic provides a harsh environment to survive in. Cold temperatures, ice, snow, and freezing fog pose additional factors for ship design, construction, and operation. Hull monitoring system is an integral part to ensure the vessels' capability to survive and to give a measure of safety when navigating in ice water.

The ice load monitoring system is always tailor-made according to the particular needs of each project. The system is based on large amount of previous field-testing. The system shows ice load, trend, severity measure, and prediction for the crew. The measured data is stored in the system, and it can be downloaded from the storage for further analysis.

For offshore structures, there are many locations that are difficult to access, and therefore it is not an easy task, if not impossible, to inspect all the damage-prone locations. Besides, they operate at very harsh environment. Also, it is complicated to distinguish the structural response between those due to wave loading and damages. Therefore, health monitoring technique would be an ideal candidate for this type of application since the frequency range of interest is much higher than the environmental excitation.

With the increasing market demand for vessels operating in cryogenic environment, such as LNG and arctic ships, new and more effective health monitoring techniques should be developed.

6 SUMMARY AND CONCLUSIONS

In this article, various techniques for monitoring structural integrity of ship and offshore structures are discussed. Typical HMS with common sensor locations are introduced as well as measurement parameters to be monitored are discussed. LBSG is typically installed at the deck of midship and at quarter length of vessels for monitoring hull girder strength. Ship motion and slamming pressures are measured by accelerometers and pressure sensors located at the bow of vessels. Different requirements for hull monitoring system are summarized for each classification societies. This is followed by a review on the introduction of emerging new technologies that are applicable for ship and offshore structures. The basic principles and the application of FOS, AE sensors, and crack detection sensors (CD) are reviewed. Finally, new requirements and future trends in terms of structural health monitoring in marine industries are introduced. In particular, the importance of structural health monitoring technology that is applicable in cryogenic environment is presented. In conclusion, structural health monitoring technique is an integral system for ensuring the safety of ships and offshore structures as well as for protecting environment.

RELATED ARTICLES

Fiber Bragg Grating Sensors

Monitoring Marine Structures

Fatigue Monitoring in Nuclear Power Plants

REFERENCES

[1] Det Norske Veritas (DNV), *Rule for Ships*, July 2006; Part 6, Chapter 11, pp. 1–7.

[2] Wang G, Pran K, Sagvolden G, Havsgard GB, Jensen AE, Johnson GA, Vohra ST. Ship hull structure monitoring using fibre optic sensors. *Smart Materials and Structures* 2001 **10**:472–478.

FURTHER READING

American Bureau of Shipping (ABS), *Guide for Hull Condition Monitoring Systems*, December 2003, 1–19.

http://hullmon.marintek.sintef.no/ 2003.

http://www.smartfibres.com/ 2007.

International Maritime Organization (IMO), *Recommendations for the Fitting of Hull Stress Monitoring Systems*, MSC/Circ.646, June 1994.

Kersey AD. A review of resent developments in fiber optic sensor technology. *Optical Fiber Technology* 1996 **2**:291–317.

Kim DH, Kim MH, Kang SW. Development of ship structure health monitoring system using fiber optic sensors. *Proceedings of the Annual Autumn Meeting*. SNAK, 2004; pp. 230–235.

Kim MH. Simultaneous health monitoring and vibration control of adaptive structures using smart sensors. *Shock and Vibration* 2002 **9**:329–339.

Lee B. Review of the present status of optical fiber sensors. *Optical Fiber Technology* 2003 **9**:57–79.

Li HN, Li DS, Song GB. Recent applications of fiber optic sensors to health monitoring in civil engineering. *Engineering Structures* 2004 **26**:1647–1657.

Lloyd's Register (LR), *Ship Event Analysis*, May 2004, pp. 1–4.

McKenzie I Jones R, Marshall IH, Galea S. Optical fibre sensors for health monitoring of bonded repair systems. *Composite Structures* 2000 **50**:405–416.

Qing XP, Chana HL, Beard SJ, Ooi TK, Marotta SA. Effect of adhesive on the performance of piezoelectric elements used to monitor structural health. *International Journal of Adhesion and Adhesives* 2006 **26**:622–628.

Roberts TM, Talebzadeh M. Acoustic emission monitoring of fatigue crack propagation. *Journal Of Constructional Steel Research* 2003 **59**:695–712.

Rogers LM. *"Structural and Engineering Monitoring by Acoustic Emission Methods—Fundamentals and Applications"* Report of LR Technical Investigation Department, September 2001.

Staveley C. Application of optical fibre sensors to structural health monitoring, optimisation and life-cycle cost control for oil and gas infrastructures. *Business Briefing, Exploration and Production, the Oil and Gas Review*, 2004; Vol. 2, pp. 72–77.

Yuan S, Wang L, Peng G. Neural network method based on a new damage signature for structural health monitoring. *Thin-Walled Structures* 2005 **43**:553–563.

Chapter 64
Directed Energy Sensors/Actuators

James L. Blackshire
Air Force Research Laboratory, Wright Patterson Air Force Base, OH, USA

1 INTRODUCTION

The interaction of electromagnetic energy with matter represents one of the most common and useful methods for inspecting and assessing a material or structure. Visual inspections, in fact, continue to be one of the most widely used and effective methods for characterizing the surface properties of a material or system [1]. By studying the reflective properties of a surface, for example, structural damage in the form of surface-breaking cracks, corrosion, and disbonds can be distinguished from undamaged areas based on surface roughness or topographic variations, which tend to reflect visible light differently, providing a simple means for detecting damage easily with the naked eye.

 ISBN: 978-0-470-05822-0.

Of course, the characterization of surface properties alone represents a somewhat limited capability for understanding the structural health state of a material system. With the fundamental scientific discovery in the mid-seventeenth century that the electromagnetic spectrum is actually a continuum of frequencies/wavelengths, and the subsequent discovery of X rays [2], microwaves [3], and radar [4], the material penetration capabilities of electromagnetic radiation were revealed with profound implications. In particular, the ability to "see" into or through a visibly opaque material became possible using radiography at high frequencies ($\sim 10^{18}$ Hz) and microwave imaging at low frequencies ($\sim 10^{11}$ Hz).

The present article is concerned with the use of *directed energy* as an inspection tool for use in structural health monitoring applications. Directed energy sensing can be loosely defined as "the controlled insertion of electromagnetic energy into a material or system, where the observation of electromagnetic–material interactions is used to characterize the structural health of the system". At the core of this definition is the electromagnetic–material interaction, which uses the changes induced in an electromagnetic wave/field by a material system as its measurand. These changes can involve phase, amplitude, frequency, directionality, polarization, time of flight, and many other phenomena that are usually associated with wave phenomenon. It is in this basic theme that the article is organized, where fundamental principles related to electromagnetic

fields/waves, directed energy beam characteristics, radiation–material interactions, and damage sensing are covered. An initial historical overview is first given followed by a brief description of the state-of-the-art directed energy sensing methods. A general overview of each of the major directed energy inspection methods is then provided, where physical principles of each method are covered along with specific measurement examples.

1.1 Historical background

Electromagnetic radiation has been at the forefront of scientific discovery for several millenniums. What began as a need for understanding the stars and heavens has grown to a scientific and technological discipline, which touches every part of human life and existence. Much of what we know about visible light and its interaction with materials was first studied in the Renaissance period (fourteenth to seventeenth centuries) by Descartes [5], Newton [6], Fermat [7], Snell [8], Huygens [9], Fresnel [10], and others [11–13]. The principles of light reflection, refraction, scattering, and diffraction were identified and systematically studied in much of this early work. The concept of "light waves" versus "light particles" was also debated during this period, which is currently known as the *wave-particle duality* of electromagnetic waves and photons of energy [14]. Both of these concepts play an important role in directed energy sensing methods.

For the next 200 years, theoretical and experimental work defined the inner workings of electromagnetism, culminating in the work of Maxwell [15] who demonstrated mathematically that electric and magnetic fields travel through space, in the form of waves, and at the constant speed of light. In 1861, Maxwell wrote his four-part publication in the *Philosophical Magazine* called "On Physical Lines of Force", where he first proposed that light was a form of energy composed of both electric and magnetic phenomena [15]. This fundamental discovery made the basic connection between electromagnetic waves and electronic material states possible, where electromagnetic fields exert a force (the Lorentz force) on the charged particles in a material system. The set of four equations known as *Maxwell's equations* describes this interrelationship between electric fields, magnetic fields, electric charges, and electric currents, forming the foundation of classical electromagnetism and electrodynamics. The material property concepts of permittivity, permeability, and complex dielectric were also defined, providing a means for characterizing electromagnetic–material interactions, where Maxwell's equations can be used in conjunction with the dielectric properties of a material to understand how light interacts with a material in a classical sense.

Around this same time (1860), the term *blackbody* was introduced by Kirchhoff [16], which relates the electromagnetic radiation emission properties of an object at increasing temperatures. In effect, the amount of electromagnetic radiation (and its corresponding wavelength) emitted by a black body source is directly related to its temperature. Black bodies above ~700 K (430 °C) produce radiation at visible wavelengths starting at red, going through orange, yellow, and white before ending up at blue as the temperature increases. This discovery represented a key connection between electromagnetic and thermodynamic physics (*see* **Thermal Imaging Methods**).

The early work of Hertz [17] and Roentgen [2] in the 1880s–1890s extended the useful range of the electromagnetic spectrum to radio frequencies and to X-ray frequencies, respectively. Earlier in 1886, Hertz developed the dipole antenna receiver, which helped to establish the photoelectric effect, and represented the first practical instrument for producing and receiving electromagnetic wave energy (at ultra high frequency (UHF) radio frequencies). At the other end of the electromagnetic spectrum, Roentgen discovered X rays on November 8, 1895 when he observed a fluorescent glow from crystals near a cathode-ray tube. His systematic characterization of the penetrating nature of the radiation emitted by cathode-ray tubes ushered in the scientific field of radiography. The first practical use of electromagnetic energy as an inspection tool soon followed when X-ray fluoroscopes were used in 1896 to inspect postal packages, porcelain materials, precious stones, and simple metal welds [2] (*see* **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**).

In 1900, Planck presented a paper at the German Physical Society in which he proposed the revolutionary scientific thought of quantum

mechanics, which describes submicron particles as being composed of discrete states of energy [18]. In that work, Planck used the blackbody radiation concepts of Kirchhoff [16] and Wien [19] to mathematically describe the distribution of electromagnetic energy emitted by the different modes of charged oscillators in matter—what is known today as *Planck's law* for blackbody radiation. Building on these original ideas, in 1905 Einstein proposed that light can act as individual, discrete energy states—what we now know to be *quantum* of radiant energy or photons [20]. Einstein further described a new relationship between the energy and frequency of a photon as $E = h\nu$, where h is known as *Planck's constant* [20]. The foundation of electromagnetic–material interactions at a quantum mechanical level was based on these early ideas, which were verified by the end of the 1920s by Bohr, Born, Heisenberg, Schrödinger, De Broglie, Pauli, Dirac, and others. The work of Bohr, in particular, set forth in 1913 the fundamental quantum mechanical relationship between atomic states and electromagnetic energy emission and absorption, where he predicted the wavelengths of emission for a hydrogen atom [21]. The light emitted (and absorbed) by an atom, molecule, or material is now understood to arise from the transition of electrons between discrete energy states in a material system, where electromagnetic energy is either given to or taken from the system. The fundamental nature of electromagnetic–material interactions was, therefore, found to be both discrete in nature—quantum mechanics and photon energies—and continuous in nature—electromagnetic wave interactions with dielectric material properties (*see* **Electric and Electromagnetic Properties Sensing**).

1.2 State of the art

The use of directed energy methods for material inspections and structural health monitoring has seen significant progress and advancement in the past decade. The state of the art now permits detailed characterization of numerous materials with unprecedented spatial resolutions and damage sensitivity levels [22]. Perhaps, the most impressive are the three-dimensional imaging approaches (X-ray computed tomography [23], optical coherence tomography [24], microwave holography [25], and terahertz time-domain spectroscopy [26]), which have recently shown capabilities for probing materials with thicknesses of more than 25.4 cm (10 in.) (e.g., ceramic foams) and at spatial resolutions approaching 1 μm [27] (*see* **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**).

The detailed characterization of surface and near-surface damage (e.g., surface-breaking cracks) using laser ultrasound [28], thermosonics [29], and optical coherence tomography [24] methods have also produced damage sensitivity levels, which now permit damage initiation, damage growth, and damage precursor indications to be measured. The whole-field imaging of structural microcracks using thermosonics, in particular, has been shown to be a powerful tool for imaging cracks, disbonds, and delaminations in complex geometry aerospace structures with damage features approaching 10 μm in size [29] (*see* **Full-field Sensing: Three-dimensional Computer Vision and Digital Image Correlation for Noncontacting Shape and Deformation Measurements**).

The detection of hidden damage has also been reported recently using penetrating, nonionizing radiation in the terahertz [26], microwave [30], and millimeter-wave frequency bands, where damage hidden beneath aerospace coatings, insulation, composite, and ceramic materials has been reported [26, 30]. Significant advances in active and passive thermography have also been made in a large part due to improvements in focal plane array camera technologies, which have dramatically improved spatial resolution, thermal sensitivity, and noise rejection levels [31–33] (*see* **Thermal Imaging Methods**).

2 PHYSICAL PRINCIPLES

Directed energy sensing fundamentally involves the interaction of electromagnetic energy with a material system. The sensing process can involve a number of different physical processes depending on the type of electromagnetic radiation used and the specific characteristics of the materials involved (*see* **Electric and Electromagnetic Properties Sensing**). A decrease in X-ray beam intensity, for example, can be attributed to an increase in material density or sample thickness based on material absorption

principles, while a phase shift in a reflected laser beam at visible wavelengths may be due to surface motions of a vibrating material surface. In this section, a description of the physical principles involved in directed energy sensing is given, where similarities and differences between the electromagnetic frequencies are highlighted.

2.1 Electromagnetic radiation

Electromagnetic radiation propagates through space as an oscillating electric and magnetic field. The two field components are in phase, and oscillate at right angles to each other and to the direction of propagation. In 1861, Maxwell derived what is referred to as the electromagnetic *wave equation*, which governs the propagation of electromagnetic energy in free space and within a given material [34]. In general, electromagnetic radiation can be classified according to the frequency of the wave: radio waves, millimeter waves, microwaves, terahertz radiation, infrared radiation, visible light, ultraviolet radiation, and X rays. Figure 1 shows a diagram of the electromagnetic spectrum, where key frequency ranges are highlighted along with technologies associated with each frequency range.

An electromagnetic wave, like other wave phenomena, can be characterized by its wavelength (the distance from a point on one cycle to the corresponding point on the next cycle) or its frequency (the number of oscillations per second). In a vacuum, all electromagnetic waves travel at the same speed, the speed of light, $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$. The wavelength, λ, and frequency, ν, of an electromagnetic wave are related by the equation

$$\nu = \frac{c}{\lambda} \tag{1}$$

which holds true for all forms of electromagnetic radiation. The energy of an electromagnetic wave is related to its frequency and wavelength by the relationship

$$E = h\nu = h\left(\frac{c}{\lambda}\right) \tag{2}$$

where h is a constant known as *Planck's constant*.

2.2 Radiation and matter

Depending on the frequency used, electromagnetic energy can interact with matter in very different ways. In a classical sense, electromagnetic energy interacts with materials as a localized field or wave phenomenon through material properties like the complex dielectric, the index of refraction, and the absorption coefficient. In a quantum mechanical sense, photons of electromagnetic energy interact with specific energy states in a material through

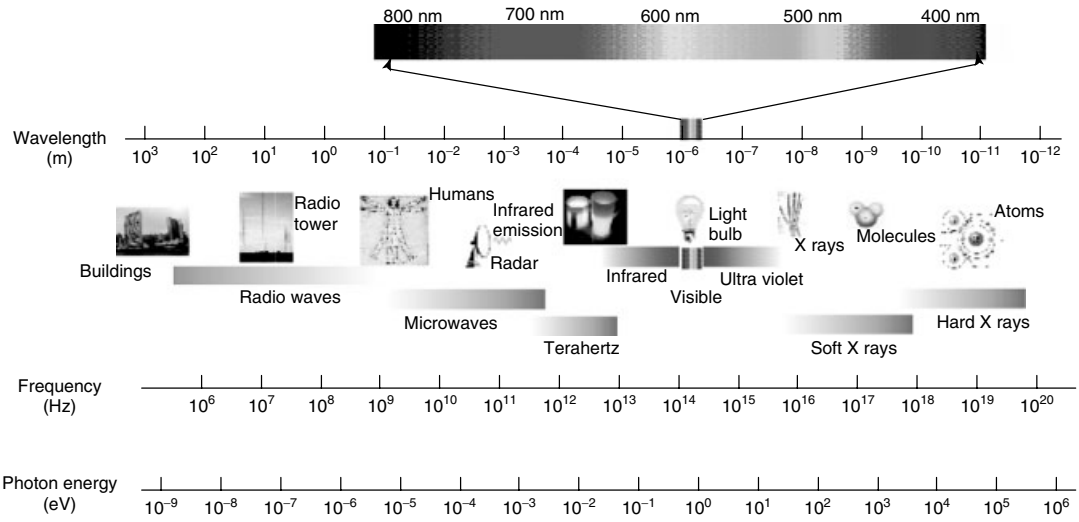


Figure 1. Electromagnetic spectrum with associated length scales and technologies.

probability functions and allowable transitions. Both of these concepts play a role in directed energy sensing (*see* **Electric and Electromagnetic Properties Sensing**).

In the classical approach, matter consists of a positive charge centers surrounded by clouds of negative charge [35]. In the presence of an electromagnetic field, the charge cloud distorts, which constitutes a simple dipole through the dipole moment. Mathematically, this process can be described using the dielectric properties of the material through a complex quantity denoted by

$$\varepsilon = \varepsilon' - \mathrm{j}\varepsilon'', \varepsilon_\mathrm{r} = \varepsilon_\mathrm{r}' - \mathrm{j}\varepsilon_\mathrm{r}'', \varepsilon = \frac{\varepsilon}{\varepsilon_0} \qquad (3)$$

where ε is called the absolute complex dielectric constant of the medium, ε_r is the relative complex dielectric constant, ε' is the permittivity (the ability of the material to store electromagnetic energy), ε'' is the dielectric loss factor (the ability of the material to absorb electromagnetic energy), and $\varepsilon_0 = 8.85419 \times 10^{-12}\,(\mathrm{F\,m^{-1}})$ is the permittivity of free space [30].

For a plane electromagnetic wave propagating in the z direction and polarized along the x axis inside a medium with a dielectric constant of $\varepsilon_\mathrm{r} = \varepsilon_\mathrm{r}' - \mathrm{j}\varepsilon_\mathrm{r}''$, the electric field intensity at any point can be described by the relationship [30]

$$\vec{E}(z) = E_0 \mathrm{e}^{-(\alpha+\mathrm{j}\beta)z}\,\hat{a}_x = E_0 \mathrm{e}^{-\gamma z}\hat{a}_x,$$
$$\alpha = k_0 |Im\{\sqrt{\varepsilon_\mathrm{r}}\}| \quad \beta = k_0 Re\{\sqrt{\varepsilon_\mathrm{r}}\} \qquad (4)$$

where E_0 is the electric field intensity at $z = 0$, α is the absorption constant, β is the phase constant, $\gamma = \alpha + j\beta$ is the propagation constant, and $k_0 = 2\pi/\lambda_0$ is the wave number in free space, and λ_0 is the wavelength in free space. For a plane electromagnetic wave traveling through a thickness, d, of material, the product $\gamma \cdot d$ determines the amount of attenuation and phase shift that the wave will experience [30].

When an electromagnetic wave traverses a boundary between two media, the concept of index of refraction becomes useful. Similar to the complex dielectric concept, the complex refractive index of a material determines how much the wave speed is reduced inside a medium and how much absorption loss occurs through the expression

$$\tilde{n} = n - \mathrm{i}k \qquad (5)$$

where n is the index of refraction, k is the extinction coefficient, and n is related to the material's relative permittivity, ε_r, and relative permeability, μ_r, through the expression $n = \sqrt{\varepsilon_\mathrm{r}}\sqrt{\mu_\mathrm{r}}$. The mathematical expressions describing the electromagnetic wave reflection and transmission at a boundary are referred to as the *law of reflection* and *Snell's law*, which are depicted schematically in Figure 2, and are given by the expressions: $\theta_\mathrm{i} = \theta_\mathrm{r}$ and $n_\mathrm{i} \sin\theta_\mathrm{i} = n_\mathrm{t} \sin\theta_\mathrm{t}$, respectively, and where $\theta_{\mathrm{i,r,t}}$ are the angles of incidence, reflectance, and transmission.

In addition to the reflection, refraction, and transmission of light, the absorption and scattering of electromagnetic energy by a material represents a key aspect of most directed energy measurement approaches (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). In particular, the scattering of wave energy becomes important when the wavelength of an electromagnetic wave approaches (Mie scattering), or becomes larger than (Rayleigh scattering), the size of a scattering object.

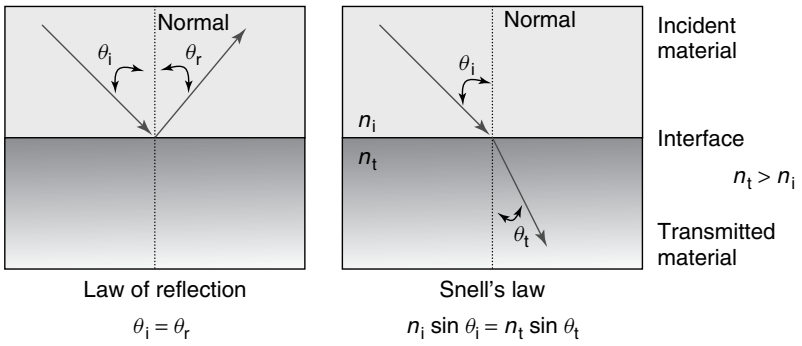


Figure 2. Reflection and refraction of electromagnetic radiation at a material interface.

For an electromagnetic wave incident on a single small scattering object, the intensity, I, of the scattered light can be written as follows:

$$I \cong I_0 \frac{1+\cos^2\theta}{2R^2}\left(\frac{2\pi}{\lambda}\right)^4\left(\frac{n^2-1}{n^2+2}\right)^2\left(\frac{d}{2}\right)^6 \quad (6)$$

where R is the distance from the particle, θ is the scattering angle, n is the refractive index of the scattering object, and d is the nominal size of the object. An important feature of equation (6) is the intensity of the scattered energy scaling inversely with the fourth power of the wavelength.

The absorption of electromagnetic energy by a material is traditionally described by Beer's law, which states that a logarithmic dependence exists between the transmission of light through a substance, the concentration/density of the substance, and the length of material that the light travels through. If the linear absorption (or attenuation) coefficient is defined as μ, then Beer's law can be written as follows:

$$I_z = I_0 \exp -[\mu z] \quad (7)$$

where I_z is the intensity of the beam at a distance z and I_0 is the intensity at the specimen surface. The primary factor that determines how radiation will be absorbed by a material is based on the atoms and molecules that make up the material. At a very fundamental level, the energy of an atomic and molecular system is quantized (i.e., made up of discrete levels). If a particular electromagnetic radiation energy matches one of these quantized energy states, then a strong interaction will result (e.g., absorption, reflection, refraction, and scattering), and if there are no available energy levels that match the energy of the incident radiation, then the material will typically be transparent to that radiation. In the following subsections, this basic concept is built upon, where the major features of each frequency range are covered.

2.2.1 X-ray interactions

The quantum energy of X-ray photons is ~124 eV and greater, which is much too high to be absorbed in electron transitions between states for most atoms and molecules [36–39]. Because of this fact, most X rays penetrate through materials, with only the occasional X ray knocking an electron completely out of an atom/molecular system. During this process, the X ray can give up all of its energy to the electron (photoionization), or it can give up part of its energy (Compton scattering). A third possibility also exists if the X ray has sufficient energy, resulting in the creation of an electron–positron pair. A few electron volts of photon energy are typically required to eject an electron and ionize an atom, which places the threshold for ionization somewhere in the ultraviolet region of the electromagnetic spectrum. The X-ray wavelength range is ~10 nm and shorter, while the frequency range is $\sim 3 \times 10^{16}$ Hz or greater. X-ray energies below 10 keV are typically referred to as *soft* X rays, which are used for porous or thin material measurements, while more energetic X rays above 10 keV are referred to as *hard* X rays for dense, thick material measurements.

2.2.2 Visible and ultraviolet light interactions

The quantum energy of visible and ultraviolet photons is in the range ~1.65–124 eV, which is the dominant range of energies for elevating bound electrons to higher energy levels within an atomic or molecular system [36–39]. There are typically many available states in most material systems, so visible and ultraviolet light are strongly absorbed by most materials. The shorter ultraviolet wavelengths can reach the ionization energy for some molecules, which permits them to act in a similar fashion to "soft" X rays, while the net result of an absorption event by nonionizing visible radiation is generally just to heat the material sample. The wavelength range for visible–ultraviolet radiation is from ~750 to 10 nm, while the frequency range is from $\sim 4 \times 10^{14}$ to 3×10^{16} Hz.

2.2.3 Infrared interactions

The quantum energy of infrared photons is in the range 0.001–1.7 eV, which corresponds to the range of energies required for separating the quantum states of molecular vibrations [36–39]. Infrared radiation is typically absorbed more strongly than terahertz and microwave frequencies, but less strongly than visible light. The result of an infrared absorption event is the heating of the material due to increased molecular vibrational activity. Infrared radiation can

penetrate further into most materials relative to visible light due in part to its longer wavelength and in part to the reduced number of available quantum states for infrared energy coupling with the material. The wavelength range for infrared radiation is from $\sim$30 μm to 750 nm, while the frequency range is $\sim 0.003-4 \times 10^{14}$ Hz.

In addition to the absorption of electromagnetic energy, infrared energy can flow or move within a material (thermal diffusion), and it can be radiated or emitted by a material according to Planck's law for blackbody radiation. The thermal transport or diffusion of heat energy within a material is important for infrared directed energy measurements, where active heating is used (e.g., pulsed thermography). For passive measurements, the emission of thermal radiation from a material is important. This "emissivity" depends on factors such as temperature, emission angle, and wavelength [31–33] (*see* **Thermal Imaging Methods**).

2.2.4 Terahertz, microwave, and millimeter-wave interactions

The quantum energy of terahertz, microwave, and millimeter-wave photons is in the range $\sim 0.00001-0.001$ eV, which is in the range of quantum state energies for molecular rotation and torsion [26, 30, 36–39]. Terahertz energy is absorbed more strongly than microwaves, but less strongly than infrared light. The interaction of terahertz/microwave radiation with matter results in the rotation of molecules and the production of heat as a result of that molecular motion. Conductors strongly absorb microwaves (and lower frequencies) because they cause electric currents to form, which quickly heats the material. Most dielectric materials are largely transparent to both terahertz and microwaves energies/frequencies. Because the quantum energies of terahertz and microwave radiation are approximately million times lower than those of X rays, they do not produce ionization and are, therefore, a safer method for characterizing materials when depth penetration is needed. Most microwave applications fall in the range 3000–30 000 MHz (3–30 GHz). Current microwave ovens operate at a nominal frequency of 2450 MHz, a band assigned by the Federal Communications Commission (FCC). There are also some amateur and radio navigation uses of the 3–30 GHz range. The wavelength range for terahertz radiation is from $\sim$30 μm to 1 mm, while the frequency range is from $\sim 3 \times 10^{11}$ to 1×10^{12} Hz. The wavelength range for microwave (and millimeter-waves) radiation is $\sim$1 mm and greater, while the frequency range is $\sim 3 \times 10^{11}$ Hz and smaller.

2.3 Directed energy measurements

As stated earlier, directed energy sensing methods use electromagnetic energy to probe and characterize a material or structure. In most cases, active illumination is used to irradiate the material at some standoff distance, and a light-sensitive detector is used to collect the transmitted, reflected, or scattered electromagnetic energy from the material system (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). Material characterization or damage identification is then accomplished by analyzing the collected signals and applying the appropriate physics to extract material information of interest.

Figure 3 depicts the three primary approaches used in directed energy sensing measurements. If the electromagnetic radiation is not transmissive (or partially transmissive), then single-sided measurements are typically used (Figure 3a and b), where electromagnetic energy reflected from the material surface or scattered from within the near-surface region of the material is collected and analyzed. Most ultraviolet, visible, and infrared wavelength methods fall into this category. Some terahertz time-domain spectroscopy and near-field microwave methods also use the single-sided detection method depicted in Figure 3(a). If the electromagnetic radiation is transmissive, then volumetric measurements are possible using the through-transmission measurement approach depicted in Figure 3(c).

A wide variety of sources are available from broadband radiating sources (e.g., thermal radiation blackbody sources as previously described), to X-ray tubes, to microwave horns, to laser sources. An equally broad range of detectors are also available from semiconductor diodes, photomultiplier tubes, to photosensitive films and electronic imaging arrays. In most cases, the goal of a directed energy measurement is to collect the electromagnetic radiation energy and convert it to an electrical

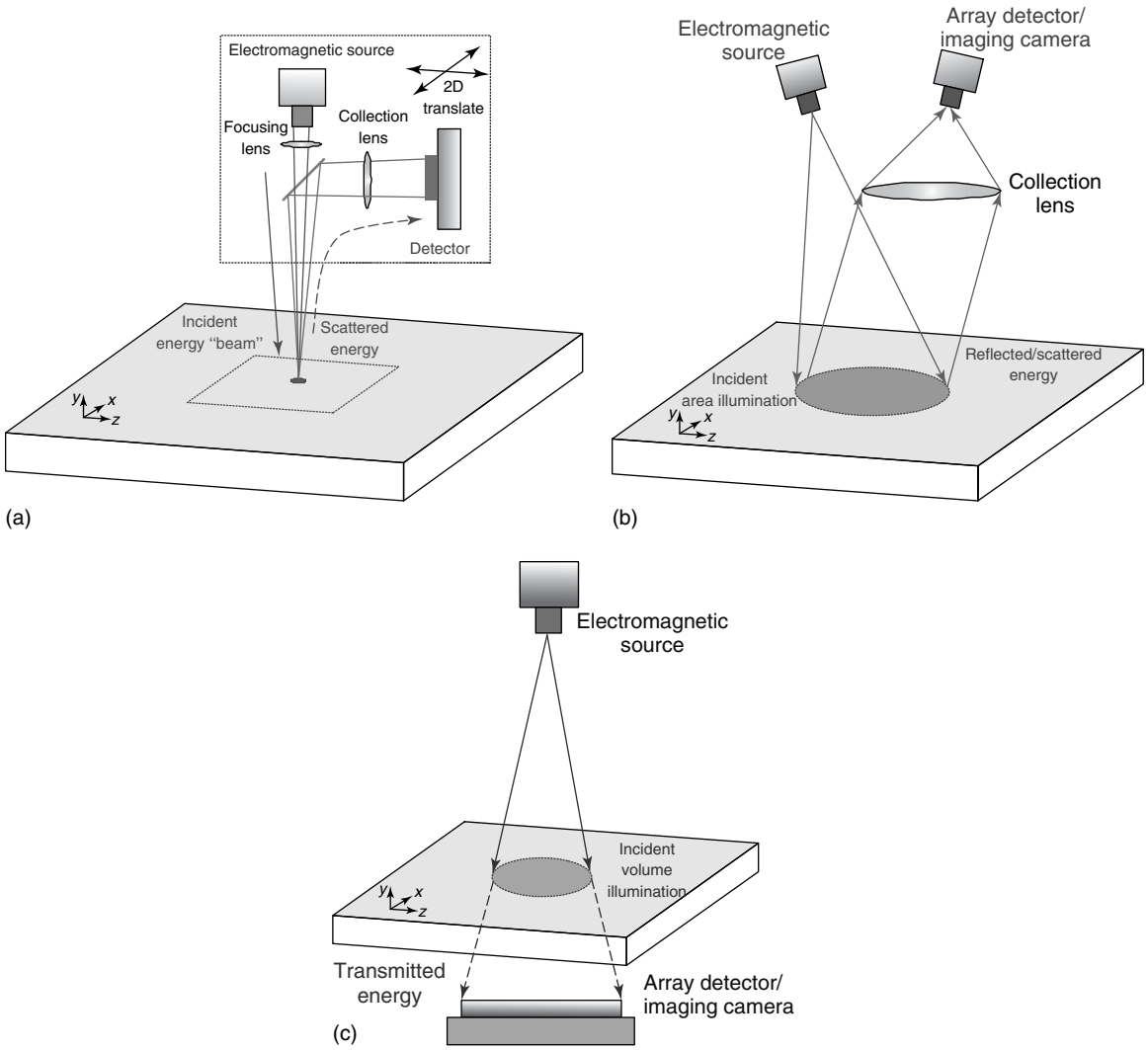


Figure 3. Directed energy measurement approaches: (a) single-sided, single-point measurement; (b) single-sided, imaging measurement; and (c) through-transmission, volumetric measurement.

signal, which can then be further analyzed. Simple measurements involve detecting raw intensity, while more sophisticated measurements keep track of phase, amplitude, and frequency. Two-dimensional imaging systems and three-dimensional computed tomographic systems also keep track of spatial position. Raster scanning of the single-point measurement system depicted in Figure 3(a) permits two-dimensional information to be obtained, while lenses and other beam manipulation systems permit whole-field images to be captured.

3 DIRECTED ENERGY SENSING METHODS

In this section, a variety of directed energy measurement examples are provided for electromagnetic energy in the X-ray, visible, infrared, terahertz, and microwave frequency bands. The examples show only a very small sampling of the available directed energy methods, which currently number in the hundreds to thousands. The interested reader is encouraged to

refer to the numerous cited articles, and also to the books by Cartz [40], Demtroder [41], Scruby and Drain [28], Mittleman [26], and Zoughi [42] for additional methods and more detailed information.

3.1 Radiographic methods

Radiography is one of the five traditional nondestructive evaluation methods [40] (*see* **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**). In recent years, technological advances in sources, detectors, and analysis methods have resulted in significant improvements in spatial resolution and sensitivity levels [43]. In addition, computed tomographic approaches have provided unmatched three-dimensional measurement capabilities for complex geometry structures and dense material samples [44]. Although different types and levels of radiographic energy can be used, X rays are most often used for most materials, while film radiography is quickly being replaced by electronic detector systems.

In most X-ray measurements, the basic setup depicted in Figure 3(c) is used, where the output signal is related to the material density according to an expression similar to equation (7). Subtle variations in material composition can also be characterized when dual-energy approaches are used [45]. Figure 4 depicts a series of measurements taken with a Phillips MGC-03 film radiography system, which represents a medium-resolution X-ray capability with a nominal 0.4-mm focal spot size [46]. The current, exposure time, and distance for the measurements depicted in Figure 4 correspond to 5 mA, 1.5 min, and 1.016 m (40 in.), respectively. The material sample (a laser-machined reference standard) was placed on the film and exposed to the X rays at a certain kilovolt energy level. The X-ray film was subsequently digitized. The image on the far left was taken at 55 kV energy level, and shows a series of laser-machined triangle features with increasing sizes and depths. The five images on the right correspond to a magnified view of the smallest (1 mm) and deepest (260 μm) triangle taken at increasing kilovolt energy levels (55–80 kV). As the kilovolt energy level increased, the signal-to-noise ratio decreased from 27.11 at 55 kV to 2.7 at 80 kV, while the image contrast remained almost constant at a ratio of $\sim 1.2 : 1.0$.

The ability to create fully three-dimensional characterizations of solid structures represents one of the most important capabilities for radiography, where X-ray computed tomography can be used. An example of an X-ray computed tomography measurement is provided in Figure 5(a), where a cross-sectional cut is depicted for the reference standard depicted in Figure 4. In this case, an ARACOR Tomoscope system was used, which uses a 225-kV, 50-μm microfocus X-ray source, and a fiber-optic scintillator/charge-coupled device (CCD)

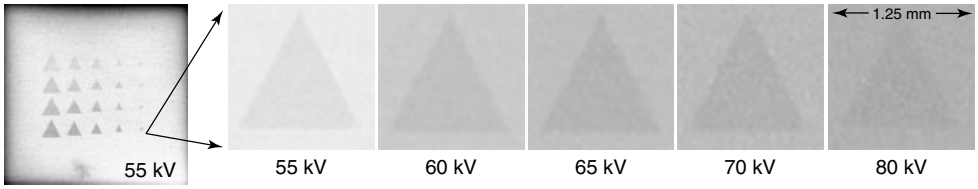


Figure 4. Radiographic through-transmission images of laser-machined triangle reference sample.

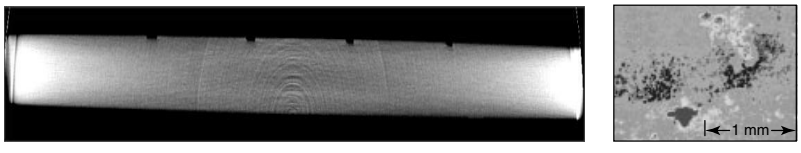


Figure 5. (a) Radiographic computer tomography measurement result showing a cross-section image through the laser-machined triangle sample and (b) X-ray measurement of corrosion.

image sensor system for detection. Thickness changes as small as 2% of the equivalent thickness were measured with an accuracy of ±1%. Figure 5(b) shows a similar measurement capability, where 1–5% corrosion material loss has been characterized [46].

3.2 Visible radiation and laser methods

Visible radiation represents one of the most prolific directed energy methods in use today. Spectroscopic measurements, in particular, provide a particularly useful characterization tool for identifying atomic and molecular states for a wide range of applications. As a material and system health monitoring capability, visible wavelength radiation has found use in a number of methods. Optical interferometry and holography, for example, provide a means for measuring surface topography and dynamic vibrations in a noncontact manner (*see* **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**). Figure 6 provides an example of an interferometric measurement system capable of measuring vibration features on a material surface with micron spatial resolution levels in the 25 kHz to 20 MHz frequency bandwidth range.

When combined with an ultrasonic wave or vibration source, the system depicted in Figure 6 can be used to characterize surface damage with microscopic precision. An example of such a measurement is provided in Figure 7, where the motion field of a surface-acoustic wave (SAW) has been imaged with the laser vibrometry system to detect and characterize a microscopic surface-breaking crack feature. The crack feature is measured as a "brightness" increase in the lower right image field due to increased vibration levels near the crack [47].

The crack image in Figure 7 was made by raster scanning the laser probe position in a two-dimensional manner, similar to the depiction in Figure 3(a). At each scan position, the out-of-plane motions were captured by the laser interferometry system, which was collected and analyzed to provide a displacement versus time signal level at each position on the material surface. Similar measurements have been made with a moderate intensity pulsed laser to excite elastic waves in the material. This type of directed energy measurement system is termed *laser ultrasound*, which has found use recently as an effective means for assessing the structural health status of composite materials. Similar to other directed energy methods, the laser ultrasound approach provides a noncontact measurement capability that can be done at a standoff distance.

3.3 Infrared methods

Thermographic imaging is a relatively new nondestructive evaluation (NDE) technology, which uses thermal differences between a material defect and its local surroundings as a noncontact and full-field NDE measurement tool (*see* **Thermal Imaging Methods**). Infrared cameras with higher sensitivities and resolutions are currently helping to transition the technology from a novelty to a comprehensive and quantitative NDE measurement tool. Materials are characterized based on variations in thermal absorption, transport, diffusion, and emission. In a passive thermography measurement, the material surface naturally radiates,

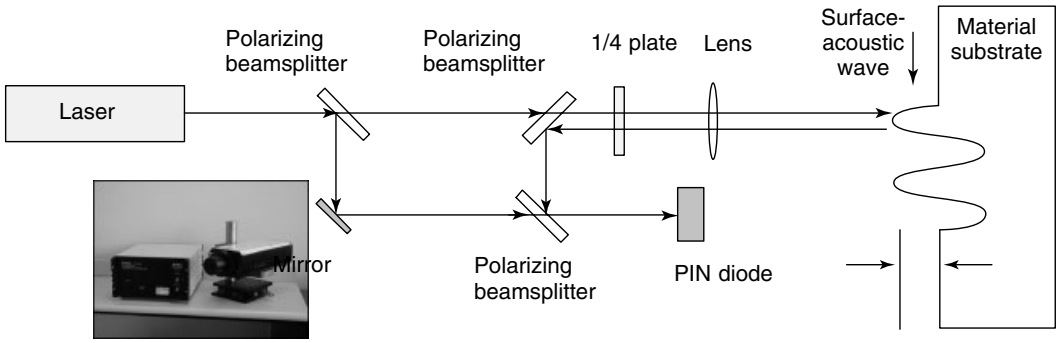


Figure 6. Laser interferometry system for measuring surface vibrations and displacements.

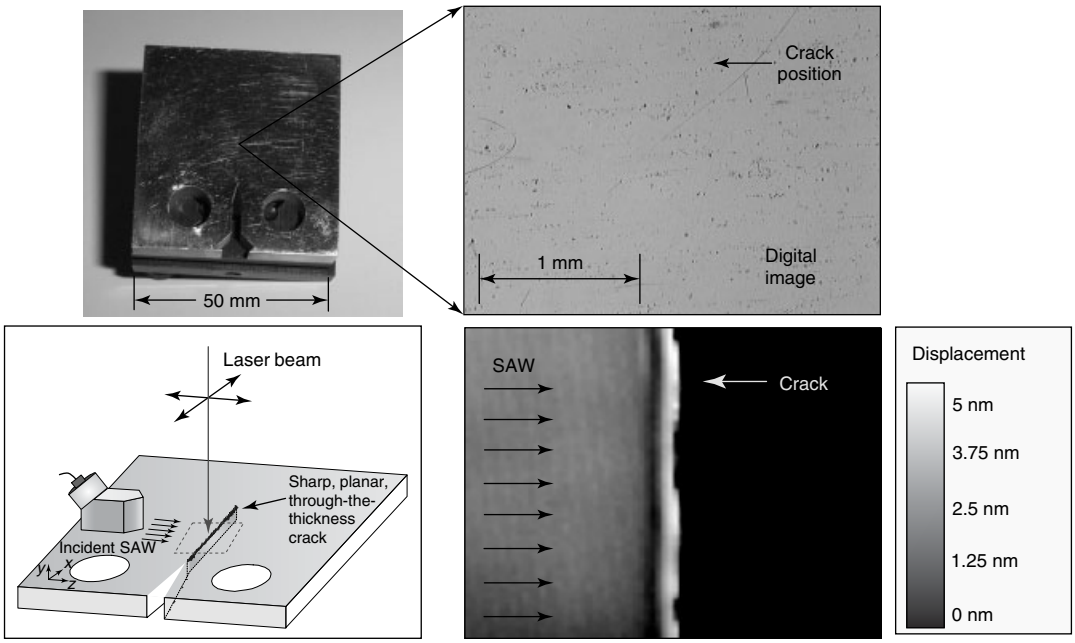


Figure 7. Laser interferometry measurement of surface-acoustic wave interaction with crack.

scatters, and reflects infrared energy that can be imaged by the thermal camera. Differences in thermal emissivity are imaged as differences in brightness by the camera, allowing corrosion, voids, and other damage to be detected and characterized.

In certain instances, thermographic measurements can be made in a spectral band-pass window that allows the infrared energy to propagate efficiently through a coating layer to probe the material substrate underneath [48]. This in fact is the case for many aerospace coating materials. Figure 8 depicts a spectral transmission window for a 3-mil-thick paint layer (∼75 microns) between the 2 and 12 μm wavelength range, which occurred between 3.5 and 5.8 μm, peaking at 5.2 μm. By using a midwave infrared camera sensitive to 3–5 μm thermal energies, and/or using band-pass filters in that wavelength range, damage (e.g., hidden corrosion) can be imaged directly through the paint with a simple mid-IR camera system (Figure 8a).

The combination of ultrasonics and thermal imaging has also recently shown promise as a new directed energy sensing method for detecting microscopic cracks, composite disbonds, and delaminations [29]. The method termed *thermosonics* or *sonic-IR* uses an infrared camera to monitor the local heating that can occur in a material when vibrational energy causes frictional rubbing of surfaces at crack boundaries or disbonded/delamination locations. Figure 9(a) depicts a typical thermosonics system, which was used to detect a microscopic fatigue crack (Figure 9c) in a turbine engine blade (Figure 9b).

3.4 Terahertz methods and microwave methods

Terahertz and microwave measurement systems, like thermal imaging systems, are relatively new technologies. For directed energy measurements, terahertz and microwave methods provide a unique ability for penetrating through most dielectric materials. When this is the case, both methods can provide a volumetric measurement capability similar to X-ray measurements without the safety hazard problems. In addition, simple and effective imaging capabilities may be possible soon, where direct imaging through various dielectric material layers may be possible with the appropriate choice of directed energy frequency/wavelength.

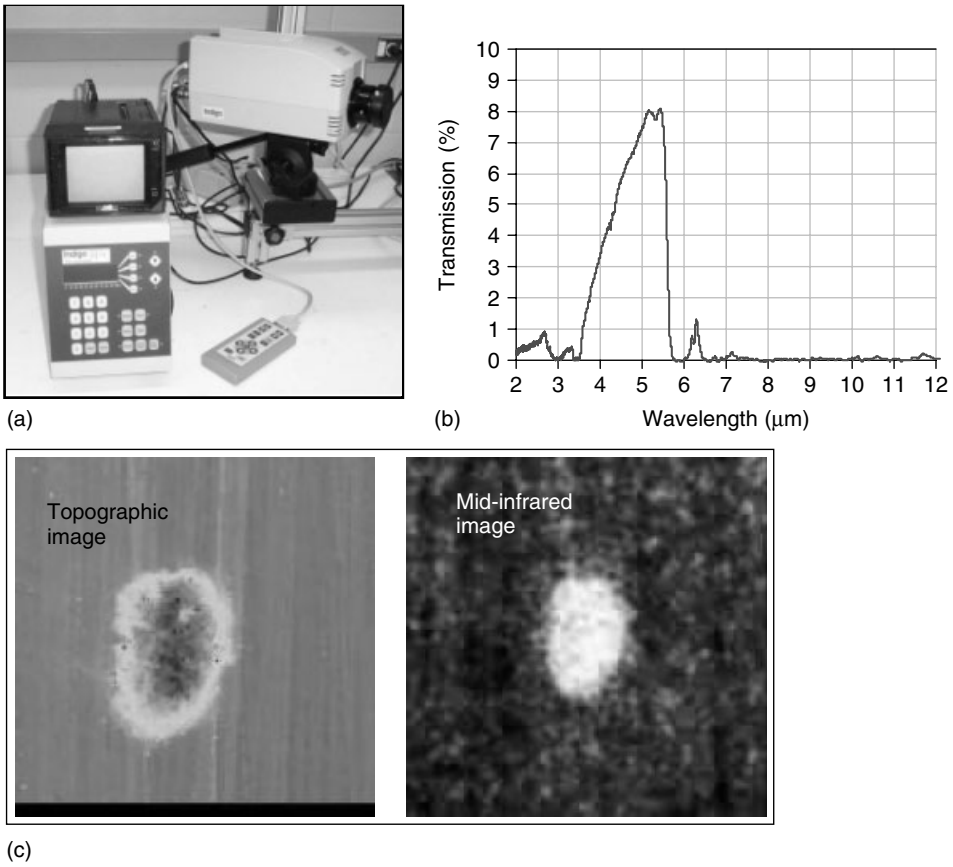


Figure 8. (a) Midinfrared camera system, (b) midinfrared transmission window through aerospace coating, and (c) passive infrared image of corrosion feature hidden under the coating.

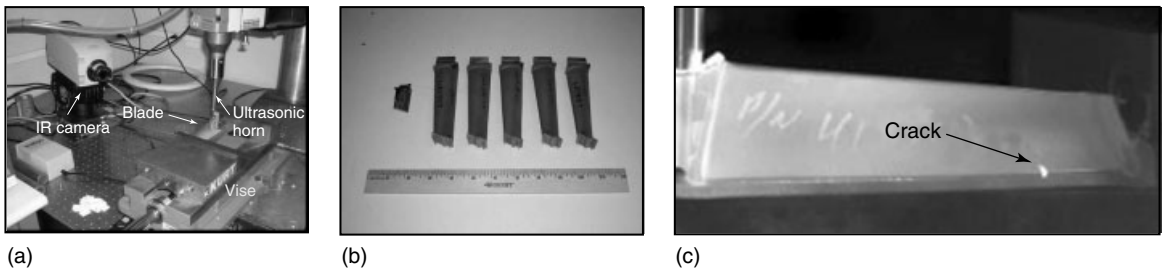


Figure 9. (a) Thermosonics system, (b) turbine engine blades, and (c) crack measurements.

An example of an evanescent microwave probe measurement taken through an aerospace coating is provided in Figure 10. An evanescent microwave probe is a coaxial resonator system that operates nominally at 1–4-GHz frequencies. The probe depicted in Figure 10(a) was resonant at ~2.485 GHz, and had a sharp coupling tip that permitted near-surface, evanescent field measurements, which enhanced spatial resolution levels and measurement sensitivity [48]. By raster scanning the probe over the sample surface, an image can be produced, which represents changes in dielectric properties of

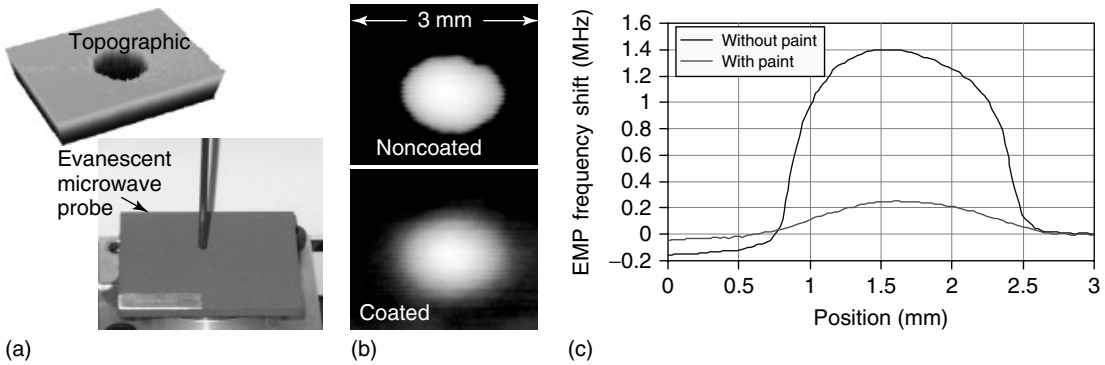


Figure 10. (a) Evanescent microwave probe system, (b) microwave measurements taken with and without coating layer present, and (c) comparison of response signals across corrosion pit.

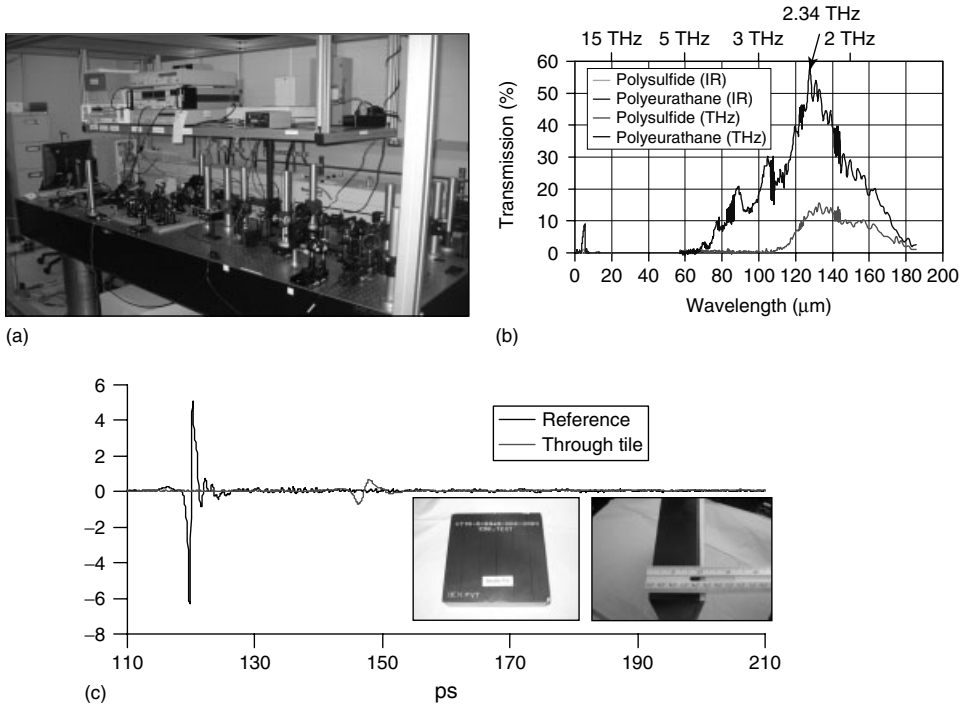


Figure 11. (a) Terahertz time-domain spectroscopy system, (b) terahertz spectral transmission through aerospace coatings near 2.3 THz, and (c) terahertz transmission through foam material.

the material according to equations (3) and (4). The presence of a hidden corrosion pit feature through an aerospace coating is depicted in Figure 10(b) and (c).

Terahertz measurement systems have also become available recently, with the promise of microwave-type penetration capabilities in reduced sizes and with improved spatial resolutions. Terahertz time-domain spectroscopy systems, in particular, have recently shown promise for measuring damage and material state through coatings, foam materials, ceramics, and composites [48]. Figure 11 depicts a typical time-domain spectroscopy system along with two measurement examples showing terahertz transmission properties through aerospace coating materials

(Figure 11b), and thermal protection system foam insulation materials (Figure 11c).

4 CONCLUSIONS

Directed energy sensing methods are becoming more capable and are finding more uses for nondestructive evaluation and structural health monitoring applications. Recent technological breakthroughs in numerous electromagnetic radiation sources and detectors are providing sensing capabilities across the entire electromagnetic spectrum from X rays to millimeter waves and beyond. In this brief review article, the key underlying principles of directed energy sensing are presented with a special emphasis on material inspections, nondestructive evaluation, and structural health monitoring applications. Specific examples are provided for directed energy sensing with X rays, laser ultrasonics, passive thermal imaging, thermosonics, evanescent microwaves, and terahertz time-domain spectroscopy. It is anticipated that directed energy sensing will continue to expand in capabilities and uses in the future, providing improved noncontact inspection capabilities for many years to come.

REFERENCES

[1] Matzkanin G, Easter J. *NDE of Hidden Corrosion—A Report Update; NTIAC Report*, NTIAC-SR-04-03. Nondestructive Testing Information Analysis Center: San Antonio, TX, 2004.

[2] Roentgen WC. On a new kind of ray. *Nature* 1896 **53**:274.

[3] Pozar DM. *Microwave Engineering*. Addison-Wesley: Boston, MA, 1993.

[4] Buderi R. *The Invention that Changed the World: The Story of Radar from War to Peace*. Simon & Schuster: New York, 1996.

[5] Descartes R. *The Geometry*. Dover Publications: New York, 1954, (1637).

[6] Newton I. *Opticks*. Dover Publications: New York, 1979, (1704).

[7] Mahoney MS. *The Mathematical Career of Pierre de Fermat*. Princeton University Press: Princeton, NJ, 1994, pp. 1601–1665.

[8] Struik DJ. Snel, Willebror. In *Dictionary of Scientific Biography XII*. Charles Scribner's Sons: New York, 1980.

[9] Huygens C. *Treatise of Light*. Dover Publications: New York, 1962, (1690).

[10] Silliman RH. Fresnel, Augustin Jean. In *Dictionary of Scientific Biography XIII*. Charles Scribner's Sons: New York, 1990.

[11] de Broglie L. *Matter and Light: The New Physics*. Dover Publications: New York, 1959.

[12] Bragg W. *The Universe of Light*. Dover Publications: New York, 1959.

[13] Sabra AI. *Theories of Light from Descartes to Newton*. Cambridge University Press: Cambridge, MA, 1981.

[14] Eisberg R, Resnick R. *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*, *Second Edition*. John Wiley & Sons: New York, 1985, pp. 59–60.

[15] Maxwell JC. On physical lines of force. *Philosophical Magazine* 1861 **21**:161–175.

[16] Boltzmann L, Kirchhoff GR. *Populäre Schriften*. Verlag von J.A. Barth: Leipzig, 1905.

[17] Buchwald JZ. *The Creation of Scientific Effects: Heinrich Hertz and Electric Waves*. University of Chicago Press: Chicago, IL, 1994.

[18] Planck M. *Treatise on Thermodynamics*. Dover Publications: New York, 1922.

[19] Rüchardt E. Zur Erinnerung an Wilhelm Wien bei der 25. Wiederkehr seines Todestages. *Naturwissenschaften* 1955 **42**(3):57–62.

[20] Einstein A. On a heuristic viewpoint concerning the production and transformation of light. *Annalen der Physik* 1905 **17**:132–148.

[21] Bohr N. On the constitution of atoms and molecules. *Philosophical Magazine* 1913 **6**(26):1–25.

[22] Morre P, McIntire P (eds). *Nondestructive Testing Handbook*, *Special Nondestructive Testing Methods*. ASNT Press: New York, 1995; Vol. 9.

[23] Morgan CL. *Basic Principles of Computed Tomography*. University Park Press: Baltimore, MD, 1983.

[24] Schmitt J. Optical coherence tomography (OCT): a review. *IEEE Selected Topics in Quantum Electronics* 1999 **5**(4):1205–1215.

[25] Case J, Randazzo A, Pastorino M, Zoughi R. Evaluation of reconstruction error in microwave holographic imaging with reduced data sets. *Proceedings*

of the Mediterranean Microwave Symposium. Genoa, 2006; pp. 137–140.

[26] Mittleman D (ed). *Sensing with Terahertz Radiation*. Springer: New York, 2003.

[27] Chen H, Kersting R, Cho G. Terahertz apertureless scanning near-field optical microscopy with nanometer resolution. *Applied Physics Letters* 2003 **83**:15.

[28] Scruby C, Drain L. *Laser Ultrasonics—Techniques and Applications*. Adam Hilger: Bristol, 1990.

[29] Favro L, Thomas R, Han X, Ouyang Z, Newaz G, Gentile D. Sonic infrared imaging of fatigue cracks. *International Journal of Fatigue* 2001 **23**:471–476.

[30] Zoughi R, Ganchev S. *Microwave Nondestructive Evaluation; NTIAC Report*, NTIAC-95-01. Nondestructive Testing Information Analysis Center: San Antonio, TX, 1995.

[31] Shull J. *Nondestructive Evaluation: Theory, Techniques, and Applications*. Marcel Dekker: New York, 2002.

[32] Maldague X. *Infrared Methodology and Technology*. CRC Press: Boca Raton, FL, 1994.

[33] Gaussorgues G. *Infrared Thermography*. Springer: Berlin, 1994.

[34] Maxwell JC. *Electricity and Magnetism*. Dover Publications: New York, 1959, (1861).

[35] Jackson JD. *Classical Electrodynamics*. John Wiley & Sons: New York, 1962.

[36] Chen S, Kotlarchyk M. *Interaction of Photons and Neutrons with Matter*. World Scientific Publishing: Hackensack, NJ, 1997.

[37] Moseley P, Crocker A. *Sensor Materials*. CRC Press: Boca Raton, FL, 1996.

[38] MacDonald N. *Nuclear Structure and Electromagnetic Interactions*. Plenum Press: New York, 1965.

[39] Weider R, Sells R. *Elementary Modern Physics*. Allyn & Bacon: Boston, MA, 1973.

[40] Cartz L. *Nondestructive Testing*. ASM International: Materials Park, OH, 1996.

[41] Demtroder W. *Laser Spectroscopy—Basic Concepts and Instrumentation*. Springer-Verlag: Berlin, 1998.

[42] Zoughi R. *Microwave Nondestructive Testing and Evaluation*. Kluwer Academic Publishers: Boston, MA, 2000.

[43] Zoofan B, Rokhlin S. Microradiographic detection of corrosion pitting. *Materials Evaluation* 1998 **56**:191–194.

[44] Hagemaier D. Aerospace radiography—the last three decades. *Materials Evaluation* 1985 **43**: 1262–1283.

[45] Guillemaud R, Robert-Coutant C, Darboux M, Gagelin J, Dinten J. Evaluation of dual-energy radiography with a digital X-ray detector. *Proceedings of SPIE* 2001 **4320**:469–478.

[46] Blackshire J, Hoffmann J, Kropas-Hughes C, Tansel I. Microscopic NDE of hidden corrosion. *Proceedings of SPIE* 2003 **5045**:93–103.

[47] Blackshire J, Sathish S. Near-field ultrasonic scattering from surface-breaking cracks. *Applied Physics Letters* 2002 **80**:3442–3444.

[48] Blackshire J, Buynak C, Steffes G, Marshall R. Nondestructive evaluation through aircraft coatings: a state-of-the-art assessment. *9th Joint FAA/DoD/NASA Aging Aircraft Conference*. Atlanta, GA, 6–9 March 2006.

Chapter 145

Noncontact Rail Monitoring by Ultrasonic Guided Waves

Piervincenzo Rizzo[1], Stefano Coccia[2], Ivan Bartoli[2] and Francesco Lanza di Scalea[2]

[1] *Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA, USA*
[2] *Department of Structural Engineering, University of California, San Diego, CA, USA*

1 INTRODUCTION

Safety statistics data from the US Federal Railroad Administration (FRA) [1] indicate that train accidents caused by track failures including rail, joint bars, and anchoring resulted in 2700 derailments and $441 million in direct costs during the decade 1992–2002. The primary cause of these accidents is the "transverse-type defects" that was found responsible for 541 derailments and $91 million in cost during the same time period. Transverse defects are cracks developing in a direction perpendicular to the rail running direction, and include transverse fissures (TFs, initiating in a location internal to the rail head) and detail fractures (DFs, initiating at the head surface as rolling contact fatigue defects) (Figure 1).

The most common methods of rail inspection are magnetic induction and contact ultrasonic testing [2, 3]. The first method is affected by environmental magnetic noise and it requires a small liftoff distance for the sensors to produce adequate sensitivity. Ultrasonic testing is conventionally performed from the top of the rail head in a pulse-echo configuration. Such an approach suffers from a limited inspection speed and from other drawbacks associated with the requirement for contact between the rail and the inspection wheel. More importantly, horizontal surface cracks such as shelling and head checks can prevent the ultrasonic beams from reaching the internal defects (IDs), resulting in false negative readings. These problems were highlighted in the derailments of Superior, WI, USA, in 1991 and Hatfield, UK, in 2000.

Guided ultrasonic waves are being considered in recent years for rail inspections as an improvement over wheel-type ultrasonic methods [4–11]. Guided

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

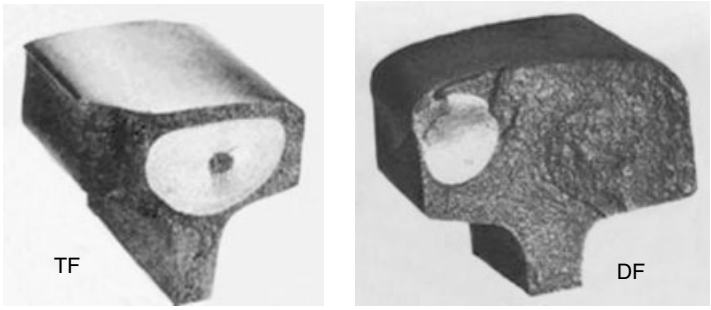


Figure 1. Transverse defects in rail: a transverse fissure (TF) and a detail fracture (DF).

waves propagate along, rather than across the rail, and are thus ideal for detecting the critical transverse defects. Guided waves are also potentially not sensitive to surface shelling since they can run underneath this discontinuity.

Techniques that do not require contact with the rail are being investigated to generate and detect guided waves. Noncontact rail testing has been demonstrated by the use of pulsed lasers and air-coupled transducers [7, 11], electromechanical acoustic transducers [7, 12, 13], and laser vibrometers [6]. However, the drawback of any noncontact ultrasonic testing is a reduced signal-to-noise ratio of the defect detection procedure when compared with conventional contact testing. The use of signal processing based on the discrete wavelet transform (DWT) helps overcoming this problem, as recently demonstrated [11, 14–16].

This article describes a rail inspection system based on guided ultrasonic waves generated and detected in a noncontact manner. The system uses a pulsed laser and an array of air-coupled sensors. Defects in the rail head are detected by monitoring changes in the energy strength of the transmitted waves. The raw ultrasonic signals are processed by the DWT in pseudo-real-time for increasing the defect detection reliability. Laboratory and field results of this system will be presented.

2 WAVE PROPAGATION MODELING

Proper modeling of guided waves propagating in the rail is important to identify modes, enhance sensitivity to certain defects, and select low-loss mode–frequency combinations. In the context of noise suppression, theoretical solutions of waves propagating in rails have been obtained below 6 kHz using beam theories [17]. For ultrasonic testing, these solutions have been extended to 50 kHz by using semianalytical finite element (SAFE) methods [18], and up to 350 kHz by using finite element analyses with cyclic symmetry [6]. The arbitrary shape of the cross section of a rail lends itself to SAFE modeling, where only a 2D mesh is required.

In the SAFE method, at each frequency ω, a discrete number of guided modes is obtained. For the given frequency, each mode is characterized by a wavenumber ξ, and by a displacement distribution over the cross section (mode shape). The SAFE algorithm used here is based on the three-dimensional theory of linear viscoelasticity. For a Cartesian reference system, the waveguide cross section is set in the y–z plane while the x axis is parallel to the waveguide length. Subdividing the cross section via finite elements, the approximated displacement at a point $\mathbf{u} = \mathbf{u}(x, y, z, t)$ is given by

$$\mathbf{u} = \mathbf{N}\mathbf{U}^{(e)} \mathrm{e}^{\mathrm{i}(\xi x - \omega t)} \tag{1}$$

where $\mathbf{N} = \mathbf{N}(y, z)$ is the matrix of the shape functions, $\mathbf{U}^{(e)}$ is the nodal displacement vector for the eth element, t is the time variable, and i the imaginary unit. It can be noted that the displacement is described by the product of an approximated finite element field over the waveguide cross section and the exact time harmonic function, $\exp[\mathrm{i}(\xi x - \omega t)]$, in the propagation direction x. The compatibility and constitutive equations can be written in synthetic matrix forms as follows:

$$\boldsymbol{\varepsilon} = \mathbf{D}\mathbf{u}, \ \boldsymbol{\sigma} = \mathbf{C}^{*}\boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}$ are the strain and stress vector, respectively, $\mathbf{D}$ is the compatibility operator, and $\mathbf{C}^*$ is the complex constitutive linear viscoelastic tensor. More details on the compatibility operator can be found in [19]. The principle of virtual works with the compatibility and constitutive laws (equation 2) leads to the following energy balance equation:

$$\int_{\Gamma} \delta \mathbf{u}^{\mathrm{T}} \mathbf{t} \, \mathrm{d}\Gamma = \int_{V} \delta \mathbf{u}^{\mathrm{T}} (\rho \ddot{\mathbf{u}}) \, \mathrm{d}V + \int_{V} \delta (\mathbf{uD})^{\mathrm{T}} \mathbf{C}^* \mathbf{Du} \, \mathrm{d}V \tag{3}$$

where Γ is the waveguide cross-section area, V is the waveguide volume, $\mathbf{t}$ is the external traction vector, and the overdot means time derivative. The finite element procedure reduces equation (3) to a set of algebraic equations:

$$[\mathbf{A} - \xi \mathbf{B}]_{2M} \mathbf{Q} = \mathbf{p} \tag{4}$$

where the subscript $2M$ indicates the dimension of the problem, with M the number of total degrees of freedom of the cross-sectional mesh. Details on the complex matrices $\mathbf{A}$,$\mathbf{B}$, and vector $\mathbf{p}$ can be found in [18]. By setting $\mathbf{p} = \mathbf{0}$ in equation (4), the associated eigenvalue problem can be solved as $\xi(\omega)$. For each frequency ω, $2M$ complex eigenvalues, ξ_m, and $2M$ complex eigenvectors, $\mathbf{Q}_m$, are obtained. The solution is symmetric, i.e., for each pair $(\xi_m - \mathbf{Q}_m)$, representing a forward guided mode, a pair exists representing the corresponding backward mode. The first M components of $\mathbf{Q}_m$ describe the cross-sectional mode shapes of the mth mode. Once ξ_m is known, the dispersion curves can be easily computed. The phase velocity can be evaluated by the expression $c_{\mathrm{ph}} = \omega/\xi_{\mathrm{real}}$, where ξ_{real} is the real part of the wavenumber. The imaginary part of the wavenumber is the attenuation, $att = \xi_{\mathrm{imag}}$, in nepers per meter. As reported in [20], for damped waveguides, the group velocity

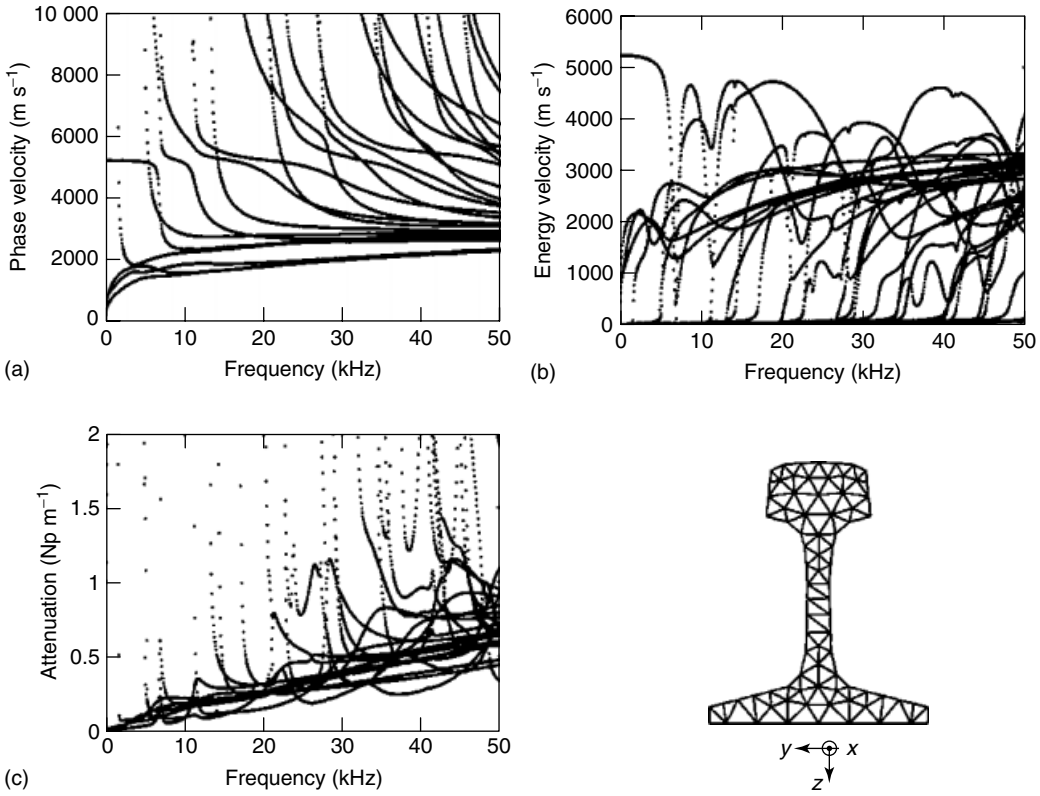


Figure 2. Dispersion results for a 115-lb AREMA viscoelastic rail for waves propagating along the rail running direction: (a) phase velocity, (b) energy velocity, and (c) attenuation.

does not coincide with the speed of propagation. In this case, the energy velocity, V_e, is the physically meaningful parameter. Energy velocity can be also readily calculated from the SAFE eigenvalues [18].

3 WAVE PROPAGATION RESULTS

The rail considered in this study is a typical 115-lb American Railway Engineering and Maintenance-of-way Association (AREMA) section, modeled as an isotropic material with hysteretic damping, and having the following properties: density $\rho = 7932\,\mathrm{kg\,m^{-3}}$, bulk longitudinal velocity $c_L = 5960\,\mathrm{m\,s^{-1}}$, bulk shear velocity $c_T = 3260\,\mathrm{m\,s^{-1}}$, longitudinal attenuation $\kappa_L = 0.003$ Np/wavelength, and shear attenuation $\kappa_T = 0.043$ Np/wavelength. The mesh of the rail cross section, shown in Figure 2 and generated by MATLAB's "pdetool", used 81 nodes for 106 triangular elements with linear interpolation displacement functions.

The dispersion results are shown in this figure up to a frequency of 50 kHz. The complexity of the modes is evident. However, these results are

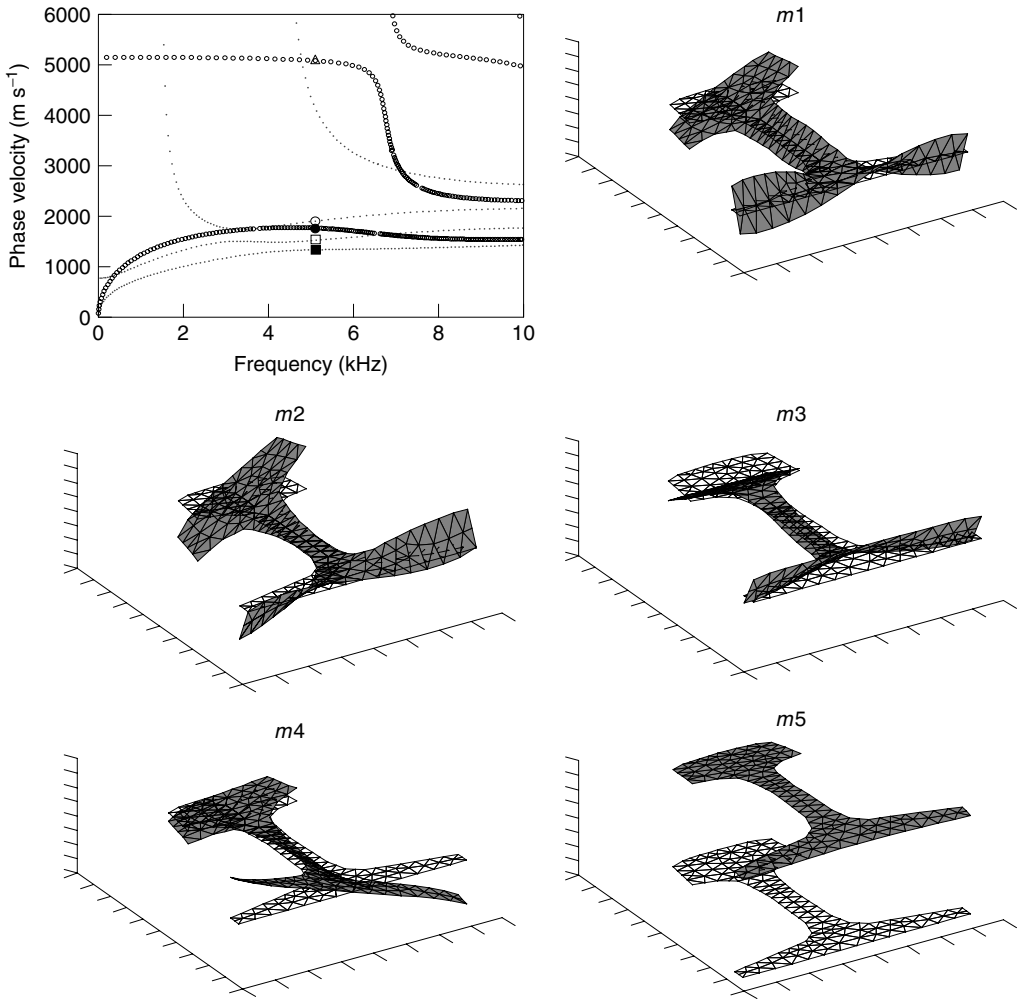


Figure 3. Dispersion results for a 115-lb AREMA viscoelastic rail: phase velocity for frequencies below 10 kHz, and first five cross-sectional mode shapes at 5 kHz, where ■, mode $m1$; □, mode $m2$; •, mode $m3$; ◯, mode $m4$; Δ, mode $m5$; ooooo, symmetric mode; · · · · · · ·, antisymmetric modes.

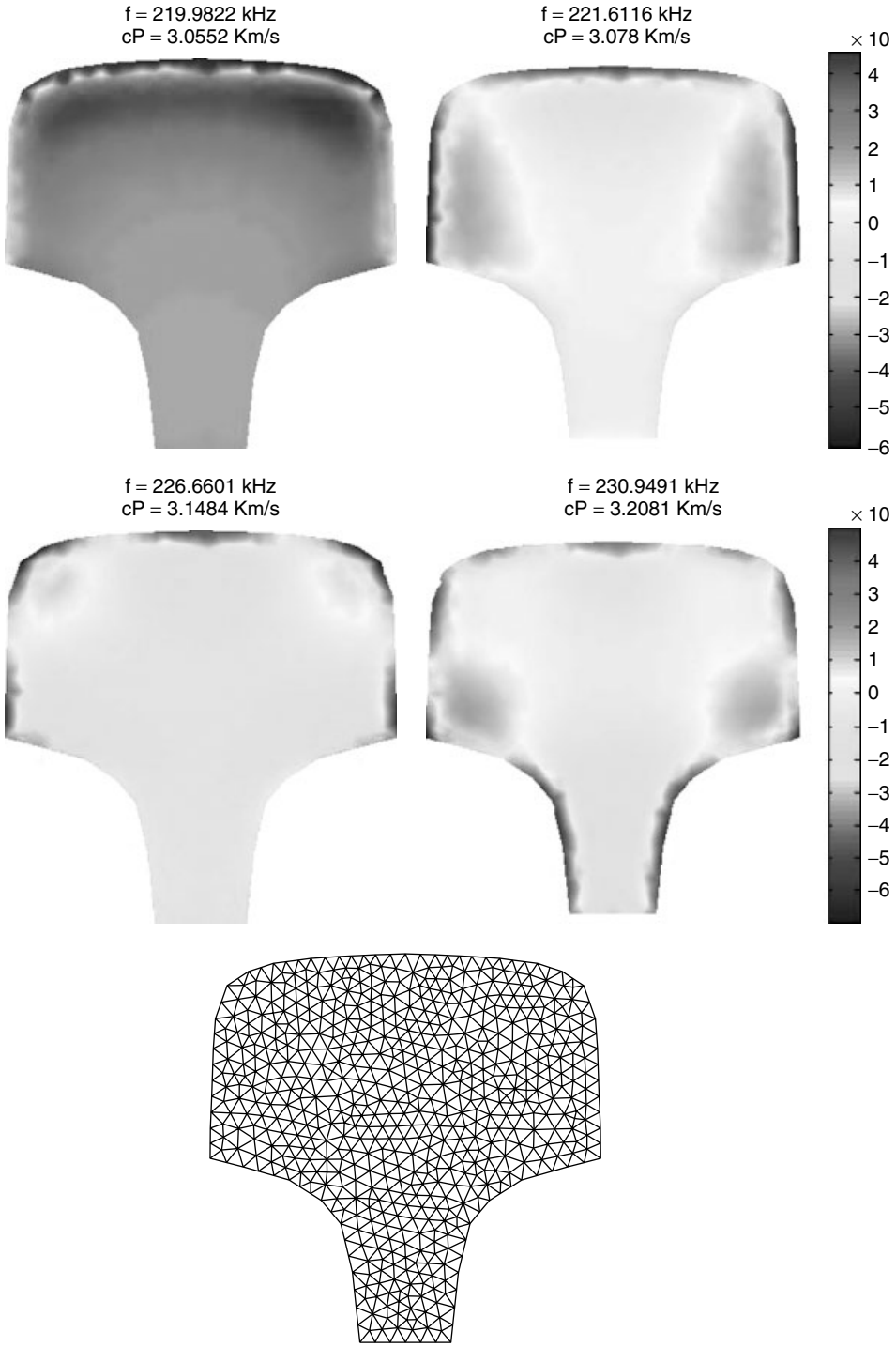


Figure 4. Axial displacement mode shapes for various symmetrical modes at frequencies around 200 kHz. The mesh used for the SAFE analysis is also shown.

useful to design an efficient NDE system, which uses mode–frequency combinations sensitive to given defects. A zoom into the low-frequency phase velocity curves is shown in Figure 3, along with the first five cross-sectional mode shapes at 5 kHz. It can be seen that modes $m1, m2$, and $m4$ are antisymmetric with respect to the x–z plane, while modes $m3$ and $m5$ are symmetric. It can also be seen that some of the modes excite preferably a certain portion of the rail, either the head or the base. Hence defects located at various locations in the rail section can be targeted selectively.

Figure 4 shows the displacement mode shapes (x component) for the first four symmetric modes at high frequency values, around 200 kHz. The character of these high-frequency modes is that of Rayleigh-surface waves, where the penetration depth is roughly equal to the wavelength. These "surface" modes in rails were examined in depth in [6]. The plots in Figure 4 were obtained using 1118 triangular elements with a total number of degrees of freedom of 1815 (see mesh in the figure).

The laser/air-coupled system presented in the following sections uses high-frequency, surface-type waves in the frequency range of 50 kHz to 1 MHz of the type shown in Figure 4 to detect transverse defects in the rail head.

4 NONCONTACT ULTRASONIC RAIL INSPECTION

4.1 Inspection prototype

The University of California at San Diego (UCSD) rail inspection prototype is based on laser excitation and air-coupled detection. The laser generator is a Nd:YAG, Q-switched type, lasing at 1064 nm with a ~10-ns pulse duration. When illuminating the rail head, it generates broadband symmetric waves of the type shown in Figure 4. An array of air-coupled sensors is used to receive the waves. The sensors are inclined at an angle of $\sim 6^{\circ}$ to maximize the sensitivity to the Rayleigh-type phase velocity owing to Snell's law. The air-coupled sensors are located at distances larger than 50.8 mm (2 in.) from the top of the rail head, thus satisfying the clearance envelope recommended by the railroad industry for inspection systems claiming noncontact operation. A portable, National Instruments unit running under LabVIEW has been assembled and programmed to perform laser control, data acquisition, processing, and reporting.

A laptop is used in conjunction with the PXI unit to form a client–server Ethernet-linked platform. The server (PXI unit) controls the laser operation and the acquisition, processing, and display of the measurements by the air-coupled sensors. Data from a positioning system are constantly acquired by the server. From the serial port (RS-232), the server triggers the laser to start the data acquisition. The ultrasonic signals detected by the air-coupled sensors are digitized through the analog input boards, processed, and logged into the server. On the client end (laptop), the user is able to start and stop the acquisition, adjust the laser power, change the lasing frequency, modify the ultrasonic and signal-processing settings, monitor the results in real time, and open a report window.

The defect detection strategy is based on "transmission" rather than on "reflection" measurements. In the "transmission mode", a defect is detected by monitoring the attenuation of the wave as it travels past the flaw. In its simplest implementation, two air-coupled sensors can be used as shown in Figure 5. In the prototype, a damage index (DI) is calculated in real time as the ratio between certain features of the signal detected by the sensor closer to the laser source $F_{\text{sens}\,1}$, over the same features from the sensor further from the laser source $F_{\text{sens}\,2}$:

$$DI = \frac{F_{\text{sens}\,1}}{F_{\text{sens}\,2}} \tag{5}$$

If a defect is located in the section of rail between the two sensors, $F_{\text{sens}\,2}$ will be smaller than $F_{\text{sens}\,1}$ and

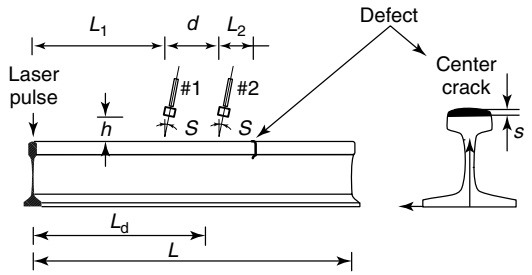


Figure 5. Laser/air-coupled system for detecting transverse defects in the rail head (sensor arrangement in the "transmission mode").

the DI will increase compared with its normal, defect-free value (nominally one). The DI in equation (5) is also expected to increase with increasing defect size.

4.2 Laboratory tests

A number of laboratory tests were conducted on 52-kg (115-lb) AREMA rail sections donated by San Diego Trolley with simulated surface defects of increasing depths. The results from the laboratory tests have been discussed extensively in recent publications [11, 15, 16]. A summary of these is presented here.

The cracks were simulated by narrow notches that were machined at depths (s in Figure 5) ranging from a minimum of 0.5 mm to a maximum of 8.5 mm, all corresponding to a cross-sectional area reduction of the rail head (% Head Area (HA) reduction) below 20%.

The ultrasonic signals were acquired at a 5-MHz sampling rate from 10 laser pulse generations at each damage condition. DWT processing (Daubechies 10 mother wavelet) was applied in real time to the raw measurements. The wavelet coefficients at detail levels 3, 4, and 5 were considered to retain the frequency band of interest. A transmission DI vector was used as a defect indicator. Thresholds equal to 70, 50, and 80% of the maximum coefficient amplitude of detail levels 3, 4, and 5, respectively, were applied in this case. The features of variance, root mean square (RMS), peak amplitude, and peak-to-peak amplitude of the thresholded wavelet coefficient vectors were computed. The reconstructions of the signal received from both sensor #1 and sensor #2 were then processed through the fast-Fourier transform (FFT) and the Hilbert transform (HT). The following additional features were finally extracted: the peak amplitude and the area below the FFT frequency spectrum, in addition to the peak and position of the amplitude of the HT.

Nine different damage conditions, ranging from pristine structure to $s = 8.5$ mm were monitored. Figure 6 shows a component of the DI vector calculated from equation (5) using the variance of the thresholded wavelet coefficients, along with a picture of the defect. Notice that for these results the inverse of the relation in equation (5) was used, thus DI decreases with increasing defect size. The mean value of 10 measurements is plotted as a function of the crack depth and the extension of the vertical line is equal to 2 SD. The monotonic decrease in the DI with increasing defect size shows the potential for crack sizing.

The defect sizes were subdivided into three classes (classes 1, 2, and 3), corresponding to % HA reduction in the ranges 0–1.1% (pristine condition), 1.5–9.9%, and 10–20%. Each class was coded with a two-digit binary number. A feed-forward, back-propagation artificial neural network with three layers was employed. Five of 10 acquisitions for each damage condition were used as training data, while the remaining data were used as testing data. The learning rate and additional momentum were equal to 0.2 and 0.5, respectively. Figure 7(a) and (b) illustrate, respectively, representative classification

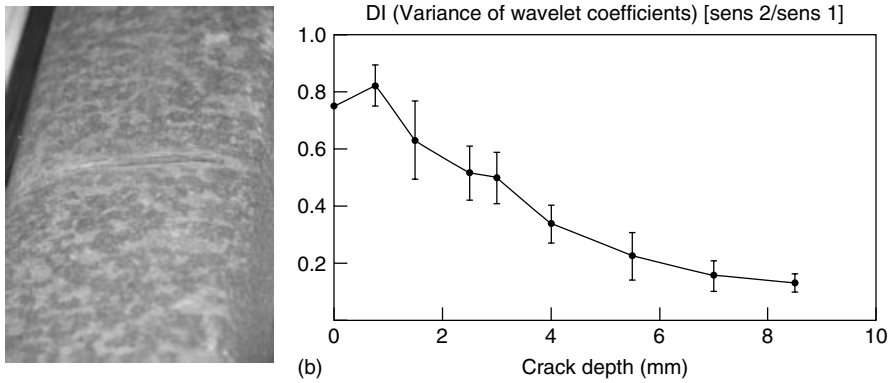


Figure 6. (a) A surface-breaking transverse head crack and (b) measured damage index (variance of wavelet coefficients) as a function of crack depth.

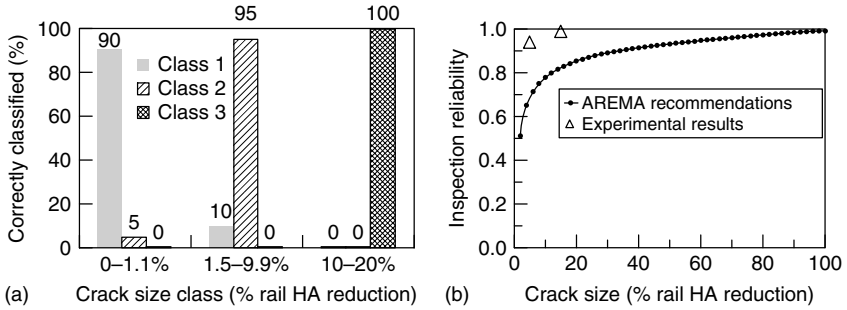


Figure 7. (a) Defect classification performance as a function of size classes; (b) inspection reliability of proposed defect detection method compared with AREMA standards.

results and a comparison with the AREMA recommendations for rail inspection reliability.

A parametric analysis identified the following four features providing the best classification performance: the variance, the RMS, the peak amplitude of the wavelet coefficient vectors, and the area below the FFT spectrum of the reconstructions. Classes 1, 2, and 3 were properly classified in 90, 95, and 100% of the cases, respectively. Thus all defects larger than 10% HA reduction were properly classified. The algorithm gave a total of 10% false positives and only a total of 5% false negatives.

5 FIELD TESTS

The first field test of the prototype was conducted in March 2006 with the technical support of ENSCO, Inc. ENSCO's support consisted of the design and construction of a cart hosting the prototype elements and managing the use of the FRA's hy-railer to tow the cart over the test track.

A picture of the laser and air-coupled sensors installed on the cart is shown in Figure 8. The black metallic box was devised to protect the laser head and the optics from accidental impacts and dust; in addition, the box avoided accidental reflections of the laser beam.

The sensor layout used in the field tests is schematized in Figure 9. The minimum distance between the first sensor (#1) and the laser was chosen so as to avoid the superposition between the air shock and the ultrasound traveling in the rail. Sensors #1 and #3 formed the first pair, hereafter indicated as sensor pair 1 (sp1); sensors #2 and #4 formed the sensor pair 2 (sp2). The sensor liftoff distance h was maintained

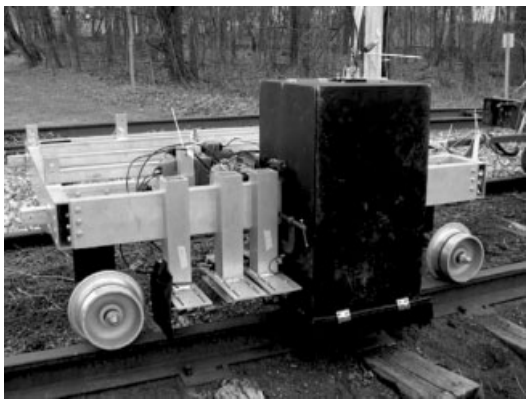


Figure 8. Laser head (inside black box) and air-coupled sensors on cart at the Gettysburg test site.

equal to 63 mm (2.5 in.). The sp1 was placed right above the center of the rail head, whereas the sp2 was placed closer to the rail gauge-side corner. The lateral offset between the two pairs was around 19 mm (0.75 in.).

5.1 Test site layout

The test site (Figure 10) was located near Gettysburg, PA, and consisted of a dismissed portion of railroad. In total, a length of 61 m (200 ft) was tested.

A total of six joints were present along the test section; some of them created a gap as large as 0.50 in. between two rail sections. Three, 1.8-m (6-ft) long, 139-lb AREMA sections with known IDs in the head were inserted in the inspected portion of the rail and fixed by joint bars. From ultrasonic hand mapping prior to the tests, the three IDs were determined to

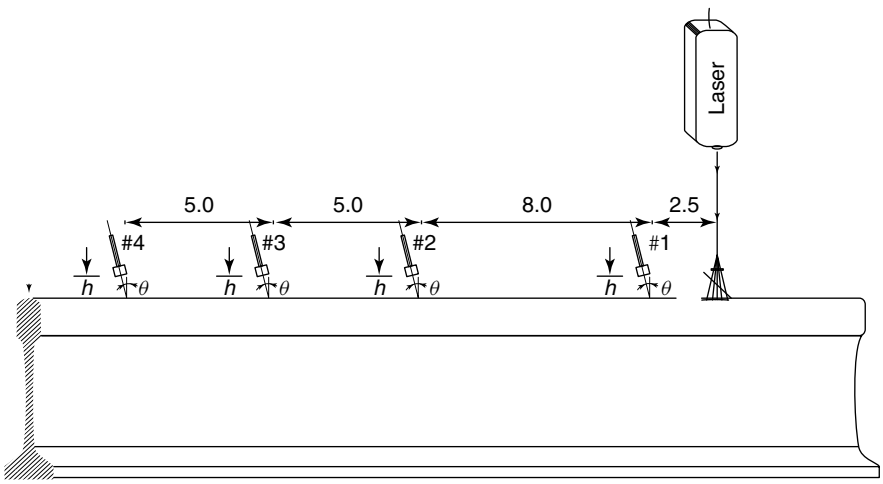


Figure 9. Laser/air-coupled sensor layout for the field tests (dimension in inches; drawing not to scale).

Figure 10. Test site near Gettysburg, PA.

extend, respectively, for 80, 10, and 40% HA. All IDs were believed to be primarily transverse to the rail running direction. Two surface cuts (SCs) were also machined perpendicular to the rail running direction causing 5 and 2% HA reductions, respectively. Two oblique SCs (45° inclination from the running direction) were also added, at a size corresponding to about 3.5% HA.

5.2 Test results

A total of 32 tests were conducted in two days. Some of the tests were performed under dry conditions, i.e., ultrasound was generated by the laser pulse irradiating the dry surface of the rail head. In the other tests, the rail was wetted by using a water pressure sprayer during the tests. Wetting was performed to examine the effect of increasing the ultrasound generation power of the laser under ablative regime.

As representative data, Figures 11 and 12 present the recordings from tests 1 and 2, respectively. The results are shown for the portion of the rail between 76 and 120 ft where the IDs and the SCs were located. Notice that no oblique cuts (OCs) were present in these runs as these were machined later in the tests. Each figure displays the DI recorded as a function of position (ft) by the two sensor pairs and for two different filtering bandwidths, namely, (a) high-frequency DI for sensor pair 1 (HF-DI sp1), (b) low-frequency DI for sensor pair 1 (LF-DI sp1), (c) high-frequency DI for sensor pair 2 (HF-DI sp2), and (d) low-frequency DI for sensor pair 2 (LF-DI sp2). The plot (e) is included to ease in the visualization of the position of each of the discontinuities. In this plot (e) the joints are placed at ordinate 5, the transverse SCs at ordinate 4, the IDs at ordinate 3, and the flaking at ordinate 1. In the four DI plots, successfully detected joints (J), SCs, or IDs are highlighted with circles. Since the two sensor pairs and filtering bandwidths cover complementary regions of the rail head, a defect should be considered successfully detected if any of the four DI plots shows an

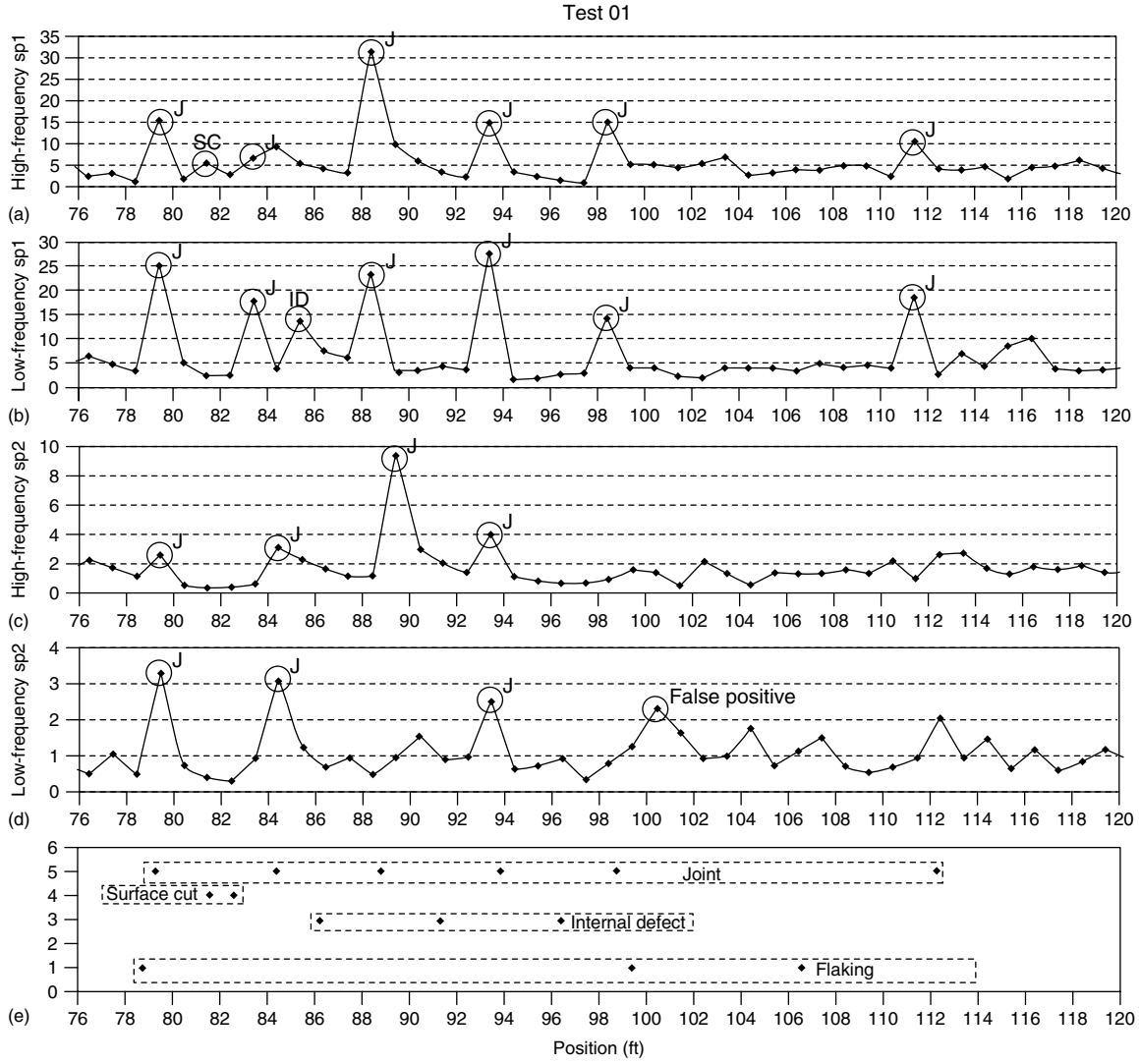


Figure 11. Results of field test no. 1 (J, joint; SC, surface cut; ID, internal defect).

index clearly above the baseline (defect-free) value. For example, Figure 12 shows that all major defects, with the exception of the third ID, were detected by at least one of the four DI plots.

The fact that the baseline DI values in Figures 11 and 12 are not necessarily those as theoretically expected from equation (5) is simply due to the practical impossibility of obtaining exactly the same sensitivities from the two sensors in each pair. This, however, does not constitute a problem since only relative changes in DI from the baseline value are monitored to detect a potential defect.

A probability of detection (POD) was estimated from a total of 30 tests. As discussed earlier, a defect was considered successfully detected if at least one of the four DI (HF-DI sp1, LF-DI sp1, HF-DI sp2, or LF-DI sp2) was activated. The results are shown in the histogram in Figure 13. In summary, it was found that the system performed very well in detecting the two SCs, two of the three IDs

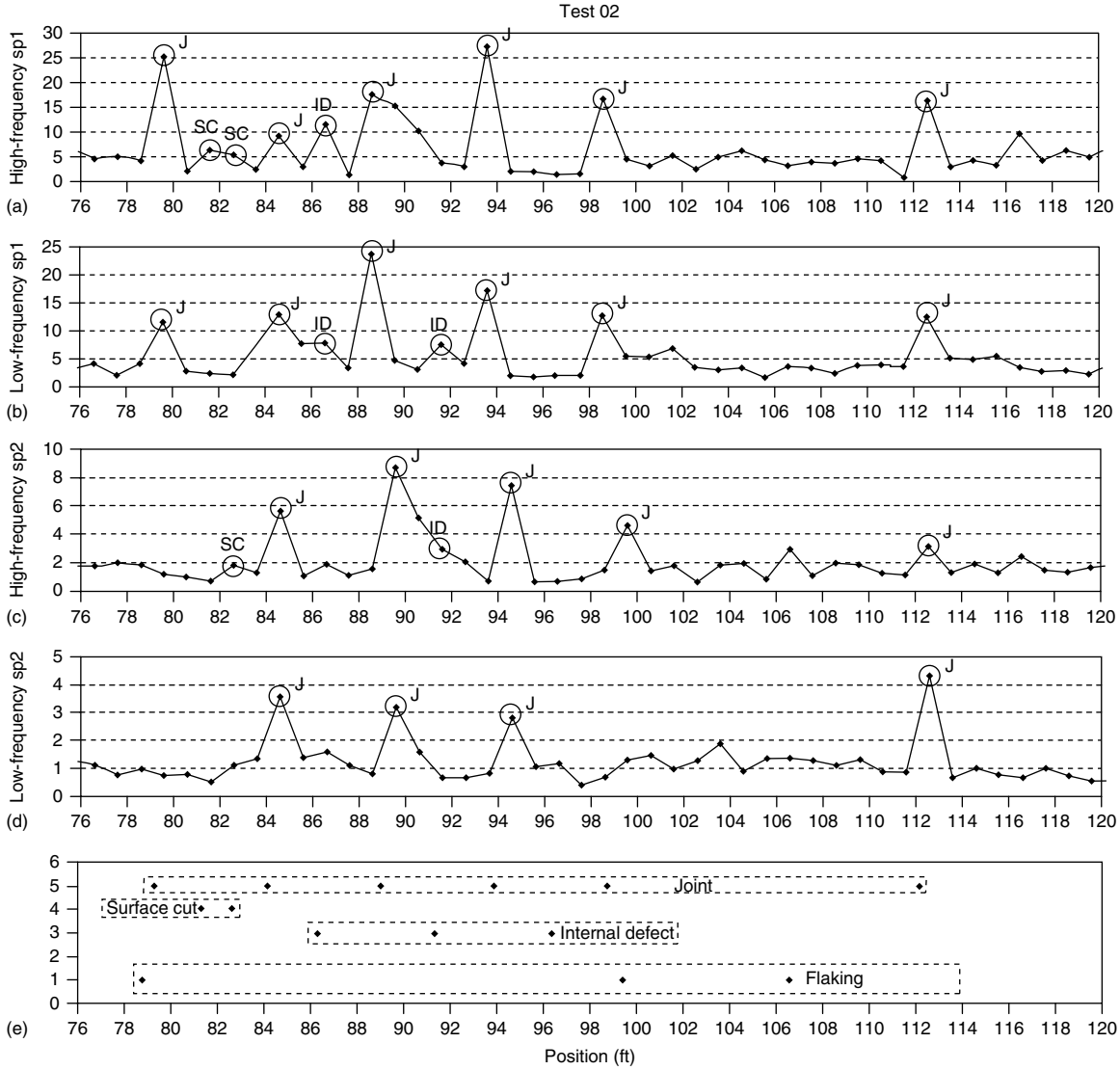


Figure 12. Results of field test no. 2 (J, joint; SC, surface cut; ID, internal defect).

(80 and 10% HA), and the second oblique SC. The detection performance was poor for the third ID (40% HA) and the first oblique SC. The fact that the SCs 1 and 2 (5 and 2% HA, respectively) were successfully detected demonstrated the ability of the system to target defects well below the 10% HA limit commonly used by railroad owners to consider removing the rail from service. The fact that the IDs 1 and 2 (80 and 10% HA) were successfully detected demonstrated that the 10% HA detectability limit also applies to internal flaws. It is unclear why ID 3 (40% HA) was seldom detected. Hand remapping of this flaw is being scheduled to verify its dimensions prior to the second field test.

In terms of dry versus wet surface, it was found that wetting the rail surface increased the POD for all defects (with the exception of oblique SC 2). The improvement from the water, although significant,

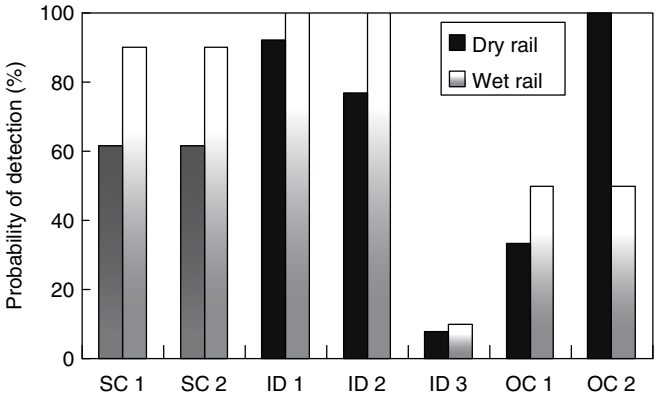


Figure 13. Probability of detection results obtained during field test under either dry or wet conditions of the rail surface (SC, surface cut; ID, internal defect; OC, oblique cut).

was not dramatic. In the end, a compromise must be struck between the burden of carrying a water supply and an acceptable level of POD.

The defect discrimination based on detection frequency band was generally successful. The SCs were predominantly detected by the HF-DIs mostly sensitive to surface features. Similarly, ID 1 was predominantly detected by the LF-DIs mostly sensitive to features deeper inside the rail head. Interestingly, ID 2 was primarily detected by the HF-DIs, suggesting that the flaw was very close to breaking the surface. This hypothesis was confirmed by an independent hand mapping performed on this defect prior to the tests.

6 CONCLUSIONS

This article discusses the topic of guided-wave defect detection in rails, and it presents in particular a prototype under development for noncontact testing. Guided waves are excellent candidates for interacting with the dangerous transverse defects in rails. However, their multimode and dispersive behavior complicates signal interpretation and inspection design. A SAFE method is an ideal tool for modeling these waves because of the lack of high-frequency (above 6 kHz) theoretical solutions for waveguides of arbitrary cross section such as rails. This modeling was used to generate dispersion curves, attenuation curves, and mode shapes of 115-lb AREMA rails. It was confirmed that high-frequency waves (~200 kHz) are predominant Rayleigh-wave type, whose penetration depth is related to their wavelength. Hence, defect sizing is possible by monitoring different frequency bands of the propagating waves.

Waves from 50 kHz to 1 MHz range are used in the prototype, which also employs a pulsed laser and air-coupled sensors for noncontact probing of the rail head. Software has been developed to denoise the measurements by DWT processing, and then compute a DI based on transmission measurements. The DI is related to the size and the position of transverse defects within the rail head (whether at the surface or internal) through its magnitude and associated frequency band. Laboratory tests showed that the prototype yields defect detection reliability for surface-breaking cracks comparable to or better than industry (AREMA) standards. Field tests also showed large probabilities of detection for both surface and IDs, with some exceptions. A second field test of an improved version of the prototype will be carried out to validate the defect detection performance, particularly at normal testing speeds.

ACKNOWLEDGMENTS

This work was performed under research grant DTFR53-02-G-00011 awarded by the Federal Railroad Administration (FRA) of the US Department of Transportation. The authors are grateful to Mahmood Fateh, the FRA's Technical Representative, for his technical guidance and valuable comments during all phases of this research, and to Gary Carr, Senior

Mechanical Engineer of ENSCO during the field tests, and now FRA's Chief of Track Research Division, for providing technical and logistic support during the prototype assembling and field testing.

REFERENCES

[1] Federal Railroad Administration. *Safety Statistics Data: 1992–2002*. U.S. Department of Transportation, 2002.

[2] Clark R. Rail flaw detection: overview and needs for future developments. *Nondestructive Testing and Evaluation International* 2004 **37**:111–118.

[3] Lanza di Scalea F. Ultrasonic testing in the railroad industry. In *Non-destructive Testing Handbook, Third Edition*, American Society for Nondestructive Testing: Columbus, OH, 2007, pp. 9–16.

[4] Wilcox P, Evans M, Pavlakovic B, Alleyne D, Vine K, Cawley P, Lowe MJS. Guided wave testing of rail. *Insight—Non-destructive Testing and Condition Monitoring* 2003 **45**:413–420.

[5] Cawley P, Lowe MJS, Alleyne D, Pavlakovic B, Wilcox P. Practical long range guided wave testing: applications to pipes and rail. *Materials Evaluation* 2003 **61**:66–74.

[6] Hesse D, Cawley P. Surface wave modes in rails. *The Journal of the Acoustical Society of America* 2006 **120**:733–740.

[7] Rose JL, Avioli MJ, Mudge P, Sanderson R. Guided wave inspection potential of defects in rail. *Nondestructive Testing and Evaluation International* 2004 **37**:153–161.

[8] Rose JL, Avioli MJ, Song WJ. Application and potential of guided wave rail inspection. *Insight—Non-destructive Testing and Condition Monitoring* 2002 **44**:353–358.

[9] McNamara J, Lanza di Scalea F, Fateh M. Automatic defect classification in long-range ultrasonic rail inspection using a support vector machine-based smart system. *Insight—Non-destructive Testing and Condition Monitoring* 2004 **46**: 331–337.

[10] Bartoli I, Lanza di Scalea F, Fateh M, Viola E. Modeling guided wave propagation with application to the long-range defect detection in railroad tracks. *Nondestructive Testing and Evaluation International* 2005 **38**:325–334.

[11] Lanza di Scalea F, Bartoli I, Rizzo P, Fateh M. High-speed defect detection in rails by non-contact guided ultrasonic testing. *Journal of the Transportation Research Board, Transportation Research Record* 2005 **1961**:66–77.

[12] Alers G. *Railroad Rail Flaw Detection System Based on Electromagnetic Acoustic Transducers*, Report DOT/FRA/ORD-88/09. U.S. Department of Transportation, 1988.

[13] Edwards RS, Dixon S, Jian X. Characterisation of defects in the railhead using ultrasonic surface waves. *Nondestructive Testing and Evaluation International* 2006 **39**:468–475.

[14] McNamara J, Lanza di Scalea F. Improvements in non-contact ultrasonic testing of rails by the discrete wavelet transform. *Materials Evaluation* 2004 **62**:365–372.

[15] Lanza di Scalea F, Rizzo P, Coccia S, Bartoli I, Fateh M, Viola E, Pascale G. Non-contact ultrasonic inspection of rails and signal processing for automatic defect detection and classification. *Insight—Non-destructive Testing and Condition Monitoring* 2005 **47**:346–353.

[16] Rizzo P, Lanza di Scalea F. Wavelet-based unsupervised and supervised learning algorithms for ultrasonic structural monitoring of waveguides. *Progress in Smart Materials and Structures Research*. Nova Science Publishers: Hauppauge, NY, 2006; Chapter 8.

[17] Thompson DJ. Wheel-rail noise generation, part III: rail vibration. *Journal of Sound and Vibration* 2003 **161**:421–446.

[18] Bartoli I, Marzani A, Lanza di Scalea F, Viola E. Modeling wave propagation in damped waveguides of arbitrary cross-section. *Journal of Sound and Vibration* 2006 **295**:685–707.

[19] Hayashi T, Song WJ, Rose JL. Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example. *Ultrasonics* 2003 **41**: 175–183.

[20] Bernard A, Lowe MJS, Deschamps M. Guided waves energy velocity in absorbing and non-absorbing plates. *The Journal of the Acoustical Society of America* 2001 **110**:186–196.

Chapter 21
Signal Processing for Damage Detection

Wieslaw J. Staszewski and Keith Worden
Department of Mechanical Engineering, University of Sheffield, Sheffield, UK

1 INTRODUCTION

Structural health monitoring (SHM) has become an important element of all maintenance activities in operation of aerospace, mechanical, and civil infrastructures. *Damage, health*, and *monitoring* of structures can be described using various definitions. In general, *health* is the ability to function/perform and maintain the structural integrity throughout the entire lifetime of the structure; *monitoring* is the process of diagnosis and prognosis, and *damage* is a material, structural, or functional failure. Also, in this context, structural integrity is the boundary condition between safety and failure of engineering components and structures.

SHM, damage detection/monitoring, and nondestructive testing (NDT) are often used interchangeably to describe the process of nondestructively evaluating the structural condition. However, only SHM defines the entire process of implementing a strategy, which includes the five most important identification elements (or levels). These are [1]; detection, classification, localization, assessment, and prognosis, as illustrated in Figure 1. In this context, *detection* gives a qualitative indication that damage might be present, *classification* gives information about the damage type, *localization* gives information about the probable position of damage, *assessment* estimates the severity of damage, and finally *prognosis* estimates the residual structural life. All these elements require various levels of data, signal, and/or information processing. What distinguishes SHM from NDT is the global and on-line implementation of various damage-detection technologies, which require periodically spaced measurements (or observations), as accurately pointed out in [2]. This process of on-line implementation needs more advanced signal processing for reliable damage detection than classical NDT techniques.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

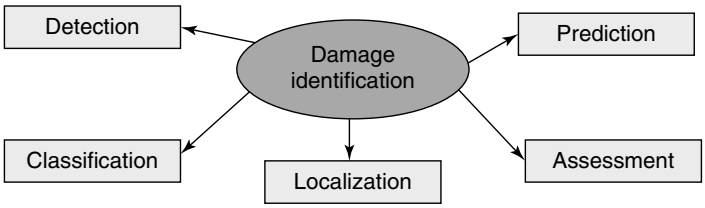


Figure 1. Damage identification levels in SHM.

Most SHM methods can be classified into model-based and signal-based (data-based) approaches. Vibration-based methods often utilize physical and/or modal parameters, obtained from physical models, for damage detection. Models are also essential when loads are monitored to obtain information about structural usage. Signal-based methods rely on various types of direct measurements such as noise, vibration, ultrasound, and temperature. Both model-based and signal-based approaches require signal-processing techniques, the former to develop appropriate models and to analyze changes in these models that are relevant to damage, and the latter to extract features and establish a relationship between these features and possible damage.

In summary, it appears that signal processing is essential for implementation of any SHM system. There exist various signal-processing techniques that are currently used and are considered for application in damage detection [3, 4]. Figure 2 shows examples of advanced methods and techniques developed over many years of research investigations. Some of these methods are well established and widely used, whereas others are more advanced and recently developed. There are six major areas where signal processing has an important role to play in SHM. These are as follows:

- measurements and data acquisition;
- data preprocessing;
- feature extraction, selection, and compression;
- pattern recognition and machine learning;
- reliability and statistical analysis;
- uncertainties and information gaps.

All these elements are briefly discussed below and relevant references are given to subsequent articles in the Encyclopedia for more reading.

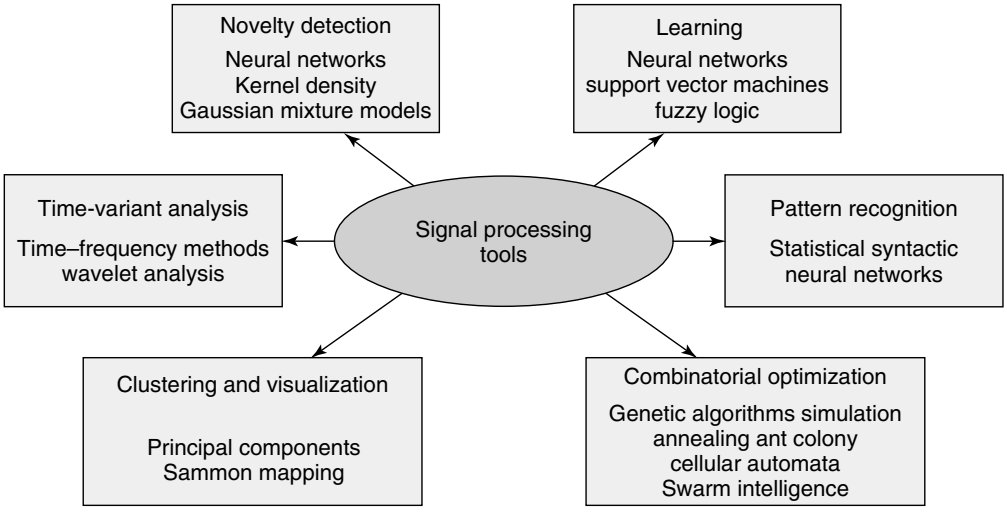


Figure 2. Selection of advanced signal-processing tools used for damage detection.

2 MEASUREMENTS AND DATA ACQUISITION

Various types of signals can be used for health monitoring, as illustrated throughout different sections of this encyclopedia. Measurements are performed using contact and noncontact approaches, which require selecting the number and locations of measuring points. Signal processing can be used to enhance this selection process. Numerous optimization techniques have been developed for optimal sensor location. This includes classical optimization techniques and various combinatorial approaches, as illustrated in **Optimization Techniques for Damage Detection**. The information content of features can be also used for assessing different types of sensors, their performance regarding required features, and their location regarding desired information, as discussed in **Sensor Placement Optimization**.

A number of measures have been developed to assess informativeness. For example, mutual information, based on Shannon entropy, can reduce the dimensionality of the space by eliminating sensors (or locations of sensors) with low information content or with high redundancy with respect to other sensors.

3 DATA PREPROCESSING

Once data from various sensors are acquired (and optionally stored), preprocessing is the first important element of the entire signal-processing analysis. It involves two major operations, namely, normalization and cleansing. The former is required to simplify and present the data collected under varying conditions to separate effects related to possible damage from operational and environmental conditions. The latter involves (i) various procedures to remove undesired features from data such as trends or noise; (ii) filtering and resampling; and (iii) selection procedures that either accept or reject data before further analysis is performed.

Trends show unwanted temporal relationships in the data. Trend analysis and removal requires some kind of model of the data. Noise is unwanted, a meaningless spectral part of the data, which does not carry any desired information. It is usually associated with higher frequencies. The level of noise in the data can be reduced by local and/or global averaging, smoothing, and denoising procedures. Filtering removes certain parts of the data within specific frequency ranges. The operations can be performed in the time, frequency, or spatial domains. Some filters can be used for smoothing and denoising. Resampling is used to reduce the frequency range of the acquired data. Finally, bad-quality data can be rejected before or even after preprocessing on the basis of individual experience and judgments and may be due to, for example, loose sensor mounting, poor connections, failure of acquisition hardware, or human error. Signals that are far from expected observations are often called *outliers*. Outliers can often be eliminated using standard statistical analysis.

A variety of different signal preprocessing methods are illustrated in Figure 3. More details related to different procedures can be found in **Data Preprocessing for Damage Detection**.

4 FEATURE EXTRACTION, SELECTION, AND COMPRESSION

Time histories acquired from sensors are very rarely used for damage detection. Most health monitoring techniques are indirect methods; in other words, measurements used in the form of different types of signals need to be transformed into features that can indicate possible damage. Various approaches can then be used. Trending analysis employing different parameters—often called *damage indices*—can be applied. The value of *damage index* is then plotted against lifetime to reveal gradual degradation. The maximum amplitude of Lamb wave response plotted against fatigue cycles is an example. Alternatively, time histories are mapped directly into patterns describing various damage conditions. For example, strain, widely used in many engineering applications, needs to be mapped into loads in order to obtain information about usage. Often sets of features in both approaches are compared with baseline measurements representing the *no-damage* condition.

Various features can be used to indicate damage. Features are any parameters extracted from the measurements through signal processing. There exist different types of features used for damage detection. Dimensionality of the features and complexity

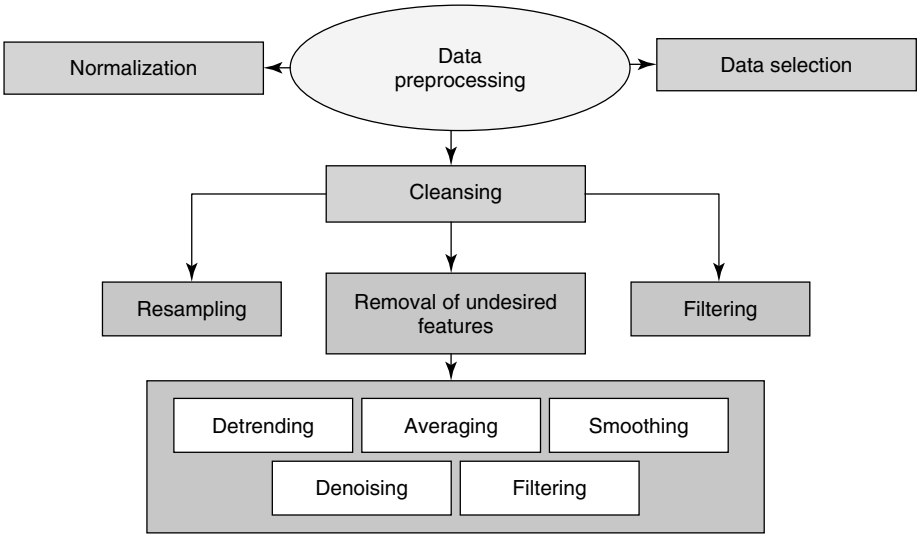


Figure 3. Data preprocessing procedures used for damage detection.

of signal processing decide the possible applications. Simple features, such as statistical parameters (e.g., maximum value, average value, root-mean-square level, and kurtosis), can be used to detect severe damage. However, information for early detection of incipient or small damage severity requires advanced signal-processing procedures. Such procedures often result in simple one-dimensional characteristics (e.g., set of principal components) or lead to more complex, multidimensional representations (e.g., timescale wavelet plots). In general, the choice of features is a trade-off between computational feasibility associated with low-level features and extensive signal processing required for high-level features.

Feature extraction, selection, and compression has received considerable attention in signal-processing research related to damage detection and has led to many different techniques and approaches. Sensor measurements can be analyzed in the time, frequency, combined time–frequency, and spatial domains. Various statistical parameters (e.g., mean and maximum value) and probability distributions can be used to analyze signal amplitudes. Time histories can be also analyzed using discrete-time models. Time-domain measurements are often transformed to another domain; various linear and nonlinear transformations are used in practice. Spectral analysis can be applied to obtain frequency content of analyzed signals, as shown in **Statistical Time Series Methods for SHM**. A variant on the spectrum—the cepstrum (*see* **Cepstral Methods of Operational Modal Analysis**)—has been applied successfully in the context of condition monitoring and may prove to be of value for SHM. This is discussed in **Cepstral Methods of Operational Modal Analysis**. Higher order spectral analysis can reveal various nonlinear interactions in data that prove to be of value in indicating damage (*see* **Higher Order Statistical Signal Processing**). Instantaneous phase and amplitude characteristics are often useful and can be obtained using the Hilbert transform, as discussed in **Hilbert Transform, Envelope, Instantaneous Phase, and Frequency**. The combination of the Hilbert transform with so-called empirical mode decomposition has led to the Hilbert–Huang transform, which has applications in nonstationary signals (*see* **Damage Detection Using the Hilbert–Huang Transform**). The analysis of nonstationary signals, in general, requires specific techniques that go beyond the classical Fourier approach. There have been major developments in the area of time–variant analysis over the past 20 years. Currently, there exist many different methods, which can be classified into two major groups, i.e., those based on time–frequency analysis (*see* **Time–frequency Analysis**) and those based on timescale analysis, as discussed in **Wavelet Analysis**. The latter approach leads to wavelet

analysis, which has received major attention in recent mathematical and signal-processing developments. Various procedures have been developed for the analysis of nonlinear systems (*see* **Nonlinear Features for SHM Applications**). Altogether, transformation of data can not only reveal specific characteristics/features related to damage but can also lead to data reduction and preferable statistical properties/parameters.

Transformed signals are often inappropriate for analysis owing to redundancy or high degree of correlation. Additionally, such analysis does not necessarily lead to reduced dimensionality. There exist different procedures that are used to deal with this problem. Altogether these procedures condense information, i.e., combine existing features into new features, select subsets of features, and map signals (features) into smaller sets of features or compress signals (features).

A number of measures have been developed to assess the informativeness of features. Mutual information, based on entropy, is an example. The method based on Fisher's linear discriminant performs an optimal linear dimensionality reduction based on the magnitude of the discriminant vector. It leads to maximum class separation in the output space. The classification binary tree uses a binary decision tree based on Kolmogorov–Smirnov minimization criterion of the Bayesian probability. The "goodness" of features for damage identification can also be estimated on the basis of the Fisher information matrix. This matrix is the inverse of the covariance matrix associated with the analyzed features. Maximizing the trace or determinant and minimizing the condition number of the Fisher information matrix corresponds to the process of feature selection. Applications of this principle can be found in the section on sensor optimization (*see* **Sensor Placement Optimization**).

Mapping techniques are capable of reducing the point dimensionality of data. Principal component analysis (PCA) is one of the most commonly used linear techniques for dimensionality reduction. PCA is a classical multivariate statistics technique that maps the original data to a space of reduced dimension using a linear transformation. The procedure seeks to retain the variance in the analyzed data as much as possible. Sammon mapping is a nonlinear equivalent of the PCA technique. The method projects the N-dimensional data into a two-dimensional (2-D) space preserving all interpoint distances, i.e., two samples that are close to each other in the original space remain close in the new space. The objective is to reveal the underlying structure of the data preserving the original topology. The general problem of dimensionality reduction is discussed in the context of SHM in **Dimensionality Reduction Using Linear and Nonlinear Transformation**.

Dimensionality reduction can be also achieved using data compression. Recent developments in this area include applications of wavelets. The wavelet transform can give a compression basis, which is independent of the data set, and can reveal local temporal correlations in the data.

5 PATTERN RECOGNITION AND MACHINE LEARNING

Time histories obtained through various measurements and features, which result from signal processing, form data sets that can be seen as patterns. This leads to damage detection based on pattern recognition. Patterns are continuous, discrete or binary, or combined variables, which are formed in vector or matrix notation. In a damage detection approach, the assumption is that patterns represent different damage conditions. Various pattern-recognition techniques can be used to indicate whether a structure is undamaged or damaged and then to assess the severity of possible deterioration. Classical approaches to pattern recognition can be categorized as statistical and syntactic methods. Syntactic pattern recognition classifies data according to its structural description, whereas statistical pattern recognition (SPR) assigns features to different classes using statistical density functions. Applications of structural pattern recognition to SHM vary and SPR is the general approach as discussed in **Statistical Pattern Recognition**. Recent advancements in pattern recognition include applications of artificial neural networks (ANN), which have proved fruitful in the context of SHM as described in **Artificial Neural Networks**.

Pattern recognition itself forms a subdiscipline of the broader field of machine learning, which is currently receiving considerable attention for SHM, and **Machine Learning Techniques** gives a brief overview of the field. The discipline of machine

learning is characterized by the fact that computational rules are inferred or *learned* on the basis of observational evidence. This contrasts with the "classical" view of computation where the algorithmic rules are imposed in the form of a sequence of serially enacted instructions. It is sometimes stated that learning theory is designed to address three main problems [5]:

- Classification, i.e., the association of a class or set label with a set or vector of measured quantities. The set of observations may be sparse and/or noisy. This is very often applied in the context of damage location, where the structure of interest is divided into a number of substructures and a class label is assigned to each.
- Regression, i.e., the construction of a map between a group of continuous input variables and a continuous output variable on the basis of a set of (again, potentially noisy) samples. This can also be used in damage localization by attempting to predict the real-valued coordinates of the damage; it can also be used for the problem of severity assessment.
- Density estimation, i.e., the estimation of probability density functions from samples of measured data. This is one of the approaches to damage detection via novelty detection as discussed below.

A further division of learning algorithms may be made between *unsupervised* and *supervised* learning. The former is concerned with characterizing a set on the basis of measurements and perhaps determining underlying structure. The latter requires examples of input and output data for a postulated relationship so that associations might be learnt and errors corrected. This approach presents serious problems in the SHM context as it is very rarely possible to acquire sufficient data covering the various possible damage states. Although it is not universally so, regression and classification problems usually require supervised learning, while density estimation can be conducted in an unsupervised framework. In general, novelty detection can be carried out in an unsupervised context. This is the basic process of distinguishing data from data measured in the normal condition of the structure of interest and thus serves as a method of detection. A number of approaches to novelty detection are discussed in **Novelty Detection**.

A powerful approach to the handling of disparate sources of information is given by the class of *data fusion* algorithms. Data fusion can operate at a number of levels within the SHM context. At a basic level, one can use the methodology to combine features into finer features for diagnostics. At a higher level, fusion can be used to combine the results of disparate classifiers in order to refine the decision process. The basic methodology of data fusion is described in **Data Fusion of Multiple Signals from the Sensor Network**.

6 RELIABILITY AND STATISTICAL ANALYSIS

One of the basic problems in SHM is the fact that many of the quantities of interest for diagnosing damage as well as the damage states themselves are often random variables. This means that the appropriate analysis is usually within a probabilistic context. At the lowest level, even the models constructed for model-based diagnosis will have to incorporate random variables in the form of unmeasured effects like noise and environmental variations. Statistical analysis has developed a number of very powerful techniques for building models under uncertain circumstances and some of the foremost approaches are discussed in **Model-based Statistical Signal Processing for Change and Damage Detection**.

A major problem is that the features extracted from measured data to accomplish diagnosis are also random variables. This means that classifier design is best carried out within a principled statistical framework like SPR as discussed above (*see* **Statistical Pattern Recognition**). SPR is able to accommodate the fact that classes will sometimes overlap, i.e., there will be situations where a given set of features may or may not indicate damage. SPR allows one to assign a probability to the diagnosis and this is critical when one needs to make a decision on the basis of the diagnosis. In the case of regression analysis, a principled statistical approach allows one to assign a confidence interval or error band to a prediction and, again, this can prove invaluable in the decision process.

7 UNCERTAINTIES AND INFORMATION GAPS

There are essentially two types of uncertainty in general. The first is *aleatory* or *irreducible* uncertainty. This is the type of uncertainty associated with, for example, the microscopic variation of material properties within a structure. The uncertainty may be characterized by a probability distribution and is irreducible to the extent that there exists no possible program of experimentation that can reduce the uncertainty, e.g., to allow better prediction of the pointwise value of a material property. Another source of aleatory uncertainty is in the predicted future loading of a given structure or component. Both these examples of uncertainty present problems in terms of damage prognosis or level five of the damage identification hierarchy. At the lower levels of damage identification like localization, the uncertainty is partly addressed by adopting a principled framework like SPR. The second type of uncertainty is *epistemic* or *reducible* uncertainty. In this case, the uncertainty is caused by a lack of knowledge, which could potentially be acquired by, say, more careful modeling or by experimental campaign. The precision limits of instrumentation are a potential cause of epistemic uncertainty as it may be possible to acquire more precise instrumentation. It is clear that SHM damage identification will potentially suffer from the presence of both flavors of uncertainty and this must somehow be accommodated within the data-to-decision process, which leads from measurements to a decision.

The problem of uncertainty is sometimes only optimally addressed in the probabilistic context on the assumption that certain *a priori* information is available. Where such information is not available, an information gap or ignorance occurs and it may be that probability theory is suboptimal. A number of other uncertainty frameworks are available, which offer the possibility of accommodating ignorance in a potentially more conservative manner than probability theory. It is useful to have such techniques available for SHM analysis as there are many sources of uncertainty in the data-to-decision process. A number of the alternative uncertainty theories are described in the context of SHM in **Uncertainty Analysis**, which also presents case study material.

8 SUMMARY

Whatever the specialism of the SHM researcher or practitioner, few would argue that signal processing is a crucial ingredient in the SHM process and will be a vital ingredient in moving SHM from the research laboratory to everyday maintenance practice. The object of the article here is to present the basic facts regarding signal processing as applied in the context of SHM. While much of the material is textbook in terms of its basic development, it is included here so that the reader can see it exposed in the damage identification context. Other materials will certainly be unfamiliar as it is only very recently beginning to penetrate the SHM literature. Overall, this article is intended to be exhaustive only in the sense of *introducing* relevant technologies; however, the references within each section are intended to provide comprehensive guidance for further reading at greater depth.

REFERENCES

[1] Rytter A. *Vibration Based Inspection of Civil Engineering Structures*, Ph.D. Thesis. Department of Building Technology and Structural Engineering, University of Aalborg: Denmark, Aalborg, 1993.

[2] Adams DE. *Health Monitoring of Structural Materials and Components*. John Wiley & Sons, 2007.

[3] Staszewski WJ, Worden K. Signal processing for damage detection. In *Health Monitoring for Aerospace Structures*, Staszewski WJ, Boller C, Tomlinson GR (eds). John Wiley & Sons, 2003.

[4] Worden K, Staszewski WJ. Data fusion – the role of signal processing for smart structures and systems. In *Smart Technologies*, Worden K, Bullough WA, Haywood J (eds). World Scientific, 2003.

[5] Cherkassky V, Mulier F. *Learning from Data, Concepts, Theory and Methods*. Wiley Interscience, 1998.

Chapter 22
Data Preprocessing for Damage Detection

Andrew Halfpenny
nCode International Ltd., Innovation Technology Centre, Rotherham, UK

1 INTRODUCTION—THE SIGNAL PATH AND THE SOURCE OF MEASUREMENT ERRORS

Measurement errors and erroneous noise are introduced along the signal path between the measurement transducer and the structural health monitoring (SHM) processor. A typical signal path for SHM measurement data is illustrated in Figure 1.

A *transducer* or *sensor* takes a physically measurable quantity such as strain or temperature and converts it to an electrical signal that can be processed. The electrical signal is usually expressed as voltage, current, charge, or resistance. Examples include strain—converted to electrical resistance, or temperature—converted to electrical voltage (also see Section 14 in the encyclopedia).

The *signal conditioning* unit takes the low-level electrical signal and "conditions" it ready for analysis. Typical conditioning steps include amplifying, filtering, and any necessary conversion of current, charge, resistance, or pulse frequency to an analog voltage.

The *analog-to-digital converter* (*ADC*) takes the amplified analog signal from the signal conditioning unit and converts it to a digital form suitable for computer-based analysis. This is a two-stage process. The ADC first samples the stream of continuous analog data selecting discrete values from the stream at periodic intervals. This stage is called *sampling* and the frequency at which samples are taken is called the *sample rate*. In the second stage, the discrete analog values are converted to a digital representation. This is referred to as *digitization* or *quantization*.

In smaller SHM systems, the ADC converter is often integrated in the SHM processor unit and sensors are connected directly to the unit. However, in larger systems, this arrangement is impractical due to the number of analog cables required to connect all the transducers, and also the effect of electrical interference on low-level analog signals. A *digital communication network* requires only one cable and is far more tolerant of electrical interference. Several digital communication standards are commonly used

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

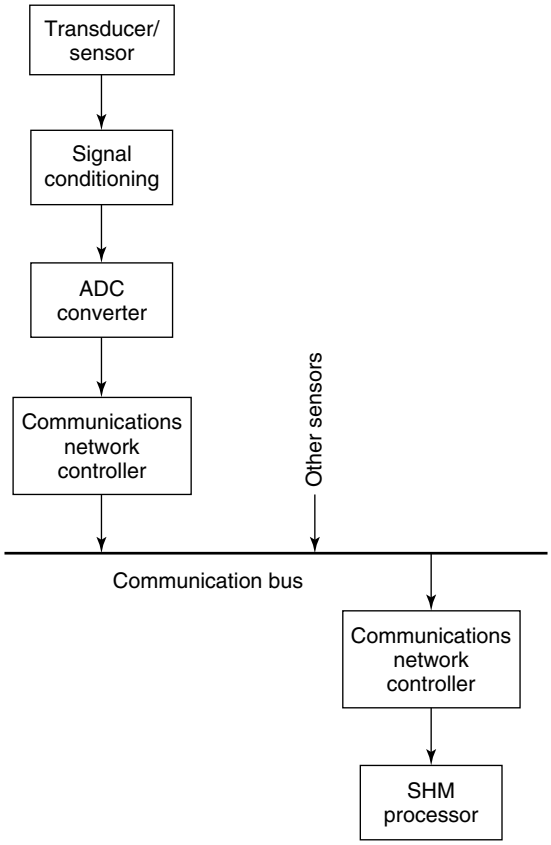


Figure 1. The signal path.

for SHM and are discussed in [1]. The controller area network (CAN) and Ethernet systems are most commonly used for wired networking although others can be found. The CAN is preferred over short distances in very noisy conditions such as engines and electromechanical machinery. Ethernet is more common for larger networks in less noisy conditions. Both standards divide the data into small "packets", which are transmitted along the cable. The packets are received by other nodes on the network and reassembled to represent the original signal. They are known as *asynchronous* communication because each sensor samples, digitizes, and transmits its signal independently, so data points in one channel are not sampled simultaneously with data points in another. Wireless networking standards are also common and include "Bluetooth" and "WiFi". Bluetooth is most commonly used for short range ($<10\,\mathrm{m}$) communication between remote sensors, whereas WiFi is more suitable for longer distances and can replace the wired network. Proprietary remote wireless telemetry systems are also common.

A good account of sensor technology and digital communication systems used in automotive engineering is presented by Hillier *et al.* [2] and also Bosch [3].

The *SHM processor* unit collates data from each sensor and performs the damage-detection analysis. This is usually done by a digital computer or, in some smaller SHM systems, a digital signal processor (DSP) or field-programmable gate array (FPGA). A good review of the various technologies is given in [4]. Data is often stored by the SHM processor for off-line analysis; this is known as *data logging*. Before performing the damage calculation, the SHM processor must validate all the measured data to ensure that errors and noise are adequately suppressed. Most cases of false damage detection are attributable to errors and noise in the measurement data. In this article, we are particularly interested in the types of noise and error present in the data, and we investigate methods of detection and cleansing. The article is concerned with the following errors:

- System errors—calibration and system faults
  - frequency sampling errors—amplitude attenuation, aliasing, phase lag, and downsampling
  - quantization error, overflow error, and underflow error
  - calibration errors—polarity, gain, offset, and linearity
  - signal clipping, saturation, and dropout errors
  - transmission lag and asynchronous channel phase lag
- Signal noise
  - Gaussian noise, tonal noise, and ac power-line interference
  - transducer resonance and mounting resonance
  - inductive coupling, capacitive coupling, and ground loops
  - low-frequency signal drift
  - signal spikes
- Noise in rotating machinery
  - pulse-transducer errors—runt pulse and pulse spike

– Gaussian noise in rotating machinery
– order analysis of rotating machinery.

2 SHM SYSTEM ERRORS—CALIBRATION AND SYSTEM FAULTS

In this section, we look at errors associated with system configuration and describe how these can be detected and avoided. It is important to eliminate these errors because they will affect the SHM analysis. We cover the following errors:

1. frequency sampling errors—amplitude attenuation, aliasing, phase lag, and downsampling;
2. quantization error, overflow error, and underflow error;
3. calibration errors—polarity, gain, offset, and linearity;
4. signal clipping, saturation, and dropout errors;
5. transmission lag and asynchronous channel phase lag.

2.1 Frequency sampling errors—amplitude attenuation, aliasing, phase lag, and downsampling

Frequency sampling occurs at the ADC stage. The ADC converter periodically samples a continuous stream of analog data and then digitizes each sampled value. The rate at which the analog signal is sampled is called the *sample rate* and is expressed in hertz (i.e., samples per second). Selecting a very high sample rate ensures a good approximation to the analog signal; however, the quantity of data collected can quickly swamp the communication network or the SHM processor and quickly fill the available memory if data logging is required. Conversely, setting a low value can introduce several serious and irrecoverable errors in the signal. It is therefore important to establish an appropriate sample rate and provide means to avoiding errors.

To establish an ideal sample rate, we should first look at the process being measured and the analysis required. For example, if we are concerned with damage accumulation on the chassis of a ground vehicle, then most of the vibration energy is in the very low frequency range (typically 0–35 Hz). Therefore, in this case, we are not interested in the higher frequency ranges. According to the Nyquist–Shannon sampling theory [5, 6], we need to set a sample rate of at least twice the desired frequency range. In our example, we therefore require a sample rate of at least 70 Hz. Although the Nyquist–Shannon theory does provide sufficient data for assessing the frequency content of the data, it is not adequate for representing the true amplitude.

Amplitude attenuation errors occur when there are insufficient samples to adequately define the peaks and troughs in the signal. Consider the simple time signal illustrated in Figure 2(a). This signal is made by summing two sinusoidal waves of similar amplitude: the first has a frequency of 5 Hz and the second 35 Hz. If the signal were sampled at only 75 Hz as in Figure 2(b), we see that the amplitude peaks are not captured by the digitized signal. Amplitude attenuation errors are reduced significantly by increasing the sample rate to 5–10 times the highest frequency component in the measured signal (i.e., 350 Hz). This ensures an adequate representation of the peaks and troughs as illustrated in Figure 2(a).

Aliasing errors are illustrated in Figure 2(c) and (d). According to the Nyquist–Shannon sampling theory, the maximum frequency range contained in a sampled signal is equal to half its sample rate. This is commonly referred to as the *Nyquist frequency*. The Nyquist frequency is the highest frequency we can see in a sampled signal. A question therefore arises, when we sample a continuous stream of data at a given sample rate, what happens to frequencies, which exceed the Nyquist frequency? This effect is called *aliasing* and the higher frequencies are effectively reflected back into the lower frequency data causing undesirable and erroneous noise. Figure 2(c) and (d) illustrate the effect of aliasing. In Figure 2(c), the signal is sampled at 70 Hz, so the Nyquist frequency coincides with the high-frequency sine wave (35 Hz) and, consequently, the digitized points miss the 35-Hz wave completely. In Figure 2(d), the high-frequency sine wave (35 Hz) is sampled at only 60 Hz. This gives a Nyquist frequency of only 30 Hz, so the 35 Hz wave is reflected back in the low-frequency region causing a false wave at 25 Hz. This is also illustrated in the frequency domain by showing a plot of the amplitude versus frequency. In this plot, we see, more

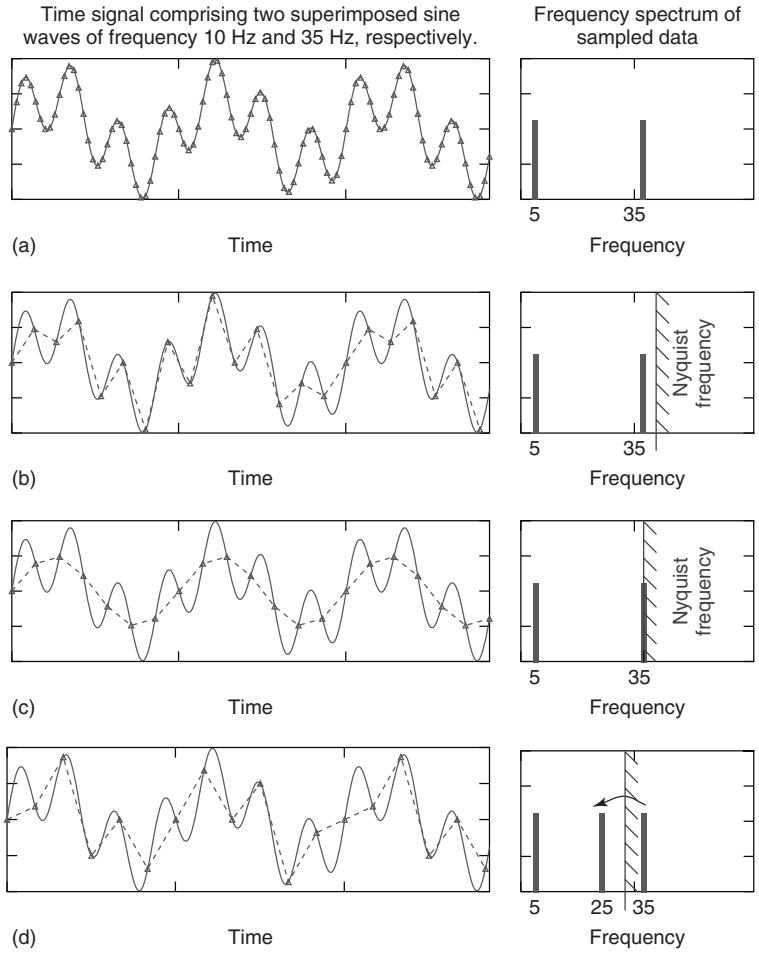


Figure 2. Illustration of amplitude attenuation and aliasing. (a) Data is sampled at 350 Hz (i.e., 10 times maximum frequency) and shows good amplitude and frequency resolution. (b) Data is sampled at 75 Hz; therefore, the Nyquist frequency is $75/2 = 37.5$ Hz. The sampled data therefore shows adequate frequency, but very poor amplitude definition around the peaks. (c) Data is sampled at 70 Hz with a Nyquist frequency equal to the 35 Hz sine wave. The sampled data only shows the 10-Hz component and misses the 35 Hz wave. (d) Data is sampled at 60 Hz therefore the Nyquist frequency is 30 Hz. The 35-Hz wave is now reflected causing an erroneous sine component of 25 Hz.

clearly, how the high-frequency data is reflected back to corrupt the low-frequency signal.

Aliasing errors cannot be corrected after the ADC stage and are very difficult to detect from the digitized data. It is, therefore, very important to filter out all frequencies greater than the Nyquist frequency in the analog signal before the ADC sampling stage. This is accomplished using a hardware "antialiasing" filter. This filter uses an analog electronic circuit that allows low frequencies to pass while higher frequencies are attenuated. It is known as a *low-pass filter*. There are several commonly used hardware filters available including Butterworth, Chebyshev, and Bessel filters. More information on data acquisition and hardware filtering is given in [7].

- **Phase lag**

Antialiasing filters often introduce a phase lag between the input signal and the filtered output. The extent of differential phase lag between channels is an important concern when developing SHM damage algorithms. For example, if we wanted to calculate

the torque in an electric motor by dividing the power by the speed, differential phase lag between the two channels could introduce errors in the calculated values. It is, therefore, important to maintain relative phase between channels by using the same antialiasing filter configuration. This implies using the same sampling rate for all measured channels. In many SHM systems, this is not practical and so phase errors are likely to occur (Section 2.5).

- **Downsampling errors**

Very often it is necessary that we further "downsample" the digitized data on the SHM unit to improve data storage or calculation efficiency. All of the errors discussed above are equally applicable when downsampling digitized data. We must, first of all, filter the data using a low-pass filter and then resample the filtered signal. In the case of software filtering, it is common to use Fourier transform–based filters, or Butterworth, or Chebyshev time-domain filters. In the latter case, we often eliminate phase-lag errors by applying a two-pass approach. The first filter introduces a phase lag as we discussed earlier, so during the second stage we pass the filtered data back through the filter in reverse order. This gives an equal and opposite phase lag, which negates the effect. This two-stage filtering process is only possible in software and is not appropriate for in-line hardware filtering of analog data; it is therefore unsuitable for the antialiasing filter.

2.2 Quantization error, overflow error, and underflow error

Quantization errors (or rounding errors) occur at the ADC stage when the analog signal is converted to a digital representation of ones and zeros. Quantization is the process of approximating a continuous range of analog values to a small set of discrete values. The quantization error is the difference between the true analog value and the approximated (or rounded) digital value. The effect is illustrated in Figure 3.

The straight line in Figure 3 represents the true relationship between voltage measured at the transducer and, in this example, the calibrated acceleration response. However, the ADC converts this into a digital representation. The quality of this representation is governed by the number of bits used. If an 8 bit representation is used over an acceleration range $\pm 50g$, then the maximum quantization resolution is given by equation (1).

$$\text{Quantisation resolution} = \frac{100g}{2^8} = 0.391g \quad (1)$$

In this case, the SHM configuration is unable to detect any change in acceleration less than $0.391g$. The resolution can be improved by increasing the number of bits (12 bit ADC gives $0.024g$ resolution) or decreasing the full-scale accelerometer range ($\pm 5g$ with an 8-bit ADC gives $0.0391g$ resolution). A

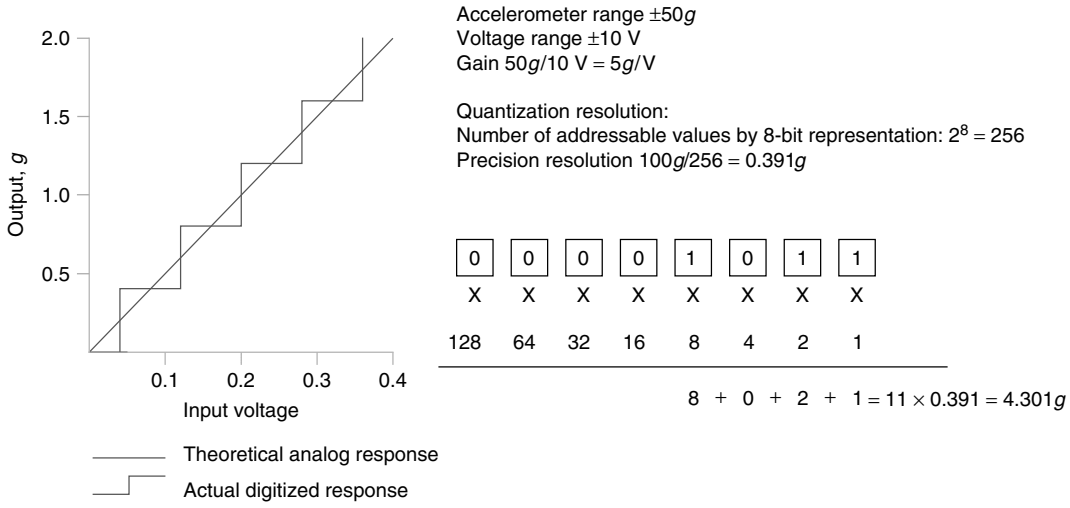


Figure 3. The effect of quantization resolution.

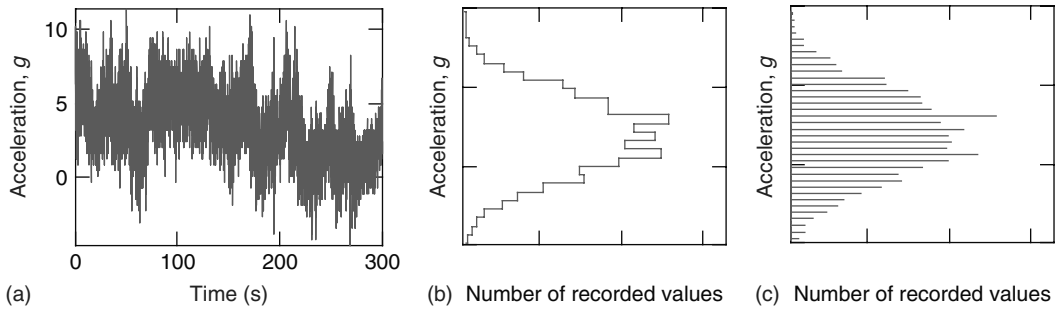


Figure 4. Checking for an adequate quantization resolution. (a) Time signal quantized with 8-bit resolution over a full-scale range $\pm 50g$. (b) Histogram of point values with 8-bit resolution. All histogram bins contain data. (c) Histogram of point values with 12-bit resolution. Many histogram bins contain no data.

process called *dithering* can also be used prior to the ADC stage to improve the apparent resolution. In this case, random noise is added to the analog signal and the signal is sampled at a much higher rate than is actually required (oversampling). The high-frequency signal is then filtered through a low-pass filter and downsampled through averaging. This randomizes the quantization errors and significantly increases the apparent resolution. A detailed review of dithering is offered in [8].

The quantization resolution should be established during the design of the SHM system. This error cannot be properly corrected by digital signal processing during service operation. Digital filtering and signal smoothing appear to improve the resolution, but they merely interpolate between the values and cannot adequately resolve peak values in the signal.

If we are worried about quantization errors, a simple histogram of point values can be used. Calculate the histogram of point values using the desired resolution; if all the histogram bins are filled, then the quantization resolution is satisfactory; if there are significant bins containing no points, then the quantization resolution is less than that required. This is illustrated in Figure 4.

Overflow and *underflow* errors are also associated with quantization. These usually occur at the digital communication or the SHM stage when the bit resolution is inadvertently reduced. For example, a 12-bit value is recoded to 8 bit. If the most significant bits are missed, then this is referred to as an *overflow* error and results in the time signal being reflected back to zero as shown in Figure 5(a). It is often difficult to identify this error from the time signal; however,

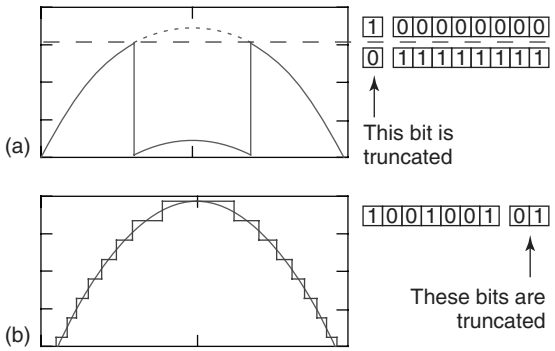


Figure 5. Illustration of overflow and underflow errors. (a) Overflow error: bit representation truncated and misses most significant bits. (b) Underflow error: bit representation truncated and misses least significant bits.

it becomes quite clear from the irregular profile of the amplitude probability distribution function (PDF) and the kurtosis statistic. If the least significant bits are missed then this is referred to as an *underflow* error and the quantization resolution is significantly reduced as illustrated in Figure 5(b). Although it is possible to correct overflow errors, it is very undesirable and so care should be taken when specifying the SHM system to avoid these. It is not possible to correct underflow errors as these lead to quantization error as discussed previously.

2.3 Calibration errors—polarity, gain, offset, and linearity

In a well-calibrated system, the digital signal reaching the SHM processor is proportional to the physical

measurement at the sensor. Most individual components along the signal path are calibrated individually, but very often we prefer to adjust this calibration at the SHM processor to account for the overall system calibration.

Polarity is the most simple calibration parameter. With the correct polarity, an increasing physical measurement should correspond with an increasing digital signal. Although this sounds trivial, it is unfortunately very common to reverse the polarity during installation. One common problem is associated with poor communication of the axis convention between the system designer and those installing the equipment. The error is usually identified during full-system calibration testing. Rather than rewiring or changing the transducer, the effect is more easily remedied at the SHM processor by negating the input signal (or multiplying the signal by -1.0).

Gain and *offset* are two factors used to establish the linear transfer function between input signal and output in the physical units (e.g., acceleration in g or temperature in K). The simple linear transfer function is expressed in equation (2). The gain and offset values are usually established from the full-scale range of the transducer.

$$\text{Output measurement} = (\text{input signal} \times \text{gain}) + \text{offset} \tag{2}$$

Nonlinearity errors occur when the transducer output is not linearly related to the desired measurement parameter. In most cases, this is handled by software at the ADC stage or analog circuitry in the signal conditioning unit. However, it can also be performed by the SHM processor through simple arithmetic manipulation. The most common nonlinear calibration is via a look-up table or a calibration curve that is usually expressed as a polynomial transfer function as illustrated in equation (3). The effect of gain, offset, and linearity is illustrated graphically in Figure 6.

$$\text{Output measurement} = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n \tag{3}$$

where x is the input voltage and a_n are the nonlinear coefficients.

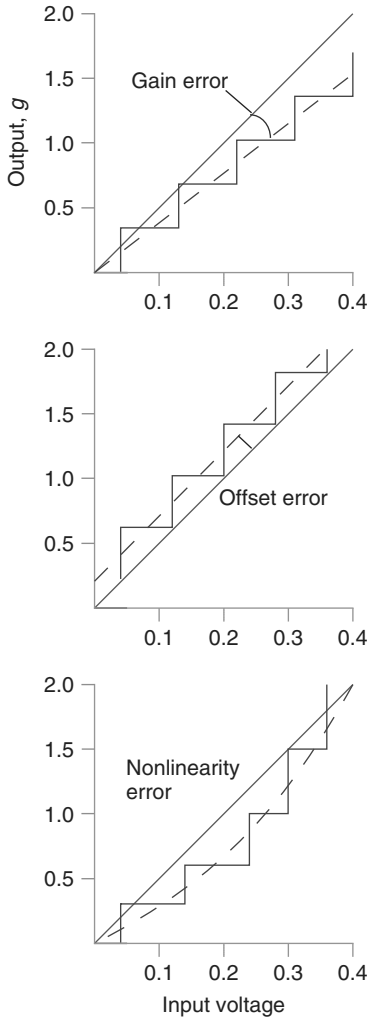


Figure 6. Calibration gain, offset, and nonlinearity.

2.4 Signal clipping, saturation, and dropout

Signal clipping occurs when a measured value exceeds the full-scale range of the SHM sensor. At this point, we are unable to predict the real physical measurement, so damage diagnosis and prognosis are impossible. It is not possible to correct these errors and a change in the hardware configuration is required. Clipping can also indicate a damaged component, so it is always important to address clipping errors unless they are attributable to an erroneous data spike (Section 3.5). Clipping can

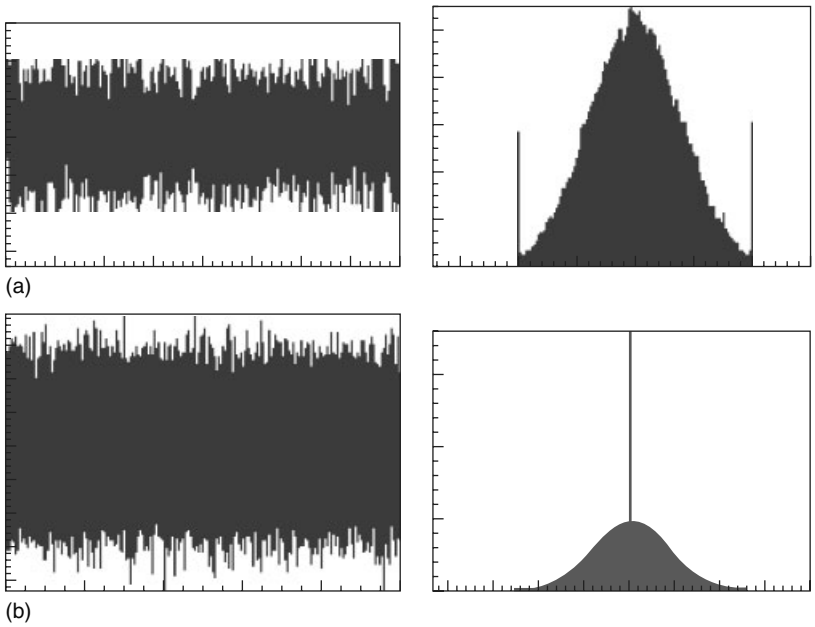


Figure 7. Illustration of clipped data and signal dropout. (a) Example of clipped data. Although it is difficult to detect by simply visualizing the whole time signal, it is clearly visible from the amplitude probability density function (PDF) as peaks at the extreme range. (b) Example of data drop out when the values periodically fall to zero. This can be difficult to detect from the time signal unless you can zoom in on the problematic area of data; however, it is clearly visible from the amplitude probability density function (PDF) as an irregular zero amplitude peak.

be detected by monitoring data values exceeding a particular threshold amplitude. They are also easily identified using the amplitude PDF as illustrated in Figure 7(a). The clipped values create a discontinuity in the PDF at the extremities.

Signal dropout can occur through circuit failure or communication failure. In this case, the measured channel can read full scale or zero depending on the type of sensor and the configuration. Intermittent dropout is usually quite easy to detect as the signal suddenly changes values (to zero or full scale) and then just as suddenly reacquires the value. In these cases, it is easy to locate the erroneous segments but it is impossible to ascertain what the correct values should have been. Damage detection and prognostic algorithms are usually designed to be tolerant of these types of error. They usually interpolate between the correct data and increase the damage parameters proportionally. A fault should be logged when dropout or saturation becomes too frequent as data analysis is unreliable. It can be difficult to determine dropout in data, which spends long periods at zero. The most revealing detection is offered by the amplitude PDF as illustrated in Figure 7(b). Dropout is indicated by an irregular zero amplitude peak.

Signal dropout can also occur due to digital communication errors such as priority override where the channel priority is low on a busy communication bus. In these cases, the SHM usually logs the time of the arriving data packet and can, therefore, asses the last valid data point. Subsequent points usually retain the last valid point.

2.5 Transmission lag errors and asynchronous channel phase lag

Transmission lag affects data transmitted by asynchronous packet-based digital communication. Each sensor periodically tries to broadcast its data along the communication bus. When the data is ready, it checks to make sure the bus is free. If it is busy, then it usually waits for the current transmission to finish before trying again. When the bus is free, the sensor starts to transmit a data packet. The data

packet usually starts with the sender's address. As the address is transmitted, the sender listens to the bus to see if another channel happened to start transmission at exactly the same instance in time. With an Ethernet-based system, both senders would stop and wait a random period before trying to retransmit. In a CAN-based system, the sender with a lower priority stops and waits for the higher priority signal to finish before trying to transmit again. The lower priority sensor notices a higher priority channel because it detects a bit being transmitted in the higher priority address when it, having a lower priority, contains a null bit. This approach to communication is known as *collision-based* and means that there is often a short delay between the data being read and the packets being received by the SHM processor. Also, the delay is not consistent and data received by each channel is not synchronized, i.e., not simultaneously sampled.

To mitigate these transmission errors, the SHM processor usually resamples the channels using a "pseudosimultaneous sample and hold" function as illustrated in Figure 8.

The sample-and-hold buffer monitors the communication bus for transmission of the required sensor channel. This is then held in the buffer awaiting a request from the trigger to present the value. If the buffer is not refreshed between triggers, then the previous value held in the buffer is returned again. If the buffer is refreshed several times between triggers, then either the final buffer value is returned or the value returned is the weighted mean of the values based on recent values having a greater relevance than the earlier ones. This offers an antialiasing capability.

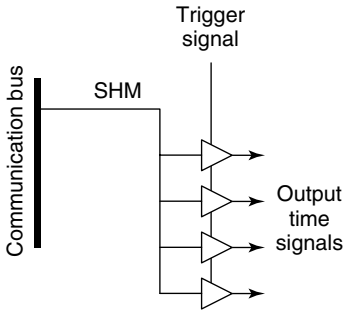


Figure 8. Illustration of pseudosimultaneous sample and hold for nonsynchronous digital communication. Buffers monitor and store communication signals as they arrive on the bus in an asynchronous fashion. The buffers are then read simultaneously based on a periodic trigger signal.

If damage-detection algorithms require simultaneous values from several channels, then this asynchronous form of communication can lead to spikes and dropout in the derived data. It is, therefore, important that these errors are also detected and cleansed before the derived data is used in diagnostic algorithms (Sections 3.5 and 2.4).

3 SIGNAL NOISE

In this section, we look at the effect of interference noise in the signal. We categorize these effects and present signal-processing techniques appropriate for denoising the signal. The types of noise considered include

1. Gaussian noise, tonal noise, and ac power-line interference
2. transducer resonance and mounting resonance
3. inductive coupling, capacitive coupling, and ground loops
4. low-frequency signal drift
5. signal spikes.

3.1 Gaussian noise, tonal noise, and ac power-line interference

The term *noise* is generically used to describe an undesirable component within the signal. Noise is very subjective and depends on what we want to find from the measured signal. For example, we might want to measure the torque on a vehicle drive shaft from the gearbox. In this case, we are interested in a fairly low frequency signal and we might consider the higher frequencies as inconvenient noise that should be removed. However, if we were looking for the progressive fretting of gear teeth, then it is the high-frequency signal that is important and the low-frequency components are now the noise. It is fairly easy to separate frequency-delimited noise using a simple low-pass or high-pass filter. In this example, the low-pass filter isolates the torque component, while the high-pass filter isolates the fretting. However, it is much more complicated when the frequencies start to overlap. Figure 9 illustrates the effect of Gaussian noise on a signal.

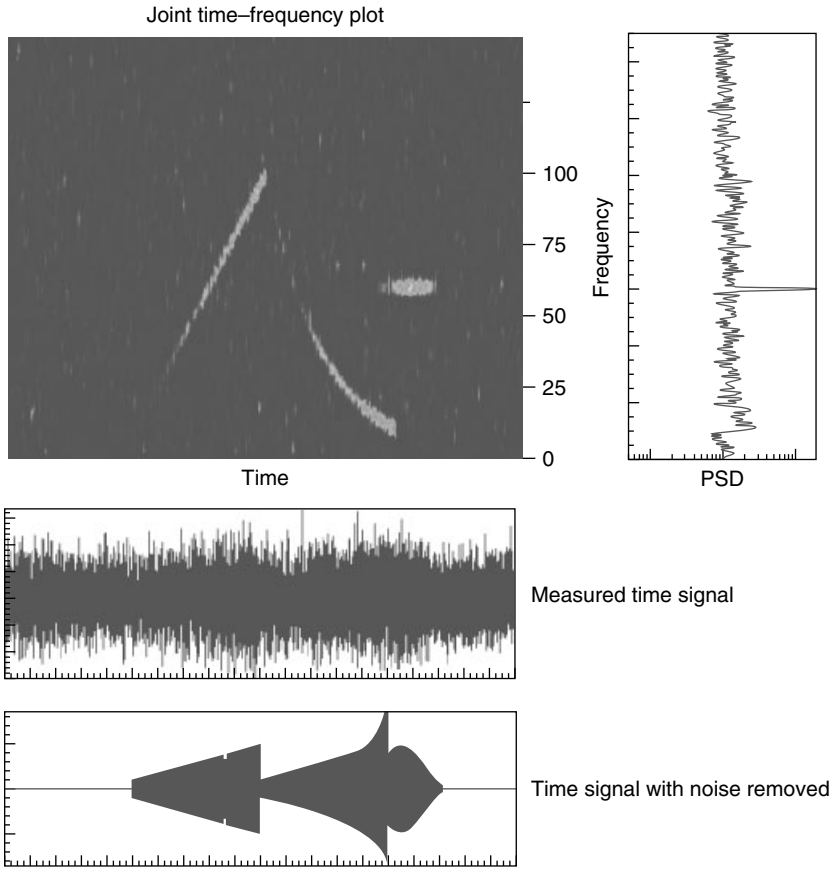


Figure 9. Illustration of Gaussian noise superimposed on an underlying signal.

Gaussian noise is a broadband random noise that is superimposed on the desired signal. It is representative of general background noise when no single source is dominant. It is usually assumed to cover the entire frequency spectrum with constant amplitude and is quantified by the signal-to-noise ratio (SNR). This is the ratio of the signal power to the noise power and is usually expressed in decibels as equation (4):

$$SNR = 10 \times \log_{10}\left(\frac{\text{signal power}}{\text{noise power}}\right) \qquad (4)$$

Tonal noise is narrow band and usually deterministic in nature. It is representative of a single source of harmonic interference. The most common form of tonal noise is due to ac power-line interference.

3.1.1 Frequency-delimited noise—high-pass, low-pass, band-pass, and band-reject filters

When the important signal and the noise occur over separate frequency ranges, we can separate one from the other using simple block filters. In the above example, we can isolate the torque signal from the high-frequency noise by using a low-pass filter, which allows only the lower frequencies to pass through while the higher frequencies are attenuated. The four common block filters are illustrated in Figure 10. In practice, it is undesirable to create very steep filters like these and the attenuation rate is rather lower than the plots suggest. It is, therefore, necessary to ensure an adequate frequency separation between the signal and noise. An excellent introduction to digital filters is offered by Hamming [9].

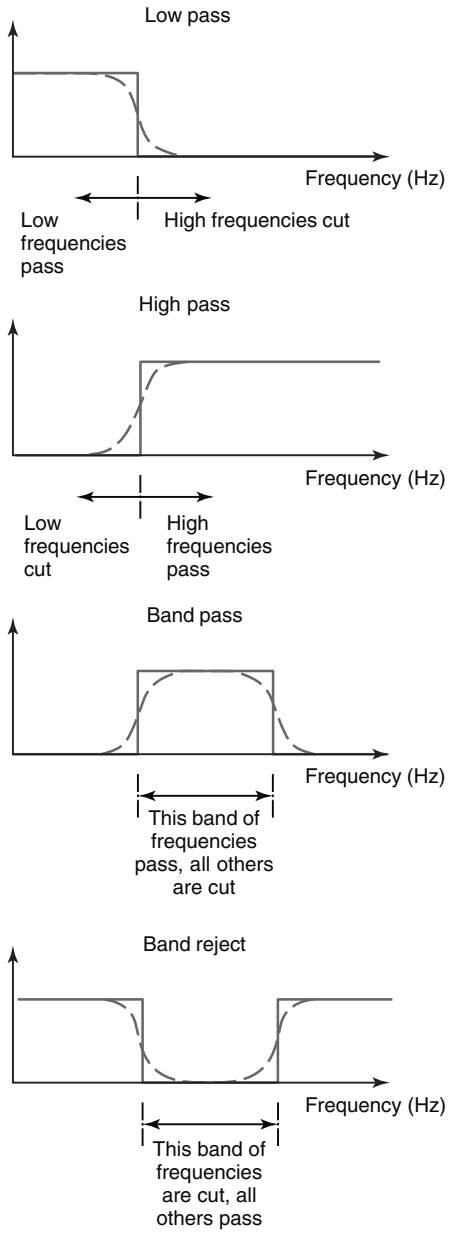


Figure 10. Standard block filters.

Frequency filtering offers a fine control on which frequencies are extracted from the data. A more generic approach to removing high-frequency noise and attenuating spikes is also offered by median smoothing, moving average, or the Loess smoothing approach.

Median smoothing involves taking a buffer of n data points and replacing the midpoint with the median value of the buffer. The buffer is then advanced by one point and the process is repeated. It is a simple and robust filter and is unlikely to be affected by spurious data points making it ideal for spike removal (Section 3.5). However, it does tend to excessively smooth over some sharp features in the data and can also give an irregular and "stepped" profile to the data. This can cause problems in subsequent analysis and is illustrated in Figure 11(a).

Moving average is similar to median smoothing except that the midpoint is now replaced with the mean or the weighted mean of n data points in a buffer. The weighted mean is often preferred as this

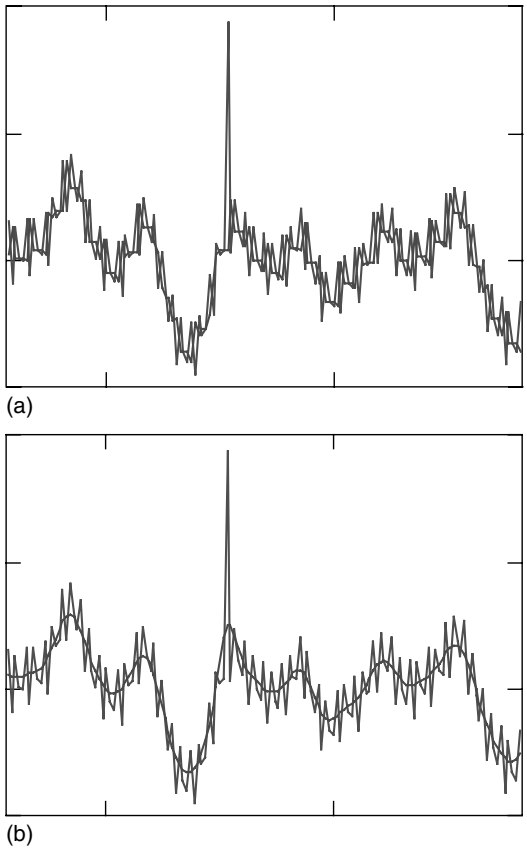


Figure 11. The effect of median smoothing and Gaussian-weighted moving averages. (a) Median smoothing on a sample of data containing high-frequency noise and spike. (b) Gaussian weighted moving average on a sample of data containing high-frequency noise and spike.

lends bias to the values in the immediate vicinity of the point. The most commonly used bias function, (also known as *kernel*), is the Gaussian distribution centered on the point of interest. This approach leads to a smooth data curve that effectively removes high-frequency noise as illustrated in Figure 11(b). It is very similar in application to a time-domain low-pass filter and, in most cases, the low-pass filter is preferred as it allows the engineer to select the noise-frequency delimitation.

Loess smoothing is based on fitting a polynomial regression curve through each local region of the data. By choosing a long period, we can successfully isolate low-frequency contributions such as mean drift. By selecting shorter intervals, we can offer an equivalent low-pass filter as illustrated in Figure 12.

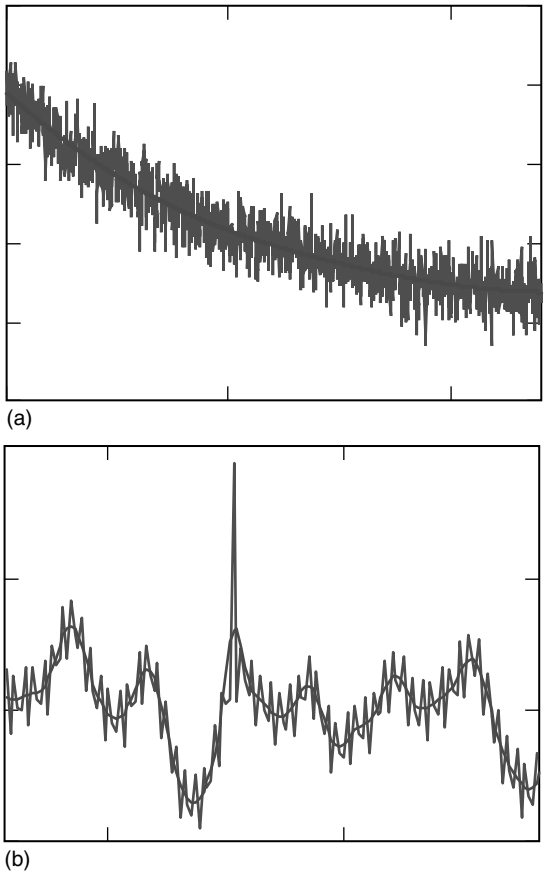


Figure 12. The effect of Loess smoothing for drift correction and high-frequency noise attenuation. (a) Loess smoothing over a long time period can be used to identify the mean drift component in a signal. (b) Loess smoothing over a short period can be used for removing high-frequency noise and attenuating spikes.

3.1.2 Frequency overlapping noise

If the frequency of the noise coincides with the underlying signal, then we cannot use the simple block filters discussed earlier. In this case, a wavelet-based approach is preferred. Using an orthogonal wavelet transform, like the Daubechies D10 wavelet, we can represent the measured time signal in the wavelet domain. We then perform the filtering operations and inverse transform back to the time domain.

There are two types of wavelet filtering known as *scale thresholding* and *magnitude thresholding*. Each wavelet scale pertains to a particular range of frequencies and scale thresholding involves setting all values in a particular scale to zero. This is directly analogous to frequency filtering. Magnitude thresholding involves looking through all the wavelet coefficients and setting those values below a given threshold to zero. This allows us to remove low-energy noise from the signal even though it coincides with the same frequency as the underlying good data. There are two approaches to magnitude thresholding called *hard* and *soft thresholding*. Hard thresholding simply sets the values below the threshold to zero and leaves the higher values untouched. Soft thresholding again sets the values below threshold to zero but also reduces the larger values by the threshold amount.

With both methods, it is important to establish a suitable threshold that extracts noise while leaving the good signal intact. Determining this threshold is usually done through trial and error by visually comparing the filtered signal. Another approach is to compare the mean square error between the filtered signal and the original. As the threshold increases, the mean square error also rises. As we approach the optimal threshold, we notice a pronounced increase in the mean square error. A more general approach is the use of the universal threshold. This offers a simple and convenient way of estimating a suitable threshold for most practical applications. The universal threshold is determined using equation (5). An excellent introduction to wavelet thresholding can be found in Addison [10], while Mallat [11] offers a thorough development of the subject (*see*

Wavelet Analysis).

$$\lambda_U = \sigma(2\ln(N))^{1/2} \tag{5}$$

where σ is the standard deviation of the noise.

3.2 Transducer resonance and mounting resonance error

Transducer resonance occurs when the frequency range of loading starts to excite the natural resonant frequencies of the transducer or the transducer mounting assembly. Resonance errors are minimized by ensuring that the resonant frequencies of the transducer or mounting assembly are at least three times greater than the maximum loading frequency.

It is seldom possible to correct a signal containing transducer resonance unless the resonant frequency and the damping ratio are known. In some cases, this can be determined with reference to the measured frequency response function (FRF). Where these are known, it is possible to construct a Fourier transform–based filter to reduce the amplitude of the resonance from the signal.

3.3 Analog conductor noise—inductive coupling, capacitive coupling, earth looping (ground looping), and cross talk

There are three types of noise associated with analog wiring: inductive coupling, capacitive coupling, and ground looping. An excellent account of these is presented by Bolton [12] and Putten [13].

Inductive coupling is caused by a changing current in a nearby circuit causing a magnetic field, which induces an electromotive force (emf) on the conductors in the analog wiring. This type of noise can be significantly reduced by using twisted pair cabling. The magnetic field induces an opposite emf in each loop, which negates the interference effect. Radio interference is also fairly common and can be significantly attenuated by using an Radio Frequency suppressor. Differential amplifiers can also be used to considerably reduce noise. This is where the required data is obtained by amplifying the difference between the two signals. If both signals contain the same interference, then the amplifier will not amplify any of the interference noise.

Earth looping (ground loop) is where multiple earth points occur and can be the cause of seemingly random drift as both earth points may be at a different potential. This source of error can be eliminated by ensuring only a single earth.

Capacitive coupling is caused by capacitance between the conductor and nearby power cables or earth. This type of noise can be reduced by adequate screening of signals using earthed enclosures and coaxial cabling. Although the screen should be earthed, this can give rise to earth loops, so care is required to ensure that cables are earthed at only one end.

Channel cross talk is where a signal transmitted in one channel has an undesirable effect on another. This can occur through poor screening. Cross-talk interference is a mixture of inductive and capacitive interference. In most cases, the cross-talk component will be low compared with the required signal and the wavelet-based denoising approach is sufficient to remove the noise component.

3.4 Low-frequency signal drift errors—mean drift and dc drift

Drift is the term used to describe a signal whose mean level is slowly changing over time as illustrated in Figure 13. Signal drift can occur for many reasons, such as the differential thermal expansion of a strain gauge with respect to the component it is measuring, or ground looping.

Drift is easily detectable by eye and also from trends in the ensemble mean. It is also often possible to detect drift using the skewness statistic and the shape of the amplitude probability distribution. Mean drift can be easily corrected through frequency filtering, regression analysis, or by using a wavelet-based approach.

Where the drift frequency is low compared with the frequency content of the measured signal, we can use a high-pass filter to remove the low-frequency drift component. Butterworth, Chebyshev, and Bessel filters are all suitable for software implementation. However, they become progressively less suitable as the drift frequency becomes very low (typically less than 1% of sample rate), or where there is a significant

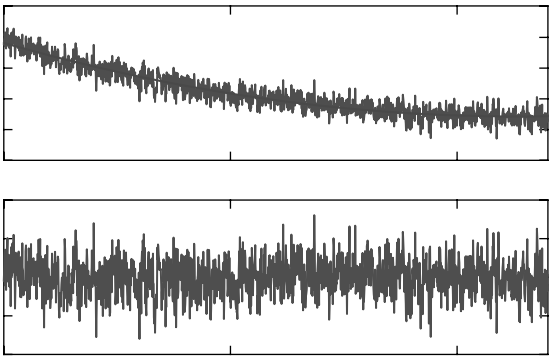


Figure 13. Illustration of mean drift. Measured strain data containing very low frequency drift due to differential thermal expansion of the strain gauge. Calculate polynomial regression line though data and subtract to two signals to recreate the corrected response.

low-frequency component in the real signal that makes it difficult to isolate the drift frequency from that of real data. Fourier-based filtering techniques should be avoided because the buffer must be larger than the drift period, which often makes them impractical.

For very low frequency drift, it is more appropriate to fit a statistical regression curve through the data and then extract the values from the original data as illustrated in Figure 13. The *Loess* smoothing technique is often most suitable in identifying the appropriate regression curve. Loess smoothing is based on fitting a polynomial regression curve through each local region of the data. By choosing a long period, we can successfully isolate low-frequency drift and then subtract this from the original measured signal.

Wavelet-scale thresholding can also be used to remove the low-frequency drift. Each wavelet scale pertains to a particular range of frequencies and scale thresholding involves setting all values in a particular scale to zero. This is directly analogous to frequency filtering. Wavelet filtering is based on the Fourier transform and will require a buffer that is longer than the period of the drift; this often renders the approach impractical.

3.5 Signal spikes

A spike is usually defined as a freak data point whose amplitude is significantly different from that of the immediate surrounding data. Spikes can originate from communication errors and capacitive and inductive noise in the analog signal path. Figure 14 illustrates three cases of spike.

- ***Extreme amplitude spikes***

In many cases, a spike will have an extreme amplitude, which approaches full scale. In this case, it is quite easy to detect and remove the spike. An amplitude threshold is established within which all data is determined as clean. Points exceeding this threshold are classified as spikes and are replaced with interpolated values based on the adjacent data points. The threshold values are easily determined using an amplitude PDF as shown in Figure 14(a). This clearly differentiates between the range of good data and the apparent spikes. The presence of high-amplitude spikes also affects the kurtosis and crest factor statistics of the data. It is, therefore, possible to monitor changes in the ensemble statistics to quickly locate areas of spikes and then process them using the spike cleansing routine described here. Extreme amplitude spikes are very similar to clipping but spikes are usually recognized as erroneous data, whereas clipping is real data.

- ***Extreme gradient spikes***

Some spikes do not have an extreme amplitude, which is distinguishable from the amplitude PDF used above. In these cases, the spike amplitude is still clearly different from that of the surrounding data; however, it is not outside the amplitude PDF of the real data. These are typical of erroneous communication packets where the erroneous data is often replaced with a zero value. This is the same phenomenon as dropout. In these cases, we can usually look at the instantaneous gradient of the data (derivative of the data). Spikes are usually classified by abnormal gradients and can be seen quite clearly from the gradient PDF as illustrated in Figure 14(b). A simple running calculation can be performed to calculate the gradient of each point and identify those exceeding the threshold. The spikes are then replaced with data based on interpolation of the clean data that surrounds them as before.

- ***Running standard deviation***

This method calculates the standard deviation over a short window of data and identifies a spike as the point that exceeds a specified multiple of the

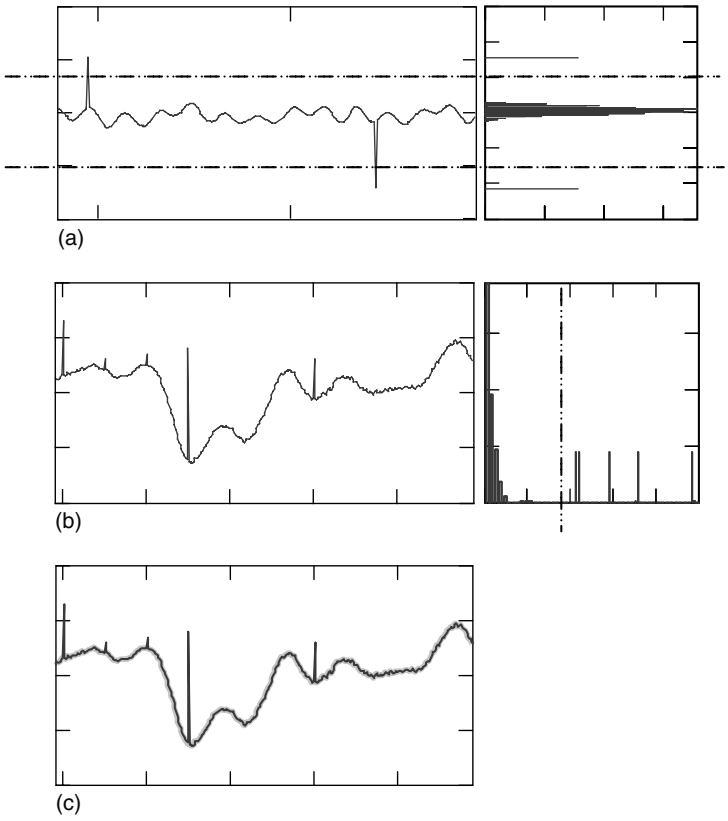


Figure 14. Illustration of spike-detection algorithms. (a) Extreme amplitude spikes: Amplitude of spikes significantly exceed amplitude of real data. Determine threshold amplitude using amplitude probability distribution. (b) Extreme gradient spikes: Gradient of spikes significantly exceed gradient of real data. Determine threshold gradient using probability distribution of gradient. (c) Running standard deviation: Calculated the running standard deviation. A spike is detected when a value passes through a specified multiple of this value.

standard deviation. The buffer is then advanced by one point and the standard deviation is recalculated. In this way, spikes can be identified on the basis of a combination of amplitude and gradient. The window varies in length and is often defined by a decaying function where neighboring points are most influential and more distant points are less influential. A Gaussian kernel function is often used in this application. A problem with this approach occurs when the signal remains at a constant value for some time: the standard deviation then tends to zero, so any departure from the constant would be falsely identified as a spike. A "gate" value is then applied, which represents the minimum amplitude change for the spike-detection algorithm. An analogous approach based on the *running crest factor* might also be used.

- ***Important note***

Data spikes are easily detectable provided that the data has not been filtered using a low-pass filter or a smoothing technique beforehand. Low-pass filtering not only attenuates the amplitude of the spike but also rescales the adjacent points and consequentially smooths the signal. The smoothed spike is undetectable using these algorithms and so it is important to remove spikes prior to using any low-pass filter. Although low-pass filtering and smoothing attenuate the spike amplitude, the amplitude can still incur significant errors in prognostic and diagnostic SHM algorithms, especially where damage is exponentially related to cycle amplitude. It is, therefore, preferable to remove the spike rather than simply smoothing it.

4 NOISE IN ROTATING MACHINERY

In this section, we look at specific techniques for removing noise and errors in signals recorded from rotating machinery. In particular, we consider

1. pulse-transducer errors—runt pulse and pulse spike
2. Gaussian noise in rotating machinery
3. order analysis of rotating machinery.

4.1 Pulse-transducer errors—runt pulse and pulse spike

Many pulse-train transducers such as Hall-effect and variable-reluctance transducers are used for rotational speed and position sensing. These transmit a digital pulse as a toothed gear approaches and then retreats from the transducer. The rotational speed can be determined by either measuring the number of pulses seen over a fixed period of time, or alternatively by measuring the time elapsed between two successive pulses. Where rotational position is also required, there are several methods of locating the reference point and these are discussed by Gyorki [7] and Hillier *et al.* [2]. A cost-effective and commonly used approach uses a missing tooth as the reference point. A typical system is illustrated in Figure 15. When rotational direction is also important, a second transducer can be placed out of phase with the first one. The phase difference between the two pulse trains then indicates direction of revolution. This is known as *quadrature position measurement.*

When rotation speed is measured by frequency, i.e., by counting the number of pulses that arrive in a specified period of time; errors in resolution are caused at low speed when only a few pulses arrive. For instance, a slowly rotating shaft gives an average pulse count of 4.5 pulses per period. This results in 4 pulses in the first period followed by 5 in the second, etc. This leads to low resolution at lower rotational speeds. The resolution problem is less problematic as the rotational speeds increase. We can improve resolution by increasing the size of the window; however, that means a longer time period and therefore a reduced time resolution. This is directly analogous to quantization error (Section 2.2).

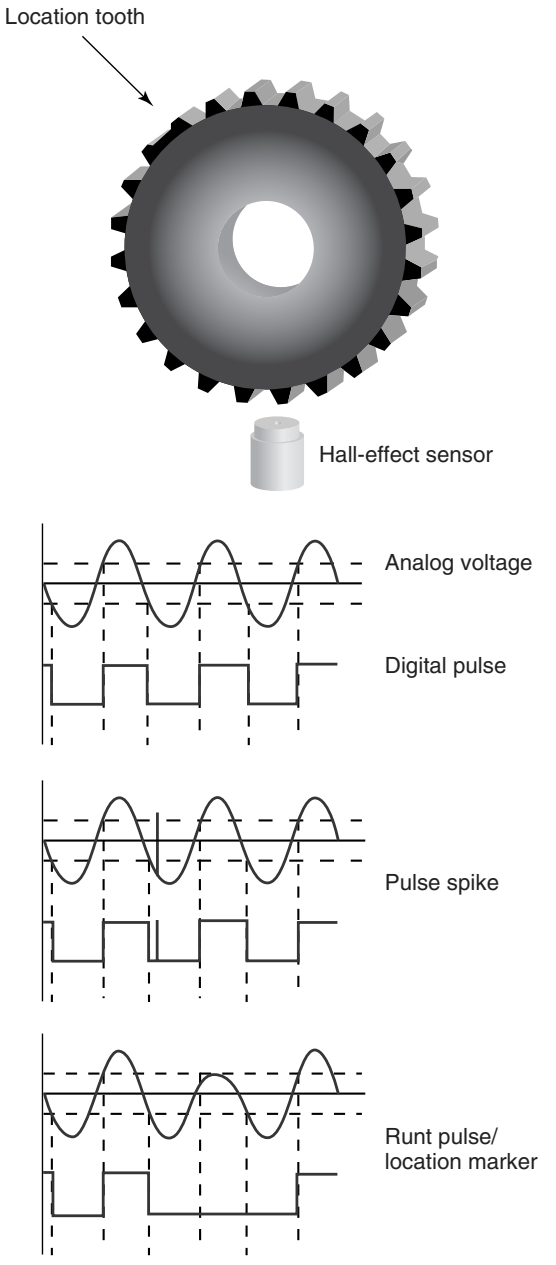


Figure 15. Illustration of pulse spike and runt pulse errors.

When rotation speed is measured by taking the time between successive pulses, a very accurate low-speed resolution is achievable; however, this is reduced as the rotation speed increases and the time between pulses is shorter. When the pulse rate approaches half

the clock rate, errors will start to increase. When the pulse rate exceeds half the clock rate, aliasing will occur Section 2.1.

A *runt pulse* is defined as a pulse that is too small to be detected by the transducer. This is fairly common in exposed sensors where debris can build up between the teeth and reduce the pulse definition. A *pulse spike* is the opposite problem where a false pulse is detected owing to electrical interference in the system or "contact bounce" in the signal.

Where speed is measured by counting the number of pulses passing over a specified period, runt pulses and spikes result in a slowly building inaccuracy that depends on the ratio of erroneous pulses to the average number of pulses counted. When speed is measured by counting the time elapsed between successive pulses, the errors result in large and immediate spikes in the observed speed. In this case, it is easy to identify and locate the erroneous pulse using the spike-detection methods described earlier (Section 3.5). This approach is hampered when rotational position is also coded using a missing tooth because this has the same effect as the runt pulse. It is, therefore, necessary to differentiate between the correct position reference and the runt pulse. If this error occurs, then the system will generally require maintenance.

4.2 Gaussian noise from rotating machinery—time synchronous averaging

Many signals measured from rotating machinery show strong correlation in the angle domain. For example, suppose we want to measure the cylinder pressure in a petrol engine; if we sample the cylinder pressure at a constant sample rate, we observe the pressure variations with respect to time along with a superimposed Gaussian noise. However, if we consider how the pressure varies over time at only one particular crank angle, say 10°, we see a much steadier variation of pressure. The Gaussian noise effectively introduces a symmetric random scatter to the steady underlying signal. We can, therefore, remove this noise by simply averaging the constant angle values. So, for any particular rotation angle, we can compute the running average of the measured values to remove the Gaussian noise component. This process is called *time synchronous averaging* (*TSA*).

TSA requires that we simultaneously measure the rotation angle of the machine as well as the actual signal required. We can then filter the measured signal as follows:

1. choose a set of desired rotation angles (e.g., 0, 1, 2,. . ., 359°);
2. interpolate the rotation angle signal to find the times at which the required angle is encountered;
3. interpolate the measured signal at the times derived above and extract the values;
4. perform a running average (or median smoothing) on the extracted values to establish the clean signal; and
5. increment the angle in step 1 and recalculate steps 2–4.

In many cases, the results in the angle domain from step 5 will be suitable for SHM analysis directly; however, if we need to reconstruct a constant sample rate time signal, then we would need to interpolate the filtered values between successive rotation angles over the constant sample rate. This stage is often computationally intensive and care is required to avoid aliasing errors introduced by resampling the data over a limited set of angles. Assuming that we want to obtain a new time signal with the same sample rate as the original time series, the minimum number of angle increments is determined from equation (6). We can then downsample the time signal by low-pass filtering and resampling if required.

$$\text{Number angle increments} > \frac{\text{Sample Rate}}{\text{Minimum rotational speed (rps)}} \tag{6}$$

4.3 Order analysis from rotating machinery—Vold–Kalman order tracking

When we review the measured data recorded off complex rotating machinery, we often observe dominant harmonics of the rotor frequency. For example, a four-cylinder, four-stroke engine "fires" twice with each rotation of the crank giving a dominant vibration at twice the engine speed. Similarly, the cam turns

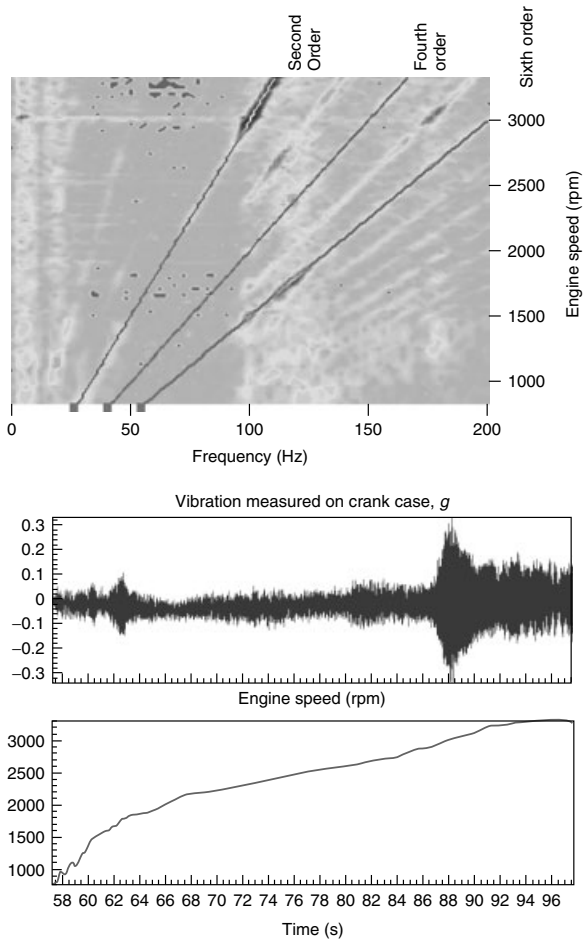


Figure 16. Rotational order analysis of a four-cylinder, four-stroke engine.

once for every two rotations of the crank and various pumps may rotate several times. It is often desirable for SHM analysis to look at trends in the harmonic response and this is called *order analysis*. The first order is the engine frequency, while the nth order is defined as $n\times$ engine speed, where the engine speed is usually expressed in hertz (or revolutions per second (rps)). The vibration response measured on a typical four-cylinder engine is illustrated in Figure 16.

In many cases, we are only interested in the signal pertaining to a particular order and ideally we should filter out the other frequencies to make any trend in the desired order more pronounced. In this case, we could use a band-pass filter but the filter frequency must continually change with respect to the engine speed. The Vold–Kalman order tracking filter is often used for this purpose. It is an adaptive band-pass filter in which the pass-band varies with respect to the engine speed and the order required.

5 CONCLUSION

This article has described how measurement errors and noise are introduced to SHM signals between the transducer and the SHM processor. The errors are classified in terms of system errors—including calibration and system configuration errors and signal noise interference along the signal path. It has described the physical cause of the errors and

identified how the symptoms can be identified within the signal, using basic signal-processing analysis. Wherever possible, algorithms are described, which can remove the erroneous artifacts and effectively denoise the signal. Many methods are discussed ranging from real-time cleansing of data spikes and frequency extraction to more complicated wavelet-based techniques.

REFERENCES

[1] Vachtsevanos G, Lewis F, Roemer M, Hess A, Wu B. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. John Wiley & Sons, 2006.

[2] Hillier V, Coombes P, Rogers D. *Hillier's Fundamentals of Motor Vehicle Technology–Powertrain Electronics*. Nelson Thornes Ltd: Cheltenham, 2006.

[3] Bosch R. *Automotive Handbook*, *Sixth Edition*. Robert Bosch GmbH, ISBN 1-86058-474-8.

[4] Stranneby D, Walker W. *Digital Signal Processing and Applications*, *Second Edition*. Elsevier Science, 2004.

[5] Nyquist H. Certain topics in telegraph transmission theory. *AIEE* 1928 **47**:617–644, Reprinted *IEEE* 2002 **90**(2).

[6] Lüke H. The origins of the sampling theorem. *IEEE Communications Magazine* 1999 **37**: 106–108.

[7] Gyorki J. *Signal Conditioning and PC-Based Data Acquisition Handbook*. IOtech Inc, 2004, ISBN 0-9656789-3-8.

[8] Schuchman L. Dither signals and their effect on quantization noise. *IEEE Transactions on Communications* 1964 **12**(4):162–165.

[9] Hamming R. *Digital Filters*. Dover Publications, 1998.

[10] Addison P. *The Illustrated Wavelet Handbook*. Institute of Physics, 2002.

[11] Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[12] Bolton W. *Instrumentation and Measurement Pocket Book*. Newnes, 2001.

[13] Putten A. *Electronic Measurement Systems*. Prentice-Hall, 1988.

Chapter 23
Statistical Time Series Methods for SHM

Spilios D. Fassois and John S. Sakellariou
Department of Mechanical and Aeronautical Engineering, University of Patras, Patras, Greece

Important Conventions

Bold-face upper/lower case symbols designate matrix/column—vector quantities, respectively.

Matrix transposition is indicated by the superscript T.

A functional argument in parentheses designates function of a real variable; for instance, $x(t)$ is a function of analog time $t \in \Re$.

A functional argument in brackets designates function of an integer variable; for instance, $x[t]$ is a function of normalized discrete time ($t = 1, 2, \ldots$). The conversion from discrete normalized time to analog time is based upon $(t-1)T_s$, with T_s designating the sampling period.

A functional argument including the imaginary unit j designates complex function; for instance, $X(j\omega)$ is a complex function of ω.

A hat designates estimator/estimate of the indicated quantity; for instance, $\widehat{\boldsymbol{\theta}}$ is an estimator/estimate of $\boldsymbol{\theta}$.

1 INTRODUCTION

Statistical time series methods for structural health monitoring utilize (i) random excitation and/or vibration response signals (time series), (ii) statistical model building, and (iii) statistical decision making for inferring the health state of a structure. This includes fault detection, identification (localization), and magnitude estimation, which may be collectively referred to as *fault diagnosis*. As with all vibration-based methods, the *fundamental principle*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

upon which they are founded is that small changes (faults) in a structure cause discrepancies in its vibration response, which may be detected and associated with a specific cause (fault type and magnitude). See references [1–11] for overviews of general vibration-based methods; also see [12, 13]. Statistical time series methods are discussed in references such as [14–19]—see [20] for a recent overview.

Statistical time series methods for structural health monitoring use random excitation and/or response (displacement or velocity or acceleration) signals from the structure in its healthy state, as well as from a number of potential faulty states, identifying suitable (parametric or nonparametric) statistical time series models describing the structure in each state, and extracting a statistical quantity (*characteristic quantity*) characterizing the structural state in each case (*baseline phase*). When the structure is to be inspected for faults, the procedure is repeated under the current conditions, and the current characteristic quantity is obtained. Fault detection, identification, and estimation are then accomplished via statistical decision making consisting of "comparing", in a statistical sense, the current characteristic quantity with that of each potential state as determined in the baseline phase (*inspection phase*).

Statistical time series methods are fundamentally of the inverse type, as the models used are *data based* rather than physics based, and inherently account for uncertainty. In addition to the main features of general vibration-based methods (no need for visual inspection, "global" coverage of the structure, time and cost effectiveness, and automation capability), statistical time series methods offer further unique advantages such as (i) no need for physics-based or finite element models; (ii) no need for complete structural models—in fact, partial models and a limited number of excitation-response signals are often used; (iii) inherent accounting of (environmental, measurement, and so on) uncertainty; (iv) statistical decision making with specified performance characteristics; and (v) effective use of natural random vibration data records (no need to interrupt normal operation).

On the other hand, as complete structural models are not employed, time series methods may identify (locate) a fault only to the extent allowed by the type of model employed. Other limitations include the need for extensive "training" in the baseline phase (necessary for fault identification and estimation), adequate expertise on part of the user, potentially limited physical insight, and sensitivity expectedly lower than that of "local" nondestructive testing-type methods.

In this article, an overview of the main classes of statistical time series methods for structural health monitoring is presented. *Parametric* and *nonparametric* methods in either time or frequency domains are outlined for the *excitation–response* and *response-only* cases. The focus is on *periodic inspection*, although extensions to *continuous monitoring* are possible—a simple generalization may be based on the continuous repetition of the inspection phase with data corresponding to a moving time window. The focus is also on *time-invariant* (stationary) *linear* structures in *Gaussian* environments, although certain versions of the methods are available for alternative cases and pertinent references are provided in the bibliographical remarks. For simplicity of presentation, the *scalar* vibration response signal case is treated.

2 THE WORKFRAME

Let $\mathcal{S}_o$ designate the structure under consideration in its *nominal* (healthy) state, $\mathcal{S}_A$, $\mathcal{S}_B$, ... the structure under fault of *type (mode)* A, B, ..., and so on, and $\mathcal{S}_u$ the structure in an unknown (to be determined) state. Each fault type includes a continuum of faults, which are characterized by common nature or location but distinct magnitudes (for instance, damage of various magnitudes in a specific structural element). The structure under a fault of type, say V, and magnitude k is designated as $\mathcal{S}_V^k$, while the fault itself is designated as F_V^k.

Statistical time series methods are based on discretized excitation $x[t]$ and/or response $y[t]$ (for $t = 1, 2, \ldots, N$) random vibration data records. Note that t refers to discrete time, with the corresponding actual time instant being $(t-1)T_s$, where T_s stands for the sampling period. Let the complete excitation and response signals be represented as X and Y, respectively, or collectively as Z, that is, $Z = (X, Y)$. Like before, a subscript (o, A, B, ..., u) is used for designating the corresponding state of the structure that provided the signals.

Note that all collected signals need to be suitably preprocessed [3, 21, 22]. This may include low- or band-pass filtering within the frequency range

Table 1. Workframe setup: structural state, vibration signals used, and the estimated characteristic quantity (baseline and inspection phases)

Structural state	Vibration signals	Estimated characteristic quantity
	Baseline phase	
S_o (healthy structure)	$z_o[t] = (x_o[t], y_o[t])\ t = 1, 2, \ldots, N \quad Z_o = (X_o, Y_o)$	$\widehat{Q}_o$
S_A (fault A)[(a)]	$z_A[t] = (x_A[t], y_A[t])\ t = 1, 2, \ldots, N \quad Z_A = (X_A, Y_A)$	$\widehat{Q}_A$
S_B (fault B)[(a)]	$z_B[t] = (x_B[t], y_B[t])\ t = 1, 2, \ldots, N \quad Z_B = (X_B, Y_B)$	$\widehat{Q}_B$
$\vdots$	$\vdots$	$\vdots$
	Inspection phase	
S_u (current structure) (unknown state)	$z_u[t] = (x_u[t], y_u[t])\ t = 1, 2, \ldots, N \quad Z_u = (X_u, Y_u)$	$\widehat{Q}_u$

[(a)] Normally various fault magnitudes are considered.

of interest, signal subsampling (in case the originally used sampling frequency is too high), estimated mean subtraction, as well as proper scaling. The latter is not only used for numerical reasons but also for counteracting—to the extent possible—different operating (including excitation levels) and/or environmental conditions. In the case of linear time-invariant structures, the estimated mean for each signal is typically subtracted, and the signal is scaled (being divided) by its estimated standard deviation. In case of multiple excitations, care should be exercised in order to ensure minimal cross-correlation among them.

The obtained data are subsequently analyzed by parametric or nonparametric time series methods and appropriate models are identified and validated [23–26]. Such models are identified on the basis of data Z_o, Z_A, Z_B, ... in the *baseline* or *training phase* (normally data sets corresponding to various fault magnitudes per fault type are used), and based on data Z_u in the *inspection phase*. From each estimated model, the corresponding estimate of a characteristic quantity Q is extracted ($\widehat{Q}_o$, $\widehat{Q}_A$, $\widehat{Q}_B$, ... in the baseline phase; $\widehat{Q}_u$ in the inspection phase—see Table 1).

Fault detection is then based on proper comparison of the true (but not precisely known) Q_u to the true (but also not precisely known) Q_o via a binary composite statistical hypothesis test that uses the corresponding estimates—see Table 2 in which $\sim$ indicates a proper relationship (such as equality, inequality, and so on). *Fault identification* is similarly based on the proper comparison of Q_u to each of Q_A, Q_B, ... via a multiple statistical hypothesis test that also uses the corresponding estimates, and *fault estimation* is based on interval estimation techniques (Table 2). The workframe of a general statistical time series method is illustrated in Figure 1.

Note that the design of a binary statistical hypothesis test is generally based on the probabilities of *type I* and *type II* error, or else the *false alarm* (α) and *missed fault* (β) probabilities. The designs presented herein are based on the false alarm probability. In selecting α it should be born in mind that as α decreases (increases) β increases (decreases). The

Table 2. The fault detection, identification, and estimation subproblems

Fault detection	$H_0 : Q_u \sim Q_o$	Null hypothesis—healthy structure
	$H_1 : Q_u \not\sim Q_o$	Alternative hypothesis—faulty structure
Fault identification	$H_A : Q_u \sim Q_A$	Hypothesis A—fault type A
	$H_B : Q_u \sim Q_B$	Hypothesis B—fault type B
	$\vdots$	$\vdots$
Fault estimation		Estimate k given the fault type

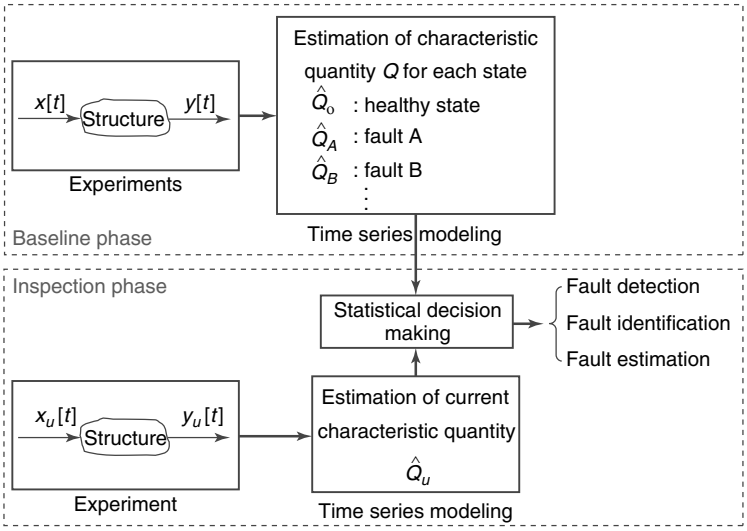


Figure 1. Workframe for statistical time series fault detection, identification, and estimation methods (in the baseline phase Q is normally estimated for various fault magnitudes per fault type).

reader is referred to references such as [27, subsection 4.2]; [28, subsection 3.3] for details on statistical hypothesis testing.

Statistical time series methods may be classified as *parametric* or *nonparametric*, depending upon whether the characteristic quantity Q is based on a parametric or nonparametric statistical model; see Section 3. Section 4 discusses model identification. Nonparametric methods tend to be simpler and computationally more efficient—they are treated in Section 5. Parametric methods may offer superior performance—they are treated in Section 6. Statistical time series methods may be also classified as *excitation–response*, where both excitation and response signals are used, or *response-only*, where only response signals are used. Both types are treated in the present article. The application of the methods for fault detection, identification, and magnitude estimation on a laboratory structure is illustrated in Section 7.

3 STATISTICAL TIME SERIES MODELS

Time series models for linear time–invariant (stationary) structures operating in a Gaussian environment are briefly reviewed in this section. Reference to models applicable to alternative cases are provided within the bibliographical remarks made within the presentation of the methods (Sections 5 and 6).

Let $h[t]$ designate the time-discretized impulse response function describing the causality relationship between an excitation and a response location on a linear time-invariant structure (Figure 2). Also let $n[t]$ designate a stationary Gaussian noise corrupting the response signal. The noise is generally assumed to be zero mean but of unknown autocovariance (or auto power spectral density) and mutually uncorrelated with the excitation $x[t]$. The convolution summation plus noise expression relating these signals is

$$y[t] = h[t] \star x[t] + n[t] = \sum_{\tau=0}^{\infty} h[\tau] \cdot x[t-\tau] + n[t] \tag{1}$$

with $\star$ designating discrete convolution. Alternatively, using the backshift operator $\mathcal{B}$ ($\mathcal{B}^i \cdot y[t] =$

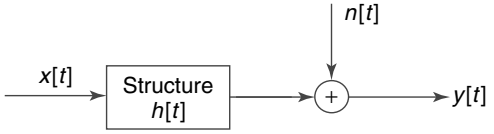


Figure 2. Time-domain representation of a linear time-invariant structure with additive response noise.

$y[t-i]$):

$$y[t] = H(\mathcal{B}) \cdot x[t] + n[t] \tag{2}$$

$$H(\mathcal{B}) = \sum_{\tau=0}^{\infty} h[\tau] \cdot \mathcal{B}^{\tau} \tag{3}$$

with $H(\mathcal{B})$ designating the discrete-time structural transfer function.

3.1 Nonparametric models

Assuming $x[t]$ to be a random stationary excitation, $y[t]$ will also be stationary in the steady state. In addition, $y[t]$ will be Gaussian if $x[t]$ and $n[t]$ are jointly Gaussian. Then each signal is fully characterized by its first two moments, *mean* and *autocovariance function (acf)*—or its normalized version, or its Fourier transform referred to as the *auto power spectral density (psd)* [24, p. 3], [25, pp. 39–40], as shown in Table 3. Note that $E\{\cdot\}$ designates statistical expectation, j the imaginary unit, τ time lag, $\omega \in [0, 2\pi/T_s]$ frequency in rad/s, and T_s the sampling period.

In the excitation–response case, a complete joint description of the excitation and response signals is given in terms of the means and the acfs/cross covariance function (ccf)—or their normalized versions, or the auto psds and the cross spectral density (csd), as indicated in Table 4.

Note that the characteristics of the response are related to those of the excitation and the noise through the expressions [25, pp. 455–456]:

$$\mu_y = H(j\omega)|_{\omega=0} \cdot \mu_x \tag{4}$$

$$\gamma_{yx}[\tau] = \gamma_{xx}[\tau] \star h[\tau]$$

$$\gamma_{yy}[\tau] = \gamma_{yx}[\tau] \star h[-\tau] + \gamma_{nn}[\tau] \tag{5}$$

$$\left.\begin{aligned} S_{yx}(j\omega) &= H(j\omega) \cdot S_{xx}(\omega) \\ S_{yy}(\omega) &= H^{\star}(j\omega) \cdot S_{yx}(j\omega) + S_{nn}(\omega) \end{aligned}\right\} \Longrightarrow$$

$$S_{yy}(\omega) = |H(j\omega)|^2 \cdot S_{xx}(\omega) + S_{nn}(\omega) \tag{6}$$

$$\text{Coherence: } \gamma^2(\omega) = \frac{|S_{yx}(j\omega)|^2}{S_{xx}(\omega) \cdot S_{yy}(\omega)} = \frac{1}{1 + \dfrac{S_{nn}(\omega)}{|H(j\omega)|^2 \cdot S_{xx}(\omega)}} \quad \in [0, 1] \tag{7}$$

where $\star$ designates discrete convolution, the superscript * complex conjugation, $|\cdot|$ complex magnitude, $H(j\omega) = H(B)|_{B=e^{-j\omega T_s}}$ is the corresponding structural frequency response function (frf), $S_{nn}(\omega)$ the noise auto psd, and the last expression defines the (squared) coherence function [29, p. 196].

3.2 Parametric models

Parametric models may be obtained via proper parametrizations of equation (1). In the response-only case it is assumed, without loss of generality,

Table 3. Nonparametric response-only models

Mean	$\mu_y = E\{y[t]\}$	Autocovariance function (acf): $\gamma_{yy}[\tau] = E\{\tilde{y}[t] \cdot \tilde{y}[t-\tau]\}$ or normalized acf: $\rho_{yy}[\tau] = \gamma_{yy}[\tau]/\gamma_{yy}[0] \in [-1, 1]$ or auto psd: $S_{yy}(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma_{yy}[\tau] \cdot e^{-j\omega\tau T_s}$

$\tilde{y}[t] = y[t] - \mu_y$.

Table 4. Nonparametric excitation–response models

Means	μ_x	μ_y	
acf's and ccf	$\gamma_{xx}[\tau]$	$\gamma_{yy}[\tau]$	$\gamma_{yx}[\tau] = E\{\tilde{y}[t] \cdot \tilde{x}[t-\tau]\}$
	$\rho_{xx}[\tau]$	$\rho_{yy}[\tau]$	$\rho_{yx}[\tau] = \gamma_{yx}[\tau]/\sqrt{\gamma_{xx}[0] \cdot \gamma_{yy}[0]} \in [-1, 1]$
auto psd and csd	$S_{xx}(\omega)$	$S_{yy}(\omega)$	$S_{yx}(j\omega) = \sum_{\tau=-\infty}^{\infty} \gamma_{yx}[\tau] \cdot e^{-j\omega\tau T_s}$

$\tilde{x}[t] = x[t] - \mu_x$, $\tilde{y}[t] = y[t] - \mu_y$.

Table 5. Parametric response-only models

ARMA model	$y[t] + \sum_{i=1}^{na} a_i \cdot y[t-i] = w[t] + \sum_{i=1}^{nc} c_i \cdot w[t-i]$	$A(\mathcal{B}) = 1 + \sum_{i=1}^{na} a_i \mathcal{B}^i$
	$\Longleftrightarrow A(\mathcal{B}) \cdot y[t] = C(\mathcal{B}) \cdot w[t]$	$C(\mathcal{B}) = 1 + \sum_{i=1}^{nc} c_i \mathcal{B}^i$
	$w[t] \sim$ i.i.d. $\mathcal{N}(0, \sigma_w^2)$	na, nc : AR, MA orders
State space (SS) model	$\boldsymbol{\psi}[t+1] = \boldsymbol{A} \cdot \boldsymbol{\psi}[t] + \boldsymbol{v}[t]$	$\boldsymbol{\psi}[t]$: state vector
	$y[t] = \boldsymbol{C} \cdot \boldsymbol{\psi}[t]$	$\boldsymbol{v}[t] \sim$ i.i.d. $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_v)$ $\boldsymbol{A}$: system matrix $\boldsymbol{C}$: output matrix

that the excitation is white (uncorrelated, that is, $\gamma_{xx}[\tau] = 0$ for $\tau \neq 0$, in which case the signal is often designated as $w[t]$) and $n[t] \equiv 0$. The signals are assumed zero mean; to comply with this requirement, the signal $\tilde{y}[t] = y[t] - \mu_y$ is typically used (also see the related comment on signal preprocessing in Section 2). This leads to the autoregressive moving average (ARMA) model [25, pp. 52–53], which may be alternatively set into state-space form [25, pp. 163–164]; [30, p. 157] consisting of a set of first-order state equations plus an output equation (Table 5). Note that in the ARMA expression, a_i and c_i are the autoregressive (AR) and moving average (MA) parameters, respectively, while $w[t]$ coincides with the model-based *one-step-ahead prediction error*, and is also referred to as the *model residual* or *innovations* [25, p. 134], [26, p. 70]. i.i.d. stands for identically independently distributed, and $\mathcal{N}(\cdot, \cdot)$ designates normal distribution with the indicated mean and variance/covariance.

In the excitation–response case, various parametrizations of equation (1) lead to a number of models, including the autoregressive with exogenous excitation (ARX) and autoregressive moving average with exogenous excitation (ARMAX) models [22], [26, pp. 81–83]; [30, p. 149–151], the Box–Jenkins model [22], [26, p. 87]; [30, p. 152], the output error (OE) model [26, pp. 85–86]; [22] which represents the structural dynamics by a rational transfer function but leaves the noise unmodeled, and the state-space model consisting of a set of first-order state equations plus an output equation (see [26, pp. 97–101] for additional versions). These are shown in Table 6. Like before, the signals are assumed zero mean; to comply with this, the signals $\tilde{x}[t] = x[t] - \mu_x$ and $\tilde{y}[t] = y[t] - \mu_y$ are typically used (also see the related comment on signal preprocessing in Section 2). $w[t]$ coincides with the model-based *one-step-ahead prediction error* and is, in this case as well, referred to as *residual* or *innovations*.

4 IDENTIFICATION OF TIME SERIES MODELS

The term identification refers to the estimation of statistical time series models based on (properly preprocessed) excitation $x[t]$ and/or response $y[t]$ (for $t = 1, 2, \ldots, N$) random vibration data records, collectively designated as $Z = (X, Y)$. This is done via *estimators*, which operate on the data records to provide *estimates* of the quantities of interest. Let $\widehat{Q}$ designate the estimator of a quantity Q. The estimator is a function of the data Z, $\widehat{Q} = g(Z)$. Considering each data point as the observed value of an underlying random variable, the estimator is a function of the data random variables, and therefore a random variable by itself.

As such, the estimator $\widehat{Q}$ is characterized by a probability density function $f_{\widehat{Q}}$, and by (at least) its first- and second-order moments $\mu_{\widehat{Q}} = E[\widehat{Q}]$, $\text{var}[\widehat{Q}] = E[(\widehat{Q} - \mu_{\widehat{Q}})^2]$. If the distribution is Gaussian, one writes $\widehat{Q} \sim \mathcal{N}(\mu_{\widehat{Q}}, \text{var}[\widehat{Q}])$. An estimator is called *unbiased* if its mean coincides with the true value of the quantity under estimation, that is, $\mu_{\widehat{Q}} = E[\widehat{Q}] = Q$. Oftentimes, the determination of the probability density function for an estimator is done asymptotically, as the data record length (in samples) tends to infinity ($N \to \infty$).

Table 6. Parametric excitation–response models

ARX model	$A(\mathcal{B}) \cdot y[t] = B(\mathcal{B}) \cdot x[t] + w[t]$	na, nb: AR, X orders $A(\mathcal{B}) = 1 + \sum_{i=1}^{na} a_i \mathcal{B}^i$ $B(\mathcal{B}) = b_0 + \sum_{i=1}^{nb} b_i \mathcal{B}^i$
ARMAX model	$A(\mathcal{B}) \cdot y[t] = B(\mathcal{B}) \cdot x[t] + C(\mathcal{B}) \cdot w[t]$	na, nb, nc: AR, X, MA orders $C(\mathcal{B}) = 1 + \sum_{i=1}^{nc} c_i \mathcal{B}^i$
Box–Jenkins model	$y[t] = \frac{B(\mathcal{B})}{A(\mathcal{B})} \cdot x[t] + \frac{C(\mathcal{B})}{D(\mathcal{B})} \cdot w[t]$	$D(\mathcal{B}) = 1 + \sum_{i=1}^{nd} d_i \mathcal{B}^i$
Output error (OE) model	$y[t] = \frac{B(\mathcal{B})}{A(\mathcal{B})} \cdot x[t] + n[t]$	$n[t]$: autocorrelated zero mean
State space (SS) model	$\boldsymbol{\psi}[t+1] = \boldsymbol{A} \cdot \boldsymbol{\psi}[t] + \boldsymbol{B} \cdot x[t] + \boldsymbol{v}[t]$ $y[t] = \boldsymbol{C} \cdot \boldsymbol{\psi}[t] + \boldsymbol{D} \cdot x[t]$	$\boldsymbol{\psi}[t]$: state vector $\boldsymbol{v}[t] \sim$ i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_v)$ $\boldsymbol{A}$: system matrix, $\boldsymbol{B}$: input matrix $\boldsymbol{C}$: output matrix, $\boldsymbol{D}$: dir. transm. matrix

$w[t] \sim$ i.i.d. $\mathcal{N}(0, \sigma_w^2)$.

Some of the estimators for nonparametric time series models, along with certain of their properties, are summarized in Table 7.

The estimation of parametric time series models involves parameter and structure estimation. The parameter vector $\boldsymbol{\theta}$ of a selected model (for example, $\boldsymbol{\theta} = [\text{coef}A, \text{coef}B, \text{coef}C]$ for an ARMAX model—coef designating the coefficients of the indicated polynomial) is estimated on the basis of an appropriate estimation method. Typical methods include prediction error (PE) and least square (LS), maximum likelihood (ML), correlation, and subspace methods [22], [26, Chapter 7]; [30, Chapters 7–8].

Model structure estimation (referring, for instance, to the determination of the AR, X, and MA orders for an ARMAX model) is more complicated, and is typically achieved by fitting increasingly higher order models to the data until no further improvement—in a proper sense—is observed. Improvement may be judged via order-selection criteria, such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC), but additional practical criteria, such as modal frequency stabilization diagrams, are useful [22], [26, pp. 505–507], [30, pp. 442–443].

5 NONPARAMETRIC TIME SERIES METHODS

Nonparametric time series methods are those in which the characteristic quantity Q is constructed on the basis of nonparametric time series models. Three such methods that operate in the frequency domain are presented. For alternative methods that employ a novelty measure, see [32–34].

5.1 Power spectral density (psd)-based method

This method's characteristic quantity is the auto psd of the vibration response signal, $Q = S_{yy}(\omega) = S(\omega)$; hence it is appropriate for the response-only case. The main idea is on the comparison of the current structure's response psd $S_{\text{u}}(\omega)$ to that of the healthy structure's, $S_{\text{o}}(\omega)$—or, in fact, to that corresponding to any other structural condition. It should be noted that response signal scaling is important in order to properly account for potentially different excitation levels.

Table 7. Estimation of nonparametric time series characteristics

Quantity	Estimator	Properties	Comments/references
Mean	$\widehat{\mu}_y = \frac{1}{N}\sum_t y[t]$	$\widehat{\mu}_y \sim \mathcal{N}\left(\mu_y, \frac{\sigma_y^2}{N^2}\sum_{\tau=1}^{N}\sum_{s=1}^{N}\rho_{yy}[s-\tau]\right)$	[23, p. 319]; [31, p. 213]
Normalized acf	$\widehat{\rho}_{yy}[\tau] = \widehat{\gamma}_{yy}[\tau]/\widehat{\gamma}_{yy}[0]$	$\widehat{\rho}_{yy}[\tau] \sim \mathcal{N}\big(\rho_{yy}[\tau],\ \mathrm{var}\big[\widehat{\rho}_{yy}[\tau]\big]\big)$	For $N \to \infty$
	$\widehat{y}[t] = y[t] - \widehat{\mu}_y$	$\mathrm{var}\big[\widehat{\rho}_{yy}[\tau]\big] \approx \frac{1}{N}\left[1 + 2\sum_{i=1}^{q}\rho_{yy}^2[i]\right]$	$\rho_{yy}[\tau] = 0$ for $\tau > q$
	$\widehat{\gamma}_{yy}[\tau] = \frac{1}{N}\sum_t \widehat{y}[t]\widehat{y}[t-\tau]$		[25, pp. 30–34]; [31, pp. 214–215]
Normalized ccf	$\widehat{\rho}_{yx}[\tau] = \widehat{\gamma}_{yx}[\tau]/\sqrt{\widehat{\gamma}_{xx}[0]\widehat{\gamma}_{yy}[0]}$	$E\{\widehat{\rho}_{yx}[\tau]\} = \rho_{yx}[\tau]$	For $N \to \infty$ [31, p. 398]
	$\widehat{\gamma}_{yx}[\tau] = \frac{1}{N}\sum_t \widehat{y}[t]\widehat{x}[t-\tau]$	$\mathrm{var}\big[\widehat{\rho}_{yx}[\tau]\big]$: see [25, pp. 411–415]	[25, pp. 411–415]; [23, p. 693]
Power spectral density (psd)	$\widehat{S}_{yy}(\omega) = \frac{1}{K}\sum_{i=1}^{K}\widehat{Y}_L^i(j\omega)\widehat{Y}_L^i(-j\omega)$	$\frac{2K\widehat{S}_{yy}(\omega)}{S_{yy}(\omega)} \sim \chi^2(2K)$	Welch method (no overlap) [24, p. 76]
	$\widehat{Y}_L^i(j\omega) = \frac{1}{\sqrt{L}}\sum_{t=1}^{L} a[t]\widehat{y}^i[t]\ \mathrm{e}^{-j\omega t T_s}$		K: number of data segments
	$\widehat{y}^i[t] = y^i[t] - \widehat{\mu}_y$		$a[t]$: time window
	(ith segment of length L)		[31, pp. 352–353]; [29, p. 309]
Cross spectral density (csd)	$\widehat{S}_{yx}(j\omega) = \frac{1}{K}\sum_{i=1}^{K}\widehat{Y}_L^i(j\omega)\widehat{X}_L^i(-j\omega)$	$E\{\lvert\widehat{S}_{yx}(j\omega)\rvert\} \approx \lvert S_{yx}(j\omega)\rvert$	Welch method (no overlap)
		$\mathrm{var}[\lvert\widehat{S}_{yx}(j\omega)\rvert] \approx \frac{\lvert S_{yx}(j\omega)\rvert^2}{\gamma^2(\omega)K}$	For $N \to \infty$, $a[t]=1$
	$\widehat{x}^i[t] = x^i[t] - \widehat{\mu}_x$	$E\{\arg\widehat{S}_{yx}(j\omega)\} \approx \arg S_{yx}(j\omega)$	$\gamma^2(\omega) \to 1$ or $K \to \infty$
	(ith segment of length L)	$\mathrm{var}[\arg\widehat{S}_{yx}(j\omega)] \approx \frac{1-\gamma^2(\omega)}{\gamma^2(\omega)2K}$	[29, pp. 318, 320, 325–326]
Frequency response function (frf)	$\widehat{H}(j\omega) = \widehat{S}_{yx}(j\omega)/\widehat{S}_{xx}(\omega)$	$E\{\lvert\widehat{H}(j\omega)\rvert\} \approx \lvert H(j\omega)\rvert$	Welch method (no overlap)
		$\mathrm{var}[\lvert\widehat{H}(j\omega)\rvert] \approx \frac{1-\gamma^2(\omega)}{\gamma^2(\omega)2K}\lvert H(j\omega)\rvert^2$	For $N \to \infty$, $a[t]=1$
			$\gamma^2(\omega) \to 1$ or $K \to \infty$ [29, pp. 329, 338]
(Squared) Coherence	$\widehat{\gamma}^2(\omega) = \lvert\widehat{S}_{yx}(j\omega)\rvert^2/\widehat{S}_{xx}(\omega)\widehat{S}_{yy}(\omega)$	$E\{\widehat{\gamma}^2(\omega)\} \approx \gamma^2(\omega) + \frac{1}{K}[1-\gamma^2(\omega)]^2$	Welch method (no overlap)
		$\mathrm{var}[\widehat{\gamma}^2(\omega)] \approx \frac{2\gamma^2(\omega)}{K}[1-\gamma^2(\omega)]^2$	For $N \to \infty$, $a[t] = 1$, $K \to \infty$ [29, pp. 333–335]

[(a)] $\widehat{x}[t]$, $\widehat{y}[t]$: sample versions of $\tilde{x}[t]$, $\tilde{y}[t]$.
[(b)] arg designates argument (angle) of the indicated complex quantity.
[(c)] $\omega \in [0, 2\pi/T_s]$ stands for frequency in radian per second. j stands for the imaginary unit. K stands for the number of segments used in Welch spectral estimation.
[(d)] The frequency-domain estimator distributions may be approximated as Gaussian for small relative errors (that is, $\gamma^2(\omega) \to 1$ or $K \to \infty$) [29, pp. 274–275].

The following hypothesis testing problem is then formulated for fault detection:

$$H_0: S_u(\omega) = S_o(\omega)$$
(null hypothesis—healthy structure)
$$H_1: S_u(\omega) \neq S_o(\omega)$$
(alternative hypothesis—faulty structure) (8)

As the true psd's, $S_u(\omega)$, $S_o(\omega)$, are unknown, their estimates $\widehat{S}_u(\omega)$, $\widehat{S}_o(\omega)$ obtained via the Welch method (with K nonoverlapping segments; refer to Table 7) are used. Then the quantity F, given in the subsequent text, follows F distribution with $(2K, 2K)$ degrees of freedom (as each of the numerator and denominator follow normalized χ^2 distribution with $2K$ degrees of freedom; refer to Table 7 and the Appendix):

$$F = \frac{\widehat{S}_o(\omega)/S_o(\omega)}{\widehat{S}_u(\omega)/S_u(\omega)} \sim F(2K, 2K) \tag{9}$$

Under the null (H_0) hypothesis the true psd's coincide, $S_u(\omega) = S_o(\omega)$, and therefore F equals $\widehat{S}_o(\omega)/\widehat{S}_u(\omega)$. This should be then in the range $[f_{\alpha/2}, f_{1-\alpha/2}]$ with probability $1-\alpha$, and decision making is as follows at the α risk level (type I error probability of α):

$$f_{\frac{\alpha}{2}}(2K, 2K) \leq F \leq f_{1-\frac{\alpha}{2}}(2K, 2K) \quad (\forall\omega)$$
$$\Longrightarrow H_0 \text{ is accepted (healthy structure)}$$
$$\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \tag{10}$$

with $f_{(\alpha/2)}$, $f_{1-(\alpha/2)}$ designating the F distribution's $\alpha/2$ and $1-(\alpha/2)$ critical points (f_α is defined such that $\text{Prob}(F \leq f_\alpha) = \alpha$; Figure 3).

Note that fault identification may be similarly achieved, while fault estimation may be achieved by possibly associating specific quantitative changes in the psd with specific fault magnitudes.

Bibliographical remarks. See [35] for the application of the method to fault detection in a railway vehicle suspension. In the nonstationary or nonlinear cases, time–frequency, polyspectra, or wavelet-based models may be used [36–42].

5.2 Frequency response function (frf)—based method

This method is similar to the psd method, the difference being in that the excitation–response frf magnitude is used as the characteristic quantity, $Q = |H(j\omega)|$. It is thus appropriate for the excitation–response case, although it may also be used in case the excitation is unavailable but more than one responses are available [43]. The main idea is on the comparison of the frf magnitude $|H_u(j\omega)|$ of the current structure to that of the healthy structure $|H_o(j\omega)|$—or, in fact, to that corresponding to any other structural condition.

The following hypothesis testing problem is then formulated for fault detection:

$$H_0: \delta|H(j\omega)| = |H_o(j\omega)| - |H_u(j\omega)| = 0$$
(null hypothesis–healthy structure)
$$H_1: \delta|H(j\omega)| = |H_o(j\omega)| - |H_u(j\omega)| \neq 0$$
(alternative hypothesis–faulty structure) (11)

As the true frf's, $H_u(j\omega)$ and $H_o(j\omega)$, are unknown, their respective estimates $\widehat{H}_u(j\omega)$ and $\widehat{H}_o(j\omega)$ obtained as indicated in Table 7 are used.

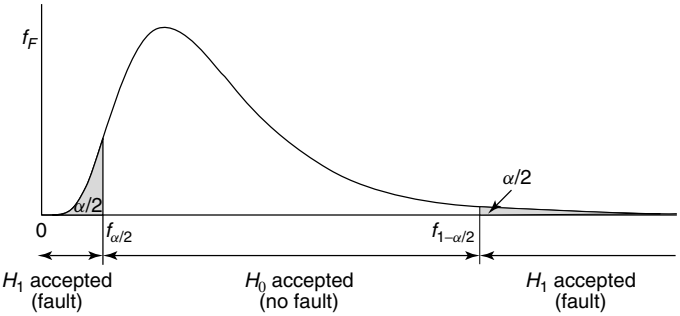


Figure 3. Statistical hypothesis testing based on an F distributed statistic (two-tailed test).

The frf magnitude estimator may, asymptotically ($N \to \infty$), be considered as approximately following Gaussian distribution (refer to Table 7). Owing to the mutual independence of the Z_u and Z_o data records, the two frf magnitude estimators are mutually independent, and thus their difference is Gaussian with mean equal to the true magnitude difference, $|H_o(j\omega)| - |H_u(j\omega)|$, and variance equal to the sum of the two variances.

Under the null (H_0) hypothesis, the true frf magnitudes coincide ($|H_u(j\omega)| = |H_o(j\omega)|$), hence

$$\text{Under } H_0\text{: } \delta|\widehat{H}(j\omega)| = |\widehat{H}_o(j\omega)| - |\widehat{H}_u(j\omega)| \sim \mathcal{N}(0, 2\sigma_o^2(\omega)) \qquad (12)$$

The variance $\sigma_o^2(\omega) = \text{var}[|\widehat{H}_o(j\omega)|]$ is generally unknown, but may be estimated in the baseline phase via the expressions of Table 7. Treating this estimator as a fixed quantity, that is a quantity characterized by negligible variability (which is reasonable for estimation based on long data records), the equality of the two frf magnitudes may be examined at the α (type I) risk level through the statistical test (Figure 4):

$$Z = |\delta|\widehat{H}(j\omega)||/\sqrt{2\widehat{\sigma}_o^2(\omega)} \le Z_{1-\frac{\alpha}{2}} \quad (\forall\omega)$$
$$\Longrightarrow H_0 \text{ is accepted (healthy structure)}$$
$$\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \qquad (13)$$

with $Z_{1-(\alpha/2)}$ designating the standard normal distribution's $1 - (\alpha/2)$ critical point. Alternatively, treating $\widehat{\sigma}_o^2(\omega)$ as a random variable leads to a *t*-distribution-based test—this approaches the above normal test for long data records (see Appendix).

Note that fault identification may be similarly achieved, while fault estimation may be achieved by possibly associating specific quantitative changes in the frf magnitude with specific fault magnitudes.

5.3 Coherence measure–based method

This method is based on the heuristic premise that, under constant experimental and environmental conditions, the integral of the coherence over frequency decreases with fault occurrence. This can be justified by the nonlinear effects introduced, or strengthened, with fault occurrence. Like the previous method, this method pertains to the excitation–response case.

The *coherence measure* defined over the discrete frequencies ω_i ($i = 1, 2, \ldots, n$; frequency resolution $\delta\omega$):

$$\Gamma = \delta\omega \cdot \sum_{i=1}^{n} \gamma^2[\omega_i] \qquad (14)$$

is used in approximating the coherence integral, and constitutes the method's characteristic quantity Q.

Fault detection is then based on the confirmation of a statistically significant reduction in the coherence measure Γ_u of the current structure (compared to Γ_o of its healthy counterpart) through the statistical hypothesis testing problem:

$$H_0\text{: } \delta\Gamma = \Gamma_o - \Gamma_u = 0$$
(null hypothesis–healthy structure)
$$H_1\text{: } \delta\Gamma = \Gamma_o - \Gamma_u > 0$$
(alternative hypothesis–faulty structure) (15)

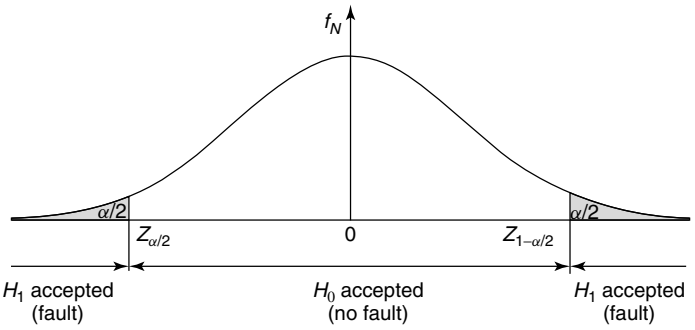


Figure 4. Statistical hypothesis testing based on a Gaussian distributed statistic (two-tailed test).

As the true Γ_u and Γ_o are unknown, their respective estimates are employed.

Owing to the properties of the Welch-based coherence estimator of Table 7 and the central limit theorem [44, p. 273] (also see Appendix), the corresponding coherence measure estimator approximately follows, for a large number of discrete frequencies ($n \to \infty$), Gaussian distribution, that is,

$$\widehat{\Gamma} \sim \mathcal{N}(\Gamma, \sigma_\Gamma^2)$$

with

$$\sigma_\Gamma^2 = (\delta\omega)^2 \sum_{i=1}^{n} \mathrm{var}[\widehat{\gamma}^2[\omega_i]] \tag{16}$$

with Γ designating the coherence measure's true value. Notice that in obtaining this expression, the bias terms present in the coherence estimator (refer to Table 7) are neglected, which is justified for either large number of segments K or for true coherence close to unity. Furthermore, the variance of $\widehat{\Gamma}$ is equal to the sum of the individual coherence estimator variances due to the fact that the coherence estimators at different frequencies are mutually independent random variables [45, p. 204].

Furthermore, as $\widehat{\Gamma}_u$ and $\widehat{\Gamma}_o$ are mutually independent and Gaussian, $\delta\widehat{\Gamma}$ will be Gaussian as follows:

$$\delta\widehat{\Gamma} = \widehat{\Gamma}_o - \widehat{\Gamma}_u \sim \mathcal{N}(\delta\Gamma, \delta\sigma_\Gamma^2)$$

with

$$\delta\Gamma = \Gamma_o - \Gamma_u, \quad \delta\sigma_\Gamma^2 = (\sigma_\Gamma^2)_o + (\sigma_\Gamma^2)_u \tag{17}$$

Under the null (H_0) hypothesis $\delta\Gamma = \Gamma_o - \Gamma_u \leq 0$ and $\delta\sigma_\Gamma^2 = 2(\sigma_\Gamma^2)_o$, thus

$$\text{Under } H_0 : \delta\widehat{\Gamma} = \widehat{\Gamma}_o - \widehat{\Gamma}_u \sim \mathcal{N}(\delta\Gamma, 2(\sigma_\Gamma^2)_o) \quad (\delta\Gamma \leq 0) \tag{18}$$

The variance $(\sigma_\Gamma^2)_o$ is unknown, but may be estimated in the baseline phase. Treating this estimator as a fixed quantity, the equality (null) hypothesis for the two coherence measures may be examined at the α (type I) risk level through the statistical test (Figure 5):

$$\begin{aligned} Z = \delta\widehat{\Gamma} / \sqrt{2(\widehat{\sigma}_\Gamma^2)_o} \leq Z_{1-\alpha} \\ \Longrightarrow H_0 \text{ is accepted (healthy structure)} \\ \text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \end{aligned} \tag{19}$$

with $Z_{1-\alpha}$ designating the standard normal distribution's $1-\alpha$ critical point.

Note that the method is not appropriate for fault identification, as different types of faults may cause similar or identical reduction in the coherence measure. Assuming that only one type of fault is possible, fault estimation may be possible by associating specific quantitative reductions in the coherence measure with specific fault magnitudes.

Bibliographical remark. The method was introduced and applied for skin damage detection and restoration quality assessment in aircraft panels [46].

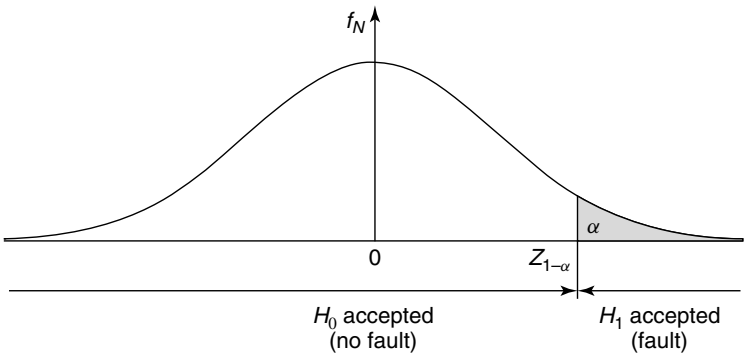


Figure 5. Statistical hypothesis testing based on a Gaussian distributed statistic (one-tailed test).

6 PARAMETRIC TIME SERIES METHODS

Parametric time series methods are those in which the characteristic quantity Q is constructed on the basis of parametric time series models. They are applicable to both the response-only and excitation–response cases, as each situation may be dealt with through the use of proper models [14], [27, subsection 6.2], [47, subsection 4.2]. Parametric methods may operate on either time or frequency domains, with the former being more frequently used and being in the focus of the present article. Parametric methods may be further classified into a number of categories, including model parameter–based methods, residual-based methods, and functional model (FM)–based methods.

6.1 Model parameter–based methods

These perform fault detection and identification by using a characteristic quantity $Q = f(\boldsymbol{\theta})$, which is a function of the parameter vector $\boldsymbol{\theta}$ of a parametric time series model ($Q = \boldsymbol{\theta}$ in the typical case).

Let $\widehat{\boldsymbol{\theta}}$ designate a proper estimator of the parameter vector $\boldsymbol{\theta}$ [22], [26, pp. 212–213]; [30, pp. 198–199]. For sufficiently long signals (N large), the estimator is (under mild assumptions) Gaussian distributed with mean equal to its true value $\boldsymbol{\theta}$ and a certain covariance $\boldsymbol{P}_{\boldsymbol{\theta}}$ [30, pp. 205–207]:

$$\widehat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{P}_{\boldsymbol{\theta}}) \tag{20}$$

Fault detection is based on testing for statistically significant changes in the parameter vector $\boldsymbol{\theta}$ between the nominal and current structures through the hypothesis testing problem:

$$\begin{aligned} &H_0 : \delta\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathrm{o}} - \boldsymbol{\theta}_{\mathrm{u}} = \mathbf{0} \\ &\quad \text{(null hypothesis—healthy structure)} \\ &H_1 : \delta\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathrm{o}} - \boldsymbol{\theta}_{\mathrm{u}} \neq \mathbf{0} \\ &\quad \text{(alternative hypothesis—faulty structure)} \end{aligned} \tag{21}$$

Toward this end observe that, owing to the mutual independence of the Z_{u} and Z_{o} data records, the difference between the two parameter vector estimators also follows Gaussian distribution:

$$\delta\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_{\mathrm{u}} \sim \mathcal{N}(\delta\boldsymbol{\theta}, \delta\boldsymbol{P})$$

with

$$\delta\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathrm{o}} - \boldsymbol{\theta}_{\mathrm{u}}, \quad \delta\boldsymbol{P} = \boldsymbol{P}_{\mathrm{o}} + \boldsymbol{P}_{\mathrm{u}} \tag{22}$$

where $\boldsymbol{P}_{\mathrm{o}}$, $\boldsymbol{P}_{\mathrm{u}}$ designate the corresponding covariance matrices. Under the null (H_0) hypothesis $\delta\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_{\mathrm{u}} \sim \mathcal{N}(\mathbf{0}, 2\boldsymbol{P}_{\mathrm{o}})$ and the quantity $\chi^2_{\boldsymbol{\theta}}$, below, follows χ^2 distribution with d (parameter vector dimensionality) degrees of freedom (as it may be shown to be the sum of squares of independent standardized Gaussian variables [26, p. 558]–also see Appendix):

$$\text{Under } H_0 : \mathcal{X}^2_{\boldsymbol{\theta}} = \delta\widehat{\boldsymbol{\theta}}^{\mathrm{T}} \cdot \delta\boldsymbol{P}^{-1} \cdot \delta\widehat{\boldsymbol{\theta}} \sim \chi^2(d)$$

with

$$\delta\boldsymbol{P} = 2\boldsymbol{P}_{\mathrm{o}} \tag{23}$$

As the covariance matrix $\boldsymbol{P}_{\mathrm{o}}$ corresponding to the healthy structure is unavailable, its estimated version $\widehat{\boldsymbol{P}}_{\mathrm{o}}$ is used. Treating it as a quantity characterized by negligible variability (which is reasonable for large N) leads to the following test constructed at the α (type I) risk level (Figure 6; see [26, p. 559] for an alternative approach based on the F distribution):

$$\begin{aligned} &\mathcal{X}^2_{\boldsymbol{\theta}} \leq \chi^2_{1-\alpha}(d) \\ &\quad \Longrightarrow H_0 \text{ is accepted (healthy structure)} \\ &\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \end{aligned} \tag{24}$$

with $\chi^2_{1-\alpha}(d)$ designating the χ^2 distribution's $1-\alpha$ critical point.

Fault identification may be based on the multiple hypotheses testing problem of Table 2, comparing the current parameter vector $\widehat{\boldsymbol{\theta}}_{\mathrm{u}}$ to those corresponding to different fault types, $\widehat{\boldsymbol{\theta}}_{\mathrm{A}}, \widehat{\boldsymbol{\theta}}_{\mathrm{B}}, \ldots$. Nevertheless, this is expected to work only for faults of specific magnitudes and cannot generally account for the continuum of fault magnitudes possible within each fault type. A geometric method that aims at circumventing this difficulty and also used for fault estimation is presented in [43, 48, 49].

Bibliographical remarks. The principles of this method have been used in a number of studies. Sohn

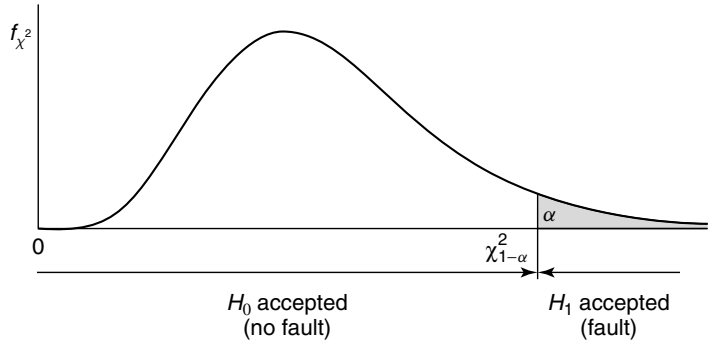


Figure 6. Statistical hypothesis testing based on a χ^2 distributed statistic (one-tailed test).

and Farrar [16] use the parameters of an AR model and statistical process control charts for fault detection in a concrete bridge column, as the column is progressively damaged. Sakellariou and Fassois [43] use the parameter vector of an OE model for fault detection in a six-story building model under earthquake excitation. Adams and Farrar [50] use ARX models in the frequency domain for fault detection and estimation in a simulated structural system and in a three-story building model. Wei *et al.* [51] use nonlinear autoregressive moving average with exogenous excitation (NARMAX) models for fault detection and identification in carbon fiber-reinforced epoxy plates based on a deterministic index, which mirrors changes incurred in the model parameters. Nair *et al.* [52] employ ARMA models, a damage-sensitive feature that depends on certain AR parameters, and statistical hypothesis testing to detect and localize (identify) faults in an ASCE benchmark structure consisting of a four-story steel-braced frame. In a follow-up study, Nair and Kiremidjian [53] employ Gaussian mixture modeling of the feature vector. Carden and Brownjohn [19] use ARMA model parameters for fault detection and localization (identification) in the IASC-ASCE benchmark four-story frame structure, in the Z24 bridge, and in the Malaysia–Singapore Second Link bridge. It should be noted that model parameter–based methods may be also used with alternative models, such as *modal models*. In this case, fault detection and identification are based on the detection of changes incurred in the structure's modal parameters [1, 2, 6, 39, 46, 54, 55]—also see Subsection 6.2.4. This topic has attracted considerable interest, although not necessarily in a statistical framework.

6.2 Model residual–based methods

These methods base fault detection, identification, and estimation on characteristic quantities that are functions of residual sequences obtained by driving the current signal(s) Z_u through predetermined—in the baseline phase—models $\mathcal{M}_o, \mathcal{M}_A, \mathcal{M}_B, \ldots$, each one corresponding to a particular state of the structure. The general idea is that the residual sequence obtained by a model that truly reflects the actual (current) state of the structure will possess some distinct properties, and will be thus possible to distinguish. An advantage of the methods is that no model identification is performed in the inspection phase. The methods have a relatively long history—mainly within the general context of engineering systems; for instance, see [14, 27, 47, 56–60].

Let $\mathcal{M}_V$ designate the model representing the structure in its V state ($V = o$ or $V = A, B, \ldots$ under specific fault magnitudes). Also, let the residual series obtained by driving the current signals Z_u through each one of the above-mentioned models be designated as $e_{ou}[t]$, $e_{Au}[t]$, $e_{Bu}[t]$, ... and be characterized by respective variances $\sigma^2_{ou}, \sigma^2_{Au}, \sigma^2_{Bu}, \ldots$—notice that the first subscript designates the model employed and the second the structural state corresponding to the current excitation and/or response signal(s) used.

Fault detection, identification, and estimation may be then based on the fact that under the H_V hypothesis (that is, the structure being in its V state, for $V = o$ or $V = A, B, \ldots$ under specific fault magnitudes), the residual series generated by driving the current signal(s) Z_u through the model $\mathcal{M}_V$ possesses the

property:

$$\text{Under } H_{\mathrm{V}}: e_{\mathrm{Vu}}[t] \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2_{\mathrm{Vu}})$$

with

$$\sigma^2_{\mathrm{Vu}} < \sigma^2_{\mathrm{Wu}} \text{ for any state W} \neq \mathrm{V} \tag{25}$$

It is tacitly assumed that under H_{V} $\sigma^2_{\mathrm{Vu}} = \sigma^2_{\mathrm{VV}}$, implying that the excitation and environmental conditions are the same in the baseline and inspection phases. Descriptions of four model residual–based methods follow.

6.2.1 Residual variance-based method

In this method, the characteristic quantity is the residual variance. Fault detection is based on the fact that the residual series $e_{\mathrm{ou}}[t]$, obtained by driving the current signal(s) Z_{u} through the model $\mathcal{M}_{\mathrm{o}}$ corresponding to the nominal (healthy) structure, should be characterized by variance σ^2_{ou}, which becomes minimal (specifically equal to σ^2_{oo}) if and only if the current structure is healthy ($\mathcal{S}_{\mathrm{u}} = \mathcal{S}_{\mathrm{o}}$).

The following hypothesis testing problem is then set up:

$$\begin{aligned} &H_0: \sigma^2_{\mathrm{oo}} = \sigma^2_{\mathrm{ou}} \text{ (null hypothesis—healthy structure)} \\ &H_1: \sigma^2_{\mathrm{oo}} < \sigma^2_{\mathrm{ou}} \text{ (alternative hypothesis—faulty structure)} \end{aligned} \tag{26}$$

Under the null (H_0) hypothesis, the residuals $e_{\mathrm{ou}}[t]$ are (just like the residuals $e_{\mathrm{oo}}[t]$) i.i.d. zero mean Gaussian with variance σ^2_{oo}. Hence the quantities $N_{\mathrm{u}}\widehat{\sigma}^2_{\mathrm{ou}}/\sigma^2_{\mathrm{oo}}$ and $(N_{\mathrm{o}} - d)\widehat{\sigma}^2_{\mathrm{oo}}/\sigma^2_{\mathrm{oo}}$ follow (central) χ^2 distributions with N_{u} and $N_{\mathrm{o}} - d$ degrees of freedom, respectively (sum of squares of independent standardized Gaussian random variables—see Appendix). Note that N_{o} and N_{u} designate the number of samples used in estimating the residual variance in the healthy and current cases, respectively (typically $N_{\mathrm{o}} = N_{\mathrm{u}} = N$), and d designates the dimensionality of the model parameter vector. Also note that N_{u} and N_{o} should be adjusted to $N_{\mathrm{u}} - 1$ and $N_{\mathrm{o}} - 1$, respectively, in case each estimated mean is subtracted from each residual series. Consequently, the following statistic follows F distribution with $(N_{\mathrm{u}}, N_{\mathrm{o}} - d)$ degrees of freedom (ratio of two independent and normalized χ^2 random variables—see Appendix):

$$\text{Under } H_0: F = \frac{\dfrac{N_{\mathrm{u}}\widehat{\sigma}^2_{\mathrm{ou}}}{\sigma^2_{\mathrm{oo}} N_{\mathrm{u}}}}{\dfrac{(N_{\mathrm{o}} - d)\widehat{\sigma}^2_{\mathrm{oo}}}{\sigma^2_{\mathrm{oo}}(N_{\mathrm{o}} - d)}} = \frac{\widehat{\sigma}^2_{\mathrm{ou}}}{\widehat{\sigma}^2_{\mathrm{oo}}} \sim F(N_{\mathrm{u}}, N_{\mathrm{o}} - d) \tag{27}$$

The following test is then constructed at the α (type I) risk level (Figure 7):

$$\begin{aligned} &F \leq f_{1-\alpha}(N_{\mathrm{u}}, N_{\mathrm{o}} - d) \\ &\Longrightarrow H_0 \text{ is accepted (healthy structure)} \\ &\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \end{aligned} \tag{28}$$

with $f_{1-\alpha}(N_{\mathrm{u}}, N_{\mathrm{o}} - d)$ designating the corresponding F distribution's $1 - \alpha$ critical point.

Fault identification may be similarly achieved via pairwise tests. An alternative possibility may be based

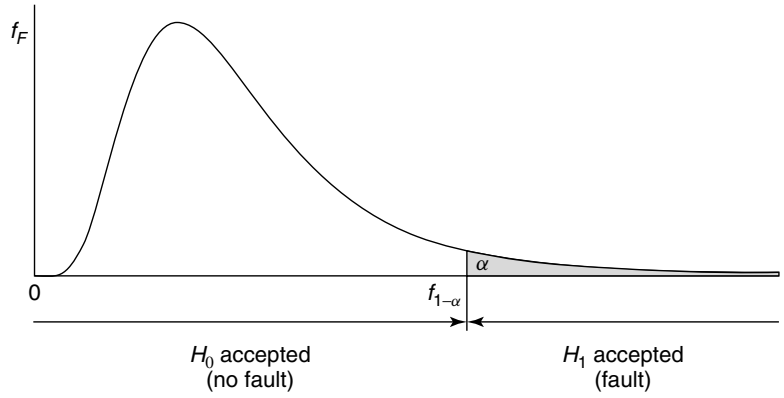


Figure 7. Statistical hypothesis testing based on an F distributed statistic (one-tailed test).

on obtaining the residual series $e_{\mathrm{Au}}[t]$, $e_{\mathrm{Bu}}[t]$,..., estimating their variances, and declaring as current fault that corresponding to minimal residual variance. Notice that by including $e_{\mathrm{ou}}[t]$, fault detection may be also treated. Fault estimation may possibly be achieved in the limited case of a single type of faults, by associating specific values of the residual variance with specific fault magnitudes.

Bibliographical remarks. Similar methods, that combine AR and ARX models, are used by Sohn and Farrar [61] and Sohn *et al.* [62] for fault detection in an eight degree-of-freedom mass-spring system and in a fast patrol boat. In a similar framework, Sohn *et al.* [63] explore the use of extreme value statistics. Fugate *et al.* [17] use an X-bar control chart for detecting changes in the mean and variance of AR model residual series for the continuous monitoring of a concrete bridge column, as the column is progressively damaged. Yan *et al.* [18] use an X-bar control chart on state-space model residuals for fault detection and identification in an aircraft skeleton and the Z24 bridge benchmark. Lu and Gao [64] use an ARX model and the standard deviation of its residuals for fault detection and localization in a two and an eight degree-of-freedom simulated mass-spring system. Zhang [65] explores data normalization procedures and a probability-based measure expressing changes in ARX residual variance for fault detection and identification in a three-span continuous girder bridge simulation model. A time-varying autoregressive with exogenous excitation (TARX) model along with statistical hypothesis testing on its residual variance is used by Poulimenos and Fassois [66] for the detection of faults in a time-varying, bridgelike, laboratory structure. Additional statistical modes (such as the mean, skewness, and kurtosis) of the residuals associated with a vector AR model are explored by Mattson and Pandit [67]. Methods employing neural network type nonlinear models and deterministic decision making based on the response error (residual) are presented in [55, 68, 69].

6.2.2 Likelihood function-based method

In this method, fault detection is based on the likelihood function under the null (H_0) hypothesis of a healthy structure [47, pp. 119–120]. The hypothesis testing problem considered is

$$\begin{aligned} &H_0 : \boldsymbol{\theta}_{\mathrm{o}} = \boldsymbol{\theta}_{\mathrm{u}} \text{ (null hypothesis—healthy structure)} \\ &H_1 : \boldsymbol{\theta}_{\mathrm{o}} \neq \boldsymbol{\theta}_{\mathrm{u}} \text{ (alternative hypothesis—faulty structure)} \end{aligned} \tag{29}$$

with $\boldsymbol{\theta}_{\mathrm{o}}$, $\boldsymbol{\theta}_{\mathrm{u}}$ designating the parameter vectors corresponding to the healthy and current structure, respectively. Assuming independence of the residual sequence, the Gaussian likelihood function for the data Y given X is [25, p. 226]:

$$\begin{aligned} L_{\mathrm{y}}(Y, \boldsymbol{\theta} / X) &= \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ -\frac{e^2[t, \boldsymbol{\theta}]}{2\sigma^2} \right\} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^N} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{N} e^2[t, \boldsymbol{\theta}] \right\} \end{aligned} \tag{30}$$

with $e[t, \boldsymbol{\theta}]$ designating the model residual (one-step-ahead prediction error) characterized by zero mean and variance σ^2.

Under the null (H_0) hypothesis, the residual series $e_{\mathrm{ou}}[t]$ generated by driving the current signal(s) through the nominal model $\mathcal{M}_{\mathrm{o}}$ is (just like $e_{\mathrm{oo}}[t]$) i.i.d. Gaussian with zero mean and variance σ^2_{oo}. Decision making may be then based on the likelihood function under H_0 evaluated for the current data, by requiring it to be larger or equal to a threshold l (which is to be selected) in order for the null (H_0) hypothesis to be accepted:

$$\begin{aligned} &L_{\mathrm{y}}(Y, \boldsymbol{\theta}_{\mathrm{o}} / X) \geq l \\ &\Longrightarrow H_0 \text{ is accepted (healthy structure)} \\ &\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \end{aligned} \tag{31}$$

The evaluation of the likelihood $L_{\mathrm{y}}(Y, \boldsymbol{\theta}_{\mathrm{o}} / X)$ requires knowledge of the true innovations variance σ^2_{oo}. If this quantity is known, or may be estimated with good accuracy in the baseline phase (which is reasonable for estimation using a large number of samples N), it may be treated as a fixed quantity (negligible variability). The above decision making rule may be then reexpressed as

$$\begin{aligned} &\chi^2_N = \sum_{t=1}^{N} \frac{e^2_{\mathrm{ou}}[t]}{\widehat{\sigma}^2_{\mathrm{oo}}} = \frac{N\widehat{\sigma}^2_{\mathrm{ou}}}{\widehat{\sigma}^2_{\mathrm{oo}}} \leq l^* \\ &\Longrightarrow H_0 \text{ is accepted (healthy structure)} \\ &\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \end{aligned} \tag{32}$$

where l^* designates the corresponding (to be selected) threshold.

Under the null (H_0) hypothesis, the statistic $\mathcal{X}_N^2$ follows χ^2 distribution with N degrees of freedom (sum of squares of mutually independent standardized Gaussian variables—see Appendix), and this leads to the following test at the α risk level (Figure 6):

$$\mathcal{X}_N^2 = \frac{N\widehat{\sigma}_{\text{ou}}^2}{\widehat{\sigma}_{\text{oo}}^2} \leq \chi_{1-\alpha}^2(N)$$
$$\Longrightarrow H_0 \text{ is accepted (healthy structure)}$$
$$\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \quad (33)$$

with $\chi_{1-\alpha}^2(N)$ designating the χ^2 distribution's $1-\alpha$ critical point.

Note that N should be adjusted to $N-1$ in case the estimated mean is subtracted from the residual series $e_{\text{ou}}[t]$. Also note that the above decision making is similar to that of the previous (residual variance based) method. The essential difference is that the variability of the estimator $\widehat{\sigma}_{\text{oo}}^2$ is accounted for in the residual variance–based method, thus leading to an F distribution of the pertinent statistic. Of course, the two tests coincide for $N_{\text{o}} \to \infty$ in equation (27), as the F distribution then approaches the χ^2 distribution [44, p. 523] (also see Appendix).

Fault identification may be achieved by computing the likelihood function for the current signal(s) for the various values of $\boldsymbol{\theta}$ ($\boldsymbol{\theta}_{\text{A}}, \boldsymbol{\theta}_{\text{B}}, \ldots$) and accepting the hypothesis that corresponds to the maximum value of the likelihood—by including $\boldsymbol{\theta}_{\text{o}}$ fault detection may be also treated. Fault estimation may be possibly achieved in the limited case of a single type of faults, by associating specific values of the likelihood with specific fault magnitudes.

6.2.3 Residual-uncorrelatedness-based method

This method is based on the fact that the residual sequence $e_{\text{ou}}[t]$ ($e_{\text{Vu}}[t]$) obtained by driving the current signal(s) Z_{u} through the model $\mathcal{M}_{\text{o}}$ ($\mathcal{M}_{\text{V}}$) will be *uncorrelated* (white) if and only if the current structure is in its nominal (healthy) S_{o} (S_{V}) state (see equation (25)).

Fault detection may be then based on the hypothesis testing problem:

$$H_0\colon \rho[\tau] = 0 \quad \tau = 1, 2, \ldots, r$$
(null hypothesis—healthy structure)
$$H_1\colon \rho[\tau] \neq 0 \quad \text{for some } \tau$$
(alternative hypothesis—faulty structure) (34)

with $\rho[\tau]$ designating the normalized autocovariance (or else correlation coefficient–refer to Table 7) of the $e_{\text{ou}}[t]$ residual sequence. The method's characteristic quantity thus is $Q = [\rho[1]\,\rho[2]\,\ldots\,\rho[r]]^{\text{T}}$, with r being a design variable.

Under the null (H_0) hypothesis, $e_{\text{ou}}[t]$ is i.i.d. (white) Gaussian with zero mean, and the statistic $\mathcal{X}_\rho^2$ below follows χ^2 distribution with r degrees of freedom ($r-1$ in case the estimated mean is subtracted from the residual series) [25, p. 314]:

$$\text{Under } H_0\colon \mathcal{X}_\rho^2 = N(N+2)\cdot\sum_{\tau=1}^{r}(N-\tau)^{-1}\cdot\widehat{\rho}^2[\tau]$$
$$\sim \chi^2(r) \quad (35)$$

with $\widehat{\rho}[\tau]$ designating the estimator of $\rho[\tau]$. Decision making is then based on the following test at the α (type I) risk level (Figure 6):

$$\mathcal{X}_\rho^2 \leq \chi_{1-\alpha}^2(r)$$
$$\Longrightarrow H_0 \text{ is accepted (healthy structure)}$$
$$\text{Else} \Longrightarrow H_1 \text{ is accepted (faulty structure)} \quad (36)$$

with $\chi_{1-\alpha}^2(r)$ designating the χ^2 distribution's $1-\alpha$ critical point.

Fault identification may be achieved by similarly examining which one of the $e_{\text{Vu}}[t]$ (for V = A, B, ...) residual series is uncorrelated. As with the previous methods, only faults of specific magnitudes (but not the continuum of fault magnitudes) may be considered. On the other hand, the method, as such, is not suitable for fault estimation.

6.2.4 Method based on residuals associated with subspace identification

This method is motivated by stochastic subspace identification (the response-only case is presently

treated; see [15] for details). It performs fault detection in the *modal space* by considering the parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\lambda} \\ \text{vec } \boldsymbol{\Phi} \end{bmatrix} \tag{37}$$

where $\boldsymbol{\lambda}$ designates the vector containing the eigenvalues of a discrete-time structural model (the eigenvalues of the system matrix $\boldsymbol{A}$ in the state-space representation), $\boldsymbol{\Phi}$ the matrix with columns being the vectors $\boldsymbol{\phi}_\lambda$ where $\boldsymbol{\phi}_\lambda = \boldsymbol{C}\boldsymbol{\varphi}_\lambda$, with $\boldsymbol{\varphi}_\lambda$ being the corresponding eigenvectors, and $\boldsymbol{C}$ the output matrix in the state-space model (Table 5). The vec of a matrix designates the vector formed by stacking the columns of the indicated matrix one underneath the other.

Fault detection is based on the fact that for a given state-space model (characterized by a given $\boldsymbol{\theta}$), the system's $(p+1)$th order (p selected sufficiently large) observability matrix:

$$\boldsymbol{\mathcal{O}}_{p+1}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Lambda} \\ \vdots \\ \boldsymbol{\Phi}\boldsymbol{\Lambda}^p \end{pmatrix} \tag{38}$$

with $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ (diagonal matrix with the system eigenvalues), and the structure's block Hankel matrix:

$$\begin{aligned} \boldsymbol{\mathcal{H}}_{p+1,q} &= \begin{pmatrix} \gamma_{yy}[0] & \gamma_{yy}[1] & \dots & \gamma_{yy}[q-1] \\ \gamma_{yy}[1] & \gamma_{yy}[2] & \dots & \gamma_{yy}[q] \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{yy}[p] & \gamma_{yy}[p+1] & \dots & \gamma_{yy}[p+q-1] \end{pmatrix} \\ &= \text{Hank}(\gamma_{yy}) \end{aligned} \tag{39}$$

with q selected such that $q \geq p+1$, and $\gamma_{yy}[\tau]$ designating the response's theoretical autocovariance, have the same left kernel space.

Let $\boldsymbol{\theta}_\text{o}$ designate the parameter vector corresponding to the healthy structure. Pick an orthonormal basis of the left kernel space of the matrix $\boldsymbol{W}_1\boldsymbol{\mathcal{O}}_{p+1}(\boldsymbol{\theta}_\text{o})$ ($\boldsymbol{W}_1$ is a selectable invertible weighting matrix), in terms of the columns of a matrix $\boldsymbol{S}$ of co-rank n (the system order) such that

$$\boldsymbol{S}^\text{T} \cdot \boldsymbol{S} = \boldsymbol{I} \tag{40}$$

and

$$\boldsymbol{S}^\text{T} \cdot \boldsymbol{W}_1 \cdot \boldsymbol{\mathcal{O}}_{p+1}(\boldsymbol{\theta}_\text{o}) = \boldsymbol{0} \tag{41}$$

with $s = (p+1)r - n$ (r designating the vibration response vector dimensionality). Notice that the matrix $\boldsymbol{S}$ is not unique. It may be, for instance, obtained through the singular value decomposition (SVD) of $\boldsymbol{W}_1\boldsymbol{\mathcal{O}}_{p+1}(\boldsymbol{\theta}_\text{o})$, and implicitly depends upon the parameter vector $\boldsymbol{\theta}_\text{o}$; hence it may be designated as $\boldsymbol{S}(\boldsymbol{\theta}_\text{o})$. The block Hankel matrix corresponding to the same system should then satisfy the following property:

$$\boldsymbol{S}^\text{T}(\boldsymbol{\theta}_\text{o}) \cdot \boldsymbol{W}_1 \cdot \boldsymbol{\mathcal{H}}_{p+1,q} \cdot \boldsymbol{W}_2^T = \boldsymbol{0} \tag{42}$$

with $\boldsymbol{W}_2$ being an additional selectable invertible weighting matrix.

Given data Y_u from the current system (with parameter vector designated as $\boldsymbol{\theta}_\text{u}$), the residual signal (compare with equation (42)):

$$\boldsymbol{\zeta}_N(\boldsymbol{\theta}_\text{o}) \triangleq \sqrt{N}\text{vec}\left(\boldsymbol{S}^\text{T}(\boldsymbol{\theta}_\text{o}) \cdot \boldsymbol{W}_1 \cdot \widehat{\boldsymbol{\mathcal{H}}}_{p+1,q} \cdot \boldsymbol{W}_2^\text{T}\right) \tag{43}$$

where $\widehat{\boldsymbol{\mathcal{H}}}_{p+1,q}$ is the estimated (sample) block Hankel matrix obtained from the current data Y_u, will have zero mean under H_0 ($\boldsymbol{\theta}_\text{u} = \boldsymbol{\theta}_\text{o}$) and nonzero mean under H_1 ($\boldsymbol{\theta}_\text{u} \neq \boldsymbol{\theta}_\text{o}$) (in fact, it should be zero under H_0 if the theoretical block Hankel matrix is used—see equation (42)).

Testing whether the current structure (with parameter vector $\boldsymbol{\theta}_\text{u}$ and associated with the estimated block Hankel matrix) coincides with the healthy (with parameter vector $\boldsymbol{\theta}_\text{o}$) is based on a "statistical local approach", according to which the following "close" hypotheses are considered:

$$\begin{aligned} &H_0 : \boldsymbol{\theta}_\text{u} = \boldsymbol{\theta}_\text{o} \\ &\quad \text{(null hypothesis—healthy structure)} \\ &H_1 : \boldsymbol{\theta}_\text{u} = \boldsymbol{\theta}_\text{o} + \frac{1}{\sqrt{N}}\delta\boldsymbol{\theta} \\ &\quad \text{(alternative hypothesis—faulty structure)} \end{aligned} \tag{44}$$

with $\delta\boldsymbol{\theta}$ designating an unknown but fixed error vector.

Let $\boldsymbol{M}(\boldsymbol{\theta}_o)$ be the Jacobian matrix designating the mean sensitivity of $\boldsymbol{\zeta}_N$, and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_o) = \lim_{N\to\infty} E_{\boldsymbol{\theta}_o}\{\boldsymbol{\zeta}_N \boldsymbol{\zeta}_N^T\}$, with $E_{\boldsymbol{\theta}_o}\{\cdot\}$ designating the expectation operator when the actual system parameter is $\boldsymbol{\theta}_o$ (these matrices do not depend upon the sample size N and may be estimated from the healthy structure during the baseline phase). Provided that $\boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$ is positive definite, the residual $\boldsymbol{\zeta}_N(\boldsymbol{\theta}_o)$ of equation (43) asymptotically follows Gaussian distribution:

$$\boldsymbol{\zeta}_N(\boldsymbol{\theta}_o) \xrightarrow{N\to\infty} \begin{cases} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_o)\big) & \text{under } H_0 \\ \mathcal{N}\big(\boldsymbol{M}(\boldsymbol{\theta}_o)\delta\boldsymbol{\theta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_o)\big) & \text{under } H_1 \end{cases} \tag{45}$$

Now let $\widehat{\boldsymbol{M}}$ and $\widehat{\boldsymbol{\Sigma}}$ be consistent estimates of $\boldsymbol{M}(\boldsymbol{\theta}_o)$ (assumed to be full column rank) and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$, respectively, obtained in the baseline phase (details in [15]). Under the null (H_0) hypothesis, the statistic $\mathcal{X}_\zeta^2$, below, follows χ^2 distribution with rank $\boldsymbol{M}$ degrees of freedom, that is,

$$\begin{aligned} \text{Under } H_0\colon \mathcal{X}_\zeta^2 = \boldsymbol{\zeta}_N^T(\boldsymbol{\theta}_o)\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{M}}(\widehat{\boldsymbol{M}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{M}})^{-1} \\ \widehat{\boldsymbol{M}}^T\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\zeta}_N(\boldsymbol{\theta}_o) \sim \chi^2(\text{rank } \boldsymbol{M}) \end{aligned} \tag{46}$$

and decision making may be accomplished by examining—at a certain risk level—whether the residual sequence has mean that is significantly different from zero (similar to the previous method).

Fault identification could be similarly based on consecutive tests of the above-mentioned form, each one corresponding to each one of the parameter vectors $\boldsymbol{\theta}_A, \boldsymbol{\theta}_B, \ldots$. Nevertheless, as with most methods, this may only work with faults of specific magnitudes but not for the continuum of fault magnitudes possible within each fault type. Finally, note that the method is not immediately suitable for fault magnitude estimation.

Bibliographical remark. The method has been effectively used in various structures, including the Z24 bridge benchmark [70] and a steel-quake benchmark [71].

6.3 Functional model–based methods

Functional model (FM)-based methods provide a framework for the combined treatment of the fault detection, identification, and magnitude estimation subproblems in the common and practically important case of a double continuum of fault types (locations) and magnitudes [72–74]. An important aim is overcoming the limitation of identifying faults of a limited number of prespecified types and magnitudes, while allowing for the correct identification and magnitude estimation for faults of any magnitude.

The cornerstone of the methods is their unique ability to accurately represent a structure for its double continuum of fault types (locations) and magnitudes via a *single* FM. FMs, or more precisely referred to as *functionally pooled* (FP) models, as they are based on pooling together multiple data records, assume various forms, and in the general case, are explicitly parametrized in terms of both fault type (location) and magnitude [74]. An infinite number of fault types and magnitudes can thus be accommodated by a single model. In the simpler case where only a small, finite number of fault types is considered, the parametrization may be in terms of fault magnitude alone [72]—a single model is then limited to a single fault type.

Consider, for the sake of clarity, the simpler case of fault magnitude parametrization. Letting the fault magnitude (within a particular fault mode V) be represented by $k \in \Re$ (typically $k = 0$ is selected to, without loss of generality, correspond to the healthy structure), a suitable linear model for the structure under fault mode V is the functionally pooled autoregressive with exogenous excitation (FP-ARX) model:

$$\begin{aligned} &\mathcal{M}_V(a_{ij}, b_{ij}, \sigma_w^2(k))\colon y_k[t] + \sum_{i=1}^{na} a_i(k)\cdot y_k[t-i] \\ &\quad = \sum_{i=0}^{nb} b_i(k)\cdot x_k[t-i] + w_k[t] \\ &w_k[t] \sim \text{i.i.d.}\,\mathcal{N}(0, \sigma_w^2(k)), \\ &a_i(k) = \sum_{j=1}^{p} a_{ij}\cdot G_j(k), \\ &b_i(k) = \sum_{j=1}^{p} b_{ij}\cdot G_j(k), \quad k \in \Re \end{aligned} \tag{47, 48}$$

In this model, $x_k[t]$, $y_k[t]$, and $w_k[t]$ are the excitation, response, and innovations (prediction error) signals, respectively, corresponding to a particular fault magnitude k. Note that the FP-ARX model resembles a conventional ARX model (refer to Table 6), but explicitly accounts for the continuum of fault magnitudes (within the particular fault mode) by allowing its parameters and innovations variance to be functions of the fault magnitude k. In fact, the model parameters belong to a p-dimensional functional subspace spanned by the mutually independent functions $G_1(k), \ldots, G_p(k)$ (*functional basis*), while the constants a_{ij} and b_{ij} designate the model's AR and X, respectively, coefficients of projection. A suitable FM, corresponding to each particular fault mode, is estimated in the baseline phase using data obtained from the structure under various fault magnitudes (details in [72, 74]).

In the inspection phase, given the current signals $Z_u = (X_u, Y_u)$, fault detection is based on the FP-ARX model of any (e.g. V) fault mode. This model is reparametrized in terms of the currently unknown fault magnitude k and the innovations variance $\sigma^2_{e_u}$ (the coefficients of projection being replaced by estimates available from the baseline phase; $e_u[t]$ represents the reparametrized model's innovations):

$$\mathcal{M}_V(k, \sigma^2_{e_u}) : y_u[t] + \sum_{i=1}^{na} a_i(k) \cdot y_u[t-i] = \sum_{i=0}^{nb} b_i(k) \cdot x_u[t-i] + e_u[t] \tag{49}$$

The following hypothesis testing problem is then formulated:

$$H_0 : k = 0 \text{ (null hypothesis—healthy structure)}$$
$$H_1 : k \neq 0 \text{ (alternative hypothesis—faulty structure)} \tag{50}$$

Estimates of k, $\sigma^2_{e_u}$ are obtained on the basis of the current data Z_u and the nonlinear least squares (NLLS) estimator (refer to [26, pp. 327–329] for details on NLLS estimation):

$$\widehat{k} = \arg\min_k \sum_{t=1}^{N} e_u^2[t] \tag{51}$$

$$\widehat{\sigma}^2_{e_u} = \frac{1}{N} \sum_{t=1}^{N} \widehat{e}_u^2[t] \tag{52}$$

Assuming that the structure is indeed under a fault belonging to the V fault mode (or healthy), the estimator may be shown to be asymptotically ($N \to \infty$) Gaussian distributed, with mean equal to its true value and variance equal to that provided by the Cramer–Rao lower bound [72]:

$$\widehat{k} \sim \mathcal{N}(k, \sigma_k^2) \tag{53}$$

Then t, below, follows t distribution with $N-1$ degrees of freedom (which should be adjusted to $N-2$ in case the estimated mean is subtracted from the residual series in the computation of $\widehat{\sigma}_k^2$). This is due to the fact that it is the ratio of a standardized normal random variable over the square root of an independent and normalized χ^2 random variable with $N-1$ degrees of freedom (see Appendix):

$$t = \widehat{k}/\widehat{\sigma}_k \sim t(N-1) \tag{54}$$

which leads to the following test at the α (type I) risk level (Figure 8):

$$\begin{aligned} & t_{\frac{\alpha}{2}}(N-1) \leq t \leq t_{1-\frac{\alpha}{2}}(N-1) \\ & \Longrightarrow H_0 \text{ is accepted (healthy structure)} \\ \text{Else } & \Longrightarrow H_1 \text{ is accepted (faulty structure)} \end{aligned} \tag{55}$$

with t_α and $t_{1-(\alpha/2)}$ designating the t distribution's corresponding critical points.

Once a fault is detected, its identification is achieved through successive estimation (using the current data Z_u) and validation of the reparametrized models $\mathcal{M}_V(k, \sigma^2_{e_u})$ (equation 49) for $V = A, B, \ldots$ corresponding to the various fault types. The procedure stops as soon as a particular model is successfully validated, with the corresponding fault mode being identified as current. Model validation may be based on statistical tests examining the hypothesis of excitation and residual sequence uncrosscorrelatedness and/or residual uncorrelatedness (see Subsection 6.2.3).

An interval estimate (at the α risk level) of the fault magnitude k is finally constructed on the basis

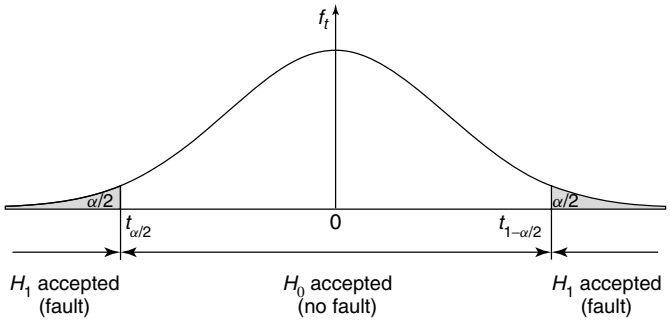


Figure 8. Statistical hypothesis testing based on a t distributed statistic (two-tailed test).

of the point estimate $\widehat{k}$ and $\widehat{\sigma}_k^2$:

k interval estimate:

$$\left[\widehat{k}+t_{\frac{\alpha}{2}}(N-1)\cdot\widehat{\sigma}_k,\ \widehat{k}+t_{1-\frac{\alpha}{2}}(N-1)\cdot\widehat{\sigma}_k\right] \quad (56)$$

7 APPLICATION OF THE METHODS TO FAULT DIAGNOSIS ON A LABORATORY STRUCTURE

Consider the scale aircraft skeleton structure of Figure 9, which has been constructed according to the Garteur SM-AG19 specifications. The structure consists of six solid beams representing the aircraft fuselage, the wings, the horizontal and vertical stabilizers, and the right and left wing tips. All elements are constructed from standard aluminum and are joined together via steel plates and screws (total skeleton mass ∼50 kg—also see [72]).

The faults considered correspond to the placement of a potentially variable number of small masses (simulating local elasticity reductions) at point B (right wing tip) or point E (horizontal stabilizer; Figure 9). Each fault is designated as F_{B}^k or F_{E}^k, with k representing the fault magnitude in gram of added mass and B or E the fault type (location). Four test cases are considered (Table 8). The structure is

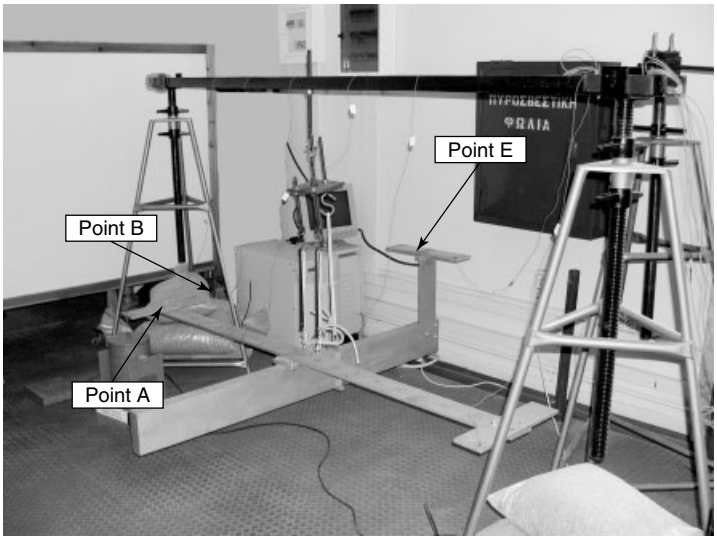


Figure 9. Aircraft scale skeleton structure and the experimental setup. The force/vibration measurement location (point A) and the fault locations (points B and E) are marked.

Table 8. The considered test cases

Test case	Structural state	Description
I	$k = 0$	Healthy structure (no mass attached)
II	$F_{\mathrm{B}}^{8.1}$	Mass (8.1 g) at point B
III	$F_{\mathrm{B}}^{24.4}$	Mass (24.4 g) at point B
IV	$F_{\mathrm{E}}^{16.3}$	Mass (16.3 g) at point E

healthy in test case I, under type B faults in test cases II and III, and under a type E fault in test case IV.

Fault detection, identification, and estimation (fault diagnosis) are based on vibration data records obtained under free–free boundary conditions. A white (uncorrelated) random Gaussian force excitation is applied vertically at the right wing tip (point A) through an electromechanical shaker equipped with a stinger. The exerted force and the resulting vertical vibration acceleration are measured via an impedance head and a lightweight accelerometer at point A (sampling frequency $f_s = 256\,\mathrm{Hz}$, signal bandwidth $3-100\,\mathrm{Hz}$, signal length $N = 5000$ samples). The estimated mean is also subtracted from each signal, plus scaling by the signal's estimated standard deviation is implemented. Note that while the type B faults occur "close" to the measurement location, the type E fault is "remote," and thus potentially harder to detect.

Results obtained by most of the presented methods are shown in the sequel—note that the coherence measure–based method is not applicable as the nature of the considered faults does not lead to nonlinear dynamics behavior.

7.1 Fault diagnosis

1. *Nonparametric methods.* The psd- and frf-based methods use Welch spectral estimates (signal length $N = 5000$ samples, segment length $L = 500$ samples, number of segments $K = 10$, zero overlap—refer to Table 7). Fault detection results are presented in Figure 10. Evidently, correct detection is obtained in each case and by both methods, as the test statistic is shown not to exceed (exceed) the critical point in the healthy (faulty) cases. This is true for the "remote" fault (test case IV) as well. Rather expectedly, the frf-based method (risk level $\alpha = 0.001$) provides clearer detection than the psd-based method (risk level $\alpha = 0.05$).
2. *Parametric methods.* The model parameter-, residual variance-, likelihood function-, and residual uncorrelatedness–based methods use excitation–response ARX modeling and a Prediction Error (PE) estimator based on $N = 5000$ sample–long signal records. The identification procedure leads to an ARX(62,62) model, which is selected as adequate. Fault detection results are presented in Figure 11, with the combined AR/X parameter vector used as the characteristic quantity in the model parameter–based method. Evidently, correct detection is obtained in each case and by all four methods, as the test statistic is shown not to exceed (exceed) the critical point (at the $\alpha = 0.01$ risk level) in the healthy (faulty) cases. As expected, the residual variance and likelihood function–based results are similar. The "remote" fault (test case IV) is clearly detected by all parametric methods, in particular, by the model parameter and residual uncorrelatedness–based ones.
3. *The functional model–based method.* This method uses $N = 4000$ sample-long excitation–response signals in both the baseline and inspection phases. The modeling of fault mode F_{B} (type B faults) is based on signals obtained from 11 experiments (for $k \in [0, 81.32]$ g; increment $\delta k \approx 8.132\,\mathrm{g}$). FP–ARX$(n, n)$ modeling leads to an FP-ARX(62,62) representation with unity excitation–response delay and functional basis consisting of the first $p = 11$ Chebyshev II polynomials [75, p. 782]. Fault detection results are presented in Figure 12(a)–(d): A fault is detected if the point $k = 0$ lies outside the shaded zone (which is equivalent to the t statistic exceeding the critical values) at the $\alpha = 0.05$ risk level. Evidently, the detection results are very accurate; no fault is detected in test case I, while a fault is detected in every other case, including very clear detection for the "remote" fault of test case IV.

Once a fault is detected, its identification (localization) is based on the residual uncorrelatedness test of Figure 12(e): the pertinent statistic must be below

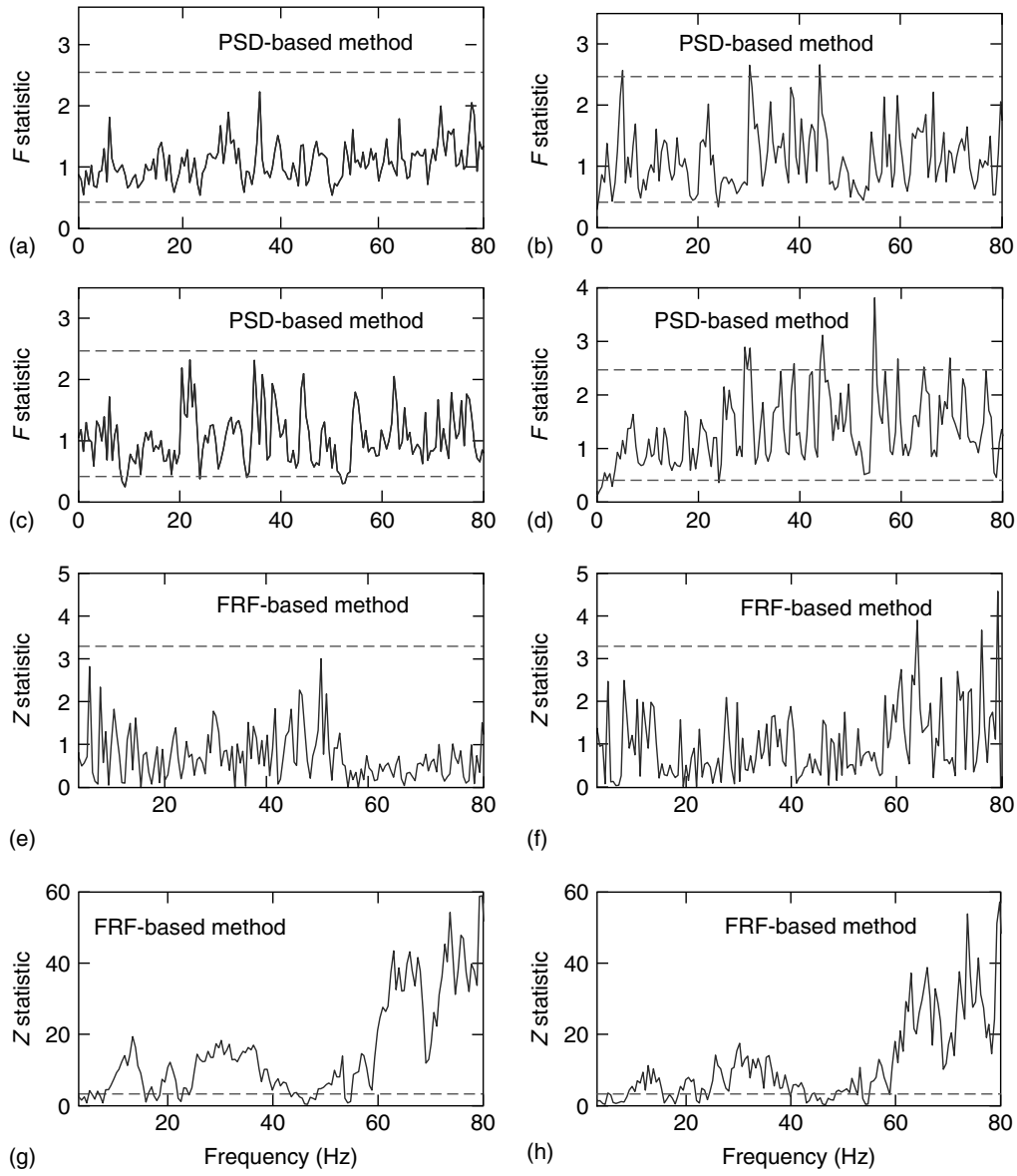


Figure 10. Fault detection results via nonparametric methods. (a–d) PSD-based method (risk level $\alpha = 0.05$); (e–h) frf-based method (risk level $\alpha = 0.001$). A fault is detected if the test statistic exceeds the critical points (dashed horizontal lines).

the critical value (dashed horizontal line; risk level $\alpha = 0.05$) for a fault to belong to fault mode (location) B. Evidently, all faults are correctly identified: the faults in test cases I–III are properly identified as belonging to fault mode B, while the fault in test case IV is very clearly and properly identified as not belonging to it (note that the healthy structure by definition belongs to any fault mode). For the faults confirmed as belonging to the modeled fault mode B (test cases I–III, but not test case IV), the shaded zones in Figure 12(a–c) then correspond to fault magnitude interval estimates (at the $\alpha = 0.05$ risk

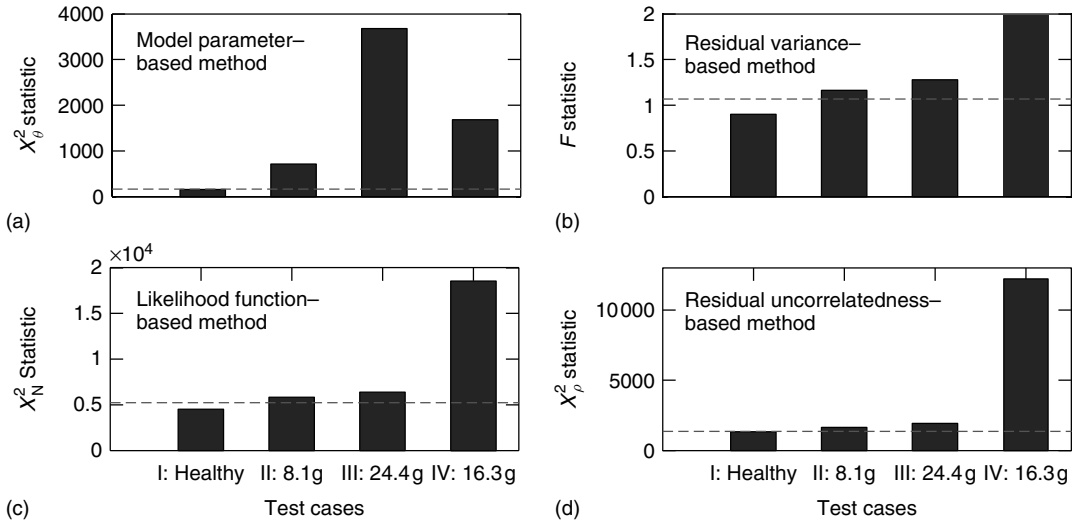


Figure 11. Fault detection results via parametric methods. (a) Model parameter; (b) residual variance; (c) likelihood function; and (d) residual uncorrelatedness-based methods. A fault is detected if the test statistic (bar) exceeds the critical point (dashed line; risk level $\alpha = 0.01$).

level), with the point estimates $(\widehat{k})$ indicated by the middle solid vertical line. The accuracy achieved in fault magnitude estimation is very good—it is worthwhile observing that the true fault magnitudes (dashed vertical lines) lie within the interval estimates.

8 CONCLUDING REMARKS

- Statistical time series methods for structural health monitoring achieve fault detection, identification (localization), and estimation based on (i) random excitation and/or vibration response signals, (ii) statistical model building, and (iii) statistical decision making under uncertainty.
- The methods are data based, inverse type, and thus of quite general applicability.
- In addition to sharing the benefits of general vibration-based methods—no visual inspection, "global" coverage, time and cost effectiveness, automation capability—statistical time series methods offer a number of unique advantages: (i) No need for physics–based or finite element models; (ii) no need for complete structural models (partial models and a limited number of responses suffice); (iii) inherent accounting of uncertainty; (iv) statistical decision making with specified performance characteristics; and (v) effective use of natural random vibration data records for in-operation structural health monitoring.
- On the other hand, statistical time series methods may identify a fault only to the extent allowed by the type of model used, require adequate expertise on part of the user, may offer limited physical insight, and sensitivity lower than that of "local" nondestructive testing-type methods.
- Statistical time series methods may be either of the nonparametric or parametric types. The latter are generally more elaborate, but offer improved capabilities particularly for the fault identification and estimation subproblems.
- Further research is necessary, in particular, for exploring the limits and limitations of the methods, as well as for tackling the less-studied fault identification (localization) and estimation subproblems.
- Research should be also devoted to attaining robustness under substantial uncertainty and varying environmental or operating conditions, achieving high levels of automation, the capability to work with large arrays of sensors, adaptation to real-time operation, as well as effective operation under nonGaussian environments and with nonlinear and time-varying structures.

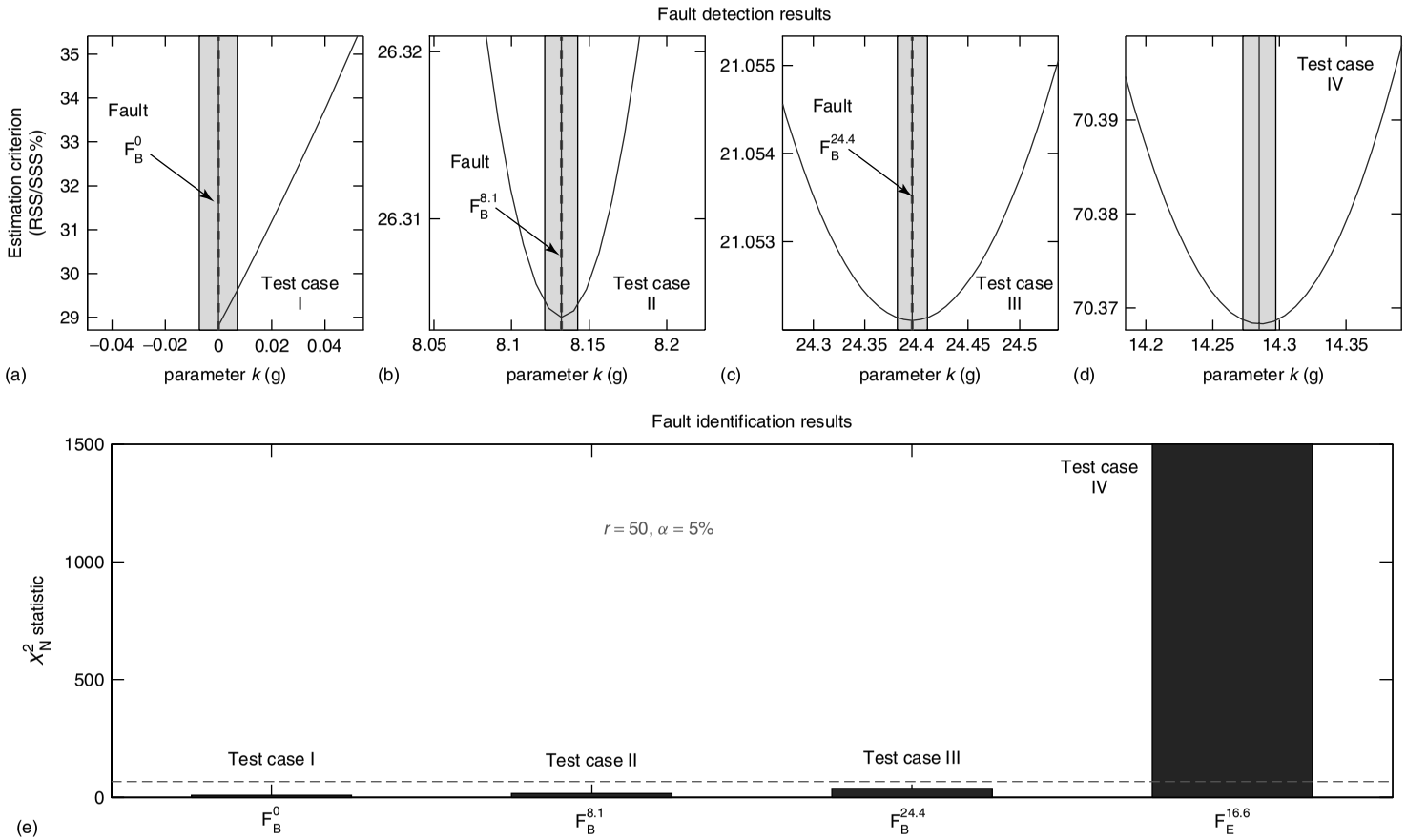


Figure 12. Fault detection, identification, and estimation results via the functional model–based method. (a–d) Fault detection results: a fault is detected if $k = 0$ lies outside the shaded zone ($\alpha = 0.05$). The curve shown is the estimation criterion as a function of the fault magnitude k. (e) Fault identification results: each fault is accepted as type B if the test statistic (bar) lies below the critical point (dashed horizontal line; $\alpha = 0.05$). For such faults, each shaded zone above then corresponds to the fault magnitude interval estimate at the $\alpha = 0.05$ risk level—the point estimate being indicated by the middle solid line and the true magnitude by the dashed line.

RELATED ARTICLES

Data Preprocessing for Damage Detection

Time–frequency Analysis

Wavelet Analysis

Higher Order Statistical Signal Processing

Statistical Pattern Recognition

Artificial Neural Networks

Nonlinear Features for SHM Applications

Novelty Detection

Model-based Statistical Signal Processing for Change and Damage Detection

Uncertainty Analysis

REFERENCES

[1] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics: A Literature Review*, Report LA-13070-MS. Los Alamos National Laboratory, 1996.

[2] Salawu OS. Detection of structural damage through changes in frequency: a review. *Engineering Structures* 1997 **19**(9):718–723.

[3] Doebling SW, Farrar CR, Prime MB, Shevitz DW. A summary review of vibration-based damage identification methods. *Shock and Vibration Digest* 1998 **30**(2):91–105.

[4] Zou Y, Tong L, Steven GP. Vibration-based model-dependent damage (delamination) identification and health monitoring for composite structures—a review. *Journal of Sound and Vibration* 2000 **230**(2):357–378, doi:10.1006/jsvi.1999.2624.

[5] Farrar CR, Doebling SW, Nix DA. Vibration-based structural damage identification. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences* 2001 **359**:131–149, doi:10.1098/rsta.2000.0717.

[6] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinmates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Report Number LA-13976-MS. Los Alamos National Laboratory, 2003.

[7] De Roeck G. The state–of–the–art of damage detection by vibration monitoring: the SIMCES experience. *Journal of Structural Control* 2003 **10**:127–134, doi:10.1002/stc.20.

[8] Carden EP, Fanning P. Vibration based condition monitoring: a review. *Structural Health Monitoring* 2004 **3**(4):355–377, doi:10.1177/1475921704047500.

[9] Fritzen C-P. Vibration–based techniques for structural health monitoring. In *Structural Health Monitoring*, Balageas D, Fritzen C-P, Guemes A (eds). ISTE, 2006; pp. 45–224.

[10] Montalvao D, Maia NMM, Ribeiro AMR. A review of vibration–based structural health monitoring with special emphasis on composite materials. *Shock and Vibration Digest* 2006 **38**(4):295–324, doi:10.1177/0583102406065898.

[11] Adams DE. *Health Monitoring of Structural Materials and Components*. John Wiley & Sons, 2007.

[12] Staszewski W, Boller C, Tomlinson G (eds). *Health Monitoring of Aerospace Structures*. John Wiley & Sons, 2004.

[13] Inman DJ, Farrar CR, Lopes V Jr., Steffen V Jr. (eds). *Damage Prognosis for Aerospace, Civil and Mechanical Systems*. John Wiley & Sons, 2005.

[14] Natke HG, Cempel C. *Model–Aided Diagnosis of Mechanical Systems: Fundamentals, Detection, Localization, Assessment*. Springer-Verlag, 1997.

[15] Basseville M, Abdelghani M, Benveniste A. Subspace-based fault detection algorithms for vibration monitoring. *Automatica* 2000 **36**(1):101–109.

[16] Sohn H, Farrar CR. Statistical process control and projection techniques for structural health monitoring. *Proceedings of the European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000.

[17] Fugate ML, Sohn H, Farrar CR. Vibration-based damage detection using statistical process control. *Mechanical Systems and Signal Processing* 2001 **15**(4):707–721, doi:10.1006/mssp.2000.1323.

[18] Yan A-M, de Boe P, Golinval J-C. Structural damage diagnosis by Kalman model based on stochastic subspace identification. *Structural Health Monitoring* 2004 **3**(2):103–119, doi:10.1177/1475921704042545.

[19] Carden EP, Brownjohn JMW. ARMA modelled time-series classification for structural health monitoring of civil infrastructure. *Mechanical Systems and Signal Processing* 2007 **22**(2):295–314, doi:10.1016/j.ymssp.2007.07.003.

[20] Fassois SD, Sakellariou JS. Time series methods for fault detection and identification in vibrating

structures. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences* 2007 **365**:411–448, doi:10.1098/rsta.2006.1929.

[21] Sohn H. Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences* 2007 **365**:539–560, doi:10.1098/rsta.2006.1931.

[22] Fassois SD. Parametric identification of vibrating structures. In *The Encyclopedia of Vibration*, Braun SG, Ewins DJ, Rao SS (eds). Academic Press, 2001; pp. 673–685.

[23] Priestley MB. *Spectral Analysis and Time Series*. Academic Press: 1981.

[24] Kay SM. *Modern Spectral Estimation: Theory and Application*. Prentice Hall: New Jersey, 1988.

[25] Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting & Control*, *Third Edition*. Prentice Hall: Englewood Cliffs, NJ, 1994.

[26] Ljung L. *System Identification: Theory for the User*, *Second Edition*. PTR Prentice Hall: Upper Saddle River, NJ, 1999.

[27] Basseville M, Nikiforov IV. *Detection of Abrupt Changes: Theory and Application*. PTR Prentice-Hall, 1993.

[28] Montgomery DC. *Introduction to Statistical Quality Control*, *Second Edition*. John Wiley & Sons, 1991.

[29] Bendat JS, Piersol AG. *Random Data: Analysis and Measurement Procedures*. *Third Edition*, Wiley–Interscience: New York, 2000.

[30] Söderström T, Stoica P. *System Identification*. Prentice Hall, 1989.

[31] Brockwell PJ, Davis RA. *Time Series: Theory and Methods*. Springer-Verlag, 1987.

[32] Worden K. Structural fault detection using a novelty measure. *Journal of Sound and Vibration* 1997 **201**(1):85–101.

[33] Worden K, Manson G, Fieller NRJ. Damage detection using outlier analysis. *Journal of Sound and Vibration* 2000 **229**(3):647–667, doi:10.1006/jsvi.1999.2514.

[34] Worden K, Manson G. Experimental validation of a structural health monitoring methodology: part I. Novelty detection on a laboratory structure. *Journal of Sound and Vibration* 2003 **259**(2):323–343, doi:10.1006/jsvi.2002.5168.

[35] Sakellariou JS, Petsounis KA, Fassois SD. Vibration analysis based on–board fault detection in railway vehicle suspensions: a feasibility study. *Proceedings First National Conference on Recent Advances in Mechanical Engineering*. Patras, 2001; paper ANG1/P080.

[36] Staszewski WJ. Advanced data pre-processing for damage identification based on pattern recognition. *International Journal of Systems Science* 2000 **31**(11):1381–1396.

[37] Staszewski WJ. Structural and mechanical damage detection using wavelets. *Shock and Vibration Digest* 1998 **30**:457–472.

[38] Peng ZK, Chu FL. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing* 2004 **18**:199–221, doi:10.1016/S0888-3270(03)00075-X.

[39] Farrar CR, Doebling SW. Using statistical analysis to enhance modal based damage identification. In *Structural Damage Assessment using Advanced Signal Processing Procedures*, Dulieu JM, Staszewski WJ, Worden K (eds). Academic Press: Sheffield: 1997; pp. 199–211.

[40] Hou Z, Noori M, Amand RSt. Wavelet–based approach for structural damage detection. *Journal of Engineering Mechanics* 2000 **126**(7):677–683.

[41] Hera A, Hou Z. Application of wavelet approach for ASCE structural health monitoring benchmark studies. *Journal of Engineering Mechanics* 2004 **130**(1):96–104.

[42] Staszewski WJ, Robertson AN. Time–frequency and time–scale analyses for structural health monitoring. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences* 2007 **365**:449–477, doi:10.1098/rsta.2006.1936.

[43] Sakellariou JS, Fassois SD. Parametric output error based identification and fault detection in structures under earthquake excitation. *Proceedings of the European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000; pp. 323–332. Extended version in the *Journal of Sound and Vibration* 2006 **297**:1048–1067. doi:10.1016/j.jsv.2006.05.009.

[44] Stuart A, Ord JK. *Kendall's Advanced Theory of Statistics: Vol 1. Distribution Theory*, *Fifth Edition*, Oxford University Press: New York, 1987.

[45] Brillinger DR. *Time Series Data Analysis and Theory*. Holden-Day: California, CA, 1981.

[46] Rizos DD, Fassois SD, Marioli-Riga ZP, Karanika AN. Vibration–based skin damage statistical detection and restoration assessment in a stiffened aircraft

panel. *Mechanical Systems and Signal Processing* 2008 **22**:315–337, http://dx.doi.org/10.1016/j.ymssp.2007.07.012.

[47] Gertler JJ. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, 1998.

[48] Sadeghi MH, Fassois SD. A geometric approach to the non–destructive identification of faults in stochastic structural systems. *AIAA Journal* 1997 **35**:700–705.

[49] Sadeghi MH, Fassois SD. Reduced–dimensionality geometric approach to fault identification in stochastic structural systems. *AIAA Journal* 1998 **36**:2250–2256.

[50] Adams DE, Farrar CR. Classifying linear and nonlinear structural damage using frequency domain ARX models. *Structural Health Monitoring* 2002 **1**(2):185–201.

[51] Wei Z, Yam LH, Cheng L. NARMAX model representation and its application to damage detection for multi–layer composites. *Composite Structures* 2005 **68**:109–117, doi:10.1016/j.compstruct.2004.03.005.

[52] Nair KK, Kiremidjian AS, Law KH. Time-series based damage detection and localization algorithm with application to the ASCE benchmark structure. *Journal of Sound and Vibration* 2006 **291**:349–368, doi:10.1016/j.jsv.2005.06.016.

[53] Nair KK, Kiremidjian AS. Time series based structural damage detection algorithm using Gaussian mixtures modeling. *Journal of Dynamic Systems Measurement and Control* 2007 **129**:285–293.

[54] Hearn G, Testa RB. Modal analysis for damage detection in structures. *Journal of Structural Engineering* 1991 **117**:3042–3063.

[55] Huang CS, Hung SL, Wen CM, Tu TT. A neural network approach for structural identification and diagnosis of a building from seismic response data. *Earthquake Engineering and Structural Dynamics* 2003 **32**:187–206, doi:10.1002/eqe.219.

[56] Mehra RK, Peschon J. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica* 1971 **7**:637–640.

[57] Willsky AS. A survey of design methods for failure detection in dynamic systems. *Automatica* 1976 **12**:601–611.

[58] Basseville M. Detecting changes in signals and systems: a survey. *Automatica* 1988 **24**(3):309–326.

[59] Frank PM. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results. *Automatica* 1990 **26**(3):459–474.

[60] Deramemaeker A, Reynders E, De Roeck G, Kullaa J. Vibration–based structural health monitoring using output–only measurements under changing environment. *Mechanical Systems and Signal Processing* 2008 **22**(1):34–56, http://dx.doi.org/10.1016/j.ymssp.2007.07.004.

[61] Sohn H, Farrar CR. Damage diagnosis using time series analysis of vibration signals. *Smart Materials and Structures* 2001 **10**:446–451.

[62] Sohn H, Farrar CR, Hunter NF, Worden K. Structural health monitoring using statistical pattern recognition techniques. *ASME Journal of Dynamic Systems Measurement and Control* 2001 **123**(4):706–711.

[63] Sohn H, Allen DW, Worden K, Farrar CR. Structural damage classification using extreme value statistics. *Journal of Dynamic Systems Measurement and Control* 2005 **127**:125–132, doi:10.1115/1.1849240.

[64] Lu Y, Gao F. A novel time-domain auto-regressive model for structural damage diagnosis. *Journal of Sound and Vibration* 2005 **283**:1031–1049, doi:10.1016/j.jsv.2004.06.030.

[65] Zhang QW. Statistical damage identification for bridges using ambient vibration data. *Computers and Structures* 2007 **85**:476–485, doi:10.1016/j.compstruc.2006.08.071.

[66] Poulimenos AG, Fassois SD. Vibration–based online fault detection in non-stationary structural systems via a statistical model based method. *Proceedings of the Second European Workshop on Structural Health Monitoring*. Munich, 2004; pp. 687–694.

[67] Mattson SG, Pandit SM. Statistical moments of autoregressive model residuals for damage localization. *Mechanical Systems and Signal Processing* 2006 **20**:627–645, doi:10.1016/j.ymssp.2004.08.005.

[68] Nakamura M, Masri SF, Chassiakos AG, Caughey TK. A method for non–parametric damage detection through the use of neural networks. *Earthquake Engineering and Structural Dynamics* 1998 **27**:997–1010.

[69] Masri SF, Smyth AW, Chassiakos AG, Caughey TK, Hunter NF. Application of neural networks for detection of changes in nonlinear systems. *American Society of Civil Engineers (ASCE) Journal of Engineering Mechanics* 2000 **126**(7):666–676.

[70] Mevel L, Goursat M, Basseville M. Stochastic subspace-based structural identification and damage detection and localization—application to the Z24 bridge benchmark. *Mechanical Systems and Signal Processing* 2003 **17**(1):143–151, doi:10.1006/mssp.2002.1552.

[71] Mevel L, Basseville M, Goursat M. Stochastic subspace-based structural identification and damage detection—application to the steel-quake benchmark. *Mechanical Systems and Signal Processing* 2003 **17**(1):91–101, doi:10.1006/mssp.2002.1544.

[72] Sakellariou JS, Fassois SD. Fault detection and identification in an aircraft skeleton structure via a stochastic functional model based method. *Mechanical Systems and Signal Processing* 2008 **22**(3):557–573, http://dx.doi.org/10.1016/j.ymssp.2007.09.002.

[73] Sakellariou JS, Petsounis KA, Fassois SD. On–board fault detection and identification in railway vehicle suspensions via a functional model based method. *Proceedings of ISMA*. Leuven, 2002.

[74] Kopsaftopoulos FP, Fassois SD. Vibration–based structural damage detection and precise assessment via stochastic functionally pooled models. *Key Engineering Materials* 2007 **347**:127–132.

[75] Abramowitz M, Stegun IA. *Handbook of Mathematical Functions*. Dover, 1970.

[76] Nguyen HT, Rogers GS. *Fundamentals of Mathematical Statistics: Vols. I and II*. Springer-Verlag: New York, 1989.

APPENDIX: CENTRAL LIMIT THEOREM AND STATISTICAL DISTRIBUTIONS ASSOCIATED WITH THE NORMAL

The central limit theorem (CLT)

([28, p. 46]; [44, p. 273], [76, Vol. I p. 420]). Let $Z_1, Z_2, \ldots Z_n$ designate mutually independent random variables each with mean μ_k and (finite) variance σ_k^2. Then, for $n \to \infty$ the distribution of the random variable $X = \sum_{k=1}^{n} Z_k$ approaches the Gaussian distribution with mean $E\{X\} = \sum_{k=1}^{n} \mu_k$ and variance $\text{var}(X) = \sum_{k=1}^{n} \sigma_k^2$.

The χ^2 distribution

Let $Z_1, Z_2, \ldots Z_n$ designate mutually independent, normally distributed, random variables, each with mean μ_k and standard deviation σ_k. Then the sum:

$$X = \sum_{k=1}^{n} \left(\frac{Z_k - \mu_k}{\sigma_k} \right)^2 \tag{A.1}$$

is said to follow a (central) χ^2 distribution with n degrees of freedom ($X \sim \chi^2(n)$). Its mean and variance are $E\{X\} = n$ and $\text{var}(X) = 2n$, respectively. Notice that imposing p equality constraints among the random variables $Z_1, Z_2, \ldots, Z_n$ reduces the set's effective dimensionality, and thus the number of degrees of freedom, by p [44, pp. 506–507].

For $n \to \infty$ the $\chi^2(n)$ distribution tends to normality [44, p. 523].

The sum $X = \sum_{k=1}^{n} Z_k^2/\sigma_k^2$ is said to follow noncentral χ^2 distribution with n degrees of freedom and noncentrality parameter $\lambda = \mu_k^2/\sigma_k^2$. This distribution is designated as $\chi^2(n;\lambda)$ [76, Vol. II p. 33].

Let $\boldsymbol{x} \in \Re^n$ follow n-variate normal distribution with zero mean and covariance $\boldsymbol{\Sigma}$ ($\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$). Then the quantity $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$ follows (central) χ^2 distribution with n degrees of freedom [30, p. 557]; [44, pp. 486–487], [47, p. 120].

The Student's t distribution

Let Z be the standard (zero mean and unit variance) normal variable. Let X follow a (central) χ^2 distribution with n degrees of freedom and be independent of Z. Then the ratio

$$T = \frac{Z}{\sqrt{X/n}} \tag{A.2}$$

is said to follow a Student's or t (central) distribution with n degrees of freedom (central because it is based on a central χ^2 distribution; [76, Vol. II p. 34]). Its mean and variance are $E\{T\} = 0$ ($n > 1$)

and $\mathrm{var}(T) = n/(n-2)$ $(n > 2)$, respectively [44, p. 513].

The (central) t distribution approaches the standard normal distribution $\mathcal{N}(0, 1)$ as $n \to \infty$ [44, p. 523].

The Fisher's F distribution

Let X_1, X_2 be mutually independent random variables following (central) χ^2 distributions with n_1, n_2 degrees of freedom, respectively. Then the ratio

$$F = \frac{X_1/n_1}{X_2/n_2} \tag{A.3}$$

is said to follow a (central) F distribution with n_1, n_2 degrees of freedom ($F \sim F(n_1, n_2)$) (central because it is based on central χ^2 distributions; [76, Vol. II p. 34]). Its mean and variance are $E\{F\} = n_2/(n_2 - 2)$ $(n_2 > 2)$ and $\mathrm{var}(F) = 2n_2^2(n_1 + n_2 - 2)/(n_1(n_2 - 2)^2(n_2 - 4))$ $(n_2 > 4)$, respectively [44, p. 518].

Note that for the distribution's $1 - \alpha$ critical point $f_{1-\alpha}(n_1, n_2) = 1/f_\alpha(n_2, n_1)$.

The (central) F distribution approaches normality as $n_1, n_2 \to \infty$. For $n_2 \to \infty$ $n_1 F$ approaches a (central) χ^2 distribution with n_1 degrees of freedom [44, p. 523].

Chapter 24

Cepstral Methods of Operational Modal Analysis

Robert B. Randall

School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, NSW, Australia

1 INTRODUCTION

The cepstrum was originally proposed by Bogert *et al.* [1] as a better alternative to the autocorrelation function for detecting echo delay times, specifically for seismic signals. At that time, it was defined as the power spectrum of the logarithm of the power spectrum. The "power cepstrum" was later redefined [2] as the inverse Fourier transform of the log power spectrum, partly because it is more logical to use the inverse transform between a function of frequency and a function of time, and partly because it is then reversible to the power spectrum (e.g., after editing). At about the same time, the "complex cepstrum" was defined as the inverse Fourier transform of the complex logarithm of the complex spectrum [2], this being reversible to a time function, and, for example, permitting removal of echoes from a time signal. This article uses the complex cepstrum, and other versions that have the same property of retaining phase information from the spectrum. It is shown below that the complex cepstrum of minimum-phase functions can be obtained from the power cepstrum, since the phase is determined by the log amplitude. Thus, for a time signal $f(t)$, its complex cepstrum is defined as

$$C(\tau) = \Im^{-1}\{\log(F(f))\} = \Im^{-1}\{\ln(A(f)) + j\phi(f)\} \quad (1)$$

where

$$F(f) = \Im\{f(t)\} = A(f)\ \mathrm{e}^{j\phi(f)} \quad (2)$$

in terms of the amplitude and phase of the spectrum of $f(t)$. The phase function $\phi(f)$ must be unwrapped to a continuous function of frequency. The independent variable τ of the cepstrum has the dimensions of time but is known as *quefrency*. It is analogous to the time variable of the autocorrelation function, and represents lag time or periodic time rather than absolute time. It is worth noting that the "complex cepstrum" is real, because the log amplitude of the spectrum is even, and the phase spectrum is odd.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

For SIMO (single input, multiple output) systems, any output signal is the convolution of the input signal with the impulse response function of the transmission path, and this convolutive relationship is converted to an addition by the cepstrum operation as follows:

If

$$y(t) = f(t) * h(t) \tag{3}$$

then

$$Y(f) = F(f) \cdot H(f) \tag{4}$$

$$\log(Y(f)) = \log(F(f)) + \log(H(f)) \tag{5}$$

and

$$\Im^{-1}\{\log(Y(f))\} = \Im^{-1}\{\log(F(f))\} + \Im^{-1}\{\log(H(f))\} \tag{6}$$

The cepstrum is said to have "homomorphic" properties since it converts a convolution into an addition, meaning that deconvolution can be achieved by linear filtering (subtraction) [3].

It is this property that makes the cepstrum interesting for determining the dynamic properties of structures (modal analysis), since in the cepstrum of the response these are added to the cepstrum of the forcing function, under certain conditions in disparate regions, thus allowing separation without full knowledge of the force.

1.1 Terminology

In their original paper [1], Bogert *et al.* coined the word "cepstrum" by reversing the first syllable of "spectrum", the justification being that it was a "spectrum of a spectrum". Similarly, the word "quefrency" was obtained from "frequency", and the authors also suggested a number of others, including

rahmonic	from	harmonic
lifter	from	filter
short-pass lifter	from	low-pass filter
long-pass lifter	from	high-pass filter
gamnitude	from	magnitude
saphe	from	phase

Of these terms, the first four are useful in clarifying that the operations or features refer to the cepstrum, rather than the spectrum or time signal, and are still regularly used in the literature as well as in this article. Note that the autocorrelation function is also a "spectrum of a spectrum", so the distinctive feature of the cepstrum is the log conversion of the spectrum.

2 THE CEPSTRUM OF STRUCTURAL TRANSFER FUNCTIONS

To determine the cepstrum of structural transfer functions, it is convenient to assume that all signals are digitized, so that the Fourier transform can be replaced by the Z transform. For a general impulse response function,

$$h(n) = \sum_{i=1}^{2N} A_i \, \mathrm{e}^{s_i n \Delta t} = \sum_{i=1}^{2N} A_i z_i^n \tag{7}$$

where the A_i are residues, s_i and z_i poles in the S plane and Z plane, respectively, and Δt the time sample spacing, the transfer function in the Z plane is given by

$$H(z) = \frac{Bz^r \prod_{i=1}^{M_i} (1 - a_i z^{-1}) \prod_{i=1}^{M_0} (1 - b_i z)}{\prod_{i=1}^{N_i} (1 - c_i z^{-1}) \prod_{i=1}^{N_0} (1 - d_i z)} \tag{8}$$

where the c_i and a_i are poles and zeros inside the unit circle in the Z plane, respectively, and the d_i and b_i are the (reciprocals of the) poles and zeros outside the unit circle, respectively (so that $|a_i|, |b_i|, |c_i|, |d_i| < 1$). Oppenheim and Schafer [3] showed that for this transfer function $H(z)$ the cepstrum can be represented as

$$\begin{aligned} C_h(n) &= \ln(B), && n = 0 \\ C_h(n) &= -\sum_i \frac{a_i^n}{n} + \sum_i \frac{c_i^n}{n}, && n > 0 \\ C_h(n) &= \sum_i \frac{b_i^{-n}}{n} - \sum_i \frac{d_i^{-n}}{n}, && n < 0 \end{aligned} \tag{9}$$

Here, the discretized time variable n represents quefrency, and the time delay represented by z^r in equation (8) should first be removed.

The development is illustrated by taking the log of the c_i terms (poles inside the unit circle) as follows:

$$-\sum_{i=1}^{N_i}\log(1-c_i z^{-1}) = \sum_{i=1}^{N_i}\sum_{n=1}^{\infty}\frac{c_i^n}{n}z^{-n}$$
$$= \sum_{n=1}^{\infty}\left(\sum_{i=1}^{N_i}\frac{c_i^n}{n}\right)z^{-n} = \mathrm{Z}\left\{\sum_{i=1}^{N_i}\frac{c_i^n}{n}\right\} \qquad (10)$$

so that the cepstrum for those terms is given by the inverse Z transform as $\sum_{i=1}^{N_i}(c_i^n/n)$, as in equation (9), and the other terms are completely analogous.

From equation (9) it can be seen that for minimum-phase functions, with no zeros or poles outside the unit circle, the cepstrum is zero for negative quefrency, and is thus causal. For this reason, the real and imaginary parts of the Fourier transform of the cepstrum of minimum-phase functions (i.e., the log amplitude and phase of the spectrum) are related by a Hilbert transform, so the complex cepstrum can be obtained from the log amplitude spectrum alone, avoiding the necessity to measure or unwrap the phase [3]. The real, even cepstrum obtained by inverse transforming the log amplitude spectrum is converted into the complex cepstrum of the minimum-phase function by doubling the value of positive quefrency components, and setting negative quefrency components to zero (i.e., by multiplying it by $2H(\tau)$ where H is the Heaviside function).

There is another related function with homomorphic properties known as the "differential cepstrum" [4], which is defined as the inverse transform of the derivative of the log spectrum; thus,

$$C_d(\tau) = \mathrm{Z}^{-1}\left\{\frac{\mathrm{d}}{\mathrm{d}z}[\log H(z)]\right\} = \mathrm{Z}^{-1}\left\{\frac{H'(z)}{H(z)}\right\} \qquad (11)$$

Since taking the derivative in one domain corresponds to multiplication by the independent variable in the other domain, the differential cepstrum can be written in terms of the poles and zeros in the Z plane (*cf* equation 9) as

$$C_{hd}(n) = -\sum_i a_i^n + \sum_i c_i^n, \quad n > 0$$

$$C_{hd}(n) = \sum_i b_i^{-n} - \sum_i d_i^{-n}, \quad n < 0 \qquad (12)$$

Note that in [4] these equations are given for quefrency displaced by one sample ($0 \rightarrow 1$), but the zero origin has been maintained here to emphasize the similarity with the complex cepstrum.

It will be recognized that the form of the differential cepstrum (sums of complex exponentials) is the same as that of the impulse response function, and this can be used for parameter extraction as shown below.

Another advantage of the differential cepstrum is that it can be directly calculated from a time signal as

$$C_{hd}(n) = \Im^{-1}\left[\frac{\Im\{n\ x(n)\}}{\Im\{x(n)\}}\right] \qquad (13)$$

which avoids the necessity to unwrap the phase of the spectrum.

Another version proposed by Antoni *et al.* [5] is the "mean differential cepstrum" (MDC), which is formally defined in terms of the partial derivative of the logarithm of the spectral correlation function, but can be calculated by a formula similar to equation (11), to which it reverts for a single realization:

$$C_{hmd}(\tau) = \Im^{-1}\left\{\frac{E[H'(f)\ H^*(f)]}{E[H(f)\ H^*(f)]}\right\} \qquad (14)$$

This permits averaging over a number of realizations, for example, of the response of a system excited by a burst random sequence and Antoni demonstrates in [5] that this gives a more repeatable result.

3 MODAL ANALYSIS USING THE CEPSTRUM OF RESPONSE SIGNALS

Where the log spectrum of the excitation is relatively smooth and flat (e.g., excitation by an impulse), its cepstrum is very short and the higher quefrency part of the cepstrum of the response is completely dominated by the structural function. It is possible to curve fit this part of the cepstrum for the poles and zeros of the transfer function as illustrated in Figure 1 (from [6]).

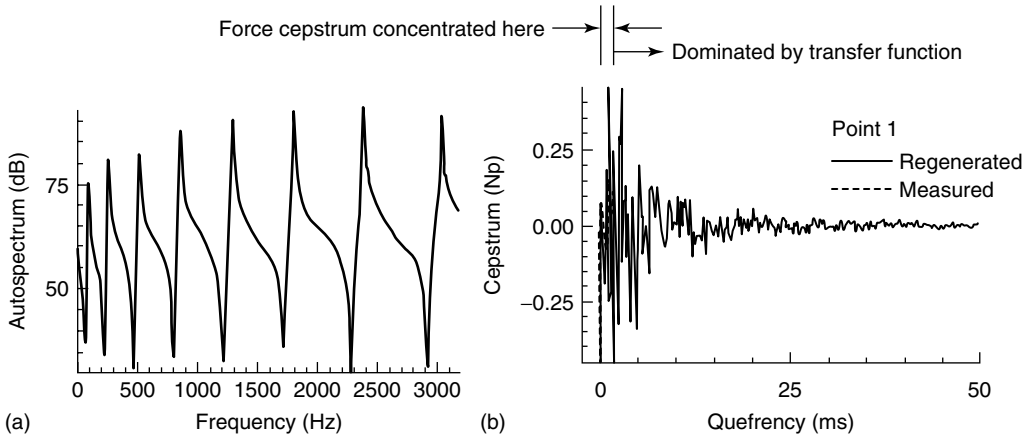


Figure 1. Extraction of the complex cepstrum of a transfer function from the response autospectrum of a beam excited by a hammer blow (a) driving point response autospectrum (b) measured and regenerated cepstra. [Reproduced from Ref. 6. © Elsevier, 1996.]

Figure 1(a) shows the response autospectrum at the driving point (one end) of a free–free beam excited by a hammer blow. Noise in the response is negligible. Figure 1(b) shows the cepstrum derived from this and the cepstrum of the transfer function obtained from it assuming minimum-phase properties. Equations of the form of (7) were curve fitted for the poles and zeros inside the unit circle using a least squares optimization (Levenberg–Marquardt) method [6]. The cepstrum of the transfer function was then regenerated using equation (7). It is seen that the regenerated and measured cepstra correspond almost perfectly at the higher quefrencies dominated by the transfer function. Equally good results were obtained in [6] by curve fitting the differential cepstrum, using both the Levenberg–Marquardt method and the ITD (Ibrahim time domain) method originally developed for extracting impulse responses from free decay vibration signals.

When an attempt was made to regenerate the frequency response functions (FRFs) from the curve-fitted poles and zeros, it was realized that there are two pieces of information that are lacking to fully restore them:

1. an overall scaling factor that is part of the zero quefrency value of the cepstrum, which is not curve fitted and
2. an equalization curve due to the lack of residual information from out-of-band modes that are excluded from the measured cepstrum.

In [7], it is described how this information can be obtained by other means from earlier or similar FRFs, for example, a finite element (FE) model of the structure. Neither of the two above factors is sensitive to small changes in the exact pole and zero positions, and so it is possible to track slow changes in the dynamic properties (such as from a developing crack) or update the FE model to have the correct (measured) in-band poles and zeros.

Figure 2 illustrates the results of both types of model updating on the free–free beam excited by hammer blows [7]. Since these FRFs have minimum-phase properties, the autospectra were used for generating the cepstra. The particular result shown is for excitation at one end and response measured near the middle, giving a number of zeros approximately half the number of poles. In [7], it is shown that the amount of correction required by the equalization function depends on the imbalance of zeros and poles in the (truncated) out-of-band region. For a driving point measurement, with the same number of poles and zeros, the correction required is minimal, since the poles and zeros can be arranged in pairs that almost balance each other in terms of amplitude and phase correction.

The advantage of the cepstral method of regenerating FRFs, compared with other methods from response measurements only, is that the zeros contain much of the information in the residues of the FRF, and when combined with the scaling and equalization

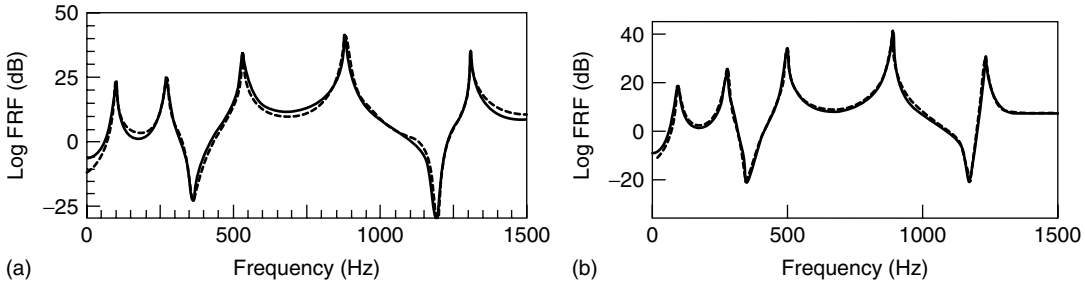


Figure 2. Use of the cepstrum to update FRFs from response measurements of a free–free beam excited by hammer blows (dashed line—measured and solid line—regenerated). [Reproduced from Ref. 7. © Elsevier, 1996.] (a) Updated FE model from cepstrum of response autospectrum and (b) model further updated to track the effects of a slot in the middle of the beam affecting the symmetric modes. Note the reduction in frequency of the odd-numbered modes.

information give the possibility of obtaining fully scaled modes that can be used for predicting forced response. Figure 2(a) was obtained by combining the in-band poles and zeros extracted from the measured autospectra with equalization and scaling information obtained from an approximate FE model of the beam as discussed below. Figure 2(b) also illustrates that these methods can be used to track changes due to increasing damage (in this case a deepening slot), and the same scaling and equalization information could be used for this further development, illustrating its insensitivity to the detailed positions of the poles and zeros. As discussed in the next section, it is possible to extract scaled mode shapes with accurate natural frequencies, and changes in either or both could be used to track the development of the "crack". This indicates a potential application to structural health monitoring.

Figure 3 (from [7]) illustrates the changes in natural frequencies due to the increasing slot, which were tracked by the cepstral method.

3.1 Obtaining scaled modes from response measurements

In [7], the equalization was done using "phantom zeros" [8], extra in-band zeros added to a transfer function to compensate for missing out-of-band poles and zeros, as commonly done in modal analysis curve-fitting software [9] to obtain a match between measured data and theoretical models. The overall scaling was done on the basis of low-frequency values, which can be inferred from static inertial properties for free–free systems or static stiffness properties for restrained systems. Later, it was realized that both corrections could be obtained by the simple expedient of comparing the amplitudes of the regenerated FRFs (on log scales) using only the in-band poles and zeros, with close estimates of the true FRF based on a similar measurement or an approximate FE model. The dB difference curve, smoothed by "short-pass liftering" to remove high-quefrency errors gives a good estimate of the scaled equalization curve. Figure 4 gives an example from [10] showing an equalization curve generated in this way, together with the two curves from which it was derived and the corrected estimate. It will be noted that the equalization curve in that example has some ripple, which represents the major source of error in the final estimate, but it should be possible to reduce this by using an improved smoothing, as well as adding an estimated "mass line" (from static properties) to the FRF calculated from the in-band poles and zeros.

Note that the compensation given by the equalization curve is purely for out-of-band modes and is virtually independent of the color of the excitation as long as the spectrum of the latter is reasonably smooth and flat. This obviates the requirement of many operational modal analysis techniques where the excitation is assumed white.

The procedure proposed in [10] for obtaining scaled mode shapes is illustrated in Figure 5.

The FE model to be used for equalization and scaling should first be updated using resonance (pole) information from any operational modal analysis (OMA) technique, including the cepstral techniques

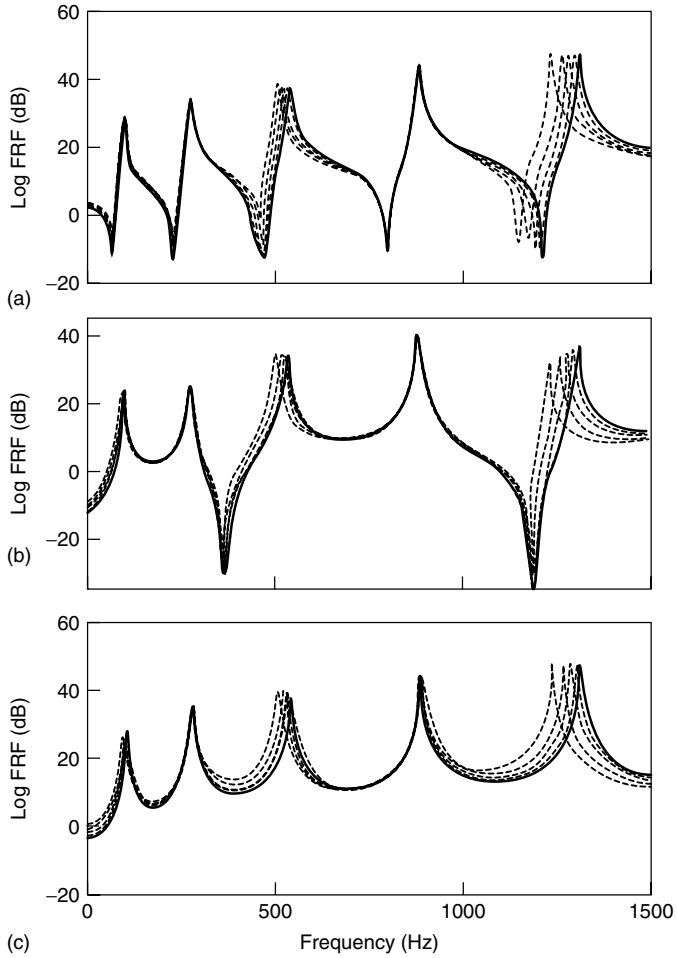


Figure 3. Changes in the FRFs between one end of a free–free beam and (a) the same point, (b) near the middle, and (c) the other end due to a slot cut progressively in the middle of the beam to a final depth of half the thickness.

proposed here. Note that most FE model updating techniques (e.g., [11]) use only resonance information. The resulting FRFs should be more accurate than the original estimates at low frequencies, which is important for the scaling. For free–free objects, the "mass" matrix should be reasonably well estimated even before updating, but for constrained objects the original stiffness matrix will usually suffer from poor estimates of joint stiffness, etc., and a good fit to the lowest modes will go a long way toward improving the stiffness matrix and the estimates of static stiffness that are important for scaling.

Otherwise, the updated model does not have to be perfect to obtain good results. Figure 6 shows the example from [10] where after updating there are still discrepancies at high frequencies, and in the estimated damping.

The equalization curves from the updated FE model are used to obtain scaled FRFs using the (in-band) pole and zero information from the cepstral curve-fitting method. Figure 7 gives two examples of the corrected FRFs from [10] using this approach. There are two main errors evident in these results:

1. Local amplitude deviations coming from the ripple referred to above, the reduction of which should be possible.

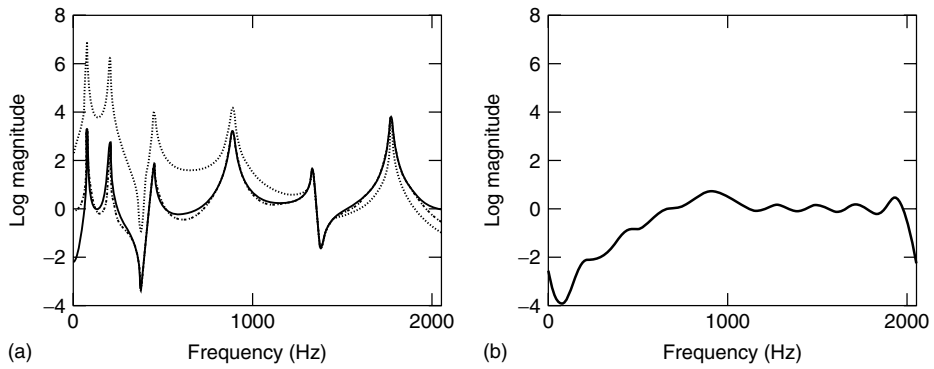


Figure 4. Typical FRFs (a) reference (solid line), regenerated (· · · ·) and scaled (– · – · –), and the corresponding equalization curve (b). [Reproduced from Ref. 10. © Elsevier, 2007.]

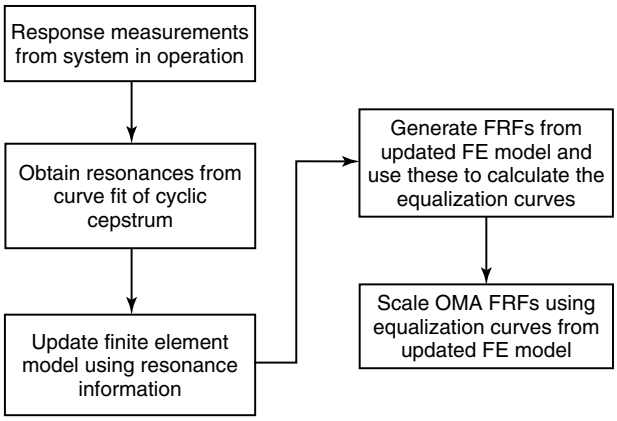


Figure 5. Procedure for obtaining scaled mode shapes from FE model updating. [Reproduced from Ref. 10. © Elsevier, 2007.]

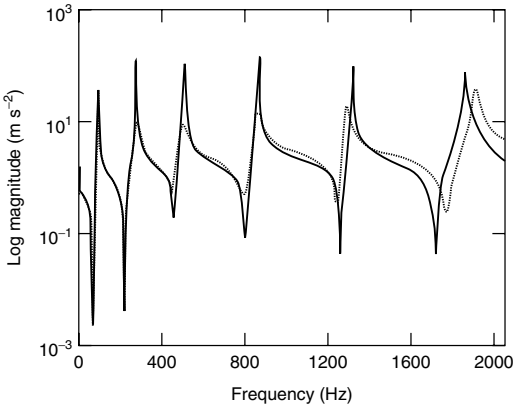


Figure 6. Comparison of the driving point FRF from measurement (solid line) and from the updated finite element model (dashed line). [Reproduced from Ref. 10. © Elsevier, 2007.]

2. Differences in frequency of the modes, most evident at the higher frequencies. These are purely the result of the fact that the measurement object in this case was a steel beam excited by two small shakers, so that the identified resonances were those of the beam plus shaker moving elements, somewhat lower than those of the beam alone. This is a problem common to any OMA technique, but exacerbated in many laboratory experiments as opposed to real practical situations where the actual excitation does not give added mass.

Despite these minor errors, the resulting mode shapes compared very well with the true measured ones, including with respect to scaling.

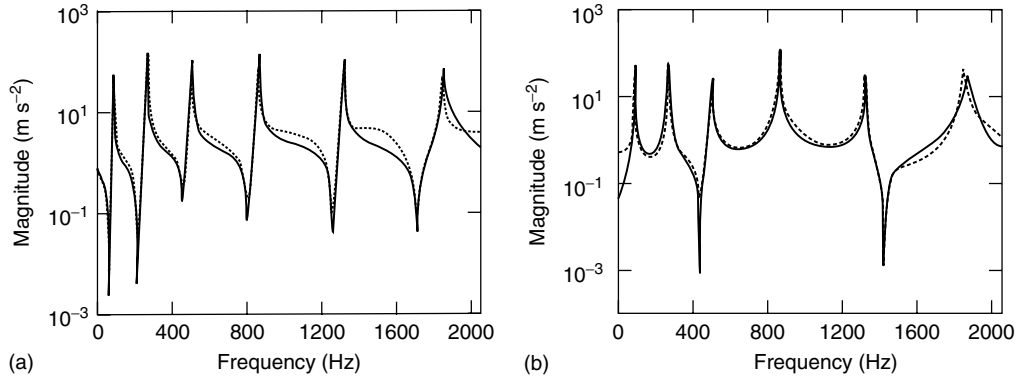


Figure 7. Comparison of correctly scaled FRFs from OMA (dashed line) and measured (solid line); driving point (a) and typical transfer measurement (b). [Reproduced from Ref. 10. © Elsevier, 2007.]

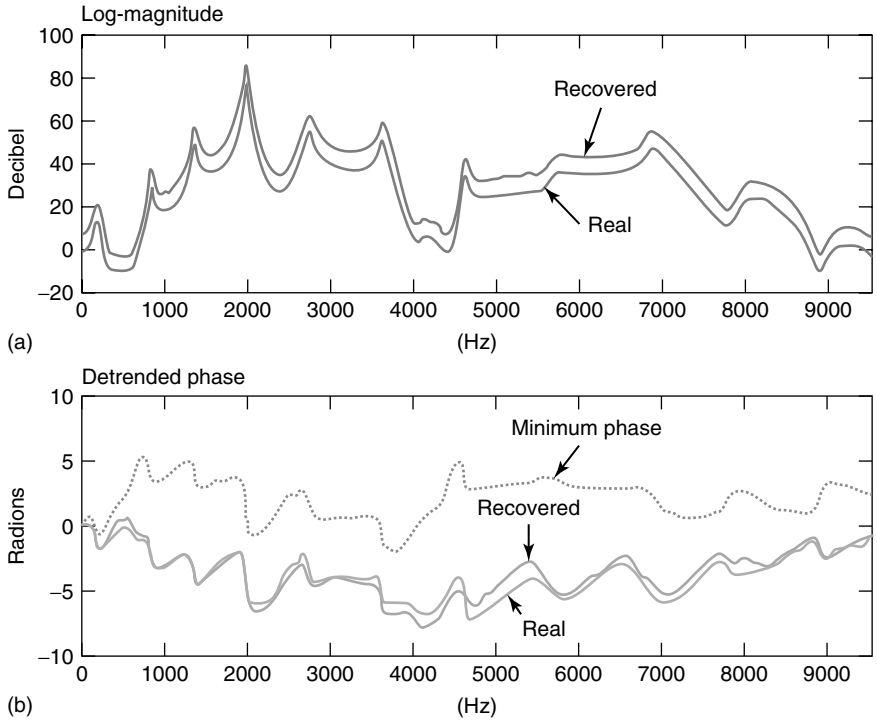


Figure 8. Nonminimum phase FRF recovered using the MDC. [Reproduced with permission from Ref. 5. © KUL, 2000.]

3.2 The mean differential cepstrum

Antoni *et al.* in [5] have shown that mixed-phase FRFs can be obtained using the MDC even in the presence of noise in the signals. A typical result is shown in Figure 8.

4 MIMO SITUATION

In the above cases, there was only one effective excitation force applied in one degree of freedom (DOF). In most cases, in practice, there are several sources acting simultaneously, which gives a MIMO (multiple

input, multiple output) situation. In that case, the response at any point is a sum of convolutions, so that equation (6) no longer applies. It is necessary to separate out the response(s) to a single source applied in a single DOF. Several ways of achieving this have been attempted.

4.1 Periodic impulsive excitation

Since the excitation must have a flat spectrum, it must either be impulsive or broadband random. If a periodic impulse can be applied in one DOF, the response to it can be extracted by synchronous averaging. Experiments were made on a free–free beam with two excitations, one periodic impulsive and the other broadband random, applied by small shakers [12]. The intended application was to the determination of the modal properties of a rail vehicle, by mounting one wheel with a flat.

Figure 9(a) shows a typical response spectrum, where the apparently shaded area is actually separate harmonic lines of the periodic part, with a separation of 1 Hz. The continuous part of the spectrum at the base is dominated by the random part. Figure 9(b) shows the spectrum of the periodic part (one period only, so that the spectrum appears continuous), which was extracted by synchronous averaging. Figure 9(c) shows the averaged power spectrum of the residual random signal. Each individual spectrum in the average has four times better resolution than that of the periodic part, as does the averaged spectrum.

Since the spectrum of the periodic part has limited resolution, determined by the fundamental frequency, it was proposed that it could be used to determine the residues (and/or zeros) of the SIMO FRF representing the response to the periodic excitation only, and possibly the frequencies of the poles, but, in particular, the accuracy of the damping would be limited by the resolution.

On the other hand, since the frequencies and damping of the poles are global values, they can be determined with more accuracy from the broadband part of the spectrum, even if they correspond to MIMO or distributed excitation. In [13], a number of methods are compared for determining the frequency and damping of the poles from the broadband residual spectrum, after the poles and zeros (and thus implicitly the residues) are obtained using the cepstrum on the periodic part.

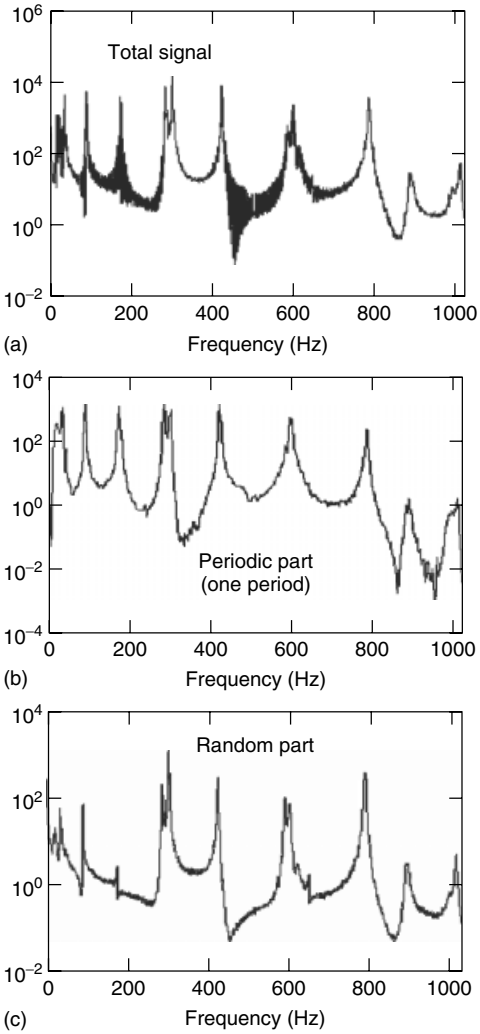


Figure 9. Response spectra with periodic impulsive and random excitation [12]. (a) Total signal (b) spectrum of periodic part (one period) (c) average spectrum of the random part after the periodic part has been subtracted. [Reproduced with permission from Ref. 5. © KUL, 2002.]

4.2 Cyclostationary input

An *n*th-order cyclostationary signal is defined as one whose *n*th-order statistics are periodic. Thus, a first-order cyclostationary signal has a periodic mean

value (e.g., a periodic signal plus noise), while a second-order cyclostationary signal has a periodic variance (e.g., a white noise modulated by a sinusoid). Many practical signals have cyclostationary properties, and if only one of the excitations in an MIMO situation has second-order cyclostationary properties at a particular cyclic frequency, it is possible to separate out the response to it at each measurement point, thus reducing the problem to SIMO, and allowing the use of cepstral techniques to separate the forcing and transfer functions.

The so-called spectral correlation function allows signals to be separated into periodic, stationary random, and cyclostationary components. It can be obtained from the generalized autocorrelation function defined as

$$R_{yy}(t, \tau) = E[y(t - \tau/2)y(t + \tau/2)] \quad (15)$$

by a two-dimensional Fourier transform with respect to lag τ (into normal frequency f) and time t (into cyclic frequency α). Thus, the Fourier transform in the τ direction gives continuous spectra from the transformation of a transient, while the Fourier transform in the t direction (actually a Fourier series) gives discrete components at the harmonics of the cyclic frequency because of the periodicity in this direction. Note that periodic (or first-order cyclostationary) signals have a spectral correlation function, which is discrete in both directions (a "bed of nails"), but if the mean value is first subtracted only second-order components are left. These can be separated from stationary random components, which only appear at zero cyclic frequency because they do not vary with time, by definition.

The spectral correlation can also be directly calculated in the frequency domain using the formula

$$S_y^\alpha(f) = \lim_{W\to\infty} \frac{1}{W} \mathrm{E}\left\{ Y_W\left(f + \frac{\alpha}{2}\right) Y_W^*\left(f - \frac{\alpha}{2}\right)\right\} \quad (16)$$

where second-order cyclostationarity is assumed at cyclic frequency α. Note that at $\alpha = 0$ the normal autospectrum is obtained.

It can be shown as in [14] that the structural dynamic properties are contained in the spectral correlation of a response signal $y(t)$ at cyclic frequency α. For minimum-phase systems, it is convenient to use the asymmetrical version of equation (16)

$$S_y^\alpha(f) = \lim_{W\to\infty} \frac{1}{W} \mathrm{E}\{Y_W(f) Y_W^*(f - \alpha)\} \quad (17)$$

Substituting the relationship $Y(f) = H(f)X(f)$ into equation (17) gives

$$S_y^\alpha(f) = H(f)H^*(f - \alpha)S_x^\alpha(f) \quad (18)$$

Taking the log and inverse Fourier transform to obtain the cepstrum

$$C_y^\alpha(\tau) = C_h(\tau) + C_h(-\tau)\ \mathrm{e}^{j2\pi\alpha\tau} + C_x^\alpha(\tau) \quad (19)$$

meaning that the response cepstrum at cyclic frequency α contains the cepstrum of the transfer function and only the input at cyclic frequency α.

Moreover, for minimum-phase systems the cepstrum is zero for negative quefrency and for broadband excitation functions has a short cepstrum equal to zero for $|\tau| > \tau_0$. Thus

$$C_y^\alpha(\tau) = C_h(\tau), \quad \tau > \tau_0 \quad (20)$$

In [14], it is shown that a somewhat more complex procedure can be used for nonminimum phase systems, where both the positive and negative quefrency parts of the cyclic cepstrum are used. The symmetric definition of the spectral correlation (equation 16) can then be used, where the cyclic cepstrum will be phase shifted for both positive and negative quefrencies.

The minimum-phase approach was applied in [14] to vehicles driven by internal combustion engines, such as diesel railcars, a result from which is presented in Figure 10, although the excitation in that case was a burst random signal applied by a shaker. Figure 10 shows the spectral correlation (cyclic spectrum) for a typical response measurement.

The spectral correlation at cyclic frequency $\alpha = 1/T$ was analyzed using the cyclic cepstrum as one of a number of SIMO responses, and gave scaled mode shapes very similar to those measured directly (see Figure 11 for an example).

Many naturally occurring signals are cyclostationary rather than periodic, so it is possible that this

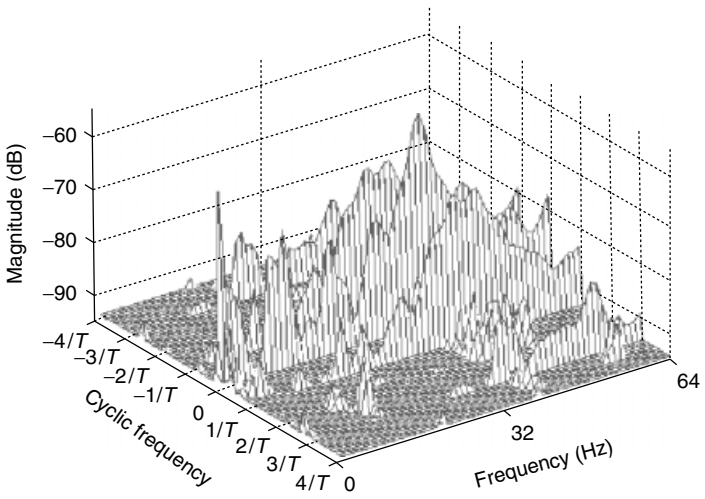


Figure 10. Cyclic spectrum of a response measurement from a passenger rail vehicle. [Reproduced from Ref. 14. © Elsevier, 2007.].

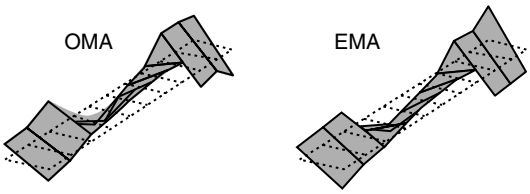


Figure 11. Comparison of mode shapes from cyclic cepstrum analysis (OMA) and direct experimental modal analysis (EMA).

technique might be applicable to OMA of structures excited by such signals.

4.3 Blind source separation

In recent years, there has been a steady development in the topic of blind source separation (BSS), initially with applications in telecommunications, but later increasingly with applications in mechanical problems [15, 16]. The idea is to separate out the responses to a number of sources at a number of measurement points, with no direct knowledge of the sources or transmission paths, but based solely on the assumption that the different sources are statistically independent.

In most mechanical problems, there is a variable transmission path from each source to each measurement point, and the mixing is convolutive, as represented by the equation:

$$\mathbf{Y}(t_n) = \mathbf{h} * \mathbf{X}(t_n) \quad \text{where} \quad \mathbf{h} = [h_{ij}]_{N\times M} \tag{21}$$

the matrix of impulse response elements h_{ij} as a mixing system convolved with the source signals.

The (Fourier transformed) FRF matrix $\mathbf{h}$ can be decomposed by SVD (singular value decomposition) as

$$\mathbf{h} = \mathbf{UDV} \tag{22}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary (rotational) matrices, and $\mathbf{D}$ is a diagonal (scaling) matrix. $\mathbf{U}$ and $\mathbf{V}$ have the property that $\mathbf{UU}^{\mathrm{H}} = \mathbf{VV}^{\mathrm{H}} = \mathbf{I}$, the identity matrix (superscript H means Hermitian transpose).

From the response vector $\mathbf{Y}$, it is possible to form the cross spectral matrix:

$$\mathbf{S} = \mathrm{E}[\mathbf{Y}\ \mathbf{Y}^{\mathrm{H}}] \tag{23}$$

and substituting equation (22) in the frequency domain (multiplicative) gives

$$\begin{aligned}\mathbf{S} &= \mathrm{E}[\mathbf{UDVX}\ \mathbf{X}^{\mathrm{H}}\mathbf{V}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}] \\ &= \mathbf{UDV}\ \mathrm{E}[\mathbf{XX}^{\mathrm{H}}]\mathbf{V}^{\mathrm{H}}\mathbf{D}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}} \\ &= \mathbf{UD}^2\mathbf{U}^{\mathrm{H}}\end{aligned} \tag{24}$$

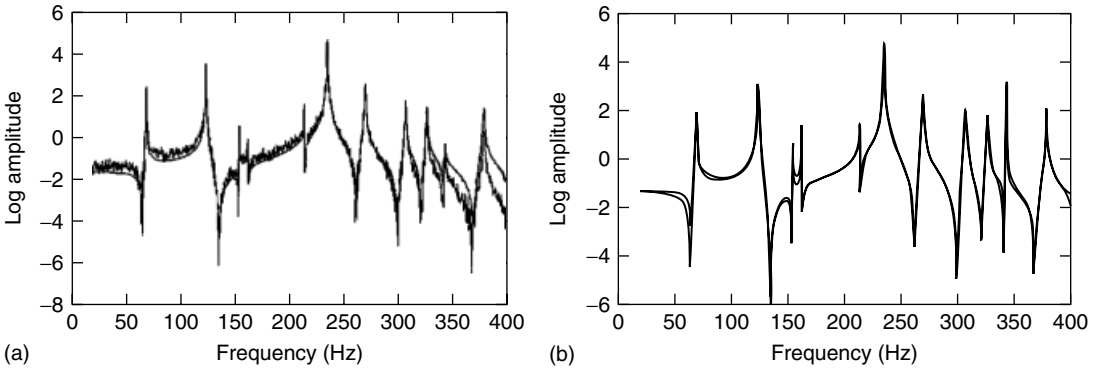


Figure 12. Reconstruction of an FRF using SVD and cepstral curve fitting. [Reproduced with permission from Ref. 17. © KUL, 1998.] (a) Comparison of first principal autospectrum with the corresponding FRF and (b) comparison of measured and regenerated FRFs.

since the source vector **X** has independent components and without loss of generality can be taken to be spatially white (any differences in gain can be assigned to the transfer functions). Thus, **U** and **D** can be found by SVD, but not the further rotational matrix **V** required to determine **h** completely.

4.3.1 Separation by SVD alone

In [17], it was found purely empirically that if one of the sources was dominant, i.e., with rms value greater than the others by a factor of at least 4, then SVD alone gave a reasonable separation, at least in terms of the zeros of the autospectrum of the first principal component obtained by SVD. Figure 12(a) compares the extracted principal autospectrum with the FRF between the dominant forcing function and the measurement point corresponding to the autospectrum. In this case, the measurement object was a steel frame excited by two broadband sources in different points, one four times stronger than the other. It is seen that the poles and zeros coincide, even though the forcing spectrum was not completely white, but rolling off by about 15 dB toward the higher frequencies. Figure 12(b) shows that the FRF reconstructed by the same method as Figure 2 corresponds very well with that measured directly, using the forces that were measured at the same time as the responses.

Note that the cepstral curve fitting was able to determine accurate pole and zero values (by least squares optimization) even in the presence of noise in the autospectrum caused by the limited number of averages. Moreover, because the reconstructed FRFs were scaled, it was possible to perform a modal analysis with scaled modes.

4.3.2 Separation by other means

Currently, a number of studies are under way to effect true blind separation using other means. It is necessary to use further information to determine the unknown rotational matrix **V**. One possibility discussed in [15] is to use higher order statistical (HOS) properties (such as different kurtosis) to effect the separation, but this limits the sources (except possibly one) to be non-Gaussian since the HOS properties of Gaussian signals are zero (*see* **Higher Order Statistical Signal Processing**). A problem with many of the methods that have been developed to date (for example, for applications in telecommunications) is that the number of sources must be known in advance, and each must be independent and applied in a single DOF. Many mechanical excitations are distributed and/or correlated at different excitation points, so the best potential appears to be the methods that separate out the response to a single source (among an unknown number), which is independent of the others and applied in a single DOF. Note that the convolutive mixing of equation (21) becomes multiplicative in the frequency domain, giving the possibility to use methods developed for the case of simple multiplicative mixtures, e.g., independent components analysis (ICA). However, these suffer from three problems:

1. An unknown gain factor between the components. This corresponds to an unknown filter in the convolutive mixture, but is not a big problem where only the responses to each source are to be separated and not the sources themselves. After separation to a SIMO situation, the cepstral methods can determine the filter functions.
2. An unknown permutation of the separated components at each frequency, which must be resolved before all frequency components can be combined to form time signals. A method is proposed in [18].
3. The fact that frequency analysis into narrow bands tends to make any signal more Gaussian and thus more difficult to be separated by HOS techniques. This is influenced by the resolution of the practical fast Fourier transform (FFT) analysis, and the spectral kurtosis can be used as a guide to the choice of resolution [18].

4.4 MIMO version of the mean differential cepstrum

In [5], Antoni proposed a MIMO version of the MDC, without developing it further. By analogy with equation (14) a matrix version of it is proposed, which does not use the logarithm explicitly. The basic derivation is as follows:

$$\begin{aligned}\mathbf{D}_{yy} &= \mathrm{E}\left(\mathbf{Y}'\mathbf{Y}^{\mathrm{H}}\right) \quad \mathrm{E}\left(\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{-1} \\ &= \mathrm{E}\left(\left[\mathbf{h}'\mathbf{X} + \mathbf{h}\,\mathbf{X}'\right]\left[\mathbf{X}^{\mathrm{H}}\mathbf{h}^{\mathrm{H}}\right]\right) \\ &\quad \times \mathrm{E}\left([\mathbf{h}\ \mathbf{X}]\left[\mathbf{X}^{\mathrm{H}}\mathbf{h}^{\mathrm{H}}\right]\right)^{-1} \\ &= \Big[\mathbf{h}'\underbrace{\mathrm{E}\left(\mathbf{X}\mathbf{X}^{\mathrm{H}}\right)}_{\mathbf{I}}\mathbf{h}^{\mathrm{H}} + \mathbf{h}\underbrace{\mathrm{E}\left(\mathbf{X}'\mathbf{X}^{\mathrm{H}}\right)}_{\left[\ddots jC\right]}\mathbf{h}^{\mathrm{H}}\Big] \\ &\quad \times \Big[\mathbf{h}\underbrace{\mathrm{E}\left(\mathbf{X}\mathbf{X}^{\mathrm{H}}\right)}_{\mathbf{I}}\mathbf{h}^{\mathrm{H}}\Big]^{-1}\end{aligned}$$

$$\mathbf{D}_{yy} = \left[\mathbf{h}'\mathbf{h}^{\mathrm{H}} + \mathbf{h}\left[\ddots jC\right]\mathbf{h}^{\mathrm{H}}\right] \quad \left[\mathbf{h}\mathbf{h}^{\mathrm{H}}\right]^{-1}$$

$$\mathbf{D}_{yy}\left[\mathbf{h}\mathbf{h}^{\mathrm{H}}\right] = \mathbf{h}'\mathbf{h}^{\mathrm{H}} + \mathbf{h}\left[\ddots jC\right]\mathbf{h}^{\mathrm{H}}$$

$$\mathbf{D}_{yy}\mathbf{h} = \mathbf{h}' + \mathbf{h}\left[\ddots jC\right]$$

$$\mathbf{h}' - \mathbf{D}_{yy}\mathbf{h} = 0 \tag{25}$$

The same assumptions are made as for equation (24) and, in addition, the purely imaginary time shift constants $\left[\ddots jC\right]$ can be arbitrarily set to zero because response signals have no zero time reference.

This is seen to result in a matrix differential equation (DE) in $\mathbf{h}$ for which there is no closed-form solution, but can be solved numerically by finite difference methods. A recent PhD thesis [19] has examined this possibility in some depth, but problems still remain. In [19], the matrix DE is developed further to give a similar matrix DE in terms of the unknown unitary matrix $\mathbf{V}$, which is somewhat more well behaved than $\mathbf{h}$, and thus easier to solve for by numerical means. Since the solution of the DE is propagative from one frequency to the next (starting from known static values near zero frequency) the permutation problem is resolved. One result from [19] is that if the matrix method is collapsed to a scalar problem, the propagative method gives an alternative way of producing the MDC, which may be found to have advantages in some situations.

Despite suffering from noise problems at this stage, these matrix methods are presented because they may be further developed in the future.

5 CONCLUSION

The cepstrum has valuable properties for determining modal properties from response measurements, since in the SIMO case the contributions of the forcing and transfer functions are additive. In addition, if the force spectrum is smooth and flat (impulsive or broadband random), its cepstrum is short, and the remainder of the response cepstrum can be curve fitted for the poles and zeros of the transfer function. With some a priori information, this gives the possibility of obtaining scaled modal properties, which is very difficult with other methods. Since operational measurements normally represent an MIMO situation, various methods are discussed to effect BSS, so as to convert the MIMO situation to a sum of SIMOs (or at least extract one SIMO system of interest) and thus permit the cepstral methods to be used more generally.

While not yet specifically applied to structural health monitoring, the successful tracking of the development of a slot in a beam (simulating a crack)

indicates the potential of these methods for this purpose.

REFERENCES

[1] Bogert BP, Healy MJR, Tukey JW. The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking. In *Proceedings of the Symposium on Time Series Analysis*, Rosenblatt M (ed). John Wiley & Sons: New York, 1963, pp. 209–243.

[2] Childers DG, Skinner DP, Kemerait RC. The cepstrum: a guide to processing. *Proceedings of the IEEE* 1977 **65**(10):1428–1443.

[3] Oppenheim AV, Schafer RW. *Discrete Time Signal Processing*. Prentice-Hall: New Jersey, 1989.

[4] Polydoros A, Fam AT. The differential cepstrum: definitions and properties. *Proceedings of the IEEE International Symposium on Circuits Systems* 1981: 77–80.

[5] Antoni J, Guillet F, Danière J. Identification of non-minimum phase transfer functions from output-only measurements. *ISMA25 Conference*. KUL: Leuven, 2000.

[6] Gao Y, Randall RB. Determination of frequency response functions from response measurements. Part I: extraction of poles and zeros from response cepstra. *Mechanical Systems and Signal Processing* 1996 **10**(3):293–317.

[7] Gao Y, Randall RB. Determination of frequency response functions from response measurements. Part II: regeneration of frequency response functions from poles and zeros. *Mechanical Systems and Signal Processing* 1996 **10**(3):319–340.

[8] Randall RB, Gao Y, Sestieri A. Phantom zeros in curve-fitted frequency response functions. *Mechanical Systems and Signal Processing* 1994 **8**:607–622.

[9] Richardson MH, Formenti DF. Global curve-fitting of frequency response measurements using the rational fraction polynomial method. *Proceedings of the 3rd IMAC*. Orlando, 1985; pp. 390–397.

[10] Hanson D, Randall RB, Antoni J, Thompson DJ, Waters TP, Ford RAJ. Cyclostationarity and the cepstrum for operational modal analysis of mimo systems—part II: obtaining scaled mode shapes through finite element model updating. *Mechanical Systems and Signal Processing* 2007 **21**(6): 2459–2473.

[11] FEMTools. *Model Updating Software*. Dynamic Design Solutions: Leuven 2007.

[12] Ford R, Randall R, Wardrop T. *Updating modal properties from response-only measurements on a rail vehicle*. ISMA2002. KUL: Leuven, 2002.

[13] Randall RB, Zurita G, Wardrop T. A comparative study of curve-fitting methods for the extraction of modal parameters from response measurements. *ICSV10*. Stockholm, 2003.

[14] Hanson D, Randall RB, Antoni J, Thompson DJ, Waters TP, Ford RAJ. Cyclostationarity and the cepstrum for operational modal analysis of mimo systems—part I: modal parameter identification. *Mechanical Systems and Signal Processing* 2007 **21**(6):2441–2458.

[15] Haykin S. *Unsupervised Adaptive Filtering Volume I: Blind Source Separation*. John Wiley & Sons: New York, 2000.

[16] Antoni J, Braun S (eds). Special issue: blind source separation. *Mechanical Systems and Signal Processing* 2005 **19**(6):1163–1380.

[17] Randall RB, Gao Y, Swevers J. Updating modal models from response measurements. *ISMA Conference*. KUL: Leuven, 1998; pp. 1153–1160.

[18] Capdevielle V, Servière C, Lacoume JL. Blind separation of wide-band sources in the frequency domain. *Proceedings of ICASSP 1995*. Detroit, MI, 1995; pp. 2080–2083.

[19] Chia WL. *Multiple-Input Multiple-Output Blind System Identification for Operational Modal Analysis using the Mean Differential Cepstrum*, Ph.D. Dissertation. University of New South Wales: Sydney, 2007; p. 2052. Electronic copy available through UNSW Library.

Chapter 25

Hilbert Transform, Envelope, Instantaneous Phase, and Frequency

Michael Feldman

Technion–Israel Institute of Technology, Haifa, Israel

1 INTRODUCTION

The Hilbert transform (HT), as a kind of integral transformation, plays a significant role in signal processing, vibration analysis, dynamic system monitoring, diagnostics, and identification. The HT was first introduced to signal theory by Denis Gabor in 1946. He defined a generalization of the well-known Euler formula $e^{iz} = \cos(z) + i\sin(z)$ in the form of the complex function $Y(t) = u(t) + iv(t)$, where $v(t)$ is the HT of $u(t)$ [1]. In signal processing, when the independent variable is time, this associated complex function is known as the *analytic signal* and the HT $v(t)$ is called *quadrature* (or *conjugate*) *function* of $u(t)$. The application of the HT to the initial signal provides some additional information about amplitude, instantaneous phase, and frequency of vibrations. This information can be useful when applied to analysis of vibration motions, including a decomposition of the signal into its simplest components and an inverse problem of vibration system identification.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 NOTATION

The HT of the function $x(t)$ is defined by an integral transform [1]:

$$H[x(t)] = \tilde{x}(t) = \pi^{-1} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} \, d\tau \qquad (1)$$

Because of the possible singularity at $t = \tau$, the integral is to be considered as a Cauchy principal value. The HT is equivalent to a special kind of filter, in which all the amplitudes of the spectral components are left unchanged, but their phases are shifted by $-\pi/2$ (Figure 1).

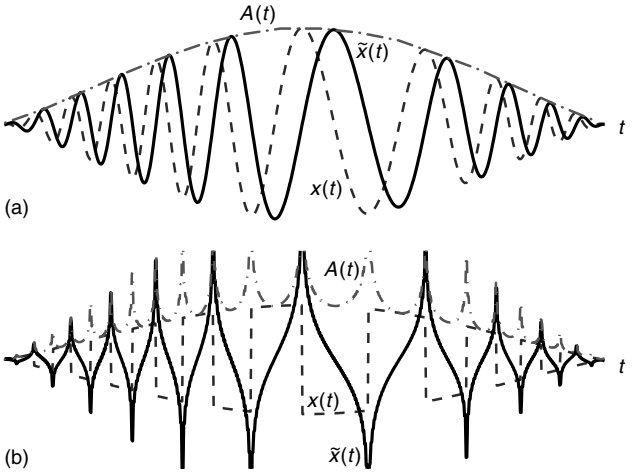


Figure 1. The initial signal $x(t)$, the Hilbert transform $\tilde{x}(t)$, and the envelope $A(t)$: quasi-harmonics (a); nonstationary square wave (b).

3 PROPERTIES OF THE HT

The HT of a real-valued function $x(t)$ extending from $-\infty$ to $+\infty$ is a real-valued function defined by equation (1). Thus, $\tilde{x}(t)$ is the convolution integral of $x(t)$ with $(\pi t)^{-1}$, written as $\tilde{x}(t) = x(t) * (\pi t)^{-1}$. The HT of a constant is zero. The double HT (the HT of a HT) yields the original function having the opposite sign, and hence it carries out a shifting of the initial signal by $-\pi$. The HT used four times on the same real function gives us the original function back. The power (or energy) of a signal and its HT are equal. A function and its HT are orthogonal over the infinite interval $\int_{-\infty}^{\infty} x(t)\tilde{x}(t)\,\mathrm{d}t = 0$. The HT of the derivative of a function is equivalent to the derivative of the HT of the function.

For $n(t)$ low-pass and $x(t)$ high-pass signals with nonoverlapping spectra, $H[n(t)x(t)] = n(t)\tilde{x}(t)$. More generally, the HT of the product of two arbitrary functions with overlapping spectra can be written in the form of a sum of two parts [1]: $H\{n(t)x(t)\} = H\{[\overline{n}(t) + n_1(t)]x(t)\} = \overline{n}(t)\tilde{x}(t) + \tilde{n}_1(t)x(t)$, where $\overline{n}(t)$ is the slow (low-pass) signal component, $n_1(t)$ is the fast (high-pass) signal component, and $\tilde{n}_1(t)$ is the HT of the fast component $n_1(t)$. The proof of the decomposition of a signal into a sum of low- and high-pass terms is based on Bedrosian's theorem for the HT of a product. For example, the HT of the cube of the harmonics $x^3 = (A\cos\varphi)^3$ is equal to $H[x^3(t)] = H[x^2(t)x(t)] = A^3(3\sin\varphi + \sin 3\varphi)/4$.

4 DIGITAL HILBERT TRANSFORMERS

We can form the HT by designing a filter with the frequency response corresponding to the impulse response $(\pi t)^{-1}$. This response is $H(\omega) = -i\,\mathrm{sgn}(\omega)$. This frequency response describes an ideal wideband 90° phase shifter, the positive frequencies of which are shifted by -90° $(-\pi/2)$ and negative frequencies by $+90^\circ$ $(\pi/2)$. There are mainly two methods for obtaining a real approximate Hilbert transformer: frequency domain and time domain.

4.1 Frequency domain

This technique is based on computing the Fourier transform of a signal. If $x(t)$ is the real input data record of length N, then the analytic signal $X(t) = x(t) + i\tilde{x}(t)$ can be obtained by: $X(t) = \mathrm{IFFT}\{B(n) \cdot \mathrm{FFT}[x(t)]\}$, where $B(n) = 2$ for $n = \{0, (N/(2-1))\}$ and $B(n) = 0$ for $n = \{N/2, (N-1)\}$. FFT is the fast Fourier transform and IFFT is the inverse fast Fourier transform. The imaginary portion of $X(t)$ contains the HT projection $\tilde{x}(t)$ and the real portion contains the real input signal $x(t)$.

Windowing and/or zero padding may have to be used to avoid ringing.

4.2 Time domain

The phase shift can be implemented by a convolution filter, obtainable by the convolution theorem relating convolution to multiplication in the frequency domain. The finite impulse response (FIR) digital Hilbert transformers (Figure 2) have the advantage of providing exactly linear phase, but at the expense of requiring a higher filter order as compared to the infinite impulse response (IIR) designs, to achieve a given negative frequency attenuation level.

5 HT AND DYNAMIC SYSTEMS

5.1 Kramers–Kronig relations

For a causal function, the HT presents interesting mathematical properties named the *Kramers–Kronig relations* [2]. The properties connect the real and imaginary parts of any complex analytic function in the upper half plane. We can associate the imaginary part of a complex frequency response function as the HT of the real part of the function. In other words, the real and imaginary parts of the transfer function form a Hilbert pair. The imaginary part of a response function describes how a system dissipates energy, since it is out of phase with the driving force. These relations imply that observing the dissipative response of a system is sufficient to determine its in-phase (reactive) response, and vice versa. Any departure from an initial linear frequency response function, i.e., distortion, can be attributed to nonlinear effects.

5.2 Hysteretic damping

Most real materials show an energy loss per cycle with a less pronounced dependency on frequency. In fact, many materials indicate force–deformation relations that are independent of the deformation rate amplitude—so-called hysteretic relations. A

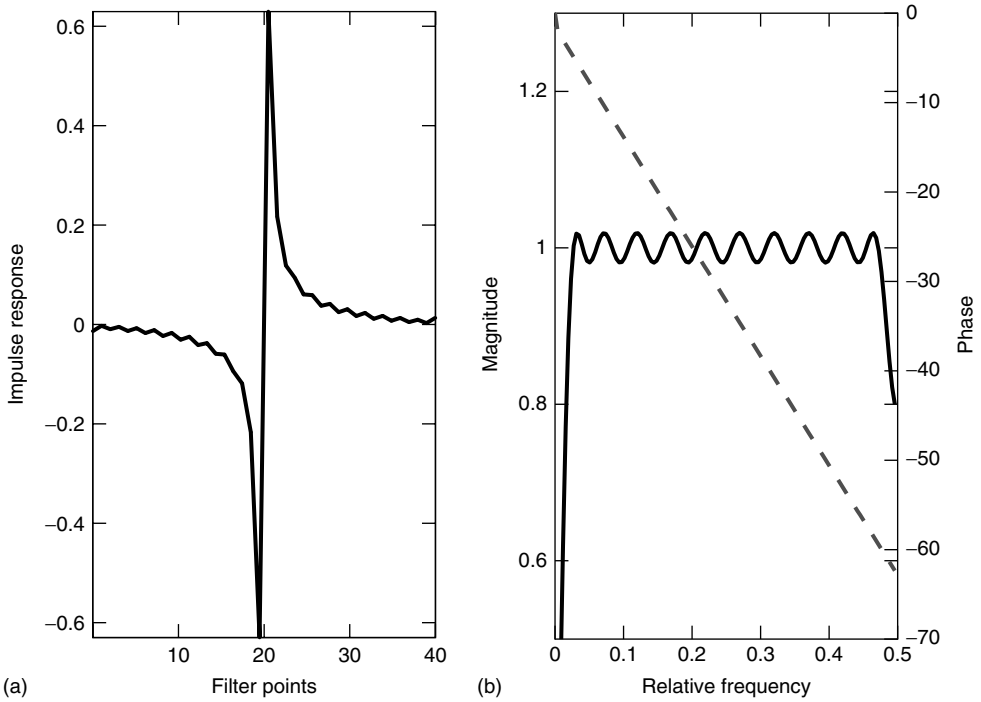


Figure 2. Length 39 Hilbert transformer (digital filter). Impulse response (a) and frequency response function (b)—magnitude (———) and phase (- - - -).

known mechanical element model shows frequency-independent storage and loss moduli. This implies that the Fourier transforms of both the element force $F(\omega)$ and the element deformation $\Delta(\omega)$ satisfy $F(i\omega) = k[1 + i\eta\text{sgn}(\omega)]\Delta(i\omega)$, where k is the stiffness and η is the loss factor (ratio of the loss and storage moduli of element). In this model, the dissipated energy in a harmonic deformation cycle of constant amplitude is independent of the frequency of deformation. Only the HT gives a correct time domain expression for the concept of linear hysteretic damping [3]. This transform can be used to replace complex-valued coefficients in differential equations modeling mechanical elements.

6 ANALYTIC SIGNAL

The complex signal whose imaginary part is the HT (equation 2) of the real part is called the *analytic* or *quadrature signal* [1]. Such a signal is a two-dimensional signal whose value at some instant in time is specified by two parts, the real and the imaginary part: $X(t) = x(t) + i\tilde{x}(t)$, where $\tilde{x}(t)$ is related to $x(t)$ by the HT. To return from a complex form of the analytic signal $X(t)$ back to the real function $x(t)$, one has to use the substitution $x(t) = 0.5[X(t) + X^*(t)]$, where $X^*(t)$ is the complex conjugate signal of $X(t)$. The analytic signal has a one-sided spectrum of positive frequencies. The conjugate analytic signal has a one-sided spectrum of negative frequencies.

6.1 Polar notation

According to analytic signal theory, a real vibration process $x(t)$, measured by, say, a transducer, is only one of possible projections (the real part) of some analytic signal $X(t)$. Then the second or quadrature projection of the same signal (the imaginary part) $\tilde{x}(t)$ will be conjugated according to the HT. The analytic signal has a geometrical representation in the form of a phasor rotating in the complex plane, as shown in Figure 3. Using the traditional representation of the analytic signal in its trigonometric or exponential form

$$\begin{aligned} X(t) &= |X(t)|[\cos\psi(t) + j\sin\psi(t)] \\ &= A(t)\ \mathrm{e}^{j\psi(t)} \end{aligned} \tag{2}$$

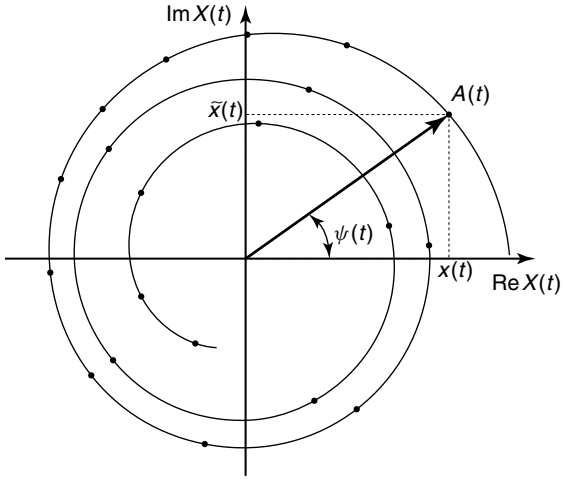


Figure 3. Analytic signal representation.

one can determine its instantaneous amplitude (envelope, magnitude, modulus)

$$A(t) = |X(t)| = \sqrt{x^2(t) + \tilde{x}^2(t)} = \mathrm{e}^{\mathrm{Re}[\ln X(t)]} \tag{3}$$

and its instantaneous phase

$$\psi(t) = \arctan\frac{\tilde{x}(t)}{x(t)} = \mathrm{Im}[\ln X(t)] \tag{4}$$

The change of coordinates from rectangular $(x, \tilde{x})$ to polar (A, ψ) gives $x(t) = A(t)\cos\psi(t)$ and $\tilde{x}(t) = A(t)\sin\psi(t)$.

7 HT AND VIBRATION SIGNALS

The analytic signal method is equally applicable to deterministic and random processes, although, generally speaking, it does not divide them into two separate groups. It enables us to investigate any oscillating time function from a general point of view. The method is also good for solving problems concerning linear and nonlinear vibration systems, stationary and nonstationary vibrations, as well as narrow- and/or wideband signals. It also allows precise analysis of the dissipation of vibration energy and vibration effects on machine durability.

7.1 Signal differentiation and integration

The analytic signal notion (equation 2) allows us to describe a relationship between an initial complex signal and its first derivative as $\dot{X} = X\left((\dot{A}/A) + i\omega\right)$. The second derivative of the signal takes the form $\ddot{X} = X\left((\ddot{A}/A) - \dot{\psi}^2 + 2i\dot{A}\psi + i\ddot{\psi}\right)$. Integration of analytic signals may also be of importance for signal processing. A complex function of a real variable t is an integral of the analytic signal if the real and the imaginary parts of the function are corresponding integrals forming a pair of the HT.

7.2 Signal demodulation

Traditional Fourier analysis simply assumes that the signal is the sum of a number of sine waves. The HT allows a complex demodulation analysis, adapted to signals of the form of a single, but modulated (perturbed), sine wave. As a vibration signal is exactly of the model that the method assumes, it is no wonder that in some cases the HT performs better than the Fourier analysis. Demodulation removes the modulation from a signal to get the original baseband signal back. For example, an envelope detector based on the HT takes a high-frequency signal as input, and provides an output that is the envelope of the original signal.

7.3 Fatigue estimation

The HT approach can be used to count fatigue cycles in an arbitrary loading waveform. It processes a time history representing the random wideband loading condition to generate the number of cycle counts with their corresponding amplitudes. The approach is general, is accurate within any desirable degree of accuracy, and is amenable to modern signal processing.

8 SIGNAL ENVELOPE

The amplitude of the oscillation can vary with time, and the shape of such time variation is called the *envelope* $A(t)$ (equation 3). The initial signal and its envelope have common tangents at points of contact (extrema points), but the signal never crosses the envelope. The plus sign of the root square corresponds to the upper positive envelope and the minus sign corresponds to the lower negative envelope, so they are always in antiphase relation. The envelope function contains important information about the energy of the signal. By using the HT, the rapid oscillations can be removed from the amplitude-modulated (AM) signal to produce a direct representation of the slow envelope alone. For example, the impulse response of a linear single-degree-of-freedom (SDOF) system is an exponentially damped sinusoid. The envelope of the signal is determined by the monotonic exponent decay rate.

AM signals can be described with the constant carrier frequency ω_0. Then the initial complex signal (equation 2) will be given by $X(t) = A_0 e^{j\psi(t)} = A_0 e^{j\psi_0(t)} e^{j\omega_0 t}$, where $A_0 e^{j\psi_0(t)}$ introduces a new complex expression as the complex envelope. The spectrum of the complex envelope can be obtained by shifting an initial signal spectrum to the left toward the origin of axes.

8.1 Average values and amplitude distribution function

The mean value of the envelope takes the form $\overline{A} = \int_{-\infty}^{\infty} A(t)\,dt$. The mean value of the square of the time-derivative of the envelope is $\overline{[\dot{A}(t)]^2} = \int_{-\infty}^{\infty} \dot{A}^2(t)\,dt$. This quantity determines the level of the envelope variation ($\overline{A} = 1$).

The envelope probability density $p(A)$ is related to the signal probability density function $p(x)$: $p(x) = \pi^{-1} \int_{|x|}^{\infty} p(A)\,dA/(A^2 - x^2)^{1/2}$. As an example of this relation, a typical classical Gaussian (normal) form of the probability density of the random vibration that conforms to the Rayleigh probability density of the vibration envelope is shown in Figure 4.

9 INSTANTANEOUS PHASE

The instantaneous phase notation (equation 4) indicates the multibranch character of the function, as shown in Figure 5, line 2, when the phase angle jumps between π and $-\pi$. These phase jumps can be unwrapped into a monotone function by an artificial changing of the phase values (Figure 5, line 1). The instantaneous relative phase shift in the case of two

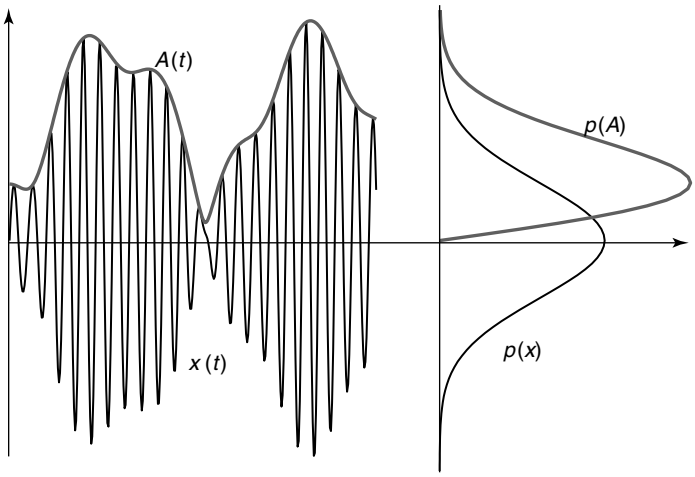


Figure 4. Random signal with envelope and their distributions.

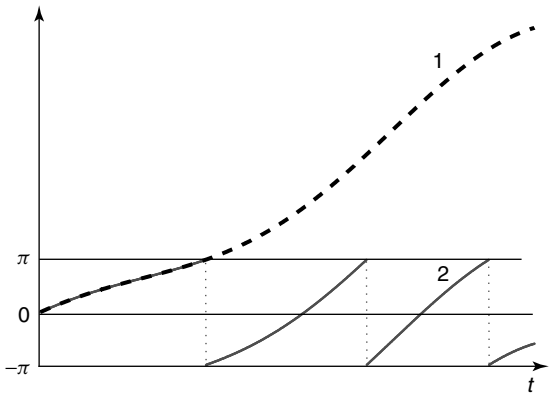


Figure 5. Instantaneous phase: unwrapped (1) and wrapped (2).

different real signals $x_1(t)$ and $x_2(t)$ can be estimated as the instantaneous relative phase between them, according to the formula $\psi_2(t) - \psi_1(t) = \arctan(x_1(t)\tilde{x}_2(t) - \tilde{x}(t)x_2(t))/(x_1(t)x_2(t) + \tilde{x}_1(t)\tilde{x}_2(t))$.

10 INSTANTANEOUS FREQUENCY

The first derivative of the instantaneous phase (equation 4) as a function of time $\omega(t) = \dot{\psi}(t)$, called the *instantaneous angular frequency*, plays an important role. The angular frequency $\omega(t) = 2\pi f(t)$ has the dimension radian per second and the cycle frequency $f(t)$ has the dimension hertz. There is a simple way to avoid the whole phase-unwrapping problem. It can be done when finding the instantaneous frequency (IF) by differentiation of the signal itself:

$$\omega(t) = \frac{x(t)\dot{\tilde{x}}(t) - \dot{x}(t)\tilde{x}(t)}{A^2(t)} = \mathrm{Im}\left[\frac{\dot{X}(t)}{X(t)}\right] \quad (5)$$

The IF $\omega(t)$ measures the rate of phasor rotation in the complex plane. Naturally, for a simple monoharmonic signal, the envelope and the IF are constant, and the phase angle increases linearly with time. In the general case, the IF of the signal is a varying function of time. Moreover, the IF in some cases may change sign in some time intervals. This corresponds to the change of rotation of the phasor from the counterclockwise to the clockwise direction. The IF always has physical meaning and is nothing more than just the varying speed (rate) of the phasor rotation in polar axes. In the time domain, the negative IF corresponds to the appearance of a complex riding cycle (complicated cycle of an alternating signal).

10.1 Average values, central frequency, and distribution function

The instantaneous amplitude and frequency of complicated vibration signals are nonconstant; they vary in time. While the IF is a positive function, the signal itself has the same numbers of zero crossings and extrema. When the IF has a negative value, the signal has one or multiple extrema between successive

zero crossings. The mean value of the IF $\omega_0 = \overline{\omega} = \int_{-\infty}^{\infty} \omega(t) A^2(t)\,\mathrm{d}t = (m_1/m_0)$, equal to the first normalized moment of the signal spectrum, is called the *central frequency* (here m_i is the ith moment of the spectrum) [4]. The mean value of the modulus of the IF given by $\overline{|\omega|} = \int_{-\infty}^{\infty} |\omega(t)| A^2(t)\,\mathrm{d}t = (m_2/m_0)^{1/2}$ is equal to the number of the signal zero crossing.

The mean value of the IF squared $\overline{\omega^2} = \int_{-\infty}^{\infty} \omega^2(t) A^2(t)\,\mathrm{d}t = (m_2/m_0) - \overline{\dot{A}^2(t)}$ determines the level of the IF variation. Some typical examples of the central frequency estimation based on the IF are shown in Table 1.

It is notable that, in the case of the SDOF vibration system, its free vibration natural frequency value equal to $(\omega_0^2 - h^2)^{1/2} \approx \omega_0 - (h^2/2\omega_0)$, where h is the viscous damping, lies between these central moments of the IF (Table 1).

As the IF may be considered to be the average of the signal frequency at a given time instant, it seems reasonable to inquire about the probability density function or spread at that time. For the random normal narrowband signal, the probability density function of the IF was derived by Bunimovich [4] $p(\omega) = (\Delta\omega^2)/(2(\omega^2 + \Delta\omega^2)^{3/2})$. From the last formula, a probabilistic prerequisite to the formation of the negative value of the IF is as follows:

$$p[\omega(t) < 0] = 0.5\left(1 - \frac{\overline{\omega}}{\overline{|\omega|}}\right) \tag{6}$$

For example, the probability of a negative value of the IF of a random signal after an ideal rectangular narrowband filter is directly proportional to the relative filter width $p[\omega(t) < 0] = \Delta\omega^2/144\omega_0^2$, where ω_0 is the central filter frequency and $\Delta\omega$ is the filter width. This indicates that after any narrowband filtering, the random signal will still have the negative value of the IF. However, for very small widths like $(\Delta\omega/\omega_0) \leq 0.01$, the probability will be less than one in a million, and the IF can be practically considered as always positive. From equation (6) it follows, for example, that the probability of a negative value of the IF of random vibration of the SDOF system (Table 1) is proportional to the system viscous damping coefficient.

Table 1. The spectral central frequencies

Vibration signal	Central frequency	
	Mean value $\overline{\omega}$	Mean modulus $\overline{\lvert\omega\rvert}$
Harmonics $\cos\omega_0 t$	ω_0	ω_0
Random vibration of the system $\ddot{y} + 2h\dot{y} + \omega_0^2 y = F(t)$	$\omega_0 - \dfrac{2h}{\pi}$	ω_0
Narrowband random noise $\omega_0 - \dfrac{1}{2}\Delta\omega < \omega < \omega_0 + \dfrac{1}{2}\Delta\omega$	ω_0	$\omega_0\left(1 + \dfrac{\Delta\omega^2}{24\omega_0^2}\right)$

10.2 Signal bandwidth

There are several techniques for estimation of the spectrum bandwidth. Probably the most familiar and simplest is the half peak level width. Also, the spectrum bandwidth could be estimated as a width of a hypothetical square with the same energy and the peak value $\Delta\omega_1 = \int_0^{\infty} S(\omega)\,\mathrm{d}\omega / S_{\max}$. In the case of the IF analysis, it is useful to introduce one more width parameter that is equal to the mean absolute value of the IF deviation from its central value plus the envelope variations. By summing up the IF variation around the mean value and the envelope variations, we obtain the average spectrum bandwidth of the signal $\Delta\omega_2$ [5]:

$$\begin{aligned}\Delta\omega_2^2 &= \int_0^{\infty} (\omega - \omega_0)^2 S(\omega)\,\mathrm{d}\omega = \overline{\omega^2} + \overline{\dot{A}^2(t)} - \overline{\omega}^2 \\ &= \frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2 \end{aligned} \tag{7}$$

where, again, m_i is the ith moment of the spectrum $S(\omega)$. Equation (7) is derived from Parseval's energy identity relation for time and frequency domains. It indicates that the signal spectrum bandwidth is equal to the sum of the IF and the envelope variations. For AM signals, the spectral bandwidth is equal to the mean square value of the velocity of the amplitude change. For signals which are only frequency modulated, the spectral bandwidth is equal to the mean square value of the IF.

Some typical examples of the spectral bandwidth estimation are shown in Table 2.

Taking into account these representations of analytic signals enables one to consider a vibration

Table 2. The spectral bandwidth

Vibration signal	Spectral bandwidth	
	Energy equivalent $\Delta\omega_1$	IF deviation $\Delta\omega_2$
Random vibration of the system $\ddot{y} + 2hy + \omega_0^2 y = F(t)$	πh	$2\left(\frac{h\omega_0}{\pi}\right)^{1/2}$
Narrowband random noise $\omega_0 - \frac{1}{2}\Delta\omega < \omega < \omega_0 + \frac{1}{2}\Delta\omega$	$\Delta\omega$	$\frac{\Delta\omega\sqrt{3}}{6}$

process, at any moment of time, as a quasiharmonic oscillation, amplitude and frequency modulated by time-varying functions $A(t)$ and $\omega(t)$: $x(t) = A(t)\cos\int_0^t \omega(t)\,\mathrm{d}t$. The instantaneous parameters are functions of time and can be estimated at any point of the vibration signal. The total number of these points, which map the vibration, is much greater than that of the peak points of the signal. This opens the way for averaging and for other statistical processing procedures, making vibration analysis more precise.

11 NARROW- AND WIDEBAND SIGNALS

Classically, the relations between the spectral bandwidth and the central frequency divide vibration signals into two groups: narrowband ($\Delta\omega \ll \overline{\omega}$) and wideband ($\Delta\omega \gg \overline{\omega}$) signals. Also a vibration-detrended (centered) signal can be considered a narrowband signal if its IF and envelope are always positive ($\omega(t) > 0$, $A(t) > 0$), and the vibration signal with zero or negative IF is considered a wideband signal ($\omega(t) \leq 0$). A slow frequency-modulated signal is a typical example of the narrowband signal. With this definition, the narrowband signal in each cycle, defined by the central frequency, involves only one mode of oscillation; no complex riding waves are allowed [6]. In effect, in this case the IF will not have the fast fluctuations induced by asymmetric waveforms.

Such a detrended vibration signal will always alternate around zero. In the narrowband case, its envelope coincides with the local maxima function. Moreover, the envelope and the opposite sign envelope for narrowband signals are always in antiphase. In the wideband case, the negative IF and the corresponding complicated riding cycle indicate the existence of the local negative maximum or the local positive minimum of the signal.

11.1 Example of two harmonics

If a signal composition is a sum of two harmonics: $x(t) = A_1(t)\cos\omega_1 t + A_2(t)\cos\omega_2 t$, the envelope $A(t)$ (equation 3) of the double-component signal composition could be written as $A(t) = [A_1^2 + A_2^2 + 2A_1A_2\cos(\omega_2 - \omega_1)t]^{1/2}$. The signal envelope $A(t)$ consists of two different parts, that is, a slow varying part, including the sum of the component amplitudes squared, and a rapidly varying part, oscillating with a new frequency equal to the difference between the component frequencies.

The IF $\omega(t)$ (equation 5) of the double-component composition is as follows:

$$\omega(t) = \omega_1 + \frac{(\omega_2 - \omega_1)[A_2^2 + 2A_1A_2\cos(\omega_2 - \omega_1)t]}{A^2(t)} \tag{8}$$

The IF of the two tones considered in equation (8) is generally time-varying and exhibits asymmetrical deviations about the frequency ω_1 of the largest harmonics. In addition, these deviations always force the IF beyond the frequency range of the signal components. The IF, in principle, consists of two different parts, that is, a frequency of the first largest component ω_1 and a rapidly varying asymmetrical oscillating part. For large amplitude of the second harmonics when $(A_2/A_1) > 1 - \sqrt{1 - (\omega_1/\omega_2)}$, the IF of the composition becomes negative.

12 MONO- AND MULTICOMPONENT SIGNALS

In the simple case, a nonstationary signal has constant or slow varying amplitude and a slow varying positive IF. This kind of signal is often referred to as a *single-* (*mono-*) *component* or *monochromatic signal*.

The term *monocomponent signal* corresponds to the "intrinsic mode function (IMF)" and is suggested in [6]. Monocomponent signals belong to the group of narrowband signals whose IF is always greater than zero. However, not every narrowband signal is a simple monocomponent signal. For example, a slow AM signal is a narrowband signal, but it is not a monocomponent, because it can be separated into three simpler components. Often, measured signals can be represented by a composition (sum) of a small number of the monocomponent signals [7]:

$$x(t) = \sum_l A_l(t) \cos\left(\int \omega_l(t)\,\mathrm{d}t\right) \tag{9}$$

where $A_l(t)$ is the instantaneous amplitude and $\omega_l(t)$ is the angular radian IF of the l component. In other words, the signal consists of l monocomponents, where each has a constant or a slowly varying amplitude $A_l(t)$ and the IF $\omega_l(t)$.

The multicomponent signal can be defined as follows: an asymptotic signal is referred to as a *multicomponent composition* if there exists even a single narrowband component such that its extraction from the composition decreases the average spectrum bandwidth of the remainder signal. This definition means that, after breaking up a signal into its simplest components and taking out any component, the envelope and the IF of a residue part will have smaller deviations in time. The signal example considered in Section 11.1 can be decomposed into pure harmonic components, each one without a deviation and with a zero spectrum bandwidth.

12.1 Upper and lower amplitude bounds (envelope of the envelope)

In the case of the multicomponent signal, its envelope sometimes exceeds the signal local maxima. To estimate the acting upper and lower bounds of the signal, consider a general l-component real signal $x(t)$ with the amplitudes and the instantaneous frequencies according to equation (9). Naturally, this multicomponent phasor as a composition $X(t) = \sum_{l=1}^{N} A_l(t)\mathrm{e}^{j\phi_l(t)}\mathrm{e}^{j\omega_l(t)t}$ will have a more complicated and faster varying envelope function. Here $X(t)$ is the signal in the analytic form and $A_l(t)\mathrm{e}^{j\phi_l(t)}$ is the complex amplitude (envelope) of every component.

For a signal formed from the summation of many components, we can invoke the generalized triangle inequality property for complex functions. The triangle inequality states that the sum of the moduli of complex numbers is greater than the modulus of the sum of these complex numbers: $|\sum_{l=1}^{N} C_l| \le \sum_{l=1}^{N} |C_l|$. In our case, the modulus on the left-hand side is the signal envelope $A(t)$ of the composition $X(t)$, and the sum of the moduli on the right-hand side is the sum of the component envelopes, thus yielding

$$A_{\mathrm{env}}(t) \le \sum_{l=1}^{N} A_l \mathrm{e}^{j\phi_l} \tag{10}$$

A new term, *envelope of the envelope* (EoE) $A_{\mathrm{env}}(t)$, is defined as the tangent curve to the local extrema, which touches only maximum points during every longest period of the composition. The EoE represents the acting maximum height (intensity) of the signal and forms the shape of the time variation of the acting extreme points. The EoE varies much more slowly with time than the signal envelope itself. To estimate the EoE, we first need to produce a decomposition of the signal and then to construct the algebraic summation of the envelopes of all the decomposed components. Such successive signal decomposition (disassembling) and the subsequent summation (reassembling) of the component envelopes generate the desirable EoE as a slow function of time. The EoE function plays an important role in the precise identification of nonlinear vibration systems by taking into consideration their high-frequency nonlinear components [8].

12.2 Average frequency and the largest energy component

For combination of two harmonics (see Example in Section 11.1) the IF consists of two different parts, that is, a slow varying frequency of the first component ω_1 and a rapidly varying asymmetrical oscillating part. However, the rapidly varying asymmetrical oscillating part of the IF has an important feature. If we now integrate the oscillating part with the integration limits corresponding to the full period

of the difference frequency $\left[0\ T = 2\pi/(\omega_2 - \omega_1)\right]$,

$$\int_0^T \frac{(\omega_2 - \omega_1)\left[a_2^2 + 2a_1a_2\cos\left(\int(\omega_2 - \omega_1)\,\mathrm{d}t\right)\right]}{a^2(t)} \times \mathrm{d}t = 0 \qquad (11)$$

we get the definite integral equal to zero. This means that the average value of the first moment of the IF (equation 8) is just equal to the frequency of the largest harmonics $\langle\omega(t)\rangle = \omega_1(t) + \int_0^T \omega(t) = \omega_1(t) + 0$. This interesting property of the IF offers the simplest and most direct way of estimating the frequency of an *a priori* unknown largest signal component [9]. In the more general case of three and more quasi-harmonics in the composition, the IF will take a more complicated form, but again, the averaging or the low-pass filtering will extract only the IF of the largest energy component.

13 HT SIGNAL DECOMPOSITION

HT decomposition methods allow any complicated data set to be decomposed into a finite and often small number of IMFs. Recently, Huang *et al.* [6, 7] proposed the empirical mode decomposition (EMD) method to extract monocomponent and symmetric components, known as *IMF*, from nonlinear and nonstationary signals (*see also* **Damage Detection Using the Hilbert–Huang Transform**). The term *empirical* chosen by the authors emphasizes the empirical essence of the proposed identification of the IMF by their characteristic timescales in the initial complicated data. Some years later, a different technique, called the *Hilbert vibration decomposition* (HVD) *method*, dedicated to the same problem of decomposition of nonstationary wideband vibration, was developed in [9]. The global HVD method is based on the HT presentation of the IF and does not involve spline fitting.

13.1 Local EMD method

The EMD method [6] automatically generates a collection of IMFs that satisfy two conditions: (i) in the complete data set, the number of extrema and the number of zero crossings must either be equal or differ at most by one; and (ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The EMD algorithm requires the following procedures at every iteration step: (i) estimation of all extrema; (ii) spline fitting of all minima and maxima, ending up with two extrema functions; (iii) computation of the average function between maxima and minima; (iv) extraction of the average from the initial signal; and (v) iteration on the residual (the sifting procedure). When applied to a nonstationary signal mixture, the EMD algorithm yields efficient estimates for the instantaneous amplitude and frequency signals of each component.

For illustration purposes, let the signal be a composition of two nonstationary harmonics, each one with a varying amplitude (Figure 6a). The EMD method successfully decomposes both the amplitude and frequency of each intrinsic function (Figure 6b, c) from the initial wideband signals, but it cannot decompose the narrowband bi-harmonics signal [10, 11].

13.2 Global HVD method

The important property of the IF (equation 11) offers the simplest and most direct way of estimating the frequency of an *a priori* unknown largest signal component. If we replace the integration of the IF by convolution with a low-pass filter instead of averaging, we will cut down the asymmetrical oscillations and leave only the slow varying frequency of the main component. Thus, the IF becomes a useful function

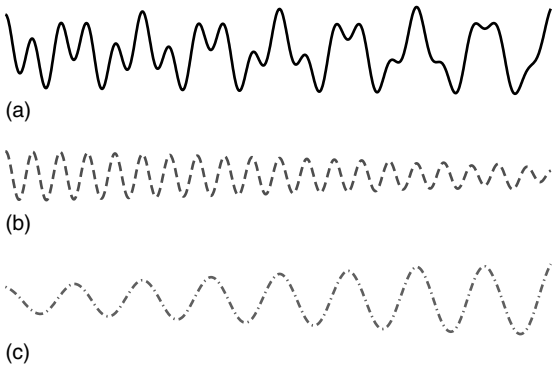


Figure 6. EMD signal decomposition: initial signal (a) and nonstationary components (b and c).

enabling detection of the largest energy component. A low-pass filter will eliminate all high-frequency components outside of the cutoff frequency, leaving the low-frequency components of the IF without modification. Such low-pass filtering of the IF corresponds to the narrowband filtering of the signal around its central frequency, and the cutoff frequency of the IF is equal to the relative bandwidth frequency of the signal.

The HVD method [9] at every iteration step requires the following procedures: (i) estimation of the IF of the largest component by low-pass filtering of the signal IF; (ii) detection of the corresponding envelope of the largest component; and (iii) subtraction of the largest component from the composition. At each iteration step, the corresponding slow varying vibration component is extracted using the low-pass filtering of the IF. Correspondingly, on each iteration step the residual contains the lower-energy components. As a result, we automatically separate the initial composition into several slowly varying oscillating components. At each iteration step after subtracting the largest component, the IF of the residual will be filtered again, and the components will be separated if the difference between their frequencies is greater than the cutoff frequency value. For illustration purposes, let the carrier signal be a harmonic $\cos t$, and let the modulation also be a single but nonstationary tone with decreasing modulation index and with increasing modulated frequency. The signal waveform is shown in Figure 7(a) and the decomposed components according to the proposed HVD method are shown together in Figure 7(b).

13.3 Comparison of the HT decomposition methods

13.3.1 Local versus global

A varying signal can be described by different signal attributes that change over time. To estimate attributes such as amplitude or frequency, any procedure will need some measurements during a definite time. Generally two approaches exist for such estimation: a local approach that measures attributes at each instant without knowing the entire function of the process and a global approach that depends on the whole

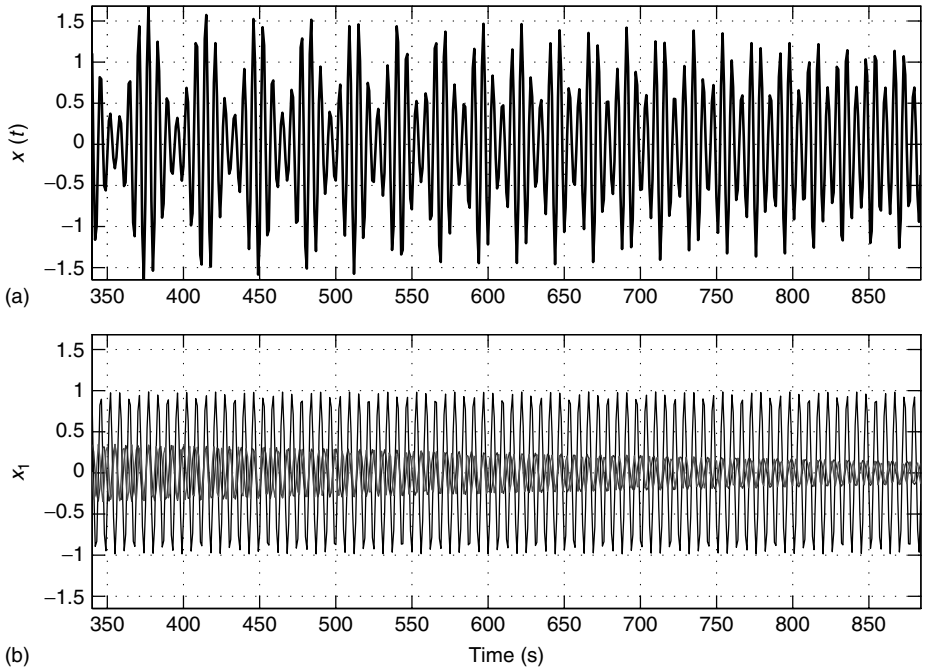


Figure 7. HVD signal decomposition: initial signal (a) and nonstationary components (b).

signal waveform during a long (theoretically infinite) measuring time. An example of the local (or differential, microscopic) approach is the estimation of a frequency by measuring the interval (spacing) between two successive zero crossings. Examples of the global (or integral, macroscopic) approach are estimating the average frequency by taking the first moment of the spectral density, estimating the envelope or the IF. In other words, local estimations consider the signal locally, i.e., in a very small interval around the instant of analysis. Quite to the contrary, global estimations have to use the whole real signal. The global versus local estimations provide different precisions and resolutions depending on many conditions, primarily noise distortions and random flicker phase modulation in a signal [4, 5].

14 HT NONLINEAR SYSTEM IDENTIFICATION

By observation (experiment), we acquire knowledge of the position and/or velocity of the object as well as the excitation at several known instants of time. The nonparametric identification will determine the initial nonlinear restoring and damping forces. In the case of free vibration, we have only the output signal—the vibration of the oscillators—whereas in the case of forced vibration we also deal with the input excitation. In modern signal processing, the HT method is more and more widely applied for analysis and identification of nonlinear dynamic structures.

14.1 Elastic nonlinearities in vibration systems

The important type of nonlinearity arises when the restoring force of a spring is not proportional to its deformation. There are several known types of static force characteristics (load–displacement curve) representing different types of nonlinearity in elastic springs: backlash, preloaded (precompressed), impact, and polynomial ones. In most nonlinear vibration systems, the natural frequency will depend decisively on amplitude of the vibrations (Figure 8). Every typical nonlinearity in a spring has its unique form of skeleton (backbone) curve [8, 12]. The unique topography of the skeleton curve is essential for the properties evaluation of the tested vibrating system, e.g., in reconstructing characteristics of the nonlinear elastic forces.

14.1.1 Small-amplitude nonlinear behavior: preloading and backlash

There are cases where vibration systems show their specific nonlinear behavior only in a small-amplitude range of vibrations. Such a system is a spring backlash (clearance). For small vibration amplitudes, the system will display its nonlinear properties

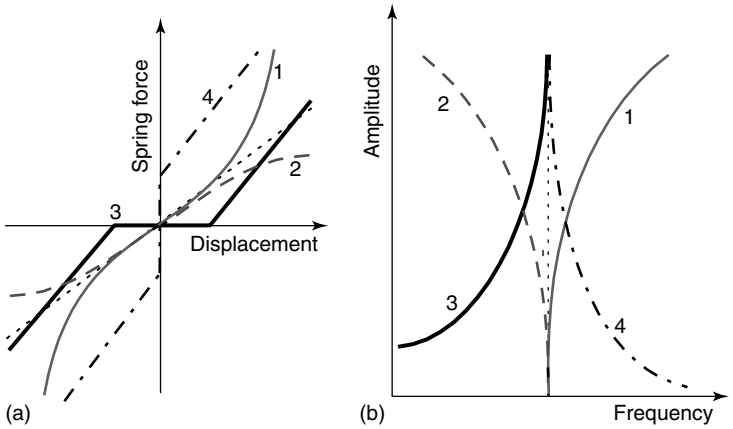


Figure 8. Characteristics of typical nonlinear vibration systems (hardening (1), softening (2), backlash (3), preloaded (4)): backbones (a) and static elastic force (b).

where natural frequency decreases with decreasing of amplitude.

Another typical example of nonlinearity in the small-amplitude range is a vibration mechanical system with preloaded (pretensioned) restoring force. Only for small amplitudes of vibration motion commensurable to a preloaded deformation the natural frequency will increase extremely with decreasing of amplitude.

14.1.2 Large-amplitude nonlinear behavior: polynomial model

Most of the known cases of nonlinear manifestation occur in the large-amplitude range of oscillation, like a nonlinear spring element with hardening or softening restoring force, or nonlinear friction quadratic or cubic force. These typical nonlinear spring elements of a mechanical vibration system could be expressed as a power series $k(x) = (\alpha_1 + \alpha_3 x^2 + \alpha_5 x^4 + \cdots)x$. In general, a second-order conservative system with a nonlinear restoring force $k(x)$ and a solution $x(t) = A\cos\omega t$ takes a simple form $\ddot{x} + k(x) = 0$. Applying the multiplication property of the HT for overlapping functions [12] to the last equation, we obtain a new form of time-varying equation of motion $\ddot{x} + j\delta(t)x + \omega_0^2(t)x = 0$, where $\omega_0(t)$ is the fast-varying natural frequency function and $\delta(t)$ is the fast-varying fictitious damping function. If we consider only the mean value of the varying natural frequency function square, we get an important result [12] $\langle\omega_0^2\rangle = T^{-1}\int_0^T \omega_0^2(t)\,\mathrm{d}t = \alpha_1 + (3/4)\alpha_3 A^2 + (5/8)\alpha_5 A^4 + \cdots$, which proves that the averaged natural frequency has, correct to the polynomial constant coefficients, the same expression as the initial nonlinear restoring force $k(x)$. This general result means that the estimated average natural frequency, and hence the system skeleton curve (backbone) $A(\langle\omega_0\rangle)$, includes the main information about the initial characteristics of nonlinear elastics and can be used for nonlinear system identification.

14.2 Damping nonlinearities in vibration systems

Vibration systems also can show their damping nonlinearities only in a small- or a large-amplitude range, depending on the type of nonlinear damping force characteristics. As an example of a small-amplitude nonlinear behavior, we mention a system with Coulomb (dry) friction, whose plot of logarithmic decrement versus vibration amplitude is a monotonic hyperbola [12]. In some other cases, like nonlinear turbulent friction, the nonlinear damping behavior appears in the large-amplitude range. But in practice, the real small damping forces practically have no effect on the mechanical system backbones.

14.3 Theoretical bases of nonlinear system identification

The time-domain techniques, based on the HT, allow direct extraction of linear and nonlinear systems parameters from a measured time signal of input and output of the vibration system [12]. The HT methods, namely, free vibration (FREEVIB) and forced vibration (FORCEVIB), of free and forced vibration analysis determine instantaneous modal parameters, even if an input signal is a high-sweep frequency signal. Such direct determination of a relationship between amplitude and natural frequency, which characterizes elastic properties, and a relationship between amplitude and damping characteristics, enables an efficient nonlinear system testing without long forced response analysis.

A second-order conservative system with a nonlinear restoring force $k(x)$, a nonlinear damping force $h(\dot{x})\dot{x}$, and a solution $x(t) = A(t)\cos[\omega(t)t]$ can be represented in a general form $\ddot{x} + h(\dot{x})\dot{x} + k(x) = 0$. In the first stage of the identification technique, the envelope $A(t)$ and the instantaneous frequency $\omega(t)$ are extracted from the vibration and excitation signals on the base of the HT signal processing. In the next stage, by applying the multiplication property of the HT for overlapping functions to the equation of motion, the instantaneous undamped natural frequency and the instantaneous damping coefficient are calculated according to formulas [12]

$$\omega_0^2(t) = \omega^2 - \frac{\ddot{A}}{A} + \frac{2\dot{A}^2}{A^2} + \frac{\dot{A}\dot{\omega}}{A\omega};$$
$$h_0(t) = -\frac{\dot{A}}{A} - \frac{\dot{\omega}}{2\omega} \tag{12}$$

where $A(t)$ and $\omega(t)$ are the envelope of the IF of the vibration.

In the last stage of low-pass filtering, the set of duplet modal parameters (the instantaneous natural frequency $\langle\omega_0^2(t)\rangle$ and instantaneous damping $\langle h(t)\rangle$) of each natural mode of vibration are defined. As a result of the HT method, the corresponded set of the duplet modal parameters as functions of the envelope (the skeleton and the damping curves) of each natural mode of vibration describe the dynamic behavior of the structure.

It is notable that for linear systems with a constant natural frequency f_0 the damping coefficient $h_0(t)$ (equation 12) depends only on the varying envelope. By estimation of the average damping coefficient as an integral in the interval of time, we will get the well-known approximate formula for the logarithmic decrement from the decay envelope:

$$\frac{\overline{h}}{f_0} = \frac{1}{(t_1 - t_2) f_0} \int_{t_1}^{t_2} \frac{\dot{A}(t)}{A(t)} \, \mathrm{d}t = n^{-1} \ln \frac{A_{i+n}}{A_i}$$

14.4 Considering high harmonics for identification

The modern vibration decomposition approaches [6, 9] divide the real multicomponent motion into a number of separated principal and high-frequency superharmonics. These approaches, considering the high superharmonics, yield more precise identification of nonlinear systems, including the nonlinear elastic and the damping static force characteristics. Theoretically, summarizing an infinitive number of partial static force characteristics will provide the exact initial static force characteristics of the nonlinear system. To precisely reconstruct the initial nonlinear force characteristics, we deduce the form of each oscillating function by using the EoE (equation 10).

In the first stage of the identification technique considering high harmonics, the signal envelope, together with the instantaneous undamped natural frequency and the instantaneous damping coefficient,

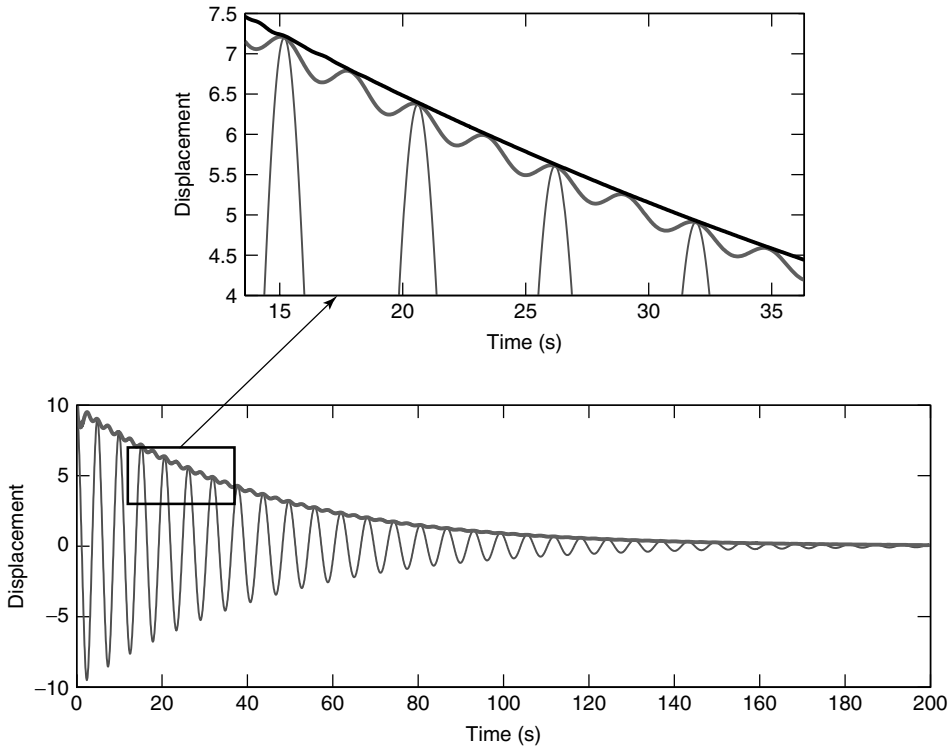


Figure 9. Nonlinear spring free vibration: the displacement (———), the envelope (▬▬), the envelope of the displacement envelope (smoothed bold line ▬▬).

is extracted from the vibration and excitation signals based on HT signal processing [8]. In the next stage, the EoE is generated by considering nonlinear high harmonics for every obtained instantaneous function, according to equation (10). In the final stage, the precise nonlinear static elastic and the damping force characteristics are constructed as a multiplication of two corresponding EoEs (for example, displacement and stiffness) according to the technique described in [8]. It is convenient to represent the final result of HT identification in a standard form that includes the skeleton curves with the initial static force characteristics of the nonlinear vibration system.

14.5 Example of nonlinear spring identification

As an example of a nonlinear elastic force, we refer to the classic Duffing equation with a hardening spring and a linear damping characteristic $\ddot{x} + 0.05\dot{x} + x + 0.01x^3 = 0$; $x_0 = 10$, $\dot{x}_0 = 0$. A simulation of the free vibration signal was carried out by using an initial displacement as shown in Figure 9. A part of the displacement envelope is shown separately with zoom to emphasize its oscillating behavior.

The obtained results of the HT identification are shown in Figure 10. The fast-varying instantaneous natural frequency (before the decomposition) is plotted against the envelope as a fast-changing spiral (Figure 10a, thin line). The same figure includes also the low-pass filtered skeleton curve (dashed line) that only approximately restores nonlinear forces correct to the first term of motion. The final precise "envelope" skeleton curve, which considers, in addition, two high superharmonics, is shown in Figure 10(a) with a bold line. The resultant identified spring data and damping force characteristics completely coincide with the initial cubic spring static force characteristics $k(x) = 1 + 0.01x^3$ (Figure 10b, dashed line)

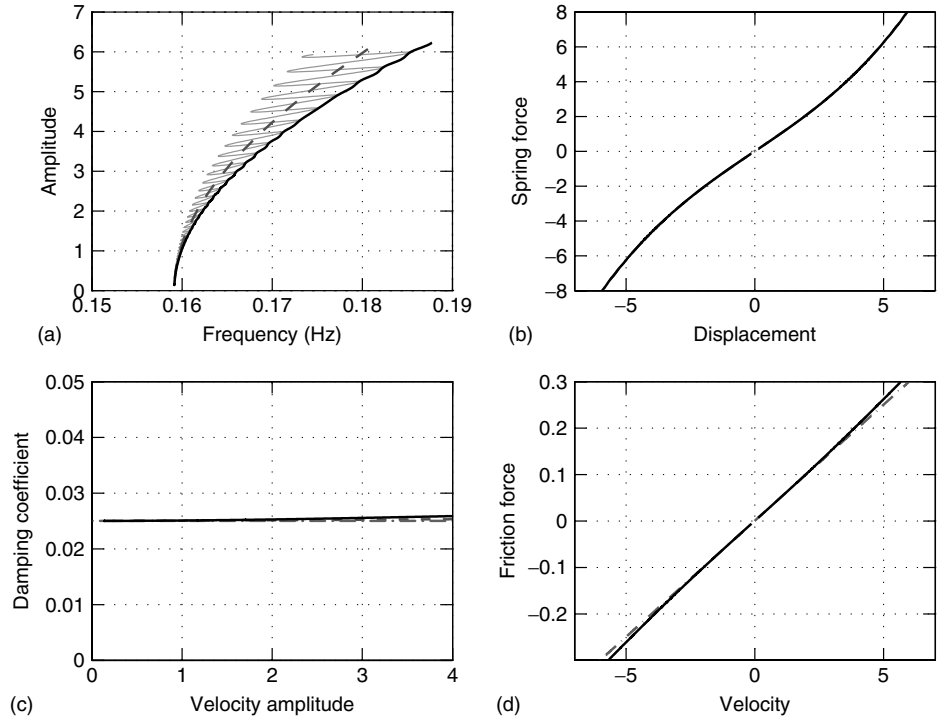


Figure 10. Identified nonlinear spring. The backbone (a): instantaneous (———), averaged (- - - -), with high harmonics (▬▬); The static spring force characteristics (b): initial (-•-), with high harmonics (▬▬); The damping curve (c): initial (-•-), averaged (- - - -), with high harmonics (▬▬); The static friction force characteristics (d): initial (-•-), with high harmonics (▬▬).

and the initial trivial straight friction force characteristics (Figure 10c and d, dashed line).

15 CONCLUSION

The HT is accepted as a standard procedure in the field of signal processing and has been used for more than 60 years. The application of the HT to the initial signal provides some additional information about instantaneous amplitude, phase, and frequency. This important information was valid when applied to monitoring and analysis of varying signals.

Two known HT signal decomposition methods (the EMD and the HVD) allow automatic and adaptive extraction of *a priori* unknown narrowband components from the nonstationary composition. The HT decomposition methods are dedicated primarily to decomposition of quasi and almost periodic oscillating-like signals. Such oscillating types could be, for example, multicomponent nonstationary modulated vibrations similar to rotor start-up or shutdown vibration, or motion of nonlinear dynamic system.

The instantaneous amplitude, phase, and frequency have valuable properties for determining changes from signal measurements. Not only this but also the instantaneous parameters are able to solve an inverse problem—the problem of identification of dynamic nonlinear systems. The HT-based technique facilitates direct estimation of system instantaneous dynamic parameters (i.e., natural frequencies and damping characteristics) as well as their dependence on vibration amplitude and frequency.

REFERENCES

[1] Hahn SL. *Hilbert Transforms in Signal Processing*. Artech House, 1996, p. 305.

[2] Pandey JN. *The Hilbert Transform of Schwartz Distributions and Applications*. John Wiley & Sons: New York, 1996.

[3] Inaudi JA, Kelly JM. Linear hysteretic damping and the Hilbert transform. *Journal of Engineering Mechanics* 1995 **121**:626–632.

[4] Vainshtein L, Vakman D. *Frequency Separation in the Theory of Vibration and Waves (in Russian)*. Nauka: Moscow, 1983, p. 288.

[5] Vakman D. *Signals, Oscillations, and Waves*. Artech House: Boston, 1998.

[6] Huang NE, Shen Z, Long SR. New view of nonlinear water waves: the Hilbert spectrum. *Annual Review of Fluid Mechanics* 1999 **31**:417–457.

[7] Huang NE, Shen SSP, (eds). *The Hilbert-Huang Transform and its Applications*. World Scientific Publishing, 2005.

[8] Feldman M. Considering high harmonics for identification of nonlinear systems by Hilbert transform. *Mechanical Systems and Signal Processing* 2007 **21/2**:943–958.

[9] Feldman M. Time-varying vibration decomposition and analysis based on the hilbert transform. *Journal of Sound and Vibration* 2006 **295/3–5**:518–530.

[10] Rilling G, Flandrin P. One or two frequencies? The empirical mode decomposition answers. *IEEE Transactions on Signal Processing* 2008 **56**(1):85–95.

[11] Feldman M. Theoretical analysis and comparison of the Hilbert transform decomposition methods. *Mechanical Systems and Signal Processing* 2008 **22**(3):509–519.

[12] Feldman M. Non-linear free vibration identification via the Hilbert transform. *Journal of Sound and Vibration* 1997 **208**(3):475–489.

Chapter 26
Time–frequency Analysis

Rosario Ceravolo

Dipartimento di Ingegneria Strutturale e Geotecnica, Politecnico di Torino, Torino, Italy

1 INTRODUCTION

Interest in time–frequency representation is motivated by the limitations of classical spectral estimation techniques in analyzing strongly nonstationary behaviors [1]. Owing to the importance of nonstationary components in vibration signals, several studies have proposed the adoption of time–frequency analysis in structural diagnostics [2, 3] and machine fault detection [4].

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Though a huge variety of plausible theories and perspectives has been proposed for time–frequency analysis, one method cannot be claimed to be superior to the others under all conditions. The benefits of each time–frequency approach should be highlighted and demonstrated by referring to specific applications. In the particular field of "structural health monitoring" (SHM), many authors have applied linear tools, such as the short-time Fourier transform and the wavelet transform, or in some cases their squared magnitudes, known as *spectrogram* (SPEC) and *scalogram* (SCAL), respectively. Others have appealed to quadratic structures, since energy is a quadratic signal representation.

The adoption of a quadratic representation makes it possible to overcome limitations due to the time–frequency resolution, since energetic and correlative transforms are not based on signal segmentation. Specifically, among the quadratic transforms, those belonging to the Cohen class [1] are characterized by their invariance to time and frequency shifts. Shift invariance is of great importance when processing signals measured on mechanical systems and justifies the interest that many researchers have shown for this approach.

The present article is intended as a comprehensive overview of some recent advances in the field of time–frequency analysis for structural identification. We focus on the shift-invariant class, including the SPEC. Other time–frequency analysis techniques include, but are not limited to, wavelet analysis (*see* **Wavelet Analysis**) and empirical mode

decomposition combined with Hilbert transform (*see* **Damage Detection Using the Hilbert–Huang Transform**).

In more detail, the exposition starts with a brief review of the main time–frequency transforms, laying a special emphasis on properties that are desirable in SHM applications. Time–frequency analysis is then discussed from the stochastic perspective, which entails the definition of time-dependent spectra that reflect the time evolution of the second-order properties of the processes. Much of the practical interest in stochastic formulation lies in defining proper time–frequency estimators for output-only system identification. Structural dynamics models are then introduced in formulations, so as to support data analysis of time series and, specifically, linear and nonlinear identification techniques.

The article concludes with a numerical example showing the identification of a simple nonlinear system, and with an experimental application to the dynamic characterization of a real reinforced concrete building.

2 JOINT TIME–FREQUENCY REPRESENTATION

The main advantage of time–frequency domain analysis is its ability to handle nonstationary waveform signals, which are very common when structural defects and machinery faults occur.

As an example, let us consider a signal that features a time localization of spectral components. The Fourier transform is not suited for the analysis of such components, since it projects the signal on infinite harmonics, which are not localized in time. If at any time instant only a single frequency is present, an instantaneous frequency may be variously defined; this quantity is commonly identified with the rate of phase change in the analytic signal (*see* **Hilbert Transform, Envelope, Instantaneous Phase, and Frequency**; **Damage Detection Using the Hilbert–Huang Transform**). Such a definition is capable of describing the time localization of a specific class of signals, but proves to be unsuitable for multicomponent ones.

In all cases where monodimensional representations are inadequate, one can turn to bidimensional (joint) functions $T_x(t, f)$ of the variables time and frequency. $T_x(t, f)$ is referred to as *time–frequency representation* (TFR) of the signal $x(t)$.

2.1 Linear time–frequency transforms

2.1.1 Short-time Fourier transform (STFT)

In order to introduce the time localization of frequency components, a simple solution is obtained by prewindowing the signal around a particular time t, as shown in Figure 1, calculating its Fourier transform, and doing that for each time instant t. Accordingly, the "short-time Fourier transform" (STFT) of a signal $x(t')$ is defined as [5, 6]

$$STFT_X^{(\gamma)}(t, f) = \int_{-\infty}^{+\infty} x(t')\gamma^*(t - t')\,\mathrm{e}^{-jft'}\,\mathrm{d}t' \quad (1)$$

where $\gamma(t)$ is a short-time analysis window centered around t. The superscript * denotes complex conjugation.

Since multiplication by the relatively short window $\gamma(t' - t)$ effectively suppresses the signal outside a neighborhood around the analysis time point $t' = t$, the STFT is a "local" spectrum of the signal $x(t')$ around t.

The STFT is evidently linear and is complex-valued, in general. Provided that the short-time window is of finite energy, the STFT is invertible through

$$x(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} STFT_X^{(\gamma)}(t', f')g(t - t') \times \mathrm{e}^{jf't}\cdot \mathrm{d}t'\cdot \mathrm{d}f' \quad (2)$$

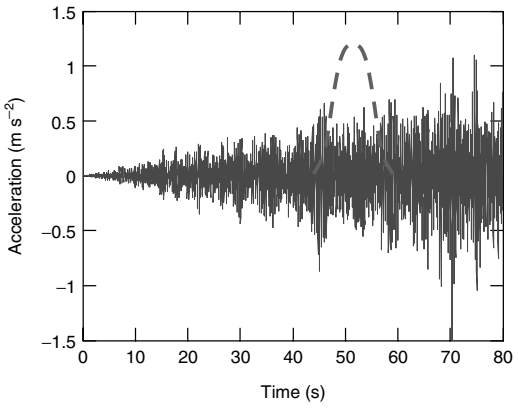


Figure 1. Prewindowing of a nonstationary signal.

with $\int g(t)\gamma^*(t)\,\mathrm{d}t = 1$. Equation (2) implies that the total signal can be decomposed as a weighted sum of elementary waveforms

$$g_{t,f}(t') = g(t'-t)\cdot \mathrm{e}^{\,jft'} \tag{3}$$

which can be interpreted as "atoms". Each atom is obtained from the window $g(t)$ by a shift in time and frequency (modulation).

The STFT may also be expressed in terms of signal and window spectra:

$$STFT_x^{(\gamma)}(t,f) = \int_{-\infty}^{+\infty} X(f')\Gamma^*(f'-f) \times \mathrm{e}^{j(f'-f)t}\,\mathrm{d}f' \tag{4}$$

where X and Γ are the Fourier transform of x and γ, respectively. Accordingly, the STFT can be interpreted as the result of passing the signal through a filter with frequency response $\Gamma(f'-f)$ and is, therefore, deduced from a mother filter $\Gamma(f)$ by a translation of f. The STFT is thus similar to a bank of band-pass filters with constant bandwidth.

The STFT preserves time–frequency shift (property P2 in Table 1). Such transform, as well as its squared magnitude, the SPEC, is frequently used in many application fields, including modal decoupling, system identification (e.g., [7, 8]), group velocity, speech recognition, etc.

2.1.2 Time–frequency resolution

The time resolution of the STFT can be obtained by considering for x a Dirac impulse:

$$x(t) = \delta(t-t_0) \Rightarrow STFT_x^{(\gamma)}(t,f) = \mathrm{e}^{-jft_0}\gamma(t-t_0) \tag{5}$$

Thus, the time resolution of the STFT is proportional to the effective duration of the analysis window γ. Similarly, to obtain the frequency resolution, we have to consider a complex sinusoid (an impulse in the frequency domain)

$$x(t) = \mathrm{e}^{j2\pi f_0 t} \Rightarrow STFT_x^{(\gamma)}(t,f) = \mathrm{e}^{-j2\pi f_0 t}\Gamma(t-t_0) \tag{6}$$

Hence, the frequency resolution of the STFT is proportional to the effective bandwidth of the analysis window γ. As a consequence, for the STFT, we have a "trade-off" between time and frequency resolutions: while a good time resolution requires a short window γ, a good frequency resolution requires a narrow-band filter i.e., a long window. This limitation is a consequence of the Heisenberg–Gabor inequality [1, 9]:

$$T\cdot B \geq 1 \tag{7}$$

where T is the signal's temporal length and B is the bandwidth. The lower bound of the product is reached for Gaussian functions.

2.1.3 Discrete STFT

Equation (1) can be sampled on a rectangular grid:

$$STFT_x^{(\gamma)}(n,m) = STFT_x^{(\gamma)}\left(nt_0, mf_0\right) = \int_{-\infty}^{+\infty} x(t')\gamma^*\left(t'-nt_0\right)\mathrm{e}^{-j2\pi m f_0 t'}\mathrm{d}t' \tag{8}$$

where n and m are integers.

The problem is then to choose the sampling period t_0 and frequency f_0 so as to minimize the STFT inherent redundancy without losing any information [6, 10]. For a sampled signal $x[n]$ whose sampling period is noted Δt, t_0 has to be chosen as a multiple of Δt such that $t_0\cdot f_0 \leq 1$. We then have the following analysis and synthesis formulae [6]:

$$STFT_x^{(w)}[n,m] = \sum_k x[k]\cdot\gamma^*[k-n]\cdot \mathrm{e}^{-j\cdot 2\pi\cdot mk} \quad \text{for } -\frac{1}{2} \leq m \leq \frac{1}{2} \tag{9}$$

$$x[k] = \sum_n \sum_m STFT_x^{(w)}[n,m]\cdot g[k-n]\cdot \mathrm{e}^{j\cdot 2\pi\cdot mk} \tag{10}$$

Equations (9) and (10) can be implemented efficiently by means of overlap "fast Fourier transform" (FFT) techniques.

Alternatively, a filter-bank implementation is possible, based on sampling equation (4).

Table 1. Properties satisfied by time–frequency transforms and corresponding kernel conditions

	Property	Condition on the kernel
P0	Nonnegativity: $T_x(t, f) \geq 0 \ \forall t \ \forall f$	$g(\nu, \tau)$ is the AF[(a)] of some $f(t)$
P1	Realness: $T_x(t, f) = T_x^*(t, f)$	$g(\nu, \tau) = g^*(\nu, \tau)$
P2	Time–frequency shift: $y(t) = x(t - t_0) \Rightarrow T_y(t, f) = T_x(t - t_0, f)$ $y(t) = x(t)\ \mathrm{e}^{2\pi f_0 t} \Rightarrow T_y(t, f) = T_x(t, f - f_0)$	$g(\nu, \tau)$ independent of t and f
P3	Time marginal: $\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}f = \lvert x(t)\rvert^2$	$g(\nu, 0) = 1 \ \forall \nu$
P4	Frequency marginal: $\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}t = \lvert X(f)\rvert^2$	$g(0, \tau) = 1 \ \forall \tau$
P5	Instantaneous frequency: $\dfrac{\int_{-\infty}^{+\infty} f T_x(t, f)\,\mathrm{d}f}{\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}f} = -\dfrac{\mathrm{d}}{\mathrm{d}t}\{\arg(x(t))\}$	$\left.\dfrac{\partial(g(\nu, \tau))}{\partial \tau}\right\rvert_{\tau=0} = 0 \ \forall \nu$
P6	Group delay: $\dfrac{\int_{-\infty}^{+\infty} t T_x(t, f)\,\mathrm{d}t}{\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}t} = -\dfrac{\mathrm{d}}{\mathrm{d}t}\{\arg(X(f))\}$	$\left.\dfrac{\partial(g(\nu, \tau))}{\partial \nu}\right\rvert_{\nu=0} = 0 \ \forall \tau$
P7	Finite time support: $x(t) = 0 \quad \text{if } \lvert t\rvert \geq T$ $\Rightarrow T_x(t, f) = 0 \text{ per } \lvert t\rvert \geq T$	$\varphi(t, \tau) = 0 \ \lvert\tau\rvert < 2\lvert t\rvert$
P8	Finite frequency support: $X(f) = 0 \quad \text{if } \lvert f\rvert \geq B$ $\Rightarrow T_x(t, f) = 0 \quad \text{if } \lvert f\rvert \geq B$	$\int_{-\infty}^{+\infty} g(\nu, \tau)\ \mathrm{e}^{-j2\pi f \tau}\,\mathrm{d}\tau \ \lvert\nu\rvert < 2\lvert f\rvert$
P9	Reduced interference	$g(\nu, \tau)$ is a low-pass filter type in (ν, τ) plane

[(a)] AF, ambiguity function.

2.1.4 Wavelet transform

Another important TFR is the time–frequency version of the Wavelet transform (WT) defined as [5, 6, 11]

$$WT_x^{(\psi)}(t, f) = \int_{t'} x(t') \cdot \sqrt{\left|\frac{f}{f_c}\right|} \cdot \psi^*\left(\frac{f}{f_c}(t' - t)\right) \cdot \mathrm{d}t' \tag{11}$$

where $\psi(t)$ is a real or a complex band-pass function centered around $t = 0$ in the time domain. The parameter f_c in equation (11) corresponds to the center frequency of $\psi(t)$. The WT was originally introduced as a timescale representation and, in fact, retains the important property of preserving time shifts and timescaling. It does not, however, preserve frequency shifts.

The WT's time and frequency resolutions are related via the Heisenberg–Gabor inequality, like in the STFT case. However, while the STFT's resolution is the same for each analysis frequency, the WT's frequency resolution (respectively time resolution) becomes poorer (respectively better) as the analysis frequency grows.

Scale shift invariance makes the WT, or its squared magnitude, the SCAL, a frequent choice in pattern recognition and many other application fields, including ridge and phase estimation (e.g., [12, 13]) and SHM (e.g., see the list of references in [2–4]). Details on the WT and its application to SHM are found in **Wavelet Analysis**.

2.2 Quadratic time–frequency transforms

Although linearity is a desirable property, quadratic TFRs [5, 14] allow for interpreting the distributions from an energy point of view. This interpretation is expressed by the so-called marginal properties:

$$\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}f = |x(t)|^2;$$
$$\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}t = |X(f)|^2 \qquad (12)$$

having defined the "instantaneous power" $|x(t)|^2$ and the "spectral energy density" $|X(f)|^2$. Consequently, the signal energy is

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2\,\mathrm{d}t = \int_{-\infty}^{+\infty} |X(f)|^2\,\mathrm{d}f$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} T_x(t, f)\,\mathrm{d}t\,\mathrm{d}f \qquad (13)$$

The marginal properties are not sufficient to identify an energy density at every point in the time–frequency plane, since the uncertainty principle does not allow such a notion. Vice versa, many quadratic TFRs may loosely support an energetic interpretation even if they do not satisfy the marginal properties, among them the SPEC and the SCAL.

$$SPEC_x^{(\gamma)}(t, f) = |STFT_x^{(\gamma)}(t, f)|^2 \qquad (14)$$
$$SCAL_x^{(\psi)}(t, f) = |WT_x^{(\psi)}(t, f)|^2 \qquad (15)$$

In equation (14) the linearity structure of the STFT is violated, and, in fact, any quadratic TFR satisfies the "quadratic superposition principle":

$$x(t) = c_1 x_1(t) + c_2 x_2(t) \Rightarrow$$
$$T_x(t, f) = |c_1|^2 T_{x_1}(t, f) + |c_2|^2 T_{x_2}(t, f)$$
$$+ c_1 c_2 T_{x_1 x_2}(t, f) + c_2 c_1 T_{x_2 x_1}(t, f) \qquad (16)$$

The last two terms in equation (16) are the cross terms or interference terms. The interference terms are oscillatory structures, which are restricted to those regions of the time–frequency plane where the autoterms (or authentic terms) overlap. In the specific cases of the SPEC and the SCAL, if two components are sufficiently far apart in the time–frequency plane, then their interference terms are virtually nil.

Quadratic representations have recently served as an effective tool in structural diagnostics and machine fault detection [2–4].

2.2.1 *The autocorrelation form and the Wigner–Ville transform*

A general approach to deriving time-dependent spectra is by generalizing the Wiener–Khintchine theorem: the correlation function and the power spectrum form Fourier transform pair (*see* **Higher Order Statistical Signal Processing**). By assuming the symmetric form for the instantaneous temporal and spectral autocorrelation, which are also functions of the time lag τ and the frequency lag ν, respectively,

$$r_x(\tau, t) = x\left(t - \frac{\tau}{2}\right) x\left(t + \frac{\tau}{2}\right) \qquad (17)$$
$$r_x(\tau, t) = X\left(f - \frac{\nu}{2}\right) X\left(f + \frac{\nu}{2}\right) \qquad (18)$$

and by transforming equations (17) and (18) one obtains [1]

$$W_x(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) \mathrm{e}^{-j2\pi f\tau}\,\mathrm{d}\tau$$
$$= \int_{-\infty}^{+\infty} X\left(f + \frac{\nu}{2}\right) X^*\left(f - \frac{\nu}{2}\right) \mathrm{e}^{j2\pi t\nu}\,\mathrm{d}\nu \qquad (19)$$

Equation (19) gives the "Wigner distribution" (WD), a transformation well known in quantum mechanics, whose signal-processing version is often referred to as *Wigner–Ville distribution*.

By taking the instantaneous cross correlation between two signals, $x_1(t)$ and $x_2(t)$, the WD assumes the following more general form:

$$W_{x_1 x_2}(t, f) = \int_{-\infty}^{+\infty} x_1\left(t + \frac{\tau}{2}\right)$$
$$\times\, x_2^*\left(t - \frac{\tau}{2}\right) \mathrm{e}^{-j2\pi f\tau}\,\mathrm{d}\tau$$

$$= \int_{-\infty}^{+\infty} X_1\left(f + \frac{\nu}{2}\right) \times X_2^*\left(f - \frac{\nu}{2}\right) \mathrm{e}^{j2\pi t\nu}\,\mathrm{d}\nu \quad (20)$$

The WD satisfies a large number of desirable properties (Tables 1 and 2), in particular, marginals, shift invariance, and real-valuedness. The instantaneous frequency (*see* **Hilbert Transform, Envelope, Instantaneous Phase, and Frequency**; **Damage Detection Using the Hilbert–Huang Transform**) and the group delay can be evaluated using the local first-order moments of the WD [1].

Among the dual class of correlative TFRs, which combine temporal and spectral correlations, an important role is played by the "ambiguity function" (AF) [5]

$$A_{x_1,x_2}(\tau,\nu) = \int_{-\infty}^{+\infty} x_1\left(t + \frac{\tau}{2}\right) x_2^*\left(t - \frac{\tau}{2}\right) \mathrm{e}^{-j2\pi\nu t}\,\mathrm{d}t = \int_{-\infty}^{+\infty} X_1\left(f + \frac{\nu}{2}\right) \times X_2^*\left(f - \frac{\nu}{2}\right) \mathrm{e}^{j2\pi\tau f}\,\mathrm{d}f \quad (21)$$

The AF may be viewed as a joint time–frequency correlation function. Along $\nu = 0$ and $\tau = 0$ axes, AF reduces to the time correlation function and the frequency correlation function, respectively (correlative marginal properties).

2.2.2 Cohen class of transforms

Among quadratic transforms, those belonging to the Cohen (or shift-invariant) class are characterized by the invariance of its members to time and frequency shifts (P2 in Table 1), a property that is desirable for correlating the signal characteristics to phenomena that take place in the mechanical system, which generates the signal. Cohen demonstrated that every member of the shift-invariant class is a filtered version of the WD, and that it is possible to use a general formula for describing all of them [1]. Indeed, equivalent formulas can be written in four different domains: temporal correlation domain (t, τ), time–frequency domain (t, f), AF domain (τ, ν), and spectral correlation domain (ν, f) [15, 16]. For instance, the following relation holds in the temporal correlation domain:

$$T_x(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} r_x(t',\tau)\varphi(t-t',\tau) \times \mathrm{e}^{-j2\pi f\tau}\,\mathrm{d}t'\,\mathrm{d}\tau$$
$$\varphi(t,\tau) = \int_{-\infty}^{+\infty} g(\nu,\tau)\ \mathrm{e}^{-j2\pi\nu t}\,\mathrm{d}\nu \quad (22)$$

where $g(\nu, \tau)$ is the kernel that uniquely identifies the specific TFR ($g(\nu, \tau) = 1$ for the WD).

2.2.3 Desirable properties of the t–f distributions

A standard set of desirable properties is usually referred [1, 5, 16] to compare the performance of different transforms. Herein, we consider a subset that contains the main characteristics that are of interest in the application context under discussion. In Table 1, properties are related to the corresponding requirements on the kernel. Table 2 reports the properties of eight important shift-invariant transforms, including the SPEC. The SPEC is, in fact, a member

Table 2. List of desirable properties satisfied by different quadratic transforms

Transforms	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9
Spectrogram (SPEC)	√	√	√							√
Wigner distribution (WD)		√	√	√	√	√	√	√	√	
"Alias-free" Wigner		√	√	√	√	√	√	√	√	
Pseudo-Wigner		√	√	√		√				
Smoothed pseudo-Wigner		√	√							√
Cone–kernel		√	√					√		√
Reduced interference		√	√	√	√	√	√	√	√	√
Choi–Williams distribution (CWD)		√	√	√	√	√	√	√[(a)]	√[(a)]	√

[(a)] Not in a strict sense, but only approximately.

of Cohen's class, but, since it does not offer independence of temporal and spectral resolutions, it is generally classed as linear [1]. It is worthy to note that only the SPEC satisfies P0, while sacrificing nonnegativity is mandatory in order to gain time–frequency resolution.

The WD does not satisfy P9 (reduced interference), which is a desirable property, as it preserves only the authentic (e.g., modal) components. Figure 2(a–f) [17] shows the results obtained by applying different transforms to three sinusoidal components and two delta functions. Figure 3 [17] reports the contour plots obtained by applying the WD and the "Choi–Williams transform" (CWT) to an accelerometric signal recorded on a simply supported alloy beam in free vibration. It clearly outlines the interference term filtering operated by the CWT.

Detailed theoretical discussions on different TFRs may be found in the specialized literature [15, 16, 18].

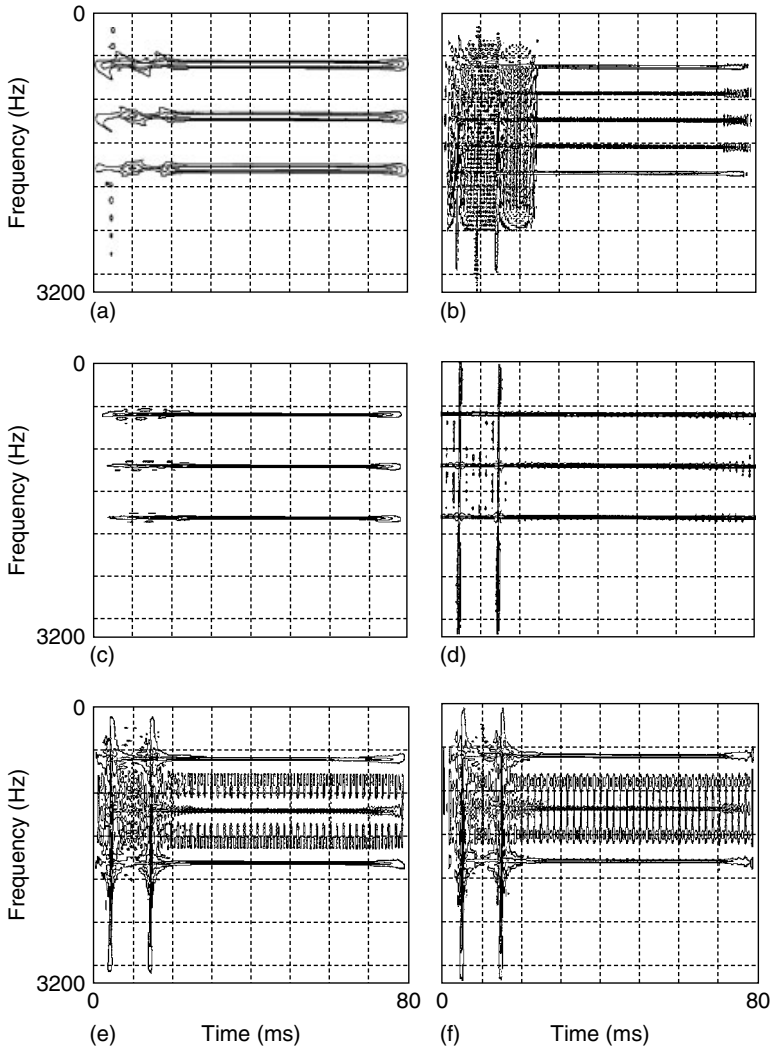


Figure 2. Comparison of different time–frequency transforms to a signal constituted by three sinusoidal components and two delta functions [17]. (a) Spectrogram, (b) Wigner, (c) smoothed pseudo-Wigner, (d) Choi–Williams, (e) cone–kernel, (f) reduced interference distribution. [Reproduced with permission from Ref. 17. © Elsevier, 1997.]

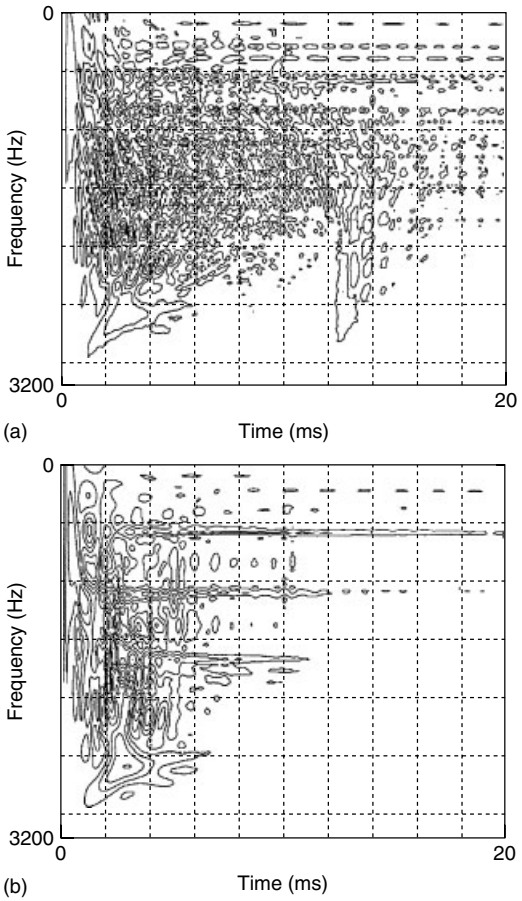


Figure 3. Comparison of (a) Wigner and (b) Choi–Williams distributions obtained by processing a signal recorded during the impulsive loading of an alloy beam. [Reproduced with permission from Ref. 17. © Elsevier, 1997.]

3 TIME–FREQUENCY REPRESENTATION OF STOCHASTIC PROCESSES

3.1 Stationary processes

The notion of spectral density, which is well consolidated in the field of stationary processes, may constitute a valuable starting point for approaching nonstationary problems.

One of the reasons of the popularity of the Wigner transform is its desirable property to preserve the instantaneous spectral information in stationary processes. In fact, introducing the autocovariance function of a process F, $C_F(t, \tau) = \mathrm{E}\{r_F(t, \tau)\}$, and taking the expectation in both sides of equation (19), we obtain the Wigner spectrum [11, 19]:

$$\mathrm{E}\{WD_F(t, f)\} = \int_{-\infty}^{+\infty} C_F\left(t + \frac{\tau}{2}, t - \frac{\tau}{2}\right) \times \exp(-j2\pi f\tau)\,\mathrm{d}\tau \qquad (23)$$

In the stationary case, the quantity expressed in equation (23) is independent of t and reduces to the usual spectral density $\mathrm{E}\{WD_F(t, f)\} = S_F(f)$ [11, 19]. In this situation, in fact, the covariance operator is a convolution operator, which is known to be diagonalized by the complex exponential functions.

Unfortunately, in practical applications, a limited number of signals is available. If the process is ergodic, a WD estimate of the spectral density may be obtained from a single realization $x(t)$ by averaging over time instantaneous spectra [20]:

$$V_B(f) = \frac{1}{B}\int_{-\frac{B}{2}}^{\frac{B}{2}} |WD_x(t, f)|\,\mathrm{d}t \qquad (24)$$

On the basis of equation (24), it also proves possible to quantify the degree of nonstationarity of a stochastic process on the given time interval via a distance measure [20]

$$d_B^2(f) = \frac{1}{B}\int_{-\frac{B}{2}}^{\frac{B}{2}} |WD_x(t, f)|^2\,\mathrm{d}t - S_x^2(t, f) \qquad (25)$$

If the process is stationary d_B reduces to zero.

In multicomponent signals, however, the Wigner representation of a single realization of the process is affected by interference terms, which can be filtered out in the AF domain but at the cost of losing resolution. There is, in fact, a tradeoff between cross-term filtering and time–frequency resolution, and hence the representation is kernel dependent. Another problem is that the Wigner transform misses the desirable property of nonnegativity over the $t-f$ plane [21]. These two aspects may have some negative implications in the definition of instantaneous estimators to be used in the analysis of deterministic signals.

Going back to linear TFRs, the square modulus of the STFT, or SPEC, of a stationary process may be

written in the form [11]

$$\mathrm{E}\left\{SPEC_F^{(\gamma)}(t, f)\right\} = \int_{-\infty}^{\infty} \left|\Gamma\left(f' - f\right)\right|^2 S_F(f)\,\mathrm{d}f' \tag{26}$$

where $\Gamma(f)$ is the spectrum of a window function $\gamma(t)$, such that $\|\Gamma(f)\| = 1$ and $S_F(f)$ is the spectral density associated with the process. Equation (26) shows that the value of the spectrum at f is a weighted average of the spectral density, when $f \sim f'$. As long as the Fourier transform of the window is still localized near the origin, the spectrogram provides, for each fixed f, information on the part of the original signal, which comes from the frequency contributions localized near f.

If one were to work with a sample realization $x(t)$, by assuming the ergodicity of the signal, an estimator of the spectral density may be defined as follows [11]:

$$V_B(f) = \frac{1}{B}\int_{-\frac{B}{2}}^{\frac{B}{2}} SPEC_x^{(\gamma)}(t, f)\,\mathrm{d}t \tag{27}$$

Owing to the weighting average operation, the spectral function expressed by equation (27) is a biased estimator. A spectral function that supports a weighted average interpretation may be also defined for the SCAL [11].

3.2 Locally stationary processes

When a process is nonstationary, the covariance operator may have complicated time-varying properties and its estimation is arduous because we do not know *a priori* how to diagonalize them. In the following, we focus on the particular class of locally stationary processes, i.e., processes whose covariance operators are approximately convolutions.

Recently, researchers have turned their attention to locally stationary processes as a tool to model systems where the behavior varies as a function of time (e.g., Mallat *et al.* [22], Dahlhaus [23]). Though in the time–frequency plane the concept of local stationarity is easily grasped, it does not exist as a universally accepted definition, to date. In order to support an intuitive idea, suppose that for any t_0 the Wigner spectrum varies very little within an interval $[t_0 - \delta, t_0 + \delta]$. Such a parameter $\delta > 0$ is called the *stationarity length* and, in general, its value depends on t_0.

The question on how to adapt the analyzing window (or the kernel, in the case of a Cohen class transform) to the stationarity length of process is still open [11]. Several techniques have been formulated to select the best strategy and often they are based on optimization procedures. For instance, Kozek [24] proposed a minimum bias optimization criterion based on support properties of the AF. In principle, these methods are conceived to deal with stochastic processes or with a proper number of sample realizations and are not suitable to deal with a single signal.

Some new ideas may arise when working with random fluctuations produced by mechanical systems, which typically change slowly with time or space. These type of signals can be generally considered as locally stationary, since they appear in the time–frequency plane as a sum of modulated harmonics concentrated at the modal frequencies. In this case, the instantaneous spectrum of a single realization may reflect and be associated with some physical parameters, whose consistency is an indirect indicator of the transform suitability.

A simple case of locally stationary process is a uniformly modulated process that is constructed as $x(t) = c(t)x_0(t)$, where $c(t)$ is a slowly varying modulation function and $x_0(t)$ some stationary process. The time evolutions of such a process are depicted correctly by the following time-varying spectrum [25]:

$$P_F(t, f) = c^2(t)S_{F_0}(f) \tag{28}$$

When dealing with more general classes of oscillatory processes, a description of temporal evolutions of spectral components, frequency by frequency, produces

$$x(t) = \int_{-\infty}^{+\infty} A_x(t)X(f)\ \mathrm{e}^{\mathrm{j}2\pi ft}\,\mathrm{d}f \tag{29}$$

thus leading to Priestley's evolutionary spectrum [25, 26]

$$P_F(t, f) = |A_F(t, f)|^2 S_F(f) \tag{30}$$

The modulation function $A_F(t, f)$ is supposed to undergo a slow time evolution, which ensures an almost orthogonal decomposition. Priestley's spectrum retains satisfactory properties (e.g., nonnegativity) but misses uniqueness.

In locally stationary conditions, evolutionary representation supports an interesting time–frequency input–output relationship [27]

$$y(t) = \int_{-\infty}^{+\infty} H(t, f) A_x(t, f) X(f)\ \mathrm{e}^{j2\pi f t}\,\mathrm{d}f \quad (31)$$

where $H(t, f)$ is the transfer function of a time-varying filter. The important result is that the output is a modulated form of the filter (or system) transfer function. This is analogous to one-dimensional filtering in the frequency domain $Y(f) = H(f)X(f)$.

Finally, it is worthwhile noting that there is a theoretical link between evolutionary spectra and members of Cohen class of transforms [26] and, what is more, for slowly varying processes the Wigner spectrum approaches the evolutionary spectrum.

4 TIME–FREQUENCY ESTIMATION AND BEST WINDOWING

An attractive idea is to extend by analogy some properties of stationarity to local stationarity. For instance, an estimator for the time-varying spectrum of a locally stationary process, $T_F(t, f)$, may be obtained from a single realization $x(t)$ by selecting a proper analysis window/kernel and posing $B = \delta$ in equations (24) and (27):

$$V_F(t, f) = |T_x(t, f)| \quad (32)$$

where $T_x(t, f)$ is any quadratic TFR. It is likewise possible to track instantaneous relationships between signals. For instance, in multichannel measurements on structures, instantaneous estimators of amplitude ratio and phase difference between channels may be defined as follows [28, 29]:

$$AR_{ij}(t, f) = \sqrt{\frac{T_{x_i}(t, f)}{T_{x_j}(t, f)}};\ PH_{ij}(t, f) = phase\left\{T_{x_i, x_j}(t, f)\right\} \quad (33)$$

In linear time-invariant systems, equation (33) can support output-only modal identification procedures, as stability of such estimators discriminates modal components from exogenous frequency components over time. In fact, modal signals are characterized by amplitude and phase relationships that are not time-dependent and, therefore, their modal shape is constant over time. The identification of modal frequencies, therefore, reduces to a search for the particular values $f = f_k$ for which the estimators remain constant with respect to the time variable, in general, by resorting to multiple criteria techniques. In frequency intervals where a single modal component is predominant, the estimators tend to lead to a constant value in time. This property is progressively more closely satisfied up to an actual constant value at the modal frequencies.

By referring to the shear-type frame reported in Figure 4 [28], modal frequencies may be identified as minima in standard deviation plots (Figure 5) defined as

$$\int_0^T \left[PH_{ij}(t, f) - \overline{PH}\right]^2 \mathrm{d}t \quad (34)$$

where T is the length of the analyzed signal and $\overline{PH}$ indicates the mean value.

Once the modal frequencies have been identified, equations (33) supply the temporal evolution of the amplitude and phase ratios. Alternatively, modal shape estimators can be taken as

$$AR_{ij}(t, f)|_{f=f_k} = \left|\frac{T_{x_i z^{(k)}}(t, f)}{T_{x_j z^{(k)}}(t, f)}\right|_{f=f_k}$$

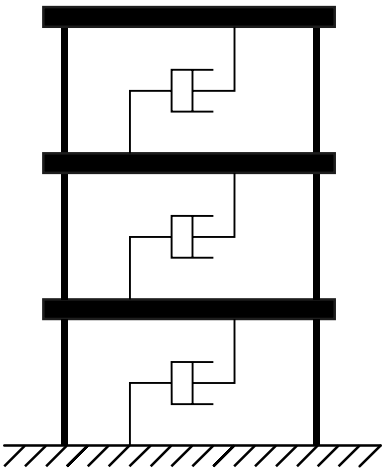


Figure 4. Model of a shear-type frame. [Reproduced with permission Ref. 28. © Elsevier, 2000.]

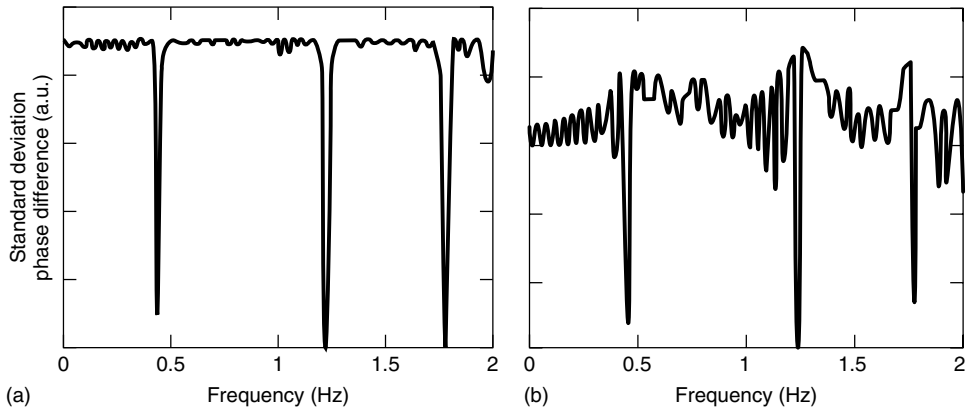


Figure 5. Standard deviation plots of the phase difference between the channel (equation 34) corresponding to the second story of the simulated structure and the reference channel (first story) as a function of frequency [28]. (a) Case with sine sweep excitation; (b) case of seismic excitation. [Reproduced with permission Ref. 28. © Elsevier, 2000.]

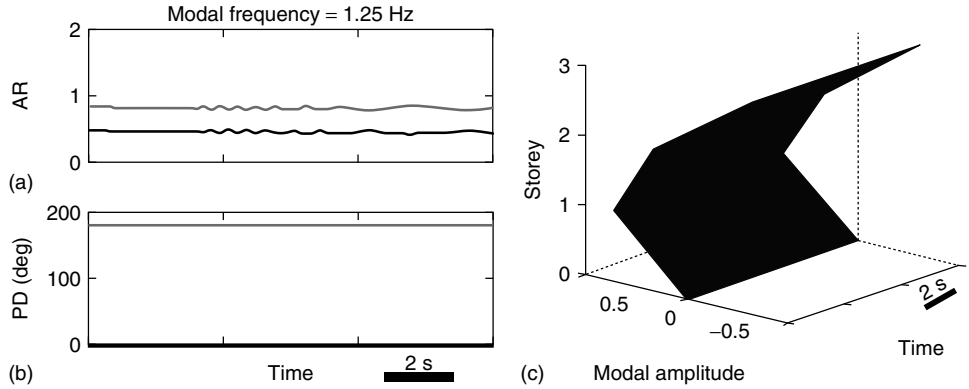


Figure 6. Evolution through time of modal amplitude ratios (a) and modal phase differences (b), as determined with respect to the reference channel (first story) for the second and third storys (sine weep excitation). Data refer to the second mode. The light and dark lines indicate the amplitude ratio and phase difference computed for the second and third story, respectively. Plot (c) shows the time evolution of the second modal shape estimate according to equation (35). [Reproduced with permission Ref. 28. © Elsevier, 2000.]

$$PH_{ij}(t, f)|_{f=f_k} = phase\left\{\frac{T_{x_i z^{(\omega)}}(t, f)}{T_{x_j z^{(\omega)}}(t, f)}\right\}\Bigg|_{f=f_k} \tag{35}$$

where $z^{(k)}(t)$ is a signal generated as a sinusoid with frequency equal to the kth modal frequency. The procedure is then repeated for all the signals at the ith and jth position. Figure 6 depicts the modal shape estimators (equation 35) for the second mode of the frame. These plots were all obtained using a CWD with $\sigma = 0.5$. This corresponds to applying the kernel $g(\nu, \tau) = \exp(-\nu^2\tau^2/\sigma)$ in equation (22) [30].

4.1 Best windowing

Typical questions about time–frequency estimation are how to select the optimal representation and window analysis and how many realizations of the process are needed to obtain an accurate estimate for $T_F(t, f)$.

While quadratic representations are very useful in analyzing strongly nonstationary signals, choosing the best kernel for a particular application appears to be a challenging task, as relationships with dynamic response characteristics are far from being trivial.

An alternative idea may consist of being satisfied with linear representations, which lend themselves to a clearer interpretation, and accepting the errors due to the fact that linear TFRs cause, in general, a distortion in the representation of the instantaneous power of stationary stochastic processes.

In principle, a suitable choice for the window of an STFT should be a function compactly supported in the interval $[t_0 - \delta, t_0 + \delta]$, and may vary according to a temporal law matched to the stationarity length of the process. Unfortunately, in most practical applications, the stationarity length δ is unknown.

Numerical studies have been performed in order to identify the influence of the analysis window on instantaneous spectral estimation via a STFT spectral function (equation 27). To this aim, an extensive set of dynamic response signals has been created numerically by exciting simple linear oscillators by means of white noise. The results reported here will focus on the following factors: type of window, window length (in samples), and decorrelation length of the process (related to damping) [8].

Equation (26) shows that the window may cause an error in the estimate and that the latter, when the number of realizations approaches infinity, decreases with increasing window temporal length. The effect of the window length (Figure 7) as well as the type of window has been examined.

Charts were then created to illustrate the accuracy of curve fitting as a function of the length of the windows in samples. The results, shown in Figure 8, indicate that, averaging over a finite number of realizations (in this case 30 demonstrated to be sufficient for a virtually exact fitting), windows of optimal length exist for the estimate of the "frequency response function" (FRF) and hence for the identification of modal parameters via linear TFRs. Such lengths are identified by minimum points in charts of the type shown in Figure 8.

In stationary conditions, the parameter that may affect the quality of time–frequency representation demonstrated to be the decorrelation length, and hence damping, while other factors such as the window shape were seen to be less relevant, at least for typical analysis windows [8]. Figure 9, which has been obtained from simulated examples, shows the evolution of the optimal length that the windows should have as a function of damping level (in this case for a Hanning-type window). This type of chart may be of practical use when representing the response of structures under ambient excitation. The choice of the optimal length for the STFT analyzing window, based on Figure 9, is conditioned by the availability of a raw estimate of damping.

5 MODEL-BASED TIME–FREQUENCY ESTIMATION

While linear representations, which permit a filter-bank interpretation, clearly allow reconstruction of the signal (e.g., equation 2), for energy decompositions the same property is less obvious. For Cohen class of transforms, the inversion is obtained from [31]

$$x(t')^* x(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{T_x(u, f)}{g(\nu, t - t')} \times \mathrm{e}^{j2\pi f(t-t') + j2\pi\nu[u-(t-t')/2]} \, \mathrm{d}u \, \mathrm{d}f \, \mathrm{d}\nu \tag{36}$$

By taking a particular value of t', for instance, zero, we have

$$x(t) = \frac{1}{x(0)^*} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{T_x(u, f)}{g(\nu, t - t')} \times \mathrm{e}^{j2\pi f t + j2\pi\nu(u - t/2)} \, \mathrm{d}u \, \mathrm{d}f \, \mathrm{d}\nu \tag{37}$$

One can easily notice that, according to equation (37), the signal can be recovered but only to within a constant phase factor.

5.1 Signal synthesis

In many practical applications, e.g., earthquake accelerogram generation, the main interest is in synthesizing signals with specific time–frequency features, rather than in signal reconstruction. Synthesis algorithms can also be used to perform time-varying filtering, multicomponent signal separation, and window and filter design. In this perspective, one might envisage filtering in the $t-f$ plane with [16, 18, 32]:

$$\tilde{T}(t, f) = \Gamma(t, f) T(t, f) \tag{38}$$

where $\Gamma(t, f)$ is referred to as *time–frequency mask*.

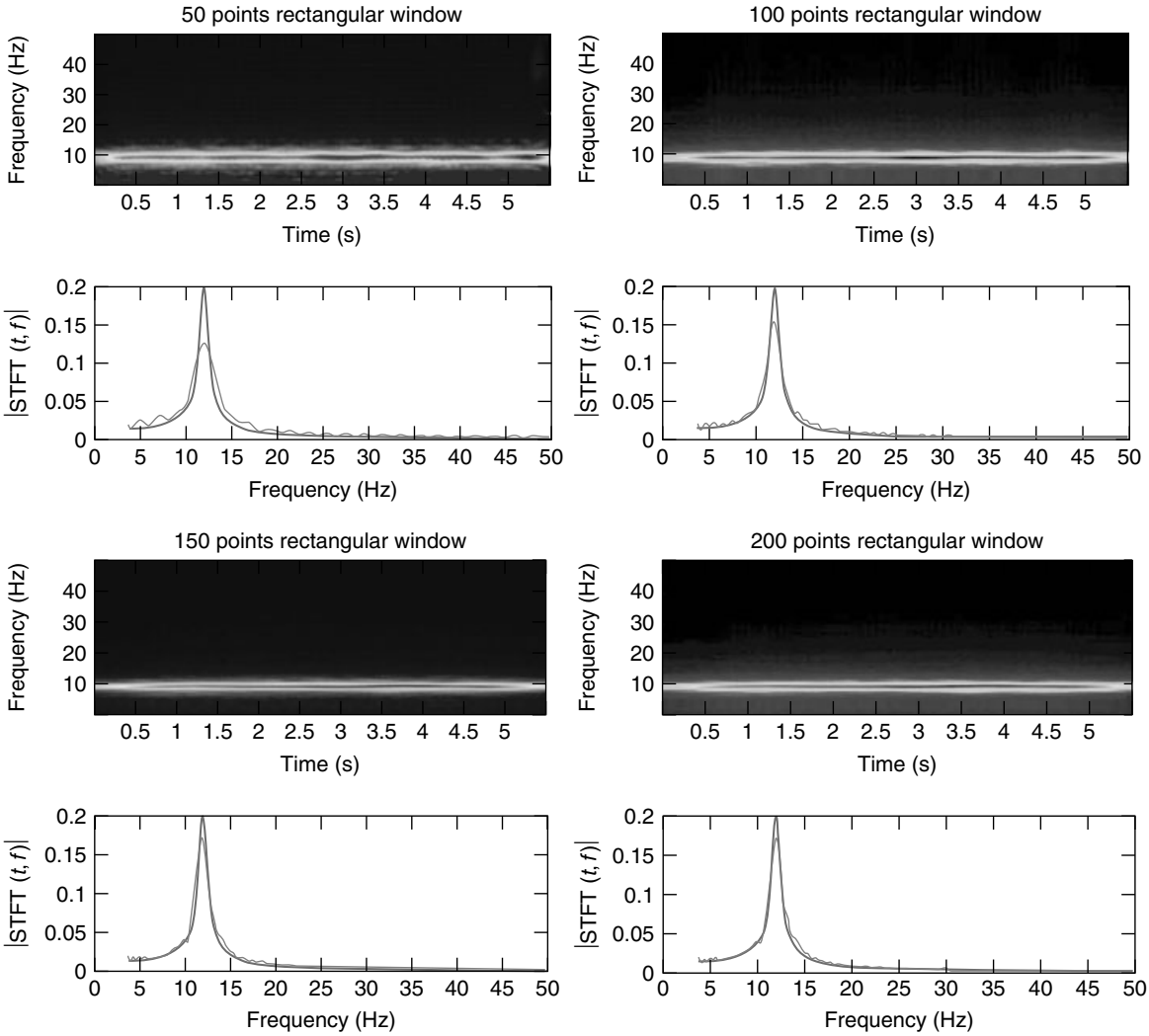


Figure 7. FRF estimation from STFT representations (average over 30 realizations) as a function of the window length (rectangular window). In sample instantaneous sections: original instantaneous spectrum (thin line) and estimated one (thick line). [Reproduced with permission from Ref. 8. © Elsevier, 2004.]

A major problem of the masking approach is that not all two-dimensional functions are valid TFRs. Instead, it is natural to resort to optimization techniques (e.g., least square methods), in order to find a time–frequency decomposition that best fits a time–frequency model.

Synthesis algorithms are usually formulated to find a signal $x(t)$ that minimizes the error ε_x between a given model, $\tilde{T}(t, f)$, and the transform of the signal $T_x(t, f)$ to be synthesized [1, 16, 32]

$$\varepsilon_x = \|T_x(t, f) - \tilde{T}(t, f)\| \longrightarrow \min_x \qquad (39)$$

Since for WD and other shift-invariant transforms the solution of the optimization process is not unique (e.g., $x(t)$ and $x(t)e^{j\alpha}$ have the same Wigner representation), the algorithm often contains a step to find the optimum phase factor.

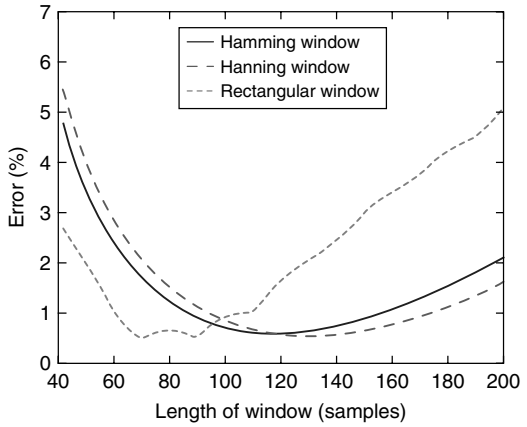


Figure 8. Error in the curve fitting performed on STFT as a function of window length (average over 30 realizations). [Reproduced with permission from Ref. 8. © Elsevier, 2004.]

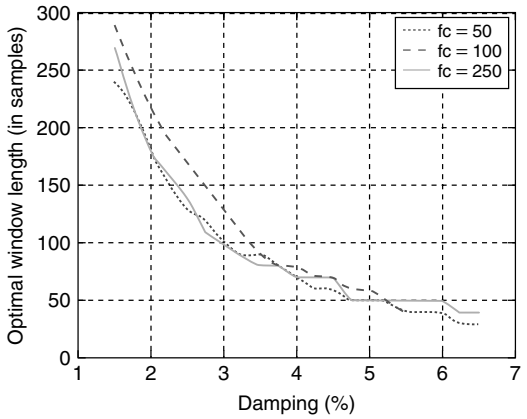


Figure 9. Optimal length in samples of the STFT (Hanning window) as a function of system's damping: curves obtained for three different values of sampling frequency. [Reproduced with permission from Ref. 8. © Elsevier, 2004.]

Under particular conditions, it proves advantageous to represent $x(t)$ in an orthonormal basis. Thus, by assuming a quadratic TFR, equation (39) becomes

$$x(t) = \sum_k \alpha_k q_k(t),$$

$$\varepsilon_x = \left\| \sum_k \sum_l \alpha_k \alpha_l^* T_{kl}(t, f) - \tilde{T}(t, f) \right\| \longrightarrow \min_{\alpha} \tag{40}$$

Such a formulation is certainly valid for linear dynamic systems, whose response is a sum of modal components.

5.2 Model identification

A formulation similar to that of equation (39) may be used also in solving the inverse problem, namely, finding a model that produces the same distribution as a measured signal.

When $x(t)$ is measured on a linear structure and we want to identify the dynamic model, a modal decomposition form is indicated for $\tilde{T}(t, f)$. As an example, a model identification can be obtained from the following unconstrained minimization:

$$\varepsilon_x = \left\| \sum_k \sum_l \alpha_k \alpha_l^* \tilde{T}_{kl}(t, f) - T_x(t, f) \right\| \longrightarrow \min_{\mathbf{p}} \tag{41}$$

where $\mathbf{p}$ is the global vector of parameters, which may contain α terms as well as other parameters of the model.

In some applications, as in structural control, the optimization can be performed on an online basis. For instance, a "block-by-block" synthesis/identification algorithm is performed on local, finite record intervals, whose length depends necessarily on the time–frequency analysis window or kernel. When parameters to be estimated retain a temporal significance (e.g., time-varying systems [33] or output-only identification), it may prove advantageous to perform an instantaneous minimization so as to obtain a punctual estimate of $\mathbf{p}$

$$\varepsilon_x(t) = \left| \int_{-\infty}^{+\infty} \left[\tilde{T}(t, f) - T_x(t, f) \right] \mathrm{d}f \right| \longrightarrow \min_{\mathbf{p}} \tag{42}$$

The most convenient minimization form and algorithm depends on the specific application. In the implementation of equations (39)–(42), the analytic signal is usually preferred to the real one, since it avoids cross-term interference between positive and negative frequencies.

5.2.1 Output-only identification of linear systems

Natural excitation of flexible structures (buildings, bridges, antennas, etc.) is characterized by slow

energy supply. Also based on the time-varying filter interpretation evoked in equation (31), in the time–frequency representation of locally stationary signals, we expect energy to concentrate around modal frequencies and be modulated according to the evolution of the TFR of the modulating waveform [2–8]. Consequently, the time–frequency model in equation (42) may be given the following (nonnegative) form [8]:

$$\tilde{T}(t, f) = \sum_k \sum_l \alpha_k(t)\alpha_l^*(t)\tilde{H}_k(f)\tilde{H}_l^*(f) \qquad (43)$$

where $\tilde{H}_k(f)$ is a scaled version the kth mode's FRF. In particular, for viscously damped dynamic systems the following expression holds:

$$\tilde{H}_k(f) = \left(f_k^2 + 2j\zeta_k f f_k - f^2\right)^{-1} \qquad (44)$$

If the assigned TFR has a stochastic nature (possibly obtained by ensemble averaging) and its modal components are uncorrelated, the cross terms vanish from equations (41) and (42). Correspondingly, also $T_F(t, f)$ tends to nonnegativity in accordance with the model in equation (43).

It is noteworthy that interference term suppression, as operated on deterministic signals by specific TFRs (property P9 in Table 2), appears to produce in equations (41) and (42) effects that are similar to those produced by a stochastic identification.

6 EXAMPLES

6.1 Identification of a system with polynomial nonlinearity in the stiffness

Let us consider a nonlinear system with a quadratic stiffness, as described by the following equation of motion:

$$m\ddot{y} + c\dot{y} + ky + k_2 y^2 = x(t) \qquad (45)$$

The input/output relationship is approximated by the following truncated Volterra series [34]:

$$\begin{aligned} Y(f) &= Y_1(f) + Y_2(f) \\ &= H_1(f)X(f) + \int_{-\infty}^{+\infty} H_2\left(f - f_1, f_1\right) \\ &\quad \times X\left(f - f_1\right) X\left(f_1\right) \mathrm{d}f_1 \end{aligned} \qquad (46)$$

In equation (46), $X(f)$ and $Y(f)$ are the Fourier transforms of the system's excitation and response, $Y_1(f)$ and $Y_2(f)$ are the Fourier transforms of the linear contribution and the quadratic contribution of the system response, and $H_1(f)$ and $H_2(f_1, f_2)$ are the system's linear and quadratic FRFs:

$$H_1(f) = \frac{1}{4\pi^2 m\left(-f^2 + i2\zeta f f_n + f_n^2\right)} \qquad (47)$$

$$\begin{aligned} H_2\left(f_1, f_2\right) &= -k_2 H_1\left(f_1\right) H_1\left(f_2\right) \\ &\quad \times H_1\left(f_1 + f_2\right) \end{aligned} \qquad (48)$$

where m is the mass, $\omega_n = 2\pi f_n = \sqrt{k/m}$ the natural frequency, $\zeta = c/(2\,\omega_n m)$ the damping coefficient, and k_2 the quadratic stiffness.

Assuming the signal $y(t)$ has been detected during N instants with a given sampling rate Δt, the response's STFT (model) has the following discrete expression [35]:

$$\begin{aligned} \sqrt{\tilde{T}[n, m]} &= \sum_{k=0}^{N-1} H_1[k]\cdot X[k]\cdot\Gamma^*[k - m]\cdot \mathrm{e}^{j\frac{2\pi kn}{N}} \\ &\quad + f_0\cdot\sum_{k=0}^{N-1}\sum_{r+s=k} H_2[r, s]\cdot X[r]\cdot X[s]\cdot\Gamma^*[k - m]\cdot\mathrm{e}^{j\frac{2\pi kn}{N}} \end{aligned} \qquad (49)$$

where $\Gamma(f)$ is a window function defined in the frequency domain, such that $\|\Gamma(f)\|^2 = 1$ and f_0 is an appropriate sampling frequency.

Let $\mathbf{p}$ be the vector of the unknown parameters of equation (45) and let us introduce the following objective function to be minimized at each time step n:

$$\begin{aligned} \varepsilon_y[n] &= \left|\sum_{m=0}^{N-1}\left[\tilde{T}[n, m] - SPEC_y^{(\gamma)}[n, m]\right]\right| \\ &\longrightarrow \min_{\mathbf{p}^T=\{f_n,\zeta,k_2\}} \end{aligned} \qquad (50)$$

The system parameters are as listed in Table 3. Assuming that the mass is known, $\mathbf{p}$ contains three parameters to be identified: f_n, ζ, and k_2. Minimization is performed on an online basis, by starting from

Table 3. Parameters of the nonlinear system to be identified [35]

m (kg)	10
ζ (%)	5
f_n (Hz)	3
k_2 (N m^{-2})	2 00 000

Reproduced with permission from Ref. 35. © Trans Tech Publications, 2005.

values obtained in the previous time step. A rough estimate of frequency and damping (e.g., obtained from a linear identification session) is needed for parameter initialization.

Let us consider the case of a measured Gaussian excitation with linearly modulated amplitude (Figure 10). Before proceeding with the identification, a parametric analysis on similar systems was performed in order to determine the best window length for the evaluation of the STFT and it was ascertained that a Hamming window with about 140 samples length was optimal. Figure 11 shows a STFT representation of system response over the first 40 s.

In order to assess the performance of the method in the presence of noisy measurements, numerical analyses are conducted by introducing white Gaussian noise into the signals. Even in the presence of noise in both input and output, instantaneous estimators give rise to a robust estimation of system parameters (Figures 12–14). The estimate of k_2 is seen to be the most adversely affected by the presence of noise.

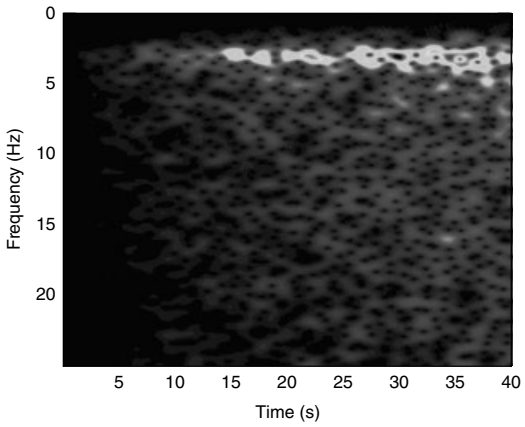


Figure 11. Spectrogram plot of the system's acceleration response (Hamming window, window length in samples: 140). [Reproduced with permission from Ref. 35. © Trans Tech Publications, 2005.]

6.2 Output-only linear identification of a real building

This paragraph examines an application of time–frequency instantaneous estimators to the identification of a real concrete building, namely, the Luciani Hospital of Caracas (Figure 15) [8].

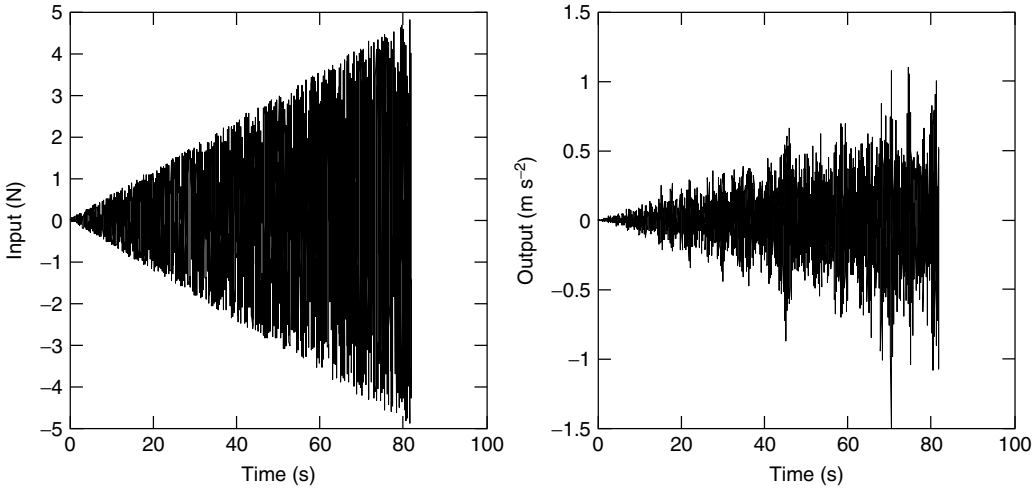


Figure 10. Excitation and response of the nonlinear system. [Reproduced with permission from Ref. 35. © Trans Tech Publications, 2005.]

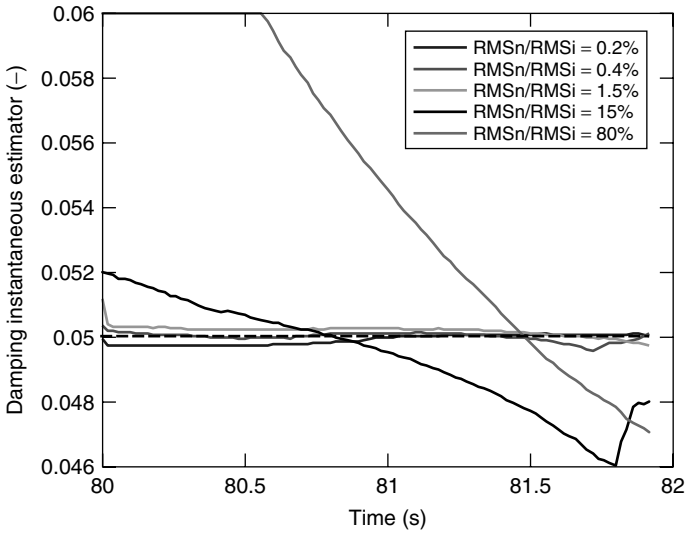


Figure 12. Nonlinear system identification: instantaneous estimator of damping for different RMS noise levels in the response. [Reproduced with permission from Ref. 35. © Trans Tech Publications, 2005.]

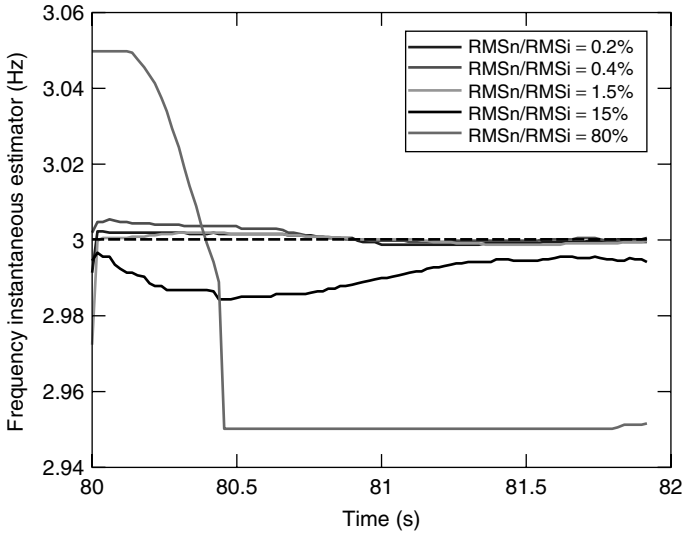


Figure 13. Nonlinear system identification: instantaneous estimator of frequency for different RMS noise levels in the response. [Reproduced with permission from aRef. 35. © Trans Tech Publications, 2005.]

On the basis of signals acquired in ambient excitation conditions, a preliminary structural identification was performed through a time-domain method, the ERA (eigensystem realization algorithm) [36] applied to random decrement functions [37] and a time–frequency instantaneous estimator (TFIE) method, based on equations (34) and (35) [28, 29]. For modal frequencies and shapes, the results yielded by the two methods were virtually the same.

Another type of dynamic test was performed by imposing an initial displacement of the structure with the aid of a hydraulic actuator positioned at roof level in an eccentric position. This type of excitation made it possible to work out, by means of current

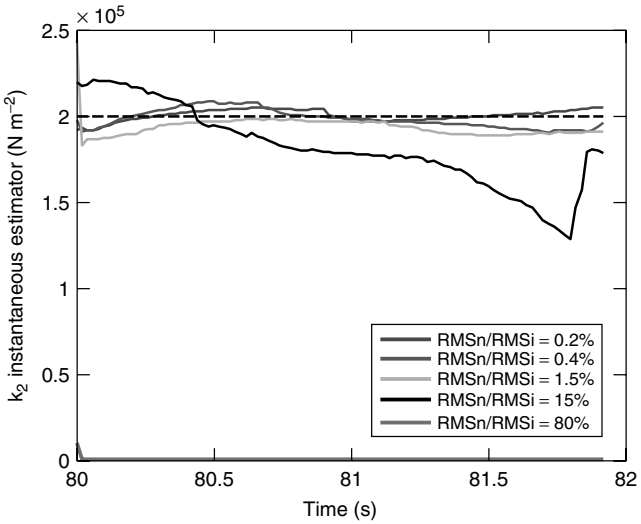


Figure 14. Nonlinear system identification: instantaneous estimator of quadratic stiffness coefficient k_2 for different rms noise levels in the response. [Reproduced with permission from Ref. 35. © Trans Tech Publications, 2005.]

Figure 15. View of Luciani Hospital in Caracas. [Reproduced with permission from Ref. 8. © Elsevier, 2004.]

techniques, a reliable assessment of damping that served as a useful term of comparison with the results yielded by the methods employed in ambient excitation conditions. It is known, in fact, that the latter situation is more critical, as it generally requires more complex identification tools leading to less accurate results.

With reference to Figure 9 and in view of the damping level expected, the analyzing window selected for a STFT analysis was a Hanning-type window with length = 100 samples. Figure 16 shows the SPEC of two signals measured along the two principal directions of the building.

The instantaneous identification was performed according to equations (42) and (43), with **p** vector containing all modal quantities α_k, f_k and ζ_k at a generic time $t = n\Delta t$

$$\varepsilon_x(n) = \left| \sum_{m=0}^{N-1} \left[\sum_k \sum_l \alpha_k \alpha_l^* H_k[m] H_l^*[m] - SPEC_x^{(\gamma)}[n,m] \right] \right| \longrightarrow \min_{\mathbf{p}} \qquad (51)$$

with $\mathbf{p} = \{\alpha_1 \quad \alpha_2 \quad .. \quad f_1 \quad f_2 \quad .. \quad \zeta_1 \quad \zeta_2 \quad ..\}$.

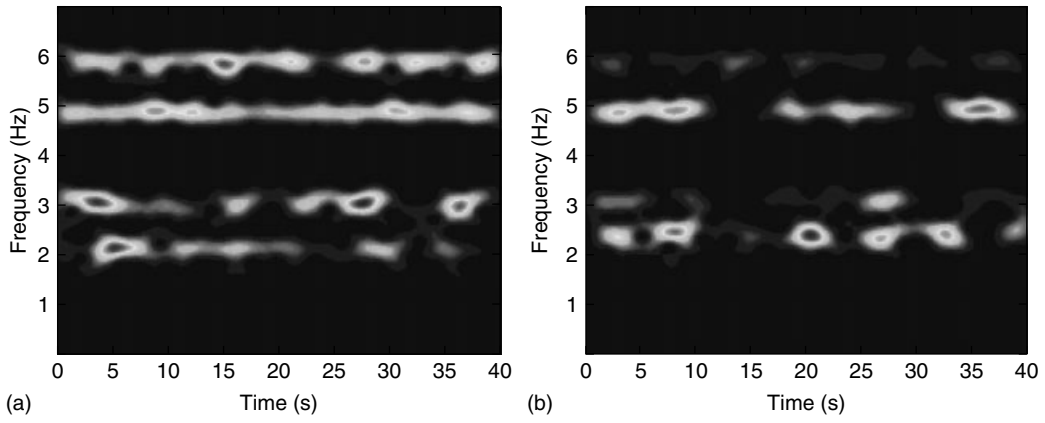


Figure 16. Structural identification of a large reinforced concrete building: spectrograms of two signals measured in (a) E–W and (b) N–S directions, respectively. [Reproduced with permission from Ref. 8. © Elsevier, 2004.]

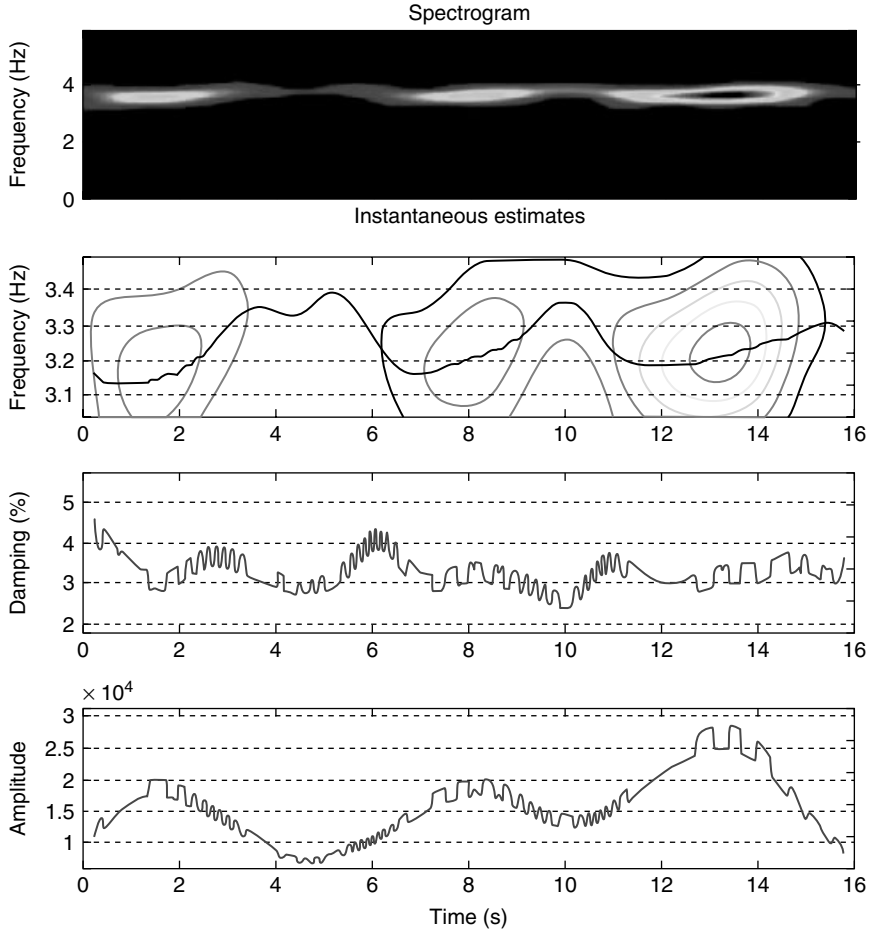


Figure 17. Instantaneous estimators calculated for the third vibration mode on a signal measured under ambient excitation. [Reproduced with permission from Ref. 8. © Elsevier, 2004.]

Figure 17, centered on the third identified mode, which involved a strong torsional component, shows the instantaneous estimators obtained from one of the signals relating to ambient excitation. In this figure, the typical SPEC representation is complemented by a contour plot showing the position of the "ridge". The identification of the ridge of a surface depends on its definition [12], in our case it is identified by the path of frequency, $f_3(n)$, as determined from equation (51). Other diagrams plot the punctual values of amplitude α_3 and equivalent viscous damping ζ_3, associated with the third vibration mode. We can see that the estimates turn out to be very good, especially in the time segments where the amplitude of the modal component to be identified is greater.

As a matter of fact, the estimation accuracy of equation (51) was seen to depend on the relative energetic importance of the modes. A possible solution can be the introduction of functions allowing for the relative energetic importance of the kth mode as a function of time. In this example, a final estimate for the kth modal damping resulted from a weighted average over different signals, based on the following weighting function:

$$w_k(t) = \frac{\alpha_k(t)|\tilde{H}_k(f_k)|}{\left[\sum_k \sum_l \alpha_k(t)\alpha_l^*(t)\tilde{H}_k(f_k)\tilde{H}_l^*(f_k)\right]} \quad (52)$$

Both in the instantaneous optimization (equation 51) and in weighted average (equation 52), parameters to be optimized were initialized with values of the preliminary identification. Results of damping identification, in ambient excitation and free decay conditions respectively, are summarized in Tables 4 and 5.

Table 4. Structural identification of Domingo Luciani Hospital building from ambient excitation signals: modal frequencies and damping identified with ERA and TFIE methods [8]

Mode	ERA		TFIE	
	Frequencies (Hz)	Damping (%)	Frequencies (Hz)	Damping (%)
1	2.37	2.51	2.35	2.76
2	2.70	2.18	2.68	2.97
3	3.25	3.79	3.30	3.21

Reproduced with permission from Ref. 8. © Elsevier, 2004.

Table 5. Results of identification procedure in free decay conditions [8]

Mode	Frequencies (Hz)	Damping (%)
1	2.35	2.71
2	2.68	2.78
3	3.21	3.15

Reproduced with permission from Ref. 8. © Elsevier, 2004.

We can conclude that in this application the accuracy afforded by the damping estimation method based on time–frequency instantaneous estimators was seen to be appreciably higher than the accuracy offered by current "output-only" identification methods. Conversely, the main drawback of time–frequency techniques is increased complexity and computational cost.

7 ACHIEVEMENTS AND PERSPECTIVES IN SHM APPLICATIONS

Difficulties encountered in SHM applications mostly arise due to the lack of accurate models to interpret the dynamic response of a system. More recently, however, diagnostic methods have been supported by data processing tools, which possess considerable potential in treating and correlating large amounts of data. Time–frequency analysis has, therefore, been used as an effective tool for practitioners, especially in dynamics, where it naturally provides information on the temporal behavior of vibrations.

Many successful applications of time–frequency tools to SHM have been reported. Most of them, including damage feature extraction, singularity detection, denoising, etc., circumvent modeling difficulties, since they do not require the system to be identified. The SHM strategy often reduces to the evaluation of symptoms reflecting the presence and the nature of the defect.

When a model or certain features of the system dynamics are known *a priori*, time–frequency algorithms admit their embedding into structural identification procedures. Accurate identification results, as those supplied by time–frequency methods, may constitute the basis for periodic and continuous SHM systems.

REFERENCES

[1] Cohen L. *Time-Frequency Analysis*. Prentice Hall: Englewood Cliffs, NJ, 1995.

[2] Hammond JK, Waters TP. Signal processing for experimental modal analysis. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 2001 **359**:41–59.

[3] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DV, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, LA-13976-MS. Los Alamos National Laboratory Report, 2003.

[4] Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 2006 **20**:1483–1510.

[5] Hlawatsch F, Boudreaux-Bartels GF. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine* 1992 **9**(2):21–67.

[6] Auger F, Flandrin P, Goncalves P, Lemoine O. *Time-Frequency Toolbox for MATLAB*. Freeware available at the URL http://tftb.nongnu.org, 1998.

[7] Spina D, Valente C, Tomlinson GR. A new procedure for detecting nonlinearity from transient data using Gabor transform. *Nonlinear Dynamics* 1996 **11**:235–254.

[8] Ceravolo R. Use of instantaneous estimators for the evaluation of structural damping. *Journal of Sound and Vibration* 2004 **274**:385–401.

[9] Gabor D. Theory of communication. *Journal of the IEE (London)* 1946 **93**:429–457.

[10] Feichtinger HH, Strohmer T (eds). *Gabor Analysis and Algorithms: Theory and Applications*. Birkhäuser, 1998.

[11] Carmona R, Hwang WL, Torrésani B. *Practical Time-Frequency Analysis*. Academic Press: New York, 1998.

[12] Newland DE. Ridge and phase identification in the frequency analysis of transient signals by harmonic wavelets. *Journal of Vibration and Acoustics* 1999 **121**:149–155.

[13] Erlicher S, Argoul P. Modal identification of linear non-proportionally damped systems by wavelet transform. *Mechanical Systems and Signal Processing* 2007 **21**:1386–1421.

[14] Loughlin PJ, Pitton JW, Atlas LE. Bilinear time-frequency representations: new insights and properties. *IEEE Transactions on Signal Processing* 1993 **41**:750–766.

[15] Claasen TA, Mecklenbrauker WF. The Wigner distribution—a tool for time-frequency analysis; part III: relations with other time-frequency signal transformations. *Philips Journal of Research* 1980 **35**:372–389.

[16] Mecklenbrauker WF, Hlawatsch F (eds). *The Wigner Distribution: Theory and Applications in Signal Processing*. Elsevier: Amsterdam, 1997.

[17] Bonato P, Ceravolo R, De Stefano A, Knaflitz M. Bilinear time-frequency trasformations in the analysis of damaged structures. *Mechanical Systems and Signal Processing* 1997 **11**:509–527.

[18] Boashash B (ed). *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*. Elsevier: Oxford, 2003.

[19] Martin W, Flandrin P. Wigner-Ville spectral analysis of non-stationary processes. *IEEE Transactions on Signal Processing* 1989 **33**:1461–1470.

[20] Flandrin P, Martin M. The Wigner-Ville spectrum of nonstationary random signals. In *The Wigner Distribution: Theory and Applications in Signal Processing*, Mecklenbrauker WF, Hlawatsch F (eds). Elsevier: Amsterdam, 1997, pp. 211–267.

[21] Janssen AJ. Positivity and spread of bilinear time-frequency distributions. In *The Wigner Distribution: Theory and Applications in Signal Processing*, Mecklenbrauker WF, Hlawatsch F (eds). Elsevier: Amsterdam, 1997, pp. 1–58.

[22] Mallat S, Papanicolaou G, Zhang Z. Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics* 1998 **28**:1–47.

[23] Dahlhaus R. Fitting time series models to nonstationary processes. *The Annals of Statistics* 1997 **25**:1–37.

[24] Kozek W. Time-frequency signal processing based on the Wigner-Weyl framework. *Signal Processing* 1996 **29**:77–92.

[25] Priestley MB. Power spectral analysis of non-stationary random processes. *Journal of Sound and Vibration* 1967 **6**:86–97.

[26] Hammond JK, White PR. The analysis of non-stationary signals using time-frequency methods. *Journal of Sound and Vibration* 1996 **190**:419–447.

[27] Dalianis SA, Hammond JK, White PR, Cambourakis GB. Simulation and identification of nonstationary systems using linear time-frequency methods. *Journal of Vibration and Control* 1998 **4**:75–91.

[28] Bonato P, Ceravolo R, De Stefano A, Molinari F. Use of cross time-frequency estimators for the

structural identification in non-stationary conditions and under unknown excitation. *Journal of Sound and Vibration* 2000 **237**:775–791.

[29] Bonato P, Ceravolo R, De Stefano A, Molinari F. Cross time-frequency techniques for the identification of masonry buildings. *Mechanical Systems and Signal Processing* 2000 **14**:91–109.

[30] Choi HI, Williams WJ. Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Transactions on Acoustics Speech and Signal Processing* 1989 **37**:862–871.

[31] Cohen L. Time-frequency distribution: a review. *Proceedings of the IEEE* 1989 **77**:941–981.

[32] Boudreaux-Bartels GF, Parks TW. Time-varying filtering and signal estimation using Wigner distribution synthesis techniques. *IEEE Transactions on Acoustics Speech and Signal Processing* 1986 **34**:442–451.

[33] Ceravolo R, Demarie GV, Erlicher S. Instantaneous identification of Bouc-Wen-type hysteretic systems from seismic response data. *Key Engineering Materials* 2007 **347**:331–338.

[34] Worden K, Tomlinson GR. *Nonlinearity in Structural Dynamics: Detection, Identification and Modelling*. Institute of Physics Publishing: Bristol, Philadelphia, PA, 2001.

[35] Demarie GV, Ceravolo R, De Stefano A. Instantaneous identification of polynomial nonlinearity based on Volterra series representation. *Key Engineering Materials* 2005 **293–294**:703–710.

[36] Juang JN. *Applied System Identification*. Prentice Hall: Englewood Cliffs, NJ, 1994.

[37] Cole HA. *On-the-line Analysis of Random Decrement Vibrations*, AIAA Paper 68–288, 1968.

Chapter 27
Wavelet Analysis

Amy N. Robertson[1] and Biswajit Basu[2]

[1] *IMTECH, Boulder, CO, USA*
[2] *Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Dublin, Ireland*

1 INTRODUCTION

Time–frequency analysis provides information on how the spectral content of a signal evolves in time, which is important when analyzing and interpreting nonstationary signals (*see also* **Time–frequency Analysis**; **Damage Detection Using the Hilbert–Huang Transform**). The wavelet transform (WT) has been gaining in popularity as the time–frequency transform of choice since its first influx into mainstream mathematics and engineering in the early 1990s. A WT is an integral transform. Hence, like Fourier transforms, it decomposes a signal into its constituent parts using a set of basis functions. However, unlike Fourier transforms, the choice of the bases is not limited to sinusoids. Sinusoids are not localized in time and thus, a Fourier transform is unable to extract time–frequency information from signals. To achieve this, a short-time Fourier transform (STFT) is used as a variation of the conventional Fourier transform. It uses a fixed window time function to obtain information on how the frequencies in the signal are changing in time. Wavelet basis functions, on the other hand, are localized both in time and frequency and can be dilated (compressed or stretched) and translated along a signal to extract both time and frequency information.

The advantage of the WT over the STFT is the ability to achieve flexible windowing time corresponding to the desired frequency information. Higher frequencies have better time resolution and lower frequencies have better frequency resolution (Figure 1). Variable resolution is often beneficial as it allows the low-frequency components, which usually give a signal its main characteristics or identity (such as modes), to be distinguished from one another in terms of frequency content, while providing an excellent temporal resolution of the high-frequency components, which add the nuances to the signal's behavior.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

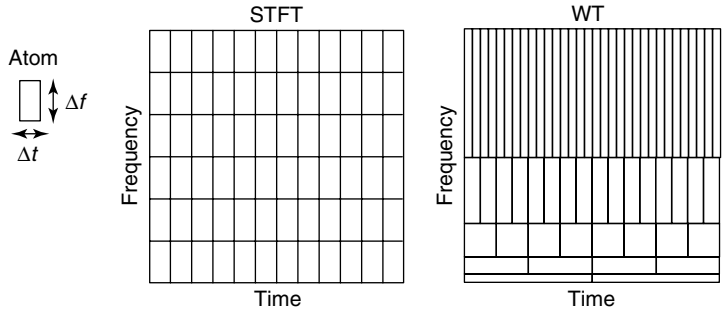


Figure 1. Time–frequency atoms of the STFT and WT.

The WT is readily applicable to the field of structural health monitoring (SHM) because of its time–frequency localization property. SHM involves the observation of a system over time/space using periodically sampled dynamic response measurements from an array of sensors. There are two different ways of assessing the health of the system: the knowledge-based approach and the signal-based approach.

The knowledge-based approach involves creating a model of the system under analysis. Damage is then identified by a change in this model. The basic premise is that damage will alter the stiffness, mass, or energy dissipation properties of the system, which, in turn, will alter the dynamic response of the system. Most of the research under this approach leads to system identification. The signal-based approach relies on the extraction of damage-sensitive features from the measured response signals. Analysis and processing of these features are then used to determine the current state of health of the system. This category includes research related to advanced signal processing, statistical analysis, and pattern recognition. The next section summarizes some of the papers that describe the use of wavelet analysis in SHM using both the knowledge-based and the signal-based approaches.

2 REVIEW OF LITERATURE

Recently, time–frequency and timescale analysis tools, particularly the wavelet analysis technique, have been proven to be powerful methods for assessment of structural health and fault monitoring. Early investigations by Staszewski and Tomlinson [1] and Wang and McFadden [2] in this area used wavelet analysis to detect local faults in machineries. A continuous wavelet transform (CWT) with a Morlet basis function was used for the studies. Several other researchers have carried out the investigations on the detection of damage based on changes in wavelet coefficients of time-domain responses. Melhem and Kim [3] carried out research to detect damage on full-scale concrete structures using a CWT and wavelet ridges. Kim and Melhem [4] have focused on reviewing the research carried out on damage detection in beams, mechanical gears, and rollers by the discrete wavelet transform (DWT) and CWT, using a Gabor, a low-oscillation Gaussian, and a Mexican hat basis function. Hou *et al.* [5] studied accumulated damage occurrences due to the San Fernando earthquake excitations, and identified the change in structural stiffness from spikes in the wavelet coefficients using multiresolution analysis (MRA). A DWT was used by Moyo and Brownjohn [6] to detect the anomalous behavior of a bridge structure from the abrupt changes in the wavelet coefficients. Damage assessment of bridges based on energy calculation was carried out by Sun and Chang [7] using a wavelet packet transform and neural networks. A CWT with a Littlewood–Paley (L–P) basis was used by Basu [8] to detect stiffness degradation in structures, and by Goggins *et al.* [9] to study the seismic response of concentrically braced frames, by correlating the wavelet coefficients at different scales. Gurley *et al.* [10] detected first and higher order correlation by using a CWT to construct filtered wavelet coherence and bicoherence maps to monitor offshore structures.

There has also been significant development, in the recent years, on wavelet-based damage detection associated with identifying singularities in the

response signal either in time or in space, or any of their derivatives. Singularity detection through wavelets has been discussed in detail by Mallat [11]. Some of the key research in this area has been carried out by Robertson *et al.* [12], who identified a rattle in a structure using wavelet-based Holder exponents. Damage detection in beams using spatial wavelet analysis is also a problem of singularity detection. Gentile and Messina [13] have investigated the selection criteria for a wavelet basis function in the presence of measurement noise and demonstrated the performance of Gaussian and Symlet wavelets. Okafor and Dutta [14], Chang and Chen [15], Loutridis *et al.* [16], Liew and Wang [17], and Wang and Deng [18] have used the spatial response data for beam structures to detect damage by applying different wavelets (both the CWT and DWT). Spanos *et al.* [19] have used a CWT modulus map for spatial wavelet analysis in damaged beams by suppressing boundary effects. Zhu and Law [20] have diagnosed damage in bridges without interrupting operations by using the CWT and relating the damage to the variation in the wavelet coefficient. Other researches related to spatial damage identification include those by Patsias and Staszewski [21] on the use of video camera–based dynamic shape identification, Lam *et al.* [22] on a Bayesian approach for detection of crack in the presence of an obstruction, Kim *et al.* [23] on the use of 2-D MRA-based Haar wavelets on plates, and Pakrashi *et al.* [24] on the wavelet–kurtosis-based damage detection and calibration, showing the effectiveness of the Coiflet 4 wavelet. Most of the research work discussed so far broadly covers the category of the signal-based approach to SHM, since it involves signal processing with pattern recognition/feature identification.

Wavelet analysis has also been used for identification of systems (knowledge-based approach) including nonstationary and nonlinear systems. Staszewski [25, 26], Kyprianou and Staszewski [27] have used the CWT to identify nonlinear systems and damping in dynamical systems. Other applications of the CWT include identification of modal parameters by Ruzzene *et al.* [28], extraction of impulse response functions by Robertson *et al.* [29], and identification of dynamic modal parameters of a bridge by Piombo *et al.* [30]. Identification of parameters of time-varying systems has been carried out by Ghanem and Romeo [31] using an orthogonal DWT. Basu *et al.* [32] have used WT packets to identify stiffness variation in structural systems.

A different approach was used by Tian *et al.* [33], who analyzed the flexural waves in beams using a CWT (Morlet wavelet), and detected damage by considering the lag of the waves. Other papers include the work by Staszewski *et al.* [34] on damage detection based on Lamb waves using both the CWT and orthogonal DWT, and a recent review on time–frequency and timescale analysis for SHM by Staszewski and Robertson [35].

3 TYPES OF WAVELET TRANSFORMS

The WT is a timescale transform, which provides information about how the spectral content of a signal is changing in time. WTs can be categorized into three main types: the continuous, discrete, and orthogonal transform.

3.1 Continuous wavelet transform

The CWT, $WT(u, s)$, is obtained by convolving the signal $f(t)$ with a set of basis functions created from the translations (u) and dilations (s) of a mother wavelet Ψ:

$$WT(u, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t)\psi^* \left(\frac{t-u}{s} \right) \mathrm{d}t \quad (1)$$

These basis functions are defined as

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi \left(\frac{t-u}{s} \right) \quad (2)$$

where u defines the time shift and s defines the scale.

The analyzing functions or "basis functions" of a WT are called *wavelets*. The wavelets are scaled to obtain a range of frequencies. They are also translated to provide the time information in the transform. A typical wavelet basis with translated and dilated versions is shown in Figure 2. This wavelet is called a *Mexican hat*, and is obtained by the second derivative of the Gaussian function. The WT works as a filter, allowing only a certain time and frequency content through. Any given atom in the time–frequency map of the WT (Figure 1) represents the correlation

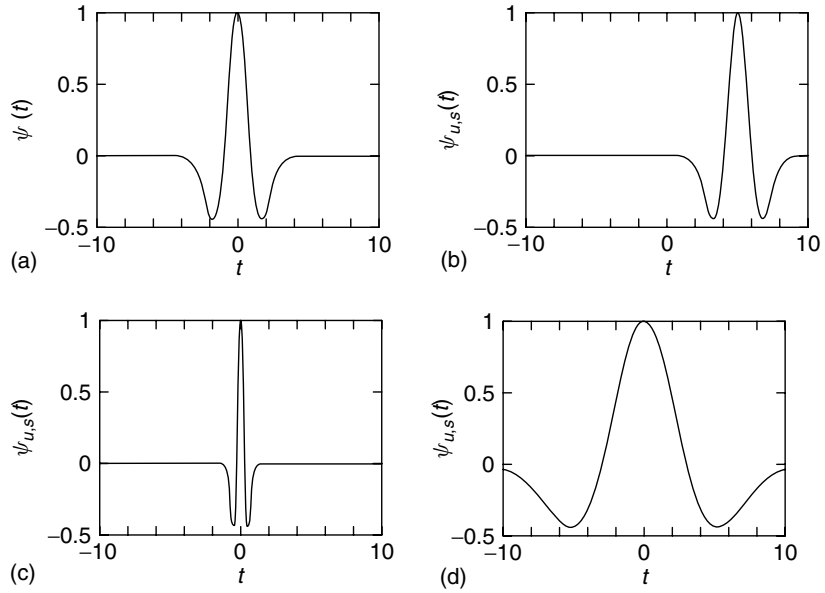


Figure 2. Dilations and translation of a mother wavelet. (a) A typical mother wavelet, (b) translated wavelet, (c) dilated wavelet (compressed) and (d) dilated wavelet (stretched).

between the wavelet basis function at that frequency dilation and the signal in that time segment. The frequency content of the WT is represented in terms of scales, which are inversely related to frequencies. The squared amplitude of the CWT is therefore called the *scalogram*. The relationship between scales and frequencies is easily found and a time–frequency map can be formed from the scalogram.

Since the WT works in a manner similar to the STFT, by convolving the signal with a function that varies both in time and frequency, it suffers from similar limitations in the resolution of the time–frequency map. Both transforms are confined by the uncertainty principle, which limits the area of a time–frequency atom in the overall time–frequency map (Figure 1). The biggest difference between the two transforms is that the atoms in the WT map are not a constant shape. In the lower frequencies, the atoms are fatter, providing a better resolution in frequency and worse resolution in time, whereas in the upper frequencies the atoms are taller, providing better time resolution and worse frequency resolution. This variable resolution can be advantageous in the analysis of structural time response data.

The CWT gets its name from the fact that the mother wavelet is continuously shifted across the length of the data being analyzed. This smooth shifting means that the time/frequency atoms shown in Figure 1 overlap one another, providing redundant information. A representation of this overlapping in the scales/frequencies is shown in Figure 3.

The variable windowing feature of wavelet analysis leads to an important property exhibiting constant-Q factor (defined as the ratio of the center frequency to bandwidth) analysis. For STFT, at an analyzing frequency ω_0, changing the window width increases or decreases the number of cycles of ω_0 inside the window. In the case of WTs, with the change in window width, mean dilation or compression of the wavelet function changes. Hence, the carrier frequency becomes ω_0/a, for a window width changing from T to aT. However, the number of cycles inside the window remains constant.

The frequency resolution is proportional to the window width both in the case of STFT and WT. However, for WT, a center frequency shift necessarily accompanies a window width change (timescaling). Thus, Q factor is invariant with respect to wavelet dilation and these dilated wavelets may be considered

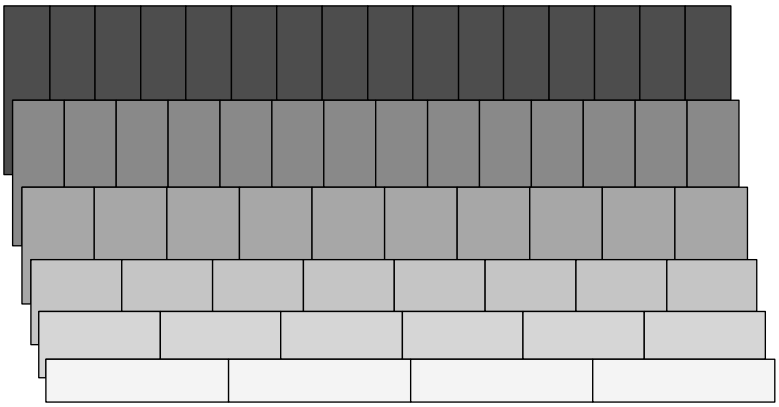


Figure 3. Continuous WT.

as constant-Q band-pass filters giving rise to the frequency selectivity of the CWT.

Since the WT is an alternative representation of a signal, it should retain the characteristics of the signal including the energy content in the signal. Thus, there should exist a similar relation to Parseval's theorem, which provides the energy relationship in the Fourier domain. The total energy of a signal in wavelet domain representation is [36]

$$E_f = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |WT(u,s)|^2 \frac{\mathrm{d}s\, \mathrm{d}u}{s^2} \qquad (3)$$

where, C_ψ is a scalar constant related to the Fourier transform of the wavelet basis (called *admissibility constant*). The wavelet basis functions can be normalized in a way such that it can attain a value of unity. The differential energy of the signal in the differential tile of scale-translation plane in wavelet domain is $|WT(u,s)|^2(\mathrm{d}s\, \mathrm{d}u/s^2)$, which leads to the construction of the scalogram.

3.2 Discrete wavelet transform

The major difference between the CWT and DWT comes in the way that the transforms are performed. With the CWT, the signal is projected on to a continuous set of frequency bands; and, the original can be reconstructed from all the frequency components. The DWT discretizes the time and scale parameters to eliminate redundancy while retaining the ability to recover the original signal without loss. This means that the discrete transform obtains only a subset of the time/scale wavelet atoms that are found using the CWT, thus providing a more compact representation of the signal.

The discrete grid on the timescale plane corresponds to a discrete set of continuous functions. The higher the subsampling of the grid, the smaller the redundancy of the WT. Dyadic discretization is one way of obtaining a countable subset, and is the most commonly used. This is the discretization shown in Figure 1.

The DWT is defined at discrete points j, k using the formula

$$DWT(j,k) = \int_{-\infty}^{\infty} f(t)\psi_{j,k}^*(t)\, \mathrm{d}t \qquad (4)$$

where

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - kus_0^j}{s_0^j}\right) \qquad (5)$$

If s_0 is chosen to be two, then the frequency sampling is called *dyadic*.

3.2.1 Multiresolution analysis

A more common way of representing the DWT is through an MRA approach. This approach was first suggested by Mallat [11], as a fast algorithm for performing the WT. The MRA approach can also be used to examine a signal at different resolutions or scales.

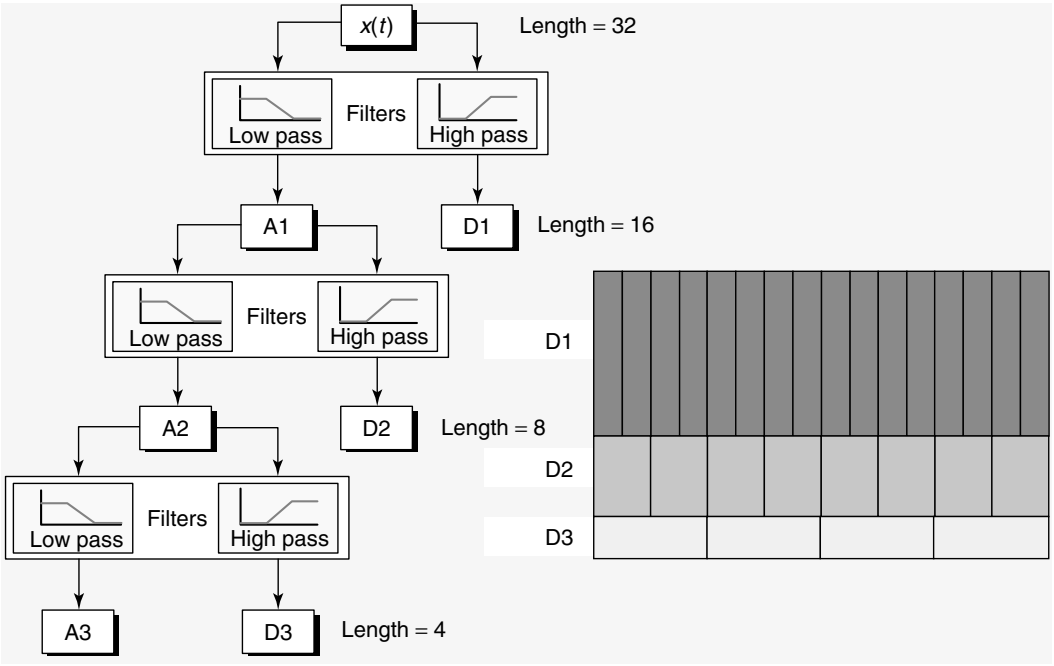


Figure 4. Filtering approach used for performing DWT.

The process works as shown in Figure 4. The original signal, $x(t)$, passes through two complementary filters, a low-pass and a high-pass filter, resulting in two signals. The signal, a, that emerges from the low-pass filter contains the low-frequency content, or approximations of the signal. The signal, d, that emerges from the high-pass filter contains the high-frequency content, or details of the signal. This procedure is then repeated on the signal a, thus removing more of the high-frequency content with each application. The result of the successive applications of the filters is a series of approximations of the signal with progressively more of the detail removed. Also obtained are a series of details of the signal, which constitutes the DWT, as demonstrated in Figure 4. Each application of the filters results in a different "level" of resolution, and these levels can be examined independently or in comparison for the extraction of damage-sensitive features.

The filters are based on both a mother and a father wavelet. The father wavelet is used as a low-pass filter or scaling function, defined as

$$\phi_{j,k} = 2^{-j/2}\phi\left(\frac{t-2^j k}{2^j}\right) \tag{6}$$

while the mother wavelet is used as the high-pass filter, and is defined as

$$\psi_{j,k} = 2^{-j/2}\psi\left(\frac{t-2^j k}{2^j}\right) \tag{7}$$

The j indexes the scale and the k indexes the translation. These functions are used to compute the approximations and details as

$$a_{j,k} = \int x(t)\phi_{j,k}(t)\,\mathrm{d}t \tag{8}$$

$$d_{j,k} = \int x(t)\psi_{j,k}(t)\,\mathrm{d}t \tag{9}$$

A signal, $x(t)$, can then be represented as a weighted sum of the scaling and wavelet functions up to a scale J:

$$x(t) = \sum_k a_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k \mathrm{d}_{j,k}\psi_{j,k}(t) \tag{10}$$

3.3 Orthogonal wavelet transform

An orthogonal WT is created from a wavelet whose dilations and translations form a set of orthogonal basis functions. When an orthonormal set of basis functions are used, there is no redundancy in the transform. A lack of redundancy means that the transformed signal can be recovered with the least information possible.

Orthogonality means that the inner product $\langle \psi_{j,k}, \psi_{j',k'} \rangle$ satisfies the condition [37]

$$\langle \psi_{j,k}, \psi_{j',k'} \rangle >= \begin{cases} 1 \text{ for } j = j' \text{ and } k = k' \\ 0 \text{ otherwise} \end{cases} \quad (11)$$

By choosing $a_0 = 2$ and $b_0 = 1$, a discretization based on a dyadic grid is obtained. The wavelets therefore assume the form

$$\psi_{j,k}(t) = 2^{-j/2} \psi \left(2^{-j} t - k\right) \quad (12)$$

The discretization on the dyadic grid leads to the dyadic WT. If there is a wavelet function that yields an orthonormal wavelet decomposition, then it can be shown that the wavelet coefficients for an orthonormal DWT are simply the samples of the CWT on the dyadic grid.

Nonorthogonal wavelet functions tend to artificially add in energy (due to the overlap) and require renormalization to conserve the information. The major drawback of using orthogonal wavelets is that localization of signal features is not possible.

3.4 Wavelet packets

While the constant-Q factor and coarser frequency resolution at high frequencies make the wavelet analysis computationally efficient, this may be a disadvantage for analysis of some signals for SHM. Better resolution at high frequencies can be obtained by wavelet packet construction.

The DWT based on MRA splits the signal into two bands, a higher band (by using a high-pass filter) and a lower band (by using a low-pass filter). The lower band is subsequently again split in two bands. This concept can be generalized by splitting the signal into several bands each time. In addition, there could be further splitting of higher bands too, not just the lower band. This generalization of MRA produces outputs called *wavelet packets*. This is a deviation from constant-Q analysis and achieves desired frequency resolution at high-frequency bands. Wavelet packets through arbitrary band splitting can choose the most suitable resolution to represent a signal.

The resolution of signals with wavelet packets is not only possible using MRA-based frequency filters in the time domain (starting with Haar wavelets) but also in the frequency domain. For the arbitrary resolution using frequency-domain-based filters, the construction for wavelet packets should be based on a modified L–P wavelet basis. The application of wavelet packets is particularly useful in system identification and damage detection for SHM where finer resolution at higher frequency is desired.

4 CHOICE OF WAVELET METHOD

4.1 Type of transform

The redundancy and time efficiency of the continuous wavelet transform and DWTs largely dictate what applications they are best used for. The CWT offers improved resolution, which creates an increase in computational time and memory required to calculate the coefficients. Therefore, the CWT is better used for analysis where one is trying to draw out features in a signal. The DWT is very time efficient, even faster than the fast Fourier transform (FFT). With improved time efficiency, however, comes a very sparse representation of the signal, which is not good for finding features in a signal. It is best used for compression (storage), denoising, and fast calculations such as matrix multiplications. Fast calculations can be very advantageous when trying to perform on-line health monitoring.

4.2 Mother wavelet

Each type of WT requires the choice of a mother wavelet from which the basis functions are created. There are an infinite number of mother wavelets that can be used. To be classified as a mother wavelet, the function must first satisfy the following conditions:

1. The function must have zero mean: $\int_{-\infty}^{\infty} |\psi(t)| \, dt = 0$.
2. Admissibility condition: $\int_0^{\infty} (|\hat{\psi}(\omega)|^2/\omega) \, d\omega < \infty$, where $\hat{\psi}(\omega)$ is the Fourier transform of ψ.

Basically, these properties mean that a wavelet is a function that must oscillate (hence the term *wave*), and that the oscillations must eventually damp out to zero.

Any given atom in the time–frequency map of the WT (Figure 1) represents the correlation between the wavelet basis function at that frequency dilation and the signal in that time segment. Therefore, if a basis function is used that is very similar to a feature of interest, the correlation between the wavelet and the signal at the point where the feature exists is very high. This is how wavelets can be used to draw out features in a signal. One can actually tailor a wavelet to resemble a feature of interest.

The most basic wavelet is the Gabor wavelet. It is simply a Fourier transform, which uses a Gaussian window to obtain time information [11]:

$$\psi(t) = g(t) \, e^{i\eta t} \tag{13}$$

where $g(t) = (1/(\sigma^2\pi)^{1/4}) e^{(-t^2/2\sigma^2)}$ is the Gaussian window.

The choice of which mother wavelet to use for analysis can greatly affect the results. For compression or denoising, it is important to choose a wavelet that closely resembles the function that is being analyzed. If this is achieved, only very few of the wavelet coefficients will be nonzero. If the function is smooth, a higher order wavelet is more appropriate. A lower order wavelet (such as the Haar wavelet) is good for identifying discontinuities in a signal. For feature extraction, one sometimes uses a wavelet that resembles the feature of interest.

When trying to obtain a good time–frequency representation of a signal, a continuous wavelet is usually the best. The most commonly used continuous wavelet is the Grossman–Morlet wavelet. This wavelet is complex valued. A complex-valued wavelet can be helpful when performing a convolution since it has the same convolution properties as the Fourier transform, meaning that a convolution becomes just a simple multiplication. A complex-valued wavelet can also provide phase information about the signal. The Grossman–Morlet wavelet is similar to a sine function with a Gaussian window:

$$\psi(t) = e^{(-\sigma^2 t^2 - i2\pi f_0 t)} \tag{14}$$

The term f_0 is the center frequency of the sinusoid and σ determines the width of the frequency band. Also, t is the time variable and i the imaginary value $\sqrt{-1}$.

Orthogonal wavelets offer a representation of the signal with as little information as is possible, and without redundancy. These features are very useful in denoising and in matrix multiplications. The best-known orthogonal wavelets are the Daubechies wavelets. Daubechies found a countable set of dyadic wavelets for the orthogonal decomposition that also has compact support. Other orthogonal wavelets include Coiflet, Symlet, and B-Spline. Battle–Lemarie and Meyer wavelets have orthonormal bases. Biorthogonal wavelets are special cases of discrete wavelets. They consist of two wavelet sets: one for decomposition and another for reconstruction.

Other features of the wavelet that are important in choosing the correct one are regularity, vanishing moments, and time–frequency localization. These features, however, are beyond the scope of this article. As stated by Staszewski [37], the proper use and interpretation are far more important in wavelet analysis than the choice of wavelet.

4.3 Wavelet features

After the type of WT has been decided upon, the task of interpreting the wavelet data arises. Many different approaches exist, and are based on the type of structure being examined and the type of damage that is being monitored. This section goes over some of the approaches and features that can be extracted from the wavelet data.

When using a CWT, damage-sensitive features are usually extracted from the scalogram. The scalogram is the squared modulus of the WT, represented in a two-dimensional plot. For a structural system, most of the energy in the scalogram is concentrated at the modes of vibration. One can also examine the power spectrum of the WT, which is the scalogram integrated over time. The power spectrum is similar

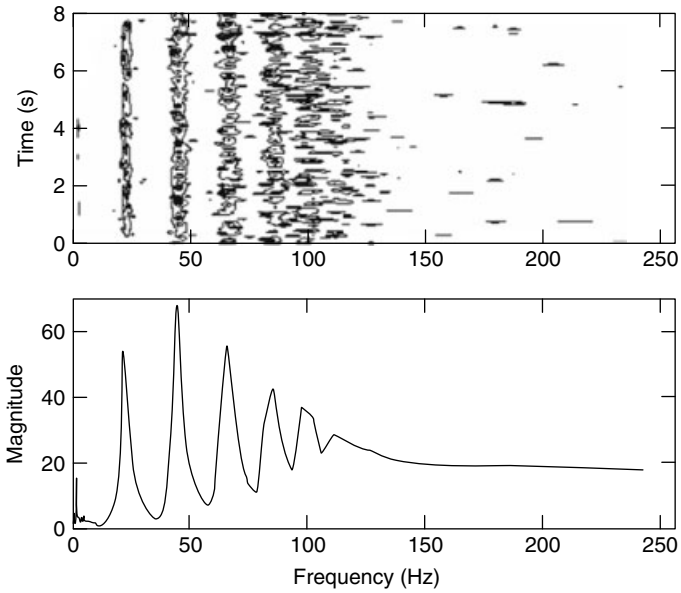


Figure 5. Wavelet transform and scalogram of an 8-DOF system.

to an FFT-based power spectrum and provides a one-dimensional view of the energy distribution in the signal across the frequencies. The variable window size of the WT means that the resolution of the power spectrum is greater at the low frequencies, which can be useful in identifying the modes of a structure. Figure 5 shows the scalogram and wavelet-based power spectrum of the time response of an 8-DOF (degree of freedom) spring-mass system subjected to random excitation. Notice that the structural modes in the lower frequencies are easily identifiable as vertical lines in the scalogram, whereas the higher frequency modes are indistinguishable owing to both the resolution of the WT in the high frequencies and the sampling rate of the data.

Sometimes the modes of a system change during time. This variation is easily seen in a time–frequency representation of the structural response. Methods have been developed to extract the needed data from the wavelet scalogram to exactly determine how the modes are changing. The concept of ridges and skeletons of the WT was developed by Tchamitchian and Torresani [38], and involves the extraction of the instantaneous amplitude and frequency content from a signal. A ridge is a string of high-magnitude wavelet coefficients, such as the ones seen in Figure 5, which represent a mode of the structure. If a mode is changing in time, the ridge is no longer a straight line, but it is curved. The skeleton is then the values of the WT restricted to its ridge. By examining the ridges and skeletons, one can identify information about the amplitude and frequency modulation of a signal, which is useful for the identification of nonstationary and nonlinear systems, and for damping estimates [25, 26].

The presence of spikes in the high-frequency region of a scalogram can indicate the presence of a discontinuity in the signal. Mallat [11] first introduced a method for detecting singularities in a signal by examining the evolution of the modulus maxima (large magnitude coefficients) of the WT across the scales. The decay of this maxima line can be used to determine the regularity (Holder exponent) of the signal at the given time point. A signal with low regularity can indicate a discontinuity in the signal, which can be caused by the opening and closing of cracks or by a rattle resulting from the loss of preload in a bolted connection (see [29] for further discussion).

When using a DWT or an orthogonal WT, damage is found by examining either the scalogram or the individual levels of the WT, as shown in the multiresolution approach. The process of denoising, for example, involves retaining only those values in the DWT scalogram that are above a certain value. Noise

does correlate well with the wavelet basis functions and therefore shows up as low-level values in the scalogram. Inverse wavelet transforming the retained values results in a time signal with the noise removed. See the examples below for a demonstration of how the DWT levels can be used for damage identification.

5 EXAMPLES OF WAVELET ANALYSIS

5.1 Continuous wavelet transform

Identifying and quantifying the level of damage in composite materials is becoming a large area of research in SHM. The most common forms of damage for composites include matrix breakage, debonding between matrix and fibers, and fiber delamination. Quite often, damage is not readily visible and must be identified using nondestructive evaluation techniques (NDTs). A number of different NDT approaches exist, including acoustic emission, electrical impedance, and Lamb wave propagation.

In this example, Lamb wave propagation is used to detect the presence of a delamination in a carbon fiber composite plate. The Lamb wave is a high-frequency (megahertz)-guided longitudinal wave that can travel long distances through a material. By propagating such waves through composite materials, faults or damage in the material can be located. The Lamb wave propagation method works as follows. An array of Piezoelectric Transducers (PZT) sensors are placed on the plate as shown in Figure 6. One PZT is activated as an actuator to launch elastic waves (the input) through the plate, and responses are measured by other PZT patches acting as sensors or possibly some other device. The structure can be systematically surveyed by sequentially using each of the PZT patches as an actuator and the remaining PZT patches as sensors. The technique looks for the possibility of damage by tracking changes in the PZT responses from a baseline "undamaged" signal. Features that might indicate damage include signal attenuation, delay, or reflection. These features are difficult to locate by examining the time response signals, and so tools such as the WT are used to draw out these features.

Staszewski [37] provides a wonderful example of wavelet analysis of Lamb wave data. Lamb waves were launched into a 600-mm^2 plate using conventional PZTs, and monitored using an embedded optical fiber connected to a Mach–Zehnder interferometer [37]. Figure 7(a) shows normalized S_0 Lamb wave responses along with power spectra and

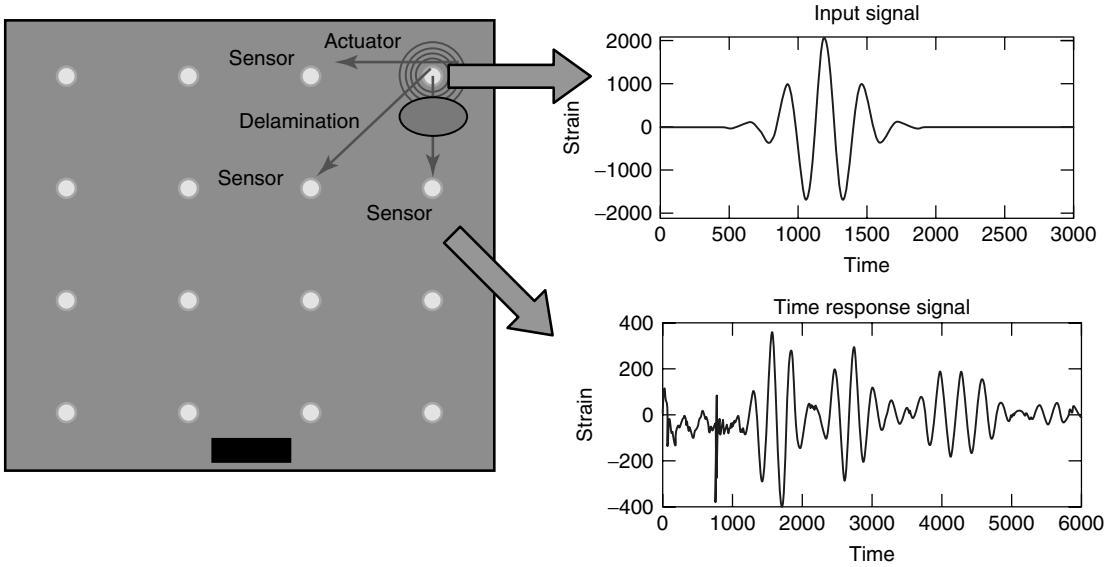


Figure 6. Example of Lamb wave propagation for damage diagnosis: composite plate under examination and input/output signals.

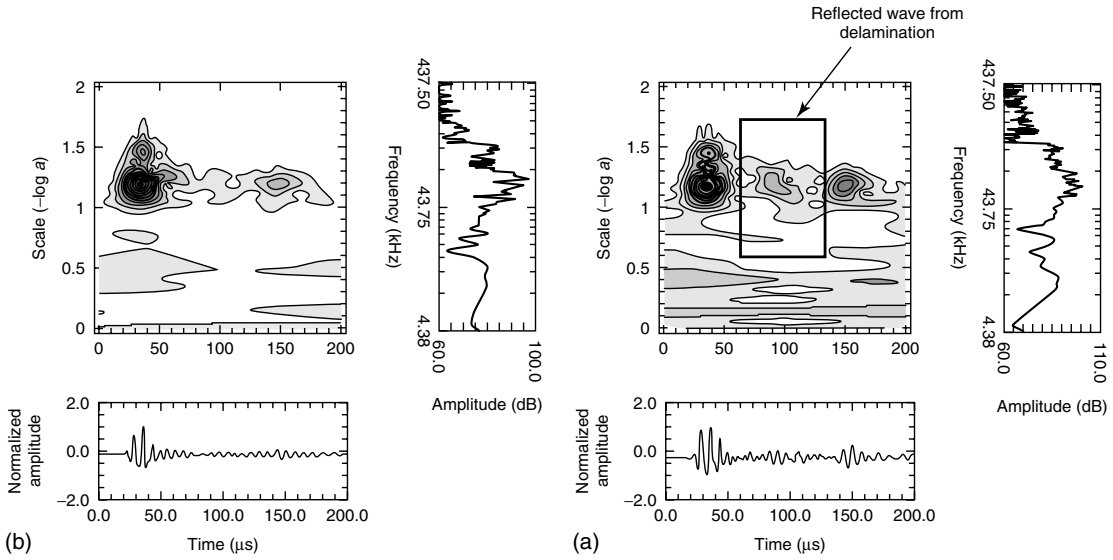


Figure 7. Continuous wavelet transform analysis of Lamb wave reflection data from a composite plate: (a) no damage present and (b) delamination present. (Staszewski [37], Reproduced by permission of Academic Press.)

contour plots of the WT modulus for a baseline "undamaged" signal. Figure 7(b) then shows the same plots for data obtained from the same composite plate with a delamination present. In Figure 7(b), a wave reflection is seen at about 90 μs as identified by an increase in the WT modulus. The reflection is due to a 400-mm^2 delamination in the plate, which is about 0.1% of the total area. The visibility of the delamination is significantly improved by the WT. A variety of methods can then be used to try to automatically identify the changes in the wavelet modulus data.

5.2 Orthogonal discrete wavelet transform

As shown previously, a signal $x(t)$ can be decomposed using a DWT into a series of resolution levels. By examining an individual level, one can identify the contribution of a given frequency band to the overall signal energy. The lower levels correspond to the lower frequencies and have the smallest bandwidth. At each increasing level, the central frequency doubles, as well as the bandwidth. The central frequency at level j is

$$f_{c_j} = 2^j \frac{f_s}{N} \tag{15}$$

where f_s is the sampling frequency and N is the number of time points in the signal.

Staszewski [37] used this decomposition to analyze the Lamb wave signal discussed in the previous section. A Daubechies fourth-order wavelet was used for analysis. Figure 8(a) shows the orthogonal wavelet decomposition of the Lamb wave response for an undamaged plate, and Figure 8(b) shows the decomposition of the Lamb wave response from the plate with a delamination present. The top row in each of the figures represents the original time response data. The following rows show the eight highest wavelet levels. Examination of these levels reveals a feature in the "damaged" data that is not present in the baseline data. On level 4, a spike in the data at 0.11 ms represents a reflection due to the delamination in the plate.

Though the applications of continuous and discrete (or orthogonal) wavelets are quite often different, this Lamb wave example has shown how both transforms can be used on the same data. The results are different, but both transforms are able to identify the presence of damage in the composite plate.

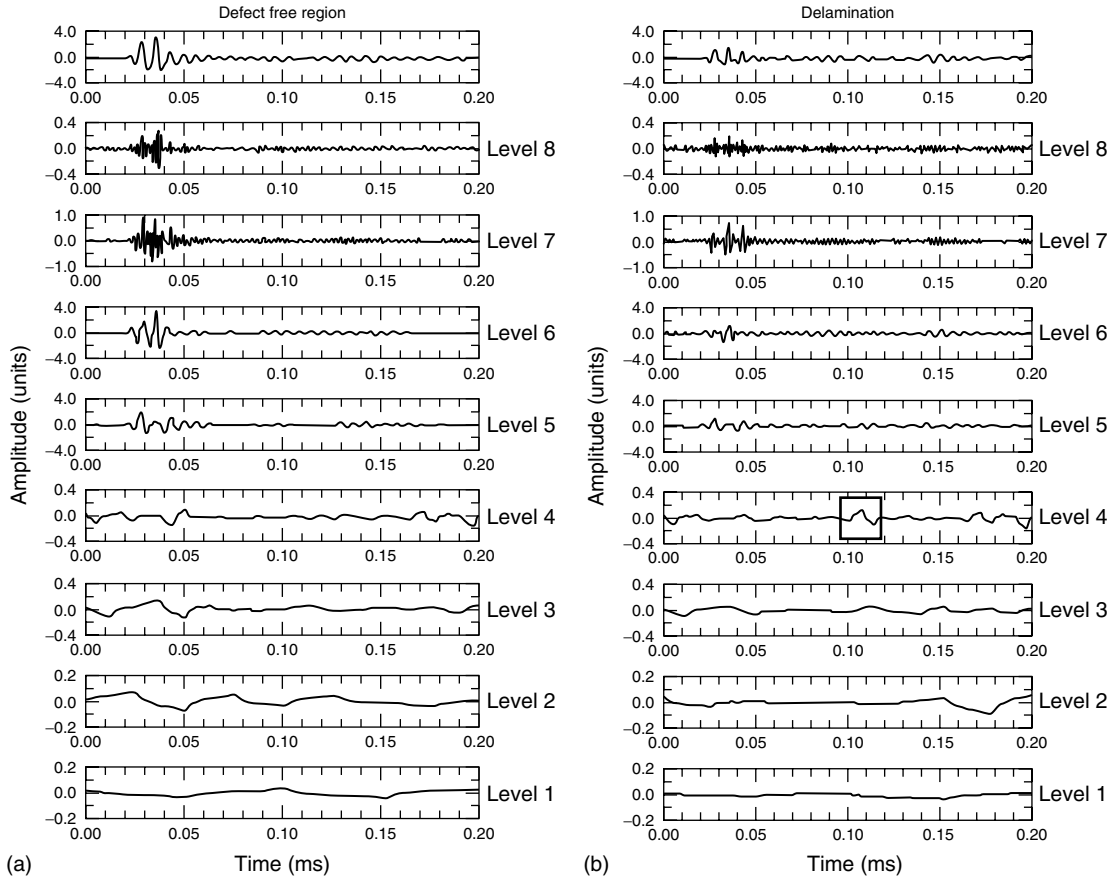


Figure 8. Orthogonal wavelet decomposition of Lamb wave data from a composite plate: (a) no damage present and (b) delamination present. (Staszewski [37], Reproduced by permission of Academic Press.)

5.3 Wavelet packets

Structural systems accumulate damage under both service load and environmental excitations. Examples include fatigue in rotating wind turbine blades or degradation in bridges with opening and closing of cracks manifested in the variation of stiffness. Under certain loading conditions, structures may also suffer significant damage that changes the dynamical and mechanical parameters. In all such cases, a time-varying representation of the stiffness is more appropriate. Identification of the modal parameters is vital in assessing the condition of the structure as well as for diagnosing its failure. Further, the information of stiffness and the variation of stiffness with time are of importance in designing control strategies e.g., for active and semiactive tuned mass dampers.

The temporal variation in the frequency content of the response signal of a structure modeled as a stiffness varying multi-degree-of-freedom (MDOF) system can be used to obtain the variation of the characteristic stiffness of the system. While wavelet analysis techniques have been successfully used mostly in off-line identification of the time-varying structural dynamic parameters of a system, it has the potential for on-line identification, using time–frequency localization properties. The on-line identification of system parameters can be achieved in a simple and computationally straightforward way by using wavelet analysis for an adaptive control application. A wavelet-based on-line identification of stiffness via the identification of modal parameters (assuming the masses to be known) of a structural system, modeled as an MDOF system (shown in Figure 9), has been

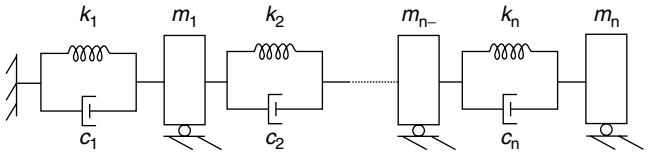


Figure 9. MDOF system.

illustrated through the following example (Basu *et al.* [39]). A modified L–P wavelet basis with wavelet packets has been used in the proposed algorithm.

It can be shown that for each of the modal equations in an MDOF system, the local energy content of the modal response z_k around the time $t = b$ can be expressed by a proportional term as follows:

$$E_j z_k(b) \propto \frac{1}{s_j} \int_{b-\varepsilon}^{b+\varepsilon} |WT(u, s_j)|^2 \, \mathrm{d}u \qquad (16)$$

where, the dilation parameter j is discretized. Since the WT localizes information, wavelet coefficients outside the range $[b - \varepsilon, b + \varepsilon]$ do not contribute to the local energy content around $t = b$. The modified L–P function (Basu *et al.* [39]) has been used as the wavelet basis for analysis, and is characterized by the Fourier transform:

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{F_1(\sigma - 1)}}, \; F_1 \leq |\omega| \leq \sigma F_1 \qquad (17)$$

where, F_1 is the initial cutoff frequency of the mother wavelet and σ is a scalar to be chosen by the user to generate nondyadic nonoverlapping (orthogonal) frequency bands for wavelet bases generated by dilations.

For an m-DOF system, the m bands with the m time-varying natural frequency parameters $\omega_k(b); k = 1, \ldots, m$ correspond to m local maxima in the variation of energy of the rth DOF, $E_j x_r(b)$ (or its proportional quantity $(1/s_j) \int_{b-\varepsilon}^{b+\varepsilon} |WT(u, s_j)|^2 \, \mathrm{d}u$) with different values of the band parameter 'j'. Wavelet packets are applied for better resolution of the modal frequency parameters desired to be obtained with better precision as shown by Chakraborty *et al.* [40]. Once the bands corresponding to the m modes with the parameters $\omega_k(b)$ are obtained, the time-varying mode shapes $\{\phi(b)\}^k$ can be found by

$$\pi_r^j(b) = \frac{WT \; x_r(b, s_j)}{WT \; x_1(b, s_j)} = \frac{\phi_r^k(b)}{\phi_1^k(b)} \qquad (18)$$

To illustrate the application of the tracking methodology, an example of a 2-DOF system has been considered (see [39]). The masses of the first and second floors are $m_1 = 10$ unit and $m_2 = 15$ unit, respectively. The stiffness of the first and second floors are $k_1 = 2500$ unit and $k_2 = 4500$ unit, respectively. These parameters lead to the first natural frequency and mode shape as $\omega_1 = 9.04 \, \mathrm{rad\,s^{-1}}$ and $\{\phi_{11} \, \phi_{21}\} = \{11.37\}$, respectively. A band-limited white-noise excitation has been simulated and applied at the base. The range of frequencies is kept wide enough to cover the frequencies of the system to be identified. The response of the system is simulated with 5% of modal damping (proportional). For the frequency-tracking algorithm, a moving window of 400 time steps equal to 4.16 s has been chosen. For the identification of the 2-DOF system, the parameters F_1 and σ are taken as $8.25 \, \mathrm{rad\,s^{-1}}$ and 1.2, respectively.

Figure 10 shows the tracking of the first natural frequency and the ratio ϕ_{21}/ϕ_{11} corresponding to the first mode shape (assuming $\phi_{11} = 1$, without loss of generality). To further observe if the proposed method can track a sudden change in the stiffness(es) of an MDOF system and follow the restoration to the original stiffness value(s), the stiffnesses k_1 and k_2 of the 2-DOF are changed to 5000 and 5200 units, respectively at an instant of 5.72 s in time. Subsequently, the stiffnesses are restored to their original value at 12.48 s. During the changed phase, the fundamental natural frequency and the mode shape are changed to $\omega_1 = 11.57 \, \mathrm{rad\,s^{-1}}$ and $\{\phi_{11} \, \phi_{21}\} = \{11.57\}$, respectively. Figure 11 shows the tracked first natural frequency and the ratio of ϕ_{21}/ϕ_{11}. As expected, there is a time lag in tracking the frequency and mode shape. The change in the

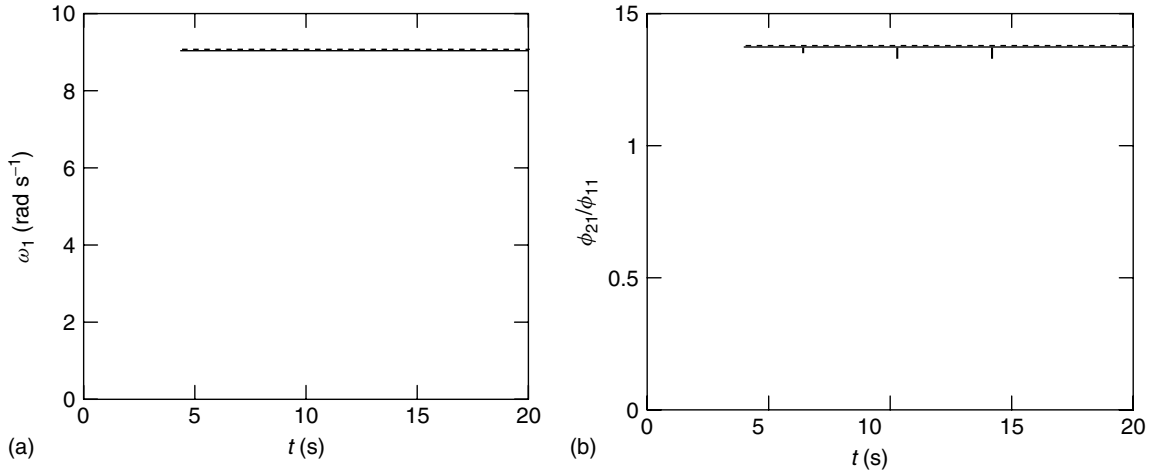


Figure 10. On-line tracking of parameters of a 2-DOF (a) fundamental frequency and (b) fundamental mode shape ("———" actual, "- - - - " estimated).

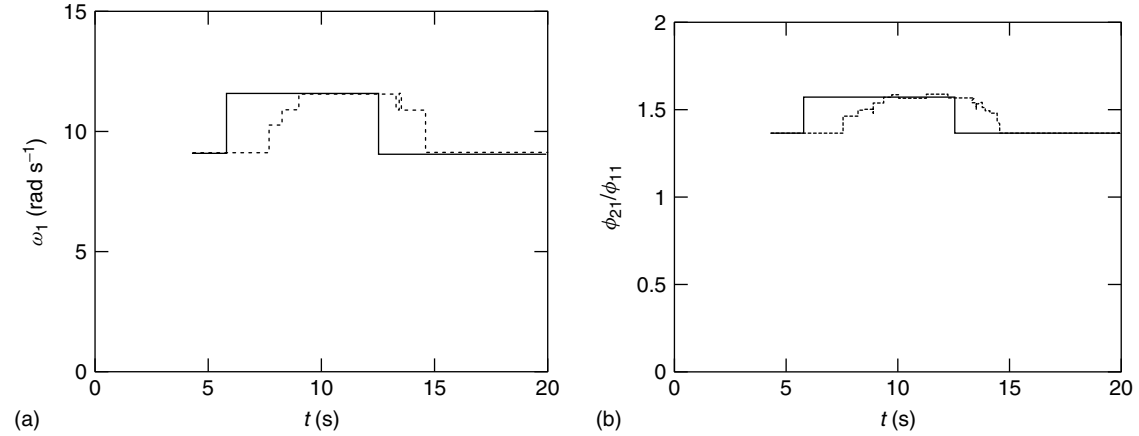


Figure 11. On-line tracking of parameters of a 2-DOF system with sudden change in stiffness (a) fundamental frequency and (b) fundamental mode shape ("———" actual, "- - - - " estimated).

frequency is tracked in (three) steps corresponding to the subbands of frequencies considered for the wavelet packets.

6 CONCLUSIONS

The application of wavelet analysis as a time–frequency tool for SHM has been discussed in this article. A brief literature review on the use of wavelet analysis as a powerful tool for assessment of structural health and fault detection has been presented.

The different types of WTs with their mathematical basis have been reviewed. The similarities and differences of wavelet analysis with STFT have been highlighted. Further, the constant-Q property of wavelet analysis and the construction of a scalogram from the CWT have been discussed. The relation between CWT and DWT has been presented. Also, the construction of an MRA-based fast algorithm for calculating the DWT has been discussed. It has been shown that the DWT can be considered as a filtering approach as evidenced by the MRA. The discussion on orthogonal WT has also revealed that the

wavelet coefficients from an orthonormal DWT are the samples of the CWT on the dyadic grid. The orthogonal wavelets offer a minimal set for the representation of a signal with no redundancy, whereas the CWT has a large number of redundant coefficients. The redundancy of CWT makes it unsuitable for exact reconstruction of signals, but it may be desirable for extracting features from a signal. While the constant-Q property of wavelets is an advantage for computational efficiency, the concept of wavelet packets has been discussed for better resolution at high frequencies. The properties of the different types of WTs decide the choice of a particular one for an application.

Two illustrative examples have been provided in this article. The first one involves the use of Lamb waves to detect a delamination in a carbon fiber composite plate. CWT and orthogonal DWT (Daubechies order 4) have been successfully used for this purpose. The second example involves the application of orthogonal wavelet packets based on a modified L–P wavelet basis. The wavelet packets are used for on-line detection of the stiffness of a structural system via the identification of modal parameters.

REFERENCES

[1] Staszewski WJ, Tomlinson GR. Application of the wavelet transform to fault detection in a spur gear. *Mechanical Systems and Signal Processing* 1994 **8**(3):289–307.

[2] Wang WJ, McFadden PD. Application of wavelets to gearbox vibration signals for fault detection. *Journal of Sound and Vibration* 1996 **192**(5):927–939.

[3] Melhem H, Kim H. Damage detection in concrete by Fourier and wavelet analyses. *Journal of Engineering Mechanics* 2003 **129**:571–577.

[4] Kim H, Melhem H. Damage detection of structures by wavelet analysis. *Engineering Structures* 2004 **26**:347–362.

[5] Hou Z, Noori M, Amand RS. Wavelet-based approach for structural damage detection. *Journal of Engineering Mechanics* 2000 **126**:677–683.

[6] Moyo P, Brownjohn JMW. Detection of anomalous structural behaviour using wavelet analysis. *Mechanical Systems and Signal Processing* 2002 **16**:429–445.

[7] Sun Z, Chang CC. Structural damage assessment based on wavelet packet transform. *Journal of Structural Engineering* 2000 **128**:1354–1361.

[8] Basu B. Identification of stiffness degradation in structures using wavelet analysis. *Construction and Building Materials* 2005 **19**:713–721.

[9] Goggins J, Broderick BM, Basu B, Elghazouli AY. Investigation of seismic response of braced frames using wavelet analysis. *Structural Control and Health Monitoring* 2007 **14**(4):627–648.

[10] Gurley K, Kijewski T, Kareem A. First- and higher-order correlation detection using wavelet transforms. *Journal of Engineering Mechanics* 2003 **129**:188–201.

[11] Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press: San Diego, CA, 2001.

[12] Robertson AN, Park KC, Alvin KF. Extraction of impulse response data via wavelet transform for structural system identification. *Journal of Vibration and Acoustics* 1998 **120**:252–260.

[13] Gentile A, Messina A. On the continuous wavelet transforms applied to discrete vibrational data for detecting open cracks in damaged beams. *International Journal of Solids and Structures* 2003 **40**:295–315.

[14] Okafor AC, Dutta A. Structural damage detection in beams by wavelet transforms. *Smart Materials and Structures* 2000 **9**:906–917.

[15] Chang CC, Chen LW. Vibration damage detection of a timoshenko beam by spatial wavelet based approach. *Applied Acoustics* 2003 **64**:1217–1240.

[16] Loutridis S, Douka E, Hadjileontiadis LJ, Trochidis A. A two-dimensional wavelet transform for detection of cracks in plates. *Engineering Structures* 2005 **27**:1327–1338.

[17] Liew KM, Wang Q. Application of wavelet theory for crack identification in structures. *Journal of Engineering Mechanics* 1998 **124**:152–157.

[18] Wang Q, Deng X. Damage detection with spatial wavelets. *International Journal of Solids and Structures* 1999 **36**:3443–3468.

[19] Spanos PD, Failla G, Santini A, Pappatico M. Damage detection in Euler-Bernoulli beams via spatial wavelet analysis. *Structural control and Health Monitoring* 2006 **13**:472–487.

[20] Zhu XQ, Law SS. Wavelet-based crack identification of bridge beam from operational deflection time history. *International Journal of Solids and Structures* 2006 **43**:2299–2317.

[21] Patsias S, Staszewski WJ. Damage detection using optical measurements and wavelets. *Structural Health Monitoring* 2000 **1**:5–22.

[22] Lam HF, Lee YY, Sun HY, Cheng GF, Guo X. Application of the spatial wavelet transform and Bayesian approach to the crack detection of a partially obstructed beam. *Thin-Walled Structures* 2003 **43**:1–21.

[23] Kim BH, Kim H, Park T. Nondestructive damage evaluation of plates using the multi-resolution analysis of two-dimensional Haar wavelet. *Journal of Sound and Vibration* 2006 **292**:82–104.

[24] Pakrashi V, Basu B, O'Connor A. Structural damage detection and calibration using a wavelet-kurtosis technique. *Engineering Structures* 2007 **29**(9):2097–2108.

[25] Staszewski WJ. Identification of damping in MDOF systems using time scale decomposition. *Journal of Sound and Vibration* 1997 **203**:283–305.

[26] Staszewski WJ. Identification of non-linear systems using multi-scale ridges and skeletons of the wavelet transform. *Journal of Sound and Vibration* 1998 **214**(4):639–658.

[27] Kyprianou A, Staszewski WJ. On the cross wavelet analysis of the duffing oscillator. *Journal of Sound and Vibration* 1999 **228**(1):119–210.

[28] Ruzzene M, Fasana A, Garibaldi L, Piombo BAD. Natural frequencies and damping identification using wavelet transform: application to real data. *Mechanical Systems and Signal Processing* 2000 **11**:207–218.

[29] Robertson AN, Farrar CR, Sohn H. Singularity detection for structural health monitoring using holder exponents. *Mechanical Systems and Signal Processing* 2003 **17**(6):1163–1184.

[30] Piombo BAD, Fasana A, Marchesiello S, Ruzzene M. Modeling and identification of the dynamic response of a supported bridge. *Mechanical Systems and Signal Processing* 2000 **14**:75–89.

[31] Ghanem R, Romeo F. A wavelet based approach for the identification of linear time-varying dynamical systems. *Journal of Sound and Vibration* 2000 **4**:555–576.

[32] Basu B, Nagarajaiah S, Chakraborty A. On-line identification of linear time varying stiffness of structural systems by wavelet analysis. *Structural Health Monitoring* 2008 **7**(1):21–36.

[33] Tian J, Li Z, Su X. Crack detection in beams by wavelet analysis of transient flexural waves. *Journal of Sound and Vibration* 2003 **261**:715–727.

[34] Staszewski WJ, Lee BC, Mallet L, Scarpa F. Structural health monitoring using scanning laser vibrometry: I. Lamb wave sensing. *Smart Materials and Structures* 2004 **13**:251–260.

[35] Staszewski WJ, Robertson AN. Time-frequency and time-scale analysis for structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2007 **365**(1851):449–477.

[36] Daubechies I. *An Introduction to Wavelets*. Academic Press: San Diego, CA, 1992.

[37] Staszewski W. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons: West Sussex, 2004.

[38] Tchamitchian P, Torresani B. *Ridge and Skeleton Extraction from the Wavelet Transform: Wavelets and Their Applications*. Jones & Bartlett Publishers: Boston, 1992.

[39] Basu B, Nagarajaiah S, Chakraborty A. Identification of stiffness variations in structural systems by modified littlewood-paley wavelets. *Third European Workshop on Structural Health Monitoring*. DEStec Publications: Pennsylvania, July 2006; pp. 715–722.

[40] Chakraborty A, Basu B, Mitra M. Identification of modal parameters of a mdof system by modified L-P wavelet packets. *Journal of Sound and Vibration* 2006 **295**(3–5):827–837.

FURTHER READING

Basu B, Gupta VK. Seismic response of SDOF systems by wavelet modelling of nonstationary processes. *Journal of Engineering Mechanics* 1998 **124**(10):1142–1150.

Chapter 34
Nonlinear Features for SHM Applications

Jonathan M. Nichols[1] and Michael D. Todd[2]

[1] *US Naval Research Laboratory, Washington, DC, USA*
[2] *Department of Structural Engineering, University of California, San Diego, CA, USA*

1 INTRODUCTION

A *feature* is a quantity extracted from measured system response data that indicates the presence, location, and/or level of damage present in the system [1]. The dominant approaches in the literature, covered in previous articles in this collection (*see* **Signal Processing for Damage Detection**), involve features derived from linear models, assumptions, or signal processing techniques [2, 3]. However, there are several situations where nonlinearity plays a role in the damage-detection problem and where the use of nonlinear features is appropriate. Perhaps the most important of such situations is where the healthy structure is appropriately modeled as linear and where damage causes a nonlinearity. In this case, the problem of damage detection can be equated with nonlinearity detection. Nonlinear system-identification techniques can further be used to indicate both the magnitude and type of nonlinearity. In other cases, the structure is appropriately modeled as nonlinear throughout the damage progression. Nonlinear features can improve the ability to track damage in this case. Still other researchers have used nonlinearity to enhance the sensitivity of a particular damage-detection scheme by tuning the excitation to elicit a specific structural response, amenable to nonlinear feature-based analysis. Each of these approaches is described in what follows, along with the appropriate citations.

It is first useful to define exactly what is meant by a "nonlinear feature". Nonlinearity is, of course, a property of the underlying system and not of the data. Rather than saying we have a "nonlinear time

 ISBN: 978-0-470-05822-0.

series", it is more appropriate to say we have "a time series produced by a nonlinear system". As a result, a "nonlinear feature" is one that is designed to take advantage of certain relationships among time series produced by a nonlinear system.

First, consider a linear, time-invariant single-degree-of-freedom (DOF) system. The input–output relationship for such a system may be modeled via the convolution

$$y(t) = \int_{-\infty}^{\infty} h(\theta)x(t-\theta)\,\mathrm{d}\theta \tag{1}$$

that is to say the response $y(t)$ is a superposition of input "impulses", $x(t)$, and the impulse response function $h(t)$. The statistical properties of the output are

$$\begin{aligned}
E[y(t)] &= \int_{-\infty}^{\infty} h(t-\theta_1)E[x(\theta_1)]\,\mathrm{d}\theta_1 \\
E[y(t)y(t+\tau_1)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t-\theta_1)h(t+\tau_1-\theta_2)E[x(\theta_1)x(\theta_2)]\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2 \\
&\vdots \\
E\left[\prod_{m=0}^{M-1} y(t+\tau_m)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{m=0}^{M-1} h(t+\tau_m-\theta_{m+1})E\left[\prod_{m=0}^{M-1} x(\theta_{m+1})\right]\prod_{m=0}^{M-1}\mathrm{d}\theta_{m+1} \quad (\tau_0=0)
\end{aligned} \tag{2}$$

If the driving signal $x(t)$ is iid, zero-mean Gaussian noise, all odd-ordered moments vanish, e.g., $E[y(t)y(t+\tau_1)y(t+\tau_2)] = 0$, while the even-ordered expectations become functions of the autocovariance $E[y(t)y(t+\tau_1)]$ [4]. Thus, for a linear, Gaussian-excited system, the output relationship $E[y(t)y(t+\tau_1)]$ is sufficient to characterize the response. However, if nonlinearities are present equation (1) does not hold and it is likely that there exist nonzero components in the higher-order correlations, e.g., $E[y(t)y(t+\tau_1)y(t+\tau_2)] \neq 0$. A feature that captures this relationship is, therefore, taking advantage of the fact that the underlying system is nonlinear. One feature that targets this third moment, the bispectrum, is explored later (the use of other higher-order correlations in detecting damage-induced nonlinearities in a structure is also discussed). Even if the driving term $x(t)$ is non-Gaussian, second-order statistics are still sufficient to describe the linear response provided that the distribution of $y(t)$ is also known. It can be shown, for example, that normalized versions of the higher-order correlations can be used effectively to detect nonlinearity in non-Gaussian processes [5].

Additionally, for the systems described by equation (1), certain quantities like natural frequency are invariant to the amplitude of excitation. For nonlinear systems, this is not true; in fact, tracking the amplitude varying properties is one way to diagnose the type of nonlinearity. It should be pointed out that the focus here is on linear *time-invariant* systems. In much of what follows, the features presented implicitly assume that the practitioner is dealing with stationary structural response data. For nonstationary response data, the estimation of many of these features becomes difficult, and fewer methods are available.

In short, there is a large body of literature concerning the use of nonlinear features in structural health monitoring (SHM). In the sections that follow, an effort is made to describe these works and place them in context. The list is not exhaustive, but may hopefully help the reader appreciate the various techniques available and the situations in which they have been used successfully.

The main reason for wanting to use nonlinear features in SHM is, of course, that often there exists a significant amount of nonlinearity in the dynamics of the structure being monitored. A prime example is the aforementioned scenario whereby damage causes a healthy structure that is well described by a linear model to exhibit nonlinearity. This fact has been exploited by a number of researchers in the field.

A few examples of nonlinear behavior in damaged structures are cited here (*see* **Civil Infrastructure Load Models for Structural Health Monitoring**; **Failure Modes of Aerospace Materials**).

1. Cracking
A cracked structure, for example, is often modeled as a nonlinearity. For example, Brandon [6] considers both crack and clearance nonlinearities as the damage mechanism. Friswell and Penny consider the modeling of cracks using a bilinear stiffness term [7]. Zhang and Testa [8] also explored the nonlinearity due to the opening/closing of a crack in an experimental beam structure.

2. Impacts and/or rattling
Impacts, or clearance nonlinearities also present an obvious situation where damage equates with nonlinearity. Brown and Adams [9], Rutherford *et al.* [10], and Nichols *et al.* [11] have all explored loosening connections as a nonlinear damage mechanism. This type of nonlinearity is also referred to in may works as *backlash* (for example, [12, 13]).

3. Delamination
Delamination has also been modeled as a damage-induced nonlinearity in composite structures. The work of Schwarts-Givli *et al.*, for example, considers both contact and geometric nonlinearity in modeling damaged sandwich panels. Earlier work by Hunt *et al.* [14] considered a low-dimensional model of a delaminated strut. The delamination was modeled as a local buckling resulting in a nonlinear response. A similar model was proposed by Murphy *et al.* and was shown to agree well with experiment [15]. A separate nonlinear model for delaminated beams was also given in [16].

4. Stick/slip, rub
In the analysis of rotating machinery, rotor–stator "rub" is viewed as a damage mechanism. An experimental study illustrating rub as a source of nonlinearity can be found in [17]. Detection of this type of damage was studied in [18, 19]. Additionally, jointed structures often exhibit stick–slip behavior. Both modeling and simulation of this type of behavior in bolted joints were studied by Song *et al.* [20], among others. "Fretting" wear is another damage mechanism arising from stick–slip behavior and has been studied by a number of authors (for example, [21]).

Additionally, there are a number of structures that exhibit significant nonlinearity regardless of damage. Monitoring such structures can therefore be aided by the use of nonlinear features. A comprehensive look at nonlinearity in structural dynamics has been presented by Worden and Tomlinson [22] and also in a recent review article by Kerschen *et al.* [23].

The following sections in this article describe some broad classes of where the role of nonlinearity in system response is either exploited or imposed to extract damage-sensitive features. These classes may be grouped as (i) direct tests of the definition of nonlinearity, (ii) phase space methods, (iii) higher-order spectra, (iv) the method of surrogate data, and (v) imposed nonlinearity.

2 TESTING THE DEFINITION OF NONLINEARITY

Probably the most well-known methods of capturing nonlinearity are adaptations to linear approaches. For example, traditional linear signal processing (e.g., the power spectral density) provides information about second-order statistics in a time series. However, if higher-order correlations (nonlinearity) are present, they still show up by violating expected properties of the linear system. Any nonlinear system H is fundamentally defined by a violation of the following rule of homogeneity (scalability) and superposition:

$$H\left(\sum_n a_n x_n\right) = \sum_n a_n H\left(x_n\right) \tag{3}$$

where the x_n are inputs to the system, weighted by a_n. Linear systems must fundamentally obey this equation. Considering a single input x_1 that results in a system response y_1 and a scaled input $x_2 = ax_1$ that results in an output y_2, equation (1) becomes

$$y_1(t) = \int_{-\infty}^{\infty} h(\theta) x_1(t-\theta)\,\mathrm{d}\theta$$

$$y_2(t) = \int_{-\infty}^{\infty} h(\theta) x_2(t-\theta)\,\mathrm{d}\theta$$

$$y_1(t) = \int_{-\infty}^{\infty} h(\theta) x_1(t-\theta)\,\mathrm{d}\theta$$

$$
\begin{aligned}
y_2(t) &= \int_{-\infty}^{\infty} h(\theta) a x_1(t-\theta)\,\mathrm{d}\theta \\
y_1(t) &= \int_{-\infty}^{\infty} h(\theta) x_1(t-\theta)\,\mathrm{d}\theta \\
y_2(t) &= a\int_{-\infty}^{\infty} h(\theta) x_1(t-\theta)\,\mathrm{d}\theta \\
Y_1(\omega) &= H(\omega) X_1(\omega) \\
Y_2(\omega) &= a H(\omega) X_1(\omega)
\end{aligned} \tag{4}
$$

where the Fourier transform has been taken in the last line to translate into the frequency domain (i.e., $Y_1(\omega)$, $H(\omega)$, and $X_1(\omega)$ are the Fourier transforms of $y_1(t), h(\theta)$, and $x_1(t)$, respectively). The frequency response function (FRF) is defined as the frequency domain ratio of output to input, so that for linear systems, $Y_2(\omega)/X_2(\omega) = aY_1(\omega)/\big(aX_1(\omega)\big) = Y_1(\omega)/X_1(\omega)$. This identity is readily tested in system response by merely applying two inputs of different magnitude to the system and observing whether the resulting measured FRFs are identical or not.

Superposition may also be readily tested by applying different inputs individually and then collectively to the system, since

$$
\begin{aligned}
y(t) &= \int_{-\infty}^{\infty} h(\theta)\left(x_1(t-\theta) + x_2(t-\theta)\right)\mathrm{d}\theta \\
&= \int_{-\infty}^{\infty} h(\theta) x_1(t-\theta)\,\mathrm{d}\theta \\
&\quad + \int_{-\infty}^{\infty} h(\theta) x_2(t-\theta)\,\mathrm{d}\theta \\
Y(\omega) &= H(\omega)X_1(\omega) + H(\omega)X_2(\omega) \\
&= Y_1(\omega) + Y_2(\omega)
\end{aligned} \tag{5}
$$

The SHM community has long exploited, particularly in applications involving rotating machinery, a consequence of superposition violations known as *harmonic generation*. The response of a linear, single-DOF system to broadband Gaussian inputs shows a single resonant peak when viewed in the frequency domain; similarly, a linear system excited by a pure tone results in a pure tone at that same frequency in the output (at steady state). However, if that same system possesses a nonlinearity, those single tonal peaks distort, and additional peaks at harmonic frequencies appear. These can be used to diagnose the presence and sometimes the type of nonlinearity. In damage detection, for example, Adams and Farrar used frequency domain Auto-Regressive eXogenous input (ARX) models to capture the influence of harmonics and use them as a damage indicator [24]. Work by Bovsunovsky and Surace [25] similarly used superharmonics to diagnose cracks in a beam structure. Their work showed that changes in the harmonics were found to be more sensitive to damage than were the changes to the main resonant peak. Recent work by Parsons and Staszewski [26] also focused on harmonic generation in the detection of fatigue cracks in an aluminum plate using ultrasonic techniques. As an example, consider the composite-to-metal jointed structure shown in Figure 1. A full description of the system and experiment is given in [27]. To detect the bolt loosening, random excitation was applied near the beam center and the strain response near the joint was recorded. The bolt was fully tightened in the undamaged scenario and damage was taken as a loosened connection. The power spectral density of both the undamaged and damaged response data is also shown in Figure 1. The appearance of additional spectral peaks is a signature of nonlinearity, which, in this case, corresponds to bolt loosening. Other work by Adams has focused on modifying the traditionally "linear" concept of a transmissibility function to accommodate nonlinearity in a health monitoring context [28]. Wong and Chen accounted for nonlinearity in a similar fashion, using the concept of an "equivalent linearization stiffness matrix" to detect and locate damage in a simulated 5-DOF structure [29]. D'Souza and Epureanu also viewed nonlinearity as a perturbation to a linear problem. By identifying the size of the perturbation, they could identify both the location and magnitude of nonlinear damage. This method was referred to as *rank perturbation*, a full description of which is given in [30]. Still others have captured nonlinearity by focusing on amplitude dependence in system properties (e.g., natural frequency) that would be invariant to amplitude under the hypothesis of a linear structure obeying equation (1). Neild *et al.*, for example, looked at the frequency variations with amplitude in the response of two reinforced concrete beams [31] and were able to successfully deduce the presence of crack damage. Other work by Feldman [32] (to be described later), also showed amplitude dependencies in natural frequency in a

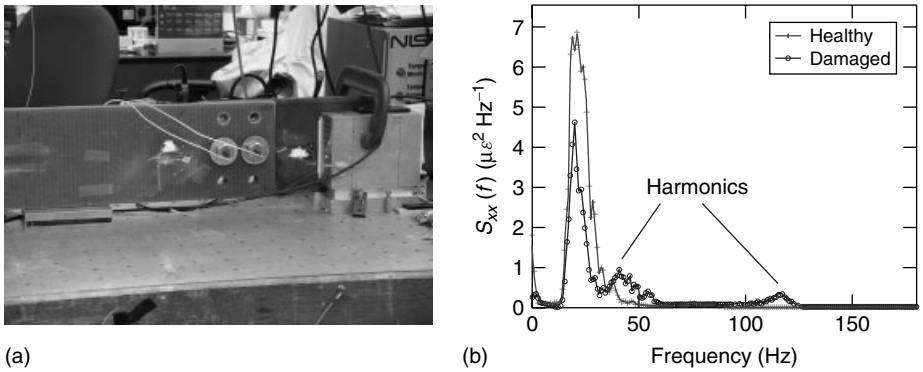


Figure 1. (a) Experimental bolted joint structure and (b) power spectral density estimated from strain response. Both the fully clamped (undamaged) and loosened (damaged) connections are shown.

cracked structure. Autoregressive, moving average models with exogenous inputs (ARMAX models) are another favorite system-identification procedure for linear systems. However, for systems possessing nonlinearity, the dynamics are more appropriately captured using a nonlinear, autoregressive, moving average model with exogenous inputs (NARMAX model). Wei *et al.* used NARMAX models, relating input/output data, to detect delamination in an experimental composite structure. Changes in the model structure were found to correlate well with damage [33]. Other tests of linearity definition violation include reciprocity checks (testing the symmetry of the FRF), nulls in the coherence function, and loss of identity in Hilbert transformations of the FRF. These methods are extensively reviewed in [22].

Finally, we mention an approach for detecting nonlinearities that cause discontinuities in structural response data. The approach is to use estimates of the Holder exponent as damage-sensitive feature. This feature measures the degree of differentiability or "smoothness" of a function; thus, a small Holder exponent indicates strong discontinuity and vice versa. In [34], a procedure for capturing the time varying nature of the Holder exponent has been developed based on wavelet transforms. This procedure has been successfully applied to detect loose components in a harmonically excited mechanical system.

3 PHASE SPACE METHODS

A number of approaches to SHM rely on a *phase space* analysis of the response data. Assume that there exists an M-DOF structure, governed by the state equations:

$$\dot{\mathbf{x}}(t) = F(\mathbf{x}(t), \boldsymbol{\mu}) \tag{6}$$

where $F(\cdot)$ is a deterministic function of both the system state and a vector of system parameters $\boldsymbol{\mu}$. In a structural dynamics context, the parameter vector might consist of stiffness, mass, and damping parameters. The state vector $\mathbf{x}(t) \equiv (x_1(t), x_2(t), \ldots, x_M(t))$ consists of all the system variables, e.g., positions and velocities of lumped mass elements. As the dynamics of the system evolve in time, they trace out a unique trajectory in this state space. Under the dissipation present in any real system (due to various sources), all trajectories eventually settle onto a dynamical *attractor*, given a deterministic input. The attractor may be viewed as a geometric object in state space. Changes in the system parameters (e.g., stiffness and damping) alter the resulting attractor, and thus various researchers have used attractor properties as a source of damage-sensitive features.

Central to these approaches is the problem of phase space reconstruction. Typically, we cannot measure all of the state variables in experiment; thus, we only have access to part of the state space. Fortunately, there exists a technique whereby the variables that are not observed can be qualitatively reconstructed such that the geometry of the underlying attractor is preserved. A full discussion of attractor reconstruction and the theorems that underlie it would be prohibitive. A comprehensive look at the reconstruction problem is found in [35] for both uni- and multivariate data. Assuming a univariate measurement (let us say only measurements of acceleration

at one point on the structure are available), the reconstructed state space is given by

$$\mathbf{x}(n) = (x(n), x(n+T), \ldots, x(n+(d-1)T)) \quad (7)$$

where T is a measure of time delay and d is the embedding dimension. Prescriptions for choosing these quantities are found in [35]. If d and T are chosen properly, the vector $\mathbf{x}(n)$ is topologically equivalent to the "true" state vector from equation (6). Figure 2 shows a two-dimensional view of attractor reconstructions, formed by embedding an undamaged and damaged response. The system of interest in this case was an 8-DOF spring–mass system driven with the output of a chaotic oscillator. Details of this approach are given in Section 6 and a full description of the simulations are documented in [36].

Once the attractor has been reconstructed, a number of metrics can be used to identify damage-induced changes to the dynamics. Most of these metrics are based on the geometry of the attractor and are not restricted to looking at a specific temporal relationship (e.g., covariance). Thus, these methods are capable of capturing correlations of any order in the data (see again equation (2)) and are appropriate for capturing the effect of structural nonlinearity on the response.

The correlation dimension is one such metric frequently used in the nonlinear time series analysis community as a descriptor of nonlinear processes. This feature effectively quantifies, on average, how local probability estimates scale with "bin size". Although a number of other measures of dimension exist (e.g., information dimension), the correlation dimension is probably the most widely used method for quantifying the geometry of a collection of points. More specifically, given a reconstructed attractor $\mathbf{x}(n)$, the correlation integral (as used in practice) reads as

$$C(r) = \frac{2}{N(N-2W-1)} \times \sum_{i=1}^{N} \sum_{j:|i-j|>W}^{N} \Theta(r - \|\mathbf{x}(i) - \mathbf{x}(j)\|) \quad (8)$$

where $\Theta(\cdot)$ is the Heaviside step function, r is a threshold parameter, and W is a Theiler window (designed to eliminate temporally correlated points from the sum [38]). It should be noted that an estimate of the probability density for point i, assuming a stationary, ergodic process, is given by

$$\hat{p}(r, \mathbf{x})(i) = \frac{1}{N} \sum_{j}^{N} \Theta(r - \|\mathbf{x}(i) - \mathbf{x}(j)\|) \quad (9)$$

Thus, the correlation integral is related to the average probability density of points on an attractor, estimated using length scale r. Probability is, in fact, one example of a measure of a collection of points. The correlation dimension quantifies how $C(r)$ scales with r. Thus, assuming the scaling law $C(r) \sim r^{d_c}$, the slope of the $\log(C(r))$ versus $\log(r)$ plot gives the correlation dimension, d_c. Often, damage causes an increase in the dimensionality of the structure's response. By monitoring these changes, one can detect damage.

This metric has been used by Logan and Mathew [39, 40] for bearing fault diagnosis. Similarly, it has been used by Craig *et al.* [41] in detecting clearance nonlinearities, as well as by Wang *et al.* for detecting damage in rotating machinery [42]. Trendafilova [43] has also made use of this feature in several works. A comparison of the above-described estimator of dimension to a different estimator (based on the

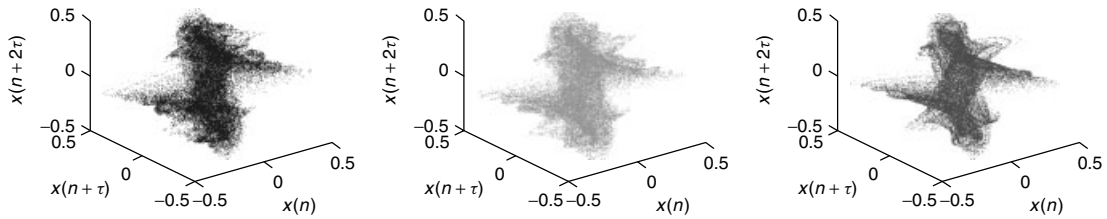


Figure 2. Reconstructed attractors for a "fixed-free" 3-DOF system described in [37]. Damage was taken as a stiffness reduction between masses 2 and 3. A number of features may be extracted from these geometric objects for comparison and thus damage tracking.

principle of maximum likelihood) in the context of health monitoring was given in [44] in diagnosing a loosening joint.

Still other researchers have focused on attractor-based measures derived straight from the local probability structure of the data. Epureanu and Yin [45] computed estimates of the local probability density as given by equation (9) and then used the resulting distribution on the attractor as the damage indicating feature. In that study, the structure of interest was a nonlinear aeroelastic plate subject to stiffness degradation.

Trendafilova also looked at the local probability structure of data in a health monitoring context [43]. Todd *et al.* used the statistics of local neighborhoods of points on an attractor to detect stiffness degradations in a simulated structure [36]. More specifically, they looked at the average ratio of local attractor variances (ALAVRs) between an undamaged and damaged attractor. This metric was designed to capture damage-induced distortions to the attractor geometry and showed greater sensitivity to the damage than some other metrics. Other metrics derived from attractors include autoprediction error and cross-prediction error [46–48]. Hively and Protopopescu [49] also used a phase space measure of probability to distinguish between healthy and damaged motor-driven systems.

Additionally, some attention has been given to using Lyapunov exponents (LEs) as a damage indicating feature. A system's LEs capture the rate at which trajectories in phase space diverge or converge in each of the phase space directions. The maximal LE was used in [43] in the analysis of a reinforced concrete plate subject to increasing static loads. This metric was also used in [13] in detecting backlash in robot joints. Golnaraghi [50] also used LEs (as well as dimension estimates) in detecting gear damage. Overbey and Todd [37] considered local LEs (and thus, local attractor stability) as a metric for comparison.

4 HIGHER-ORDER SPECTRA

When identifying changes to linear system parameters, the focus is often on the power spectral density of the response data. By the Wiener–Khintchine relations, the power spectral density is simply the Fourier transform of the second moment about the mean (first line of equation (2)). Analogously, Fourier transforms of higher-order cumulants result in higher-order spectra (*see* **Higher Order Statistical Signal Processing** for a more complete discussion). A good example of this is the bispectrum or, its normalized form, the bicoherence. The bispectrum associated with the measured signal $x(t)$ is defined as the double Fourier transform of the third moment about the mean, i.e.,

$$\begin{aligned} B(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & E[(x(t) - \bar{x}) \\ & \times (x(t+\tau_1) - \bar{x})(x(t+\tau_2) - \bar{x})] \\ & \times \mathrm{e}^{-\mathrm{i}(\omega_1\tau_1 + \omega_2\tau_2)}\, \mathrm{d}\tau_1\, \mathrm{d}\tau_2 \end{aligned} \tag{10}$$

By equation (2), this quantity is nonzero if the process is non-Gaussian. Conversely, a linear structure driven with Gaussian noise has a bispectrum which is zero everywhere. Nonlinearity results in peaks in the bispectrum, which can be used for purposes of detection (see example below). A normalized version of the bispectrum, the bicoherence can similarly be used to detect nonlinearity in non-Gaussian processes (e.g., a linear structure driven with non-Gaussian noise). Rivola and White [51] used the bicoherence to detect cracks in an experimental beam structure while Zhang *et al.* [52] focused on detecting gear faults, both using the bispectrum. The bispectrum was also used to assess the magnitude of a bilinear stiffness nonlinearity, simulating a crack, in a simple spring–mass system [53]. Other higher-order spectra (e.g., the trispectrum) have been considered for damage detection as well [54]; however, the estimation of such spectra remains a difficult challenge. An example of bispectral detection of nonlinearity is presented graphically in Figure 3. The system of interest is a 2-DOF spring–mass–damper structure with $m_1 = m_2 = 1.0$ (kg), $k_1 = k_2 = 1000$ (N m^{-1}) and $c_1 = c_2 = 3.0$ [N·s m^{-1}] and the forcing is taken as a realization of an iid Gaussian process. The governing equations are

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{Bmatrix} + \begin{bmatrix} c_1 & -c_2 \\ -c_2 & c_2 \end{bmatrix} \begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} + \begin{bmatrix} k_1 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ \mathcal{N}(0,1) \end{Bmatrix} \tag{11}$$

The stiffness matrix is bilinear in nature, described by

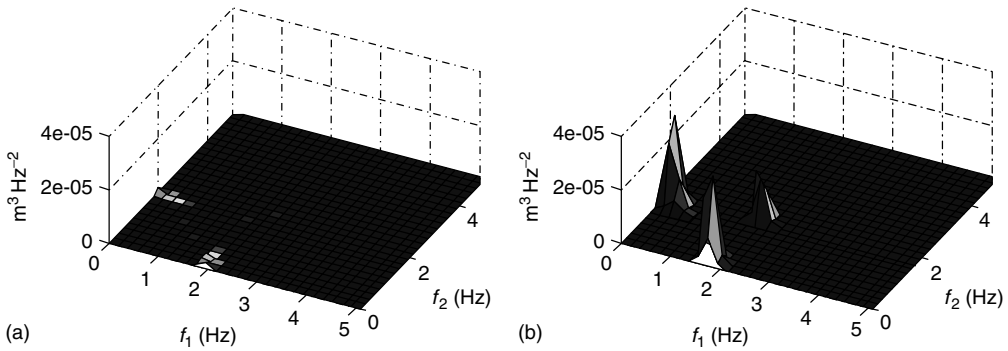


Figure 3. Magnitude bispectrum for a Gaussian driven, 2-DOF structure. (a) Result for the linear system and (b) magnitude bispectrum of the response of the first mass where a bilinear stiffness of 50% was placed in between the masses.

$$k_2(x_2 - x_1) = \begin{cases} k_2(x_2 - x_1) & (x_2 - x_1) \geq 0 \\ \delta k_2(x_2 - x_1) & (x_2 - x_1) < 0 \end{cases} \quad (12)$$

where the parameter $0 < \delta \leq 1$ controls the degree of bilinearity in the stiffness element. For a linear structure, $\delta = 0$ and the bispectrum of the mass 1 response is shown in Figure 3. Using a simple crack model, we allow the damage structure a bilinear stiffness with $\delta = 0.5$. The bispectrum for the bilinear response is also shown in Figure 3. Clearly, the nonlinearity results in nonzero values for the bispectral density, showing peaks at the systems' first resonant frequency. The locations and magnitude of these poles can, in fact, be determined analytically and used to establish the Type-I and Type-II errors of such a detector [55]. The peak height for the magnitude bispectrum is, therefore, a useful nonlinear feature, capable of capturing certain kinds of damage-induced nonlinearity.

5 REFERENCE-FREE APPROACHES TO NONLINEARITY DETECTION

As was stated earlier, damage often results in the presence of a nonlinearity in a structure whose dynamics are otherwise better described by a linear model. There is a potentially large advantage to casting damage detection as a problem of nonlinearity detection, namely, the practitioner can pose the question in the absence of a baseline or reference data set. The null hypothesis of a linear (healthy) structure can be tested at any point in the structure's life and can be accomplished without comparing newly acquired features to those extracted from a pristine structure's response. This is important because effects other than damage often cause changes to the feature being monitored (e.g., temperature and sensor drift). Methods that rely on baseline data sets tend to see these changes as damage and, as a result, suffer from false positives. In the following approaches, any changes that do not influence the underlying model (linear/nonlinear) do not influence the diagnosis.

One promising approach to nonlinearity detection in SHM is the work of Park *et al.* [56, 57]. This work uses high-frequency Lamb waves to interrogate the structure. If the structure is linear, the reflected wave has a certain shape. Damage-induced nonlinearities cause a change in the received signal. The practitioner is testing whether or not the reflected wave comes from a linear or nonlinear model. This diagnosis can, in principle, be made without referencing back to previously acquired data. This approach, therefore, has the benefits mentioned above, namely, the method is not severely influenced by many sources of environmental variability.

There are also several works that make use of a nonlinearity detection scheme referred to as the *method of surrogate data*. This approach was developed by the physics community over the last 15 years and recently applied to the problem of damage detection. The idea is to perform a randomization of the acquired data such that the statistical properties of the data associated with a linear, time-invariant model (the null hypothesis) are preserved while those associated with nonlinearity are destroyed. One can then compare an appropriately chosen feature

computed from the data to those computed from some number of surrogates. If the measured data came from a damaged structure, there will be statistically significant differences in these feature values. If the measured data came from a healthy structure, the surrogate feature values and data feature values will be statistically indistinguishable. This approach has been used by both Trendafilova and Brussel [13] and Nichols *et al.* [27, 58] as a damage-detection strategy.

Algorithms for generating surrogate data are numerous and are tailored to different null hypotheses. The most often used algorithm is the iterative amplitude-adjusted Fourier transform (IAAFT) approach of Schreiber [59]; however, this approach can sometimes lead to a larger than expected number of false rejections of the null. A separate approach that preserves the correct variance among the surrogate population is given by Dolan [60] and is referred to as the digitally filtered, shuffled (DFS) algorithm.

Figure 4 shows some sample results from the bolted-joint experiment of Figure 1. The goal was to detect the bolt loosening using ambient vibrations (excitation was a realization of a process described by the Pierson–Moskowitz distribution for wave height) and in the presence of varying temperature. All data were recorded at a sampling rate of 2 kHz using a fiber-optic strain sensor located on the beam, near the joint. Ten separate time series were collected, each consisting of 2^{15} points, for the clamped beam while six time series (also consisting of 2^{15} points) were collected from the beam in a loosened condition (either finger-tight or loose). Forty surrogates, each made using the aforementioned DFS algorithm, were then generated. The bicoherence function was estimated for both data and surrogates, and the final feature was taken as the average of the estimated bicoherence function over all frequencies. Figure 4 shows the average bicoherence feature, b_{avg}, for the data (cross) and surrogates (dot) for each of the collected time series. The ambient temperature for this experiment, as measured at the beam surface, was 40 °C. Clearly, the response is consistent with a linear model until the joint loosens at which point nonlinearity becomes evident. Figure 4 shows a repeat of the experiment, but with data collected at a surface temperature of 30 °C. This temperature shift altered the beam's natural frequencies for the clamped (undamaged) condition, potentially resulting in false-positive diagnosis for modal-based features. However, the temperature variation does not influence the ability to correctly classify the response as linear (undamaged) or nonlinear (damaged).

6 EXPLOITING NONLINEARITY IN SHM

A number of researchers have attempted to use nonlinearity to their advantage in the damage-detection problem. Yin and Epureanu [61], for example, used nonlinear feedback to enhance the

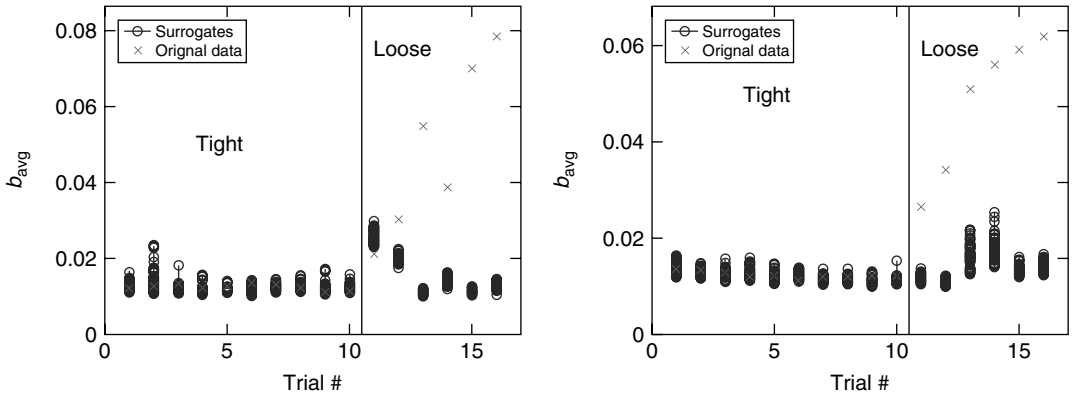


Figure 4. Average bicoherence feature computed for data and surrogates over 16 separate trials. The first 10 trials were with the beam in the fully clamped (linear) condition. The last six trials were for data collected in finger-tight or loose conditions. The data for the two figures were recorded at 40 and 30 °C, respectively. Separation of surrogates from data indicates significant amounts of nonlinearity accompanying the bolt loosening.

sensitivity of their damage-detection approach. The idea is to create a bifurcation point in the system, via feedback control, such that a small change to the parameter of interest (damage) causes a large change in system behavior. This change can then be easily detected.

Another recently developed approach works by tailoring a highly nonlinear forcing function to excite the structure's dynamics. The properties of the forcing are such that damage causes an increase in the dimension (specifically the information dimension that is closely related to the aforementioned correlation dimension) of the response. The relationship between an "undamaged" and "damaged" structure's response, in this case, is conjectured to be nonlinear and, thus, is best captured using a nonlinear feature. Assume the structure's dynamics are governed by a function $F(\cdot)$ and there exists a separate system used to force the structure governed by $G(\cdot)$ and described by the state vector $\mathbf{z}$. The coupled forcing/structure dynamics are given by

$$\begin{aligned} \dot{\mathbf{x}} &= F(\mathbf{x}, \mu) + B\mathbf{z} \\ \dot{\mathbf{z}} &= G(\mathbf{z}) \end{aligned} \tag{13}$$

where B is a constant coefficient matrix that simply couples the structure to the forcing. It can be shown that if the excitation is deterministic and has a positive LE (the dynamics are chaotic), and the Lyapunov exponents of the forcing are in the correct range, the dimension of the response $\mathbf{x}$ changes as the structure's parameter μ changes. More specifically, a change in μ is conjectured to cause a loss of continuity in the function $\phi(\cdot)$ relating an undamaged response $\mathbf{x_u}(n)$ to a "damaged" response $\mathbf{x_d}(n)$, i.e., $\mathbf{x_d}(n) = \phi(\mathbf{x_u})$. Figure 5 illustrates this approach in schematic form. A common drive signal results in two different structural responses (one undamaged, one damaged) that are related by the function ϕ. With an appropriately chosen drive signal, the properties of ϕ are altered by damage. The theorems that underlie this approach are presented in [62–64]. Any metric that can capture the degradation in the continuity of ϕ is a potential damage indicator. To date, several metrics have been tried, including Holder exponents [65], continuity statistics [66], and prediction error [47]. In some cases, this approach can produce improved results over those obtained by more standard combinations

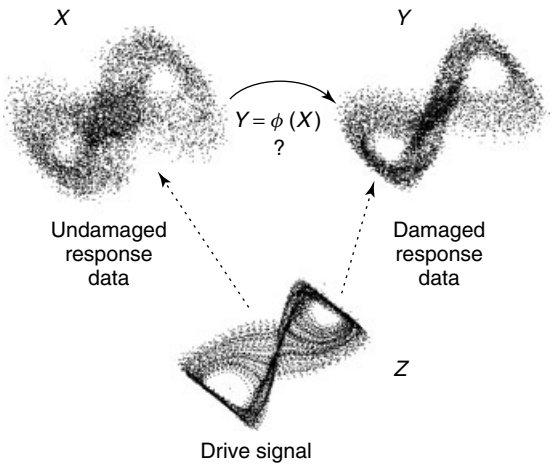


Figure 5. Illustration of the functional relationship between a "damaged" and an "undamaged" response attractor. Capturing the damage-induced changes to this function has produced several viable SHM techniques.

of excitation and feature. The approach is appropriate for linear or nonlinear damage. Other works that focus on the mapping between an undamaged and damaged attractor include those of Chelidze *et al.* [67] and Cusumano *et al.* [68]. In these works, the authors were able to track the battery discharge in a vibrating beam system by tracking changes to the mapping between attractors. More recent work by these authors has led to improvements in the approach, referred to as *phase space warping*. These works are documented in [69, 70].

7 IDENTIFYING THE TYPE OF NONLINEARITY

The problem of nonlinearity detection still poses a number of open questions and will no doubt continue to receive attention. However, some researchers have made progress in identifying the specific type and magnitude of nonlinearity. This is especially valuable if the type and degree of damage is required. Feldman [71, 72] and later Staszewski [73] described an approach for identifying nonlinearity based on the so-called "backbone" curves for a nonlinear system. These curves are obtained via time–frequency analysis and can be used to discern the *form* of the underlying nonlinearity. Feldman then used this approach to detect damage

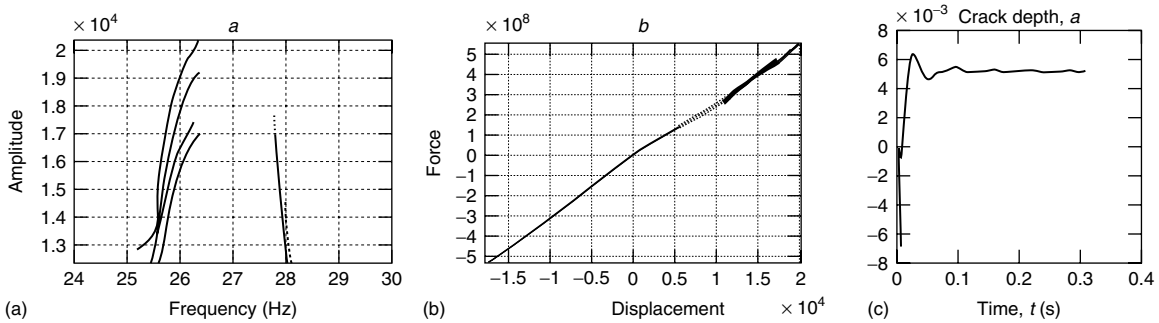


Figure 6. Estimated force characteristic of a crack and notch structure: (a) backbone curve, (b) spring force characteristic, and (c) identification of crack depth. [Reprinted with permission.]

in a rotor system, correctly identifying both the type and location of the defect [32]. This same procedure was also used recently by Tjahjowidodo *et al.* [12] in identifying backlash in a mechanical system.

Figure 6 illustrates Feldman's approach to diagnosing a crack in a rotor system. The first subplot shows the amplitude dependence on frequency, indicating that there exists a nonlinearity. The second plot shows the backbone curve for the rotor system and indicates a slight bilinearity in the stiffness. The final plot demonstrates how the approach detects the size of the crack as a function of time. Relatively few of the existing nonlinear system-identification techniques have been applied to SHM. However, a comprehensive review of system identification for nonlinear structural systems can be found in Worden and Tomlinson [22].

8 SUMMARY

The use of nonlinear features in the SHM problem can be appropriate in a number of instances. These features take advantage of relationships among the response data that are not found in linear systems. The result is that in some cases these features provide a more powerful approach to detecting, quantifying, and locating damage in structures. This article reviews a number of these features, presents the motivation for using them, and describs the context in which they were used. The list is not exhaustive, but should give the reader a good feel for the types of features that have been used to date.

REFERENCES

[1] Worden K, Farrar CR, Haywood J, Todd MD. A review of nonlinear dynamics applications to structural health monitoring. *Structural Control and Health Monitoring* (to appear).

[2] Doebling SW, Farrar CR, Prime MB. A summary review of vibration-based identification mehods. *Shock and Vibration Digest* 1998 **205**(5):631–645.

[3] Sohn H, Farrar CR, Hemez FM, Czarnecki JJ, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Technical Report LA-13070-MS. Los Alamos National Laboratory, 2004.

[4] Schetzen M. *The Volterra and Wiener Theories of Nonlinear Systems*. John Wiley & Sons: New York, 1980.

[5] Collis WB, White PR, Hammond JK. Higher-order spectra: the bispectrum and trispectrum. *Mechanical Systems and Signal Processing* 1998 **12**(3):375–394.

[6] Brandon JA. Some insights into the dynamics of defective structures. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 1998 **212**:441–454.

[7] Friswell MI, Penny JET. Crack modeling for structural health monitoring. *Structural Health Monitoring* 2002 **1**(2):139–148.

[8] Zhang WZ, Testa RB. Closure effects on fatigue crack detection. *ASCE—Journal of Engineering Mechanics* 1999 **125**(10):1125–1132.

[9] Brown RL, Adams DE. Equilibrium point damage prognosis models for structural health monitoring. *Journal of Sound and Vibration* 2003 **262**:591–611.

[10] Rutherford AC, Park G, Farrar CR. Non-linear feature identifications based on self-sensing impedance measurements for structural health assessment.

Mechanical Systems and Signal Processing 2007 **21**:322–333.

[11] Nichols JM, Trickey ST, Seaver M. Detecting damage-induced nonlinearities in structures using information theory. *Journal of Sound and Vibration* 2006 **297**:1–16.

[12] Tjahjowidodo T, Al-Bender F, Brussel HV. Experimental dynamic identification of backlash using skeleton methods. *Mechanical Systems and Signal Processing* 2007 **21**:959–972.

[13] Trendafilova I, Brussel HV. Non-linear dynamics tools for the motion analysis and condition monitoring of robot joints. *Mechanical Systems and Signal Processing* 2001 **15**:1141–1164.

[14] Hunt GW, Hu B, Butler R, Almond DP, Wright JE. Nonlinear modeling of delaminated struts. *AIAA Journal* 2004 **42**(11):2364–2372.

[15] Murphy KD, Nichols JM, Motley SR. Nonlinear mechanics of delaminated beams. In *Proceedings of the 6th International Workshop on Structural Health Monitoring*, Chang F-K (ed). DEStech Publications: Lancaster, PA, 2007.

[16] Luo H, Hanagud S. Dynamics of delaminated beams. *International Journal of Solids and Structures* 2000 **37**:1501–1519.

[17] Chu FL, Lu WX. Experimental observation of nonlinear vibrations in a rub-impact rotor system. *Journal of Sound and Vibration* 2005 **283**:621–643.

[18] Nichols JM, Seaver M, Trickey ST, Bash T, Kasarda M. Use of information theory in structural monitoring applications. In *Proceedings of the 5th International Workshop on Structural Health Monitoring*, Chang F-K (ed). DEStech Publications: Lancaster, PA, 2005.

[19] Peng ZK, Chu FL, Tse PW. Detection of the rubbing-caused impacts for rotor-stator fault diagnosis using reassigned scalogram. *Mechanical Systems and Signal Processing* 2005 **19**(2):391–409.

[20] Song Y, Hartwigsen CJ, McFarland DM, Vakakis AF, Bergman LA. Simulation of dynamics of beam structures with bolted joints using adjusted Iwan beam elements. *Journal of Sound and Vibration* 2004 **273**:249–276.

[21] Jeong SH, Park JM, Lee YZ. Transition of friction and wear by stick-slip phenomenon in various environments under fretting conditions. *Key Engineering Materials* 2006 **321–323**:1344–1347.

[22] Worden K, Tomlinson GR. *Nonlinearity in Structural Dynamics*. Institute of Physics Publishing: Bristol, Philadelphia, PA, 2001.

[23] Kerschen G, Worden K, Vakakis AF, Golinval J-C. Past, present and future of nonlinear system identification in structural dynamics. *Mechanical Systems and Signal Processing* 2006 **20**:505–592.

[24] Adams DE, Farrar CR. Classifying linear and nonlinear structural damage using frequency domain ARX models. *Structural Health Monitoring* 2002 **1**(2):185–201.

[25] Bovsunovsky AP, Surace C. Considerations regarding superharmonic vibrations of a cracked beam and the variation in damping caused by the presence of the crack. *Journal of Sound and Vibration* 2005 **288**:865–886.

[26] Parsons Z, Staszewski WJ. Nonlinear acoustics with low-profile piezoceramic excitation for crack detection in metallic structures. *Smart Materials and Structures* 2006 **15**:1110–1118.

[27] Nichols JM, Trickey ST, Seaver M, Motley SR, Eisner ED. Using ambient vibrations to detect loosening of a composite-to-metal bolted joint in the presence of strong temperature fluctuations. *Journal of Vibration and Acoustics* 2007 **129**:710–717.

[28] Johnson TJ, Adams DE. Transmissibility as a differential indicator of structural damage. *Journal of Vibration and Acoustics* 2002 **124**:634–641.

[29] Wong L-A, Chen J-C. Damage identification of nonlinear structural systems. *AIAA Journal* 2000 **38**(8):1444–1452.

[30] D'Souza K, Epureanu BI. Damage detection in nonlinear systems using system augmentation and generalized minimum rank perturbation theory. *Smart Materials and Structures* 2005 **14**:989–1000.

[31] Neild SA, Williams MS, McFadden PD. Nonlinear vibration characteristics of damaged concrete beams. *Journal of Structural Engineering* 2003 **129**(2): 260–268.

[32] Feldman M, Seibold S. Damage diagnosis of rotors: application of Hilbert transform and multihypothesis testing. *Journal of Vibration and Control* 1999 **5**(3):421–442.

[33] Wei Z, Yam LH, Cheng L. Narmax model representation and its application to damage detection for multi-layer composites. *Composite Structures* 2005 **68**:109–117.

[34] Robertson AN, Farrar CR, Sohn H. Singularity detection for structural health monitoring using holder exponents. *Mechanical Systems and Signal Processing* 2003 **17**(6):1163–1184.

[35] Pecora LM, Moniz L, Nichols JM, Carroll TL. A unified approach to attractor reconstruction. *Chaos* 2007 **17**:013110.

[36] Todd MD, Nichols JM, Pecora LM, Virgin LN. Vibration-based damage assessment utilizing state space geometry changes: local attractor variance ratio. *Smart Materials and Structures* 2001 **10**:1000–1008.

[37] Overbey LA, Todd MD. Analysis of local state space models for feature extraction in structural health monitoring. *Structural Health Monitoring: An International Journal* 2007 **6**(2):145–172.

[38] Theiler J. Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A* 1986 **34**:2427.

[39] Logan D, Mathew J. Using the correlation dimension for vibration fault diagnosis of rolling element bearings—I. Basic concepts. *Mechanical Systems and Signal Processing* 1996 **10**(3):241–250.

[40] Logan D, Mathew J. Using the correlation dimension for vibration fault diagnosis of rolling element bearings—II. Selection of experimental parameters. *Mechanical Systems and Signal Processing* 1996 **10**(3):251–264.

[41] Craig C, Neilson RD, Penman J. The use of correlation dimension in condition monitoring of systems with clearance. *Journal of Sound and Vibration* 2000 **231**(1):1–17.

[42] Wang WJ, Chen J, Wu XK. The application of some non-linear methods in rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing* 2001 **15**(4):697–705.

[43] Trendafilova I. Vibration-based damage detection in structures using time series analysis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 2006 **220**(3):261–272.

[44] Nichols JM, Virgin LN, Todd MD, Nichols JD. On the use of attractor dimension as a feature in structural health monitoring. *Mechanical Systems and Signal Processing* 2003 **17**(6):1305–1320.

[45] Epureanu BI, Yin S-H. Identification of damage in an aeroelastic system based on attractor deformations. *Computers and Structures* 2004 **82**:2743–2751.

[46] Nichols JM, Todd MD, Seaver ME, Virgin LN. Use of chaotic excitation and attractor property analysis in structural health monitoring. *Physical Review E* 2003 **67**:016209.

[47] Nichols JM, Todd MD, Wait JR. Using state space predictive modeling with chaotic interrogation in detecting preload loss in a frame structure experiment. *Smart Materials and Structures* 2003 **12**(4):580–601.

[48] Nichols JM, Nichols CJ, Todd MD, Seaver M, Trickey ST, Virgin LN. Use of data-driven phase space models in assessing the strength of a bolted connection in a composite beam. *Smart Materials and Structures* 2004 **13**:241–250.

[49] Hively LM, Protopopescu VA. Machine failure forewarning via phase-space dissimilarity measures. *Chaos* 2004 **14**(2):408–419.

[50] Golnaraghi M, Lin D, Fromme P. Gear damage detection using chaotic dynamics techniques: a preliminary study. *Proceedings of the 1995 ASME Design Engineering Technical Conferences, Symposium on Time-Varying Systems and Structures*. Boston, MA, 1995; pp. 121–127.

[51] Rivola A, White PR. Bispectral analysis of the bilinear oscillator with application to the detection of cracks. *Journal of Sound and Vibration* 1998 **216**(5):889–910.

[52] Zhang GC, Chen J, Li FC, Li WH. Extracting gear fault features using maximal bispectrum. *Key Engineering Materials* 2005 **293–294**:167–174.

[53] Nichols JM. Examining structural dynamics using information flow. *Probabilistic Engineering Mechanics* 2006 **21**:420–433.

[54] Teng KK, Brandon JA. Diagnostics of a system with an interface nonlinearity using higher order spectral estimators. *Key Engineering Materials* 2001 **204–205**:271–285.

[55] Nichols JM, Milanese A, Marzocca P. Characterizing the auto-bispectrum as a detector of nonlinearity in structural systems. *Proceedings of the SPIE, Health Monitoring and Smart Nondestructive Evaluation of Structural and Biological Systems IV*. SPIE Optical Engineering Press: Bellingham, WA, 2007; Vol. 6523.

[56] Park HW, Sohn H, Law KH, Farrar CR. Time reversal active sensing for health monitoring of a composite plate. *Journal of Sound and Vibration* 2007 **302**:50–66.

[57] Kim SD, In CW, Cronin KE, Sohn H, Harries K. Reference-free NDT technique for debonding detection in CFRP-strengthened RC structures. *Journal of Structural Engineering* 2007 **133**(8):1080–1091.

[58] Nichols JM, Seaver M, Trickey ST, Salvino LW, Pecora DL. Detecting impact damage in experimental composite structures: an information-theoretic

approach. *Smart Materials and Structures* 2006 **15**:424–434.

[59] Schreiber T, Schmitz A. Improved surrogate data for nonlinearity tests. *Physical Review Letters* 1996 **77**(4):635–638.

[60] Dolan KT, Spano ML. Surrogate for nonlinear time series analysis. *Physical Review E* 2001 **64**: 0461281–0461286; Article No. 046128.

[61] Yin S-H, Epureanu BI. Enhanced nonlinear dynamics and monitoring bifurcation morphing for the identification of parameter variations. *Journal of Fluids and Structures* 2005 **21**:543–559.

[62] Davies ME, Campbell KM. Linear recursive filters and nonlinear dynamics. *Nonlinearity* 1996 **9**:487–499.

[63] Hunt BR, Ott E, Yorke JA. Differentiable generalized synchronization of chaos. *Physical Review E* 1997 **55**(4):4029–4034.

[64] Ott W, Yorke J. Learning about reality from observation. *SIAM Journal on Applied Dynamical Systems (Online)* 2003 **2**(3):297–322.

[65] Nichols JM, Trickey ST, Seaver M, Moniz L. Use of fiber optic strain sensors and holder exponents for detecting and localizing damage in an experimental plate structure. *Journal of Intelligent Material Systems and Structures* 2007 **18**(1):51–67.

[66] Moniz L, Pecora LM, Nichols JM, Todd MD, Wait JR. Dynamical assessment of structural damage using the continuity statistic. *International Journal of Structural Health Monitoring* 2004 **3**(3): 199–212.

[67] Chelidze D, Cusumano JP, Chatterjee A. A dynammical systems approach to damage evolution tracking, part 1: description and experimental application. *Journal of Vibration and Acoustics—Transactions of the ASME* 2002 **124**:250–257.

[68] Cusumano JP, Chelidze D, Chatterjee A. A dynamical systems approach to damage evolution tracking, part 2: model-based validation and physical interpretation. *Journal of Vibration and Acoustics—Transactions of the ASME* 2002 **124**:258–264.

[69] Chelidze D, Liu M. Multidimensional damage identification based on phase space warping: an experimental study. *Nonlinear Dynamics* 2006 **46**:61–72.

[70] Liu M, Chelidze D. Identifying damage using local flow variation method. *Smart Materials and Structures* 2006 **15**:1830–1836.

[71] Feldman M. Non-linear system vibration analysis using Hilbert transform—I. Free vibration analysis method "freevib". *Mechanical Systems and Signal Processing* 1994 **8**(2):119–127.

[72] Feldman M. Non-linear system vibration analysis using Hilbert transform—II. Forced vibration analysis method "forcevib". *Mechanical Systems and Signal Processing* 1994 **8**(3):309–318.

[73] Staszewski WJ. Identification of non-linear systems using multi-scale ridges and skeletons of the wavelet transform. *Journal of Sound and Vibration* 1998 **214**(4):639–658.

Chapter 33

Dimensionality Reduction Using Linear and Nonlinear Transformation

Gaëtan Kerschen and Jean-Claude Golinval

Aerospace and Mechanical Engineering Department (LTAS), University of Liège, Liège, Belgium

1 INTRODUCTION

In many domains of applied sciences, dealing with large data sets is a central issue. In structural dynamics and particularly in large civil infrastructure applications, densely distributed sensors are required for reliable health monitoring procedures. With the recent advances in communication technology, wireless monitoring systems have emerged as a promising and practical solution in this context [1–3]. There is, therefore, a clear trend toward very large experimental data sets, and dimensionality reduction has become an important step of structural health monitoring [4].

Many dimensionality reduction techniques have been proposed in the statistical literature. The present article is biased toward one popular technique, called *proper orthogonal decomposition* (*POD*). This method may serve two purposes, namely, order reduction by projecting high-dimensional data onto a lower dimensional space and feature extraction by revealing relevant but unexpected structure hidden in the data. The key idea of the POD is to reduce a large number of interdependent variables to a much smaller number of uncorrelated variables while retaining as much as possible of the variation in the original variables. An orthogonal transformation to the basis of the eigenvectors of the sample covariance matrix is performed, and the data are projected onto the subspace spanned by the eigenvectors corresponding to the largest eigenvalues. This transformation decorrelates the signal components and maximizes variance.

The most striking property of the POD is its optimality in the sense that it minimizes the average squared distance between the original signal and its reduced linear representation. Although it is frequently applied to nonlinear problems, it should be borne in mind that the POD only gives the optimal approximating linear manifold in the space represented by the data. As stated in [5], the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

POD is "a safe haven in the intimidating world of nonlinearity; it may not do the physical violence of linearization methods".

This article is organized as follows. In the next section, the POD method is briefly presented, and the different means of computing this decomposition are described. Section 3 introduces two nonlinear extensions of the POD, namely, a global approach called *nonlinear principal component analysis* (*NLPCA*) and a local approach called *vector quantization principal component analysis* (*VQPCA*). Finally, reduced-order modeling and damage detection under varying environmental conditions are discussed in Section 4.

2 PROPER ORTHOGONAL DECOMPOSITION

2.1 Brief historical perspective

The POD, also known as the *Karhunen–Loève decomposition* (*KLD*), was proposed independently by several scientists including Karhunen, Kosambi, Loève, Obukhov, and Pougachev and was originally conceived in the framework of continuous second-order processes. When restricted to a finite dimensional case and truncated after a few terms, the POD is equivalent to principal component analysis (PCA) [6]. This latter methodology originated not only with the work of Pearson [7] as a means of fitting planes by orthogonal least squares but was also proposed by Hotelling [8]. It is emphasized that the method appears in various guises in the literature and is known by other names depending on the area of application, namely, PCA in the statistical literature, empirical orthogonal function in oceanography and meteorology, and factor analysis in psychology and economics. The reader can refer to [9, 10] for a detailed discussion about the equivalence of the POD, PCA, and KLD, and their connection with the singular value decomposition (SVD).

Because of the large amount of computations required to find the POD modes, the technique was virtually unused until the middle of the twentieth century. Radical changes came with the appearance of powerful computers. The POD has now gained popularity and is being used in numerous fields (e.g., biomedical engineering, forecasting in meteorology, classification of speech data, image processing, and chemical engineering). The first applications of the POD in the field of structural dynamics date back to the 1990s [11, 12].

The POD now enjoys various applications in structural dynamics, such as active control [13], dynamic characterization [14–16], finite element model updating [17, 18], modal analysis [19, 20], model order reduction [21–24], and stochastic structural dynamics [25]. Recently, the POD has also been used for damage detection [26–35] and sensor validation [36].

2.2 Mathematical formulation

The mathematical formulation of the POD presented here closely follows the one in [37]. Let $\theta(x,t)$ be a random field on a domain Ω. This field is first decomposed into mean $\mu(x)$ and time varying parts $\vartheta(x,t)$:

$$\theta(x,t) = \mu(x) + \vartheta(x,t) \tag{1}$$

At time t_k, the system displays a snapshot $\vartheta^k(x) = \vartheta(x,t_k)$. The POD aims at obtaining the most characteristic structure $\varphi(x)$ of an ensemble of snapshots of the field $\vartheta(x,t)$. This is equivalent to finding the basis function $\varphi(x)$ that maximizes the ensemble average of the inner products between $\vartheta^k(x)$ and $\varphi(x)$:

$$\text{Maximize} \quad \langle |(\vartheta^k, \varphi)|^2 \rangle \quad \text{with } \|\varphi\|^2 = 1 \tag{2}$$

where $(f,g) = \int_\Omega f(x)g(x)\,\mathrm{d}\Omega$ denotes the inner product in Ω; $\langle\cdot\rangle$ denotes the averaging operation; $\|\cdot\| = (\cdot,\cdot)^{1/2}$ denotes the norm; and $|\cdot|$ denotes the modulus. Expression (2) means that if the field ϑ is projected along φ, the average energy content is greater than if the field is projected along any other basis function.

The constraint $\|\varphi\|^2 = 1$, imposed to make the computation unique, can be taken into account by the use of a Lagrange multiplier

$$J[\varphi] = \langle |(\vartheta, \varphi)|^2 \rangle - \lambda \left(\|\varphi\|^2 - 1 \right) \tag{3}$$

The extremum is reached when the functional derivative is equal to zero. Holmes *et al.* [37] shows that this condition reduces to the following integral

eigenvalue problem:

$$\int_{\Omega} \langle \vartheta^k(x)\vartheta^k(x')\rangle \varphi(x')\,\mathrm{d}x' = \lambda\varphi(x) \tag{4}$$

where $\langle \vartheta^k(x)\vartheta^k(x')\rangle$ is the averaged autocorrelation function.

The solution of the optimization problem (2) is thus given by the orthogonal eigenfunctions $\varphi_i(x)$ of the integral equation (4), called the *proper orthogonal modes* (POMs) or POD modes. The corresponding eigenvalues λ_i ($\lambda_i \geq 0$) are the proper orthogonal values (POVs). The POMs may be used as a basis for the decomposition of the field $\vartheta(x,t)$:

$$\vartheta(x,t) = \sum_{i=1}^{\infty} a_i(t)\varphi_i(x) \tag{5}$$

where the coefficients $a_i(t)$ are uncorrelated, i.e., $\langle a_i(t)a_j(t)\rangle = \delta_{ij}\lambda_i$, and are determined by $a_i(t) = (\vartheta(x,t), \varphi_i(x))$.

The POM associated with the greatest POV is the optimal vector to characterize the ensemble of snapshots. The POM associated with the second greatest POV is the optimal vector to characterize the ensemble of snapshots but restricted to the space orthogonal to the first POM, and so forth. The energy ε contained in the data is defined as the sum of the POVs, i.e., $\varepsilon = \sum_j \lambda_j$, and the energy percentage captured by the ith POM is given by $\lambda_i / \sum_j \lambda_j$.

We note that a physical interpretation of the POMs in a structural dynamics context has been provided in [38, 39].

2.3 Practical computation

In practice, the data are discretized in space and time. Accordingly, n observations of an m-dimensional vector $\mathbf{x}$ are collected, and an $(m \times n)$ response matrix is formed:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} x_{11} & \cdot & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{6}$$

The purpose of this section is to present several means of computing the POD.

2.3.1 Eigensolutions of the sample covariance matrix

Since the data are now discretized and do not necessarily have a zero mean, the averaged autocorrelation function is replaced by the covariance matrix $\boldsymbol{\Sigma} = E\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\right]$, where $E[\cdot]$ is the expectation, and $\boldsymbol{\mu} = E[\mathbf{x}]$ is the mean of the vector $\mathbf{x}$. Under the assumption that the process is stationary and ergodic and that the number of time instants is large, a reliable estimate of the covariance matrix is given by the sample covariance matrix

$$\boldsymbol{\Sigma}_{\mathrm{S}} = \frac{1}{n}\begin{Bmatrix} \left\{\sum_{j=1}^{n}\left(x_{1j} - \frac{1}{n}\sum_{k=1}^{n} x_{1k}\right)^2\right\} & \cdots & \left\{\sum_{j=1}^{n}\left(x_{1j} - \frac{1}{n}\sum_{k=1}^{n} x_{1k}\right)\left(x_{mj} - \frac{1}{n}\sum_{k=1}^{n} x_{mk}\right)\right\} \\ \cdots & \cdots & \cdots \\ \left\{\sum_{j=1}^{n}\left(x_{mj} - \frac{1}{n}\sum_{k=1}^{n} x_{mk}\right)\left(x_{1j} - \frac{1}{n}\sum_{k=1}^{n} x_{1k}\right)\right\} & \cdots & \left\{\sum_{j=1}^{n}\left(x_{mj} - \frac{1}{n}\sum_{k=1}^{n} x_{mk}\right)^2\right\} \end{Bmatrix} \tag{7}$$

The POMs and POVs are thus characterized by the eigensolutions of the sample covariance matrix $\mathbf{\Sigma}_S$. If the data have a zero mean, the sample covariance is merely given by the following expression:

$$\mathbf{\Sigma}_S = \frac{1}{n}\mathbf{X}\mathbf{X}^T \tag{8}$$

It should be noted that when the number of degrees of freedom is much higher than the number of snapshots (e.g., in turbulence theory), the computation of the sample covariance matrix may become expensive. In this context, the POD modes are most easily computed using the method of snapshots proposed by Sirovich [40].

2.3.2 Singular value decomposition of the response matrix

For any real $(m \times n)$ matrix $\mathbf{X}$ (i.e., for our purpose, the response matrix measured about its mean), there exists a real factorization called the *singular value decomposition* that can be written as

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{9}$$

where $\mathbf{U}$ is an $(m \times m)$ orthonormal matrix containing the left singular vectors; $\mathbf{S}$ is an $(m \times n)$ pseudodiagonal and semipositive definite matrix with diagonal entries containing the singular values σ_i and $\mathbf{V}$ is an $(n \times n)$ orthonormal matrix containing the right singular vectors.

Reliable algorithms have been developed to compute the SVD. The SVD may also be calculated by solving two eigenvalue problems, or even one if only the left or the right singular vectors are required. Indeed,

$$\begin{aligned}\mathbf{X}\mathbf{X}^T &= \mathbf{U}\mathbf{S}^2\mathbf{U}^T \\ \mathbf{X}^T\mathbf{X} &= \mathbf{V}\mathbf{S}^2\mathbf{V}^T\end{aligned} \tag{10}$$

According to equation (10), the singular values of $\mathbf{X}$ are found to be the square roots of the eigenvalues of $\mathbf{X}\mathbf{X}^T$ or $\mathbf{X}^T\mathbf{X}$. In addition, the left and right singular vectors of $\mathbf{X}$ are the eigenvectors of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$, respectively. The POMs, defined as the eigenvectors of the sample covariance matrix $\mathbf{\Sigma}_S$ (8), are thus equal to the left singular vectors of $\mathbf{X}$. The POVs, defined as the eigenvalues of matrix $\mathbf{\Sigma}_S$, are the square of the singular values divided by the number of samples m.

The main advantage in considering the SVD to compute the POD instead of the eigenvalue problem described in Section 2.3.1 is that additional information is obtained through the matrix $\mathbf{V}$. The column $\mathbf{v}_i$ of matrix $\mathbf{V}$ contains the time modulation of the corresponding POM $\mathbf{u}_i$, normalized by the singular value σ_i. This information provides important insight into the system dynamics and plays a prominent role in the model updating of nonlinear systems [18].

2.3.3 A simple learning rule

Consider the following simple iterative scheme [41]:

$$\begin{aligned}y &= \mathbf{x}^T\mathbf{w} \\ \Delta\mathbf{w} &= \mathbf{w}_{(j+1)} - \mathbf{w}_{(j)} \\ &= \alpha\left(y_{(j)}\mathbf{x}_{(j)} - {y_{(j)}}^2\mathbf{w}_{(j)}\right)\end{aligned} \tag{11}$$

where $\mathbf{x}$ is the vector containing the inputs; $\mathbf{w}$ is the vector containing the weights; y is the output; α is the learning rate controlling the speed of convergence; and $\ldots_{(j)}$ refers to the result of the jth iteration.

After convergence, the expectation of the weight update $\Delta\mathbf{w}$ is equal to zero:

$$\begin{aligned}E[\Delta\mathbf{w}] &= E[\alpha(y\mathbf{x} - y^2\mathbf{w})] \\ &= \alpha E\left[\mathbf{x}\mathbf{x}^T\mathbf{w} - \mathbf{w}^T\mathbf{x}\mathbf{x}^T\mathbf{w}\mathbf{w}\right] \\ &= \alpha\left(\mathbf{\Sigma}\mathbf{w} - \mathbf{w}^T\mathbf{\Sigma}\mathbf{w}\mathbf{w}\right) = 0\end{aligned} \tag{12}$$

and,

$$\mathbf{\Sigma}\mathbf{w} = \left(\mathbf{w}^T\mathbf{\Sigma}\mathbf{w}\right)\mathbf{w} = \lambda\mathbf{w} \tag{13}$$

where $\mathbf{\Sigma} = E[\mathbf{x}\mathbf{x}^T]$ is the covariance matrix of $\mathbf{x}$ that is assumed to be zero mean. As can be seen from equation (13), the weights converge to an eigenvector of the covariance matrix; i.e., a POM. In fact, it can be shown that they approach the eigenvector with the largest eigenvalue. This scheme can easily be generalized to extract the first k POMs.

2.3.4 Autoassociative neural networks

The POD can also be computed using autoassociative neural networks (AANN). An artificial neural

network is a computational system inspired by the learning characteristics and the structure of biological neural networks [42]. The type of neural network considered here is composed of a series of parallel layers, each of which contains a number of processing elements analogous to neurons. The neurons are tied together with weighted connections that are analogous to synapses. The network has distinguished input and output neurons, and all other neurons are called *hidden neurons*. Each neuron computes a weighted sum of its inputs, and its resulting output is a well-defined function of the weighted sum. The outputs of the ith layer are used as inputs to the $(i+1)$th layer. If $y_j^{(i)}$ is the output of the jth neuron of the ith layer, then

$$y_k^{(i+1)} = \sigma^{(i+1)}\left(\sum_j w_{kj}^{(i+1)} y_j^{(i)} + b_k^{(i+1)}\right) \quad (14)$$

is the output of the kth neuron of the $(i+1)$th layer. The elements w_{kj} and b_k are referred to as *weights* and *biases*, respectively, and $\sigma(\cdot)$ is a linear or nonlinear transfer function.

AANNs are those in which the target output pattern is identical to the input pattern. The aim is thus to approximate, as closely as possible, the input data itself. When used with a hidden layer smaller than the input and output layers, a perfect reconstruction of all input data is generally not possible.

An AANN with linear activation functions performs a compression scheme equivalent to the POD [43]. To this end, let us consider an AANN with a single hidden unit and a linear activation function as shown in Figure 1. The sum of squares error (SSE) function for this network is given by the following equation:

$$\begin{aligned} SSE &= \tfrac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}}(\mathbf{x}-\hat{\mathbf{x}}) \\ &= \tfrac{1}{2}(\mathbf{x}-y\mathbf{w})^{\mathrm{T}}(\mathbf{x}-y\mathbf{w}) \end{aligned} \quad (15)$$

The derivative of the error function is

$$\begin{aligned} \frac{\partial(SSE)}{\partial \mathbf{w}} &= -\left(\frac{\partial y}{\partial \mathbf{w}}\mathbf{w} + y\frac{\partial \mathbf{w}}{\partial \mathbf{w}}\right)(\mathbf{x}-y\mathbf{w}) \\ &= -\left(\mathbf{x}^{\mathrm{T}}\mathbf{w} + y\right)(\mathbf{x}-y\mathbf{w}) \\ &= -2y(\mathbf{x}-y\mathbf{w}) \end{aligned} \quad (16)$$

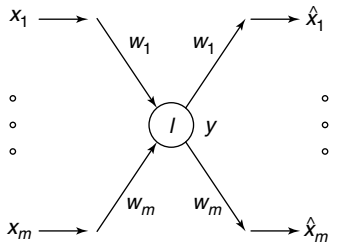


Figure 1. An autoassociative neural network with a single hidden unit with a linear activation function.

and the weight update becomes

$$\Delta\mathbf{w} = -\alpha\frac{\partial(SSE)}{\partial \mathbf{w}} = 2\alpha(y\mathbf{x} - y^2\mathbf{w}) \quad (17)$$

This latter expression is equivalent to learning rule (11), which demonstrates the ability of the neural network to perform the POD. Although neural networks offer no direct advantage over other means of computing the POD, they suggest an interesting nonlinear generalization, as detailed in the next section.

3 NONLINEAR GENERALIZATIONS OF PROPER ORTHOGONAL DECOMPOSITION

The linear nature of the POD may represent a restriction for some data sets. For instance, the covariance matrix of data sampled from a helix in $\mathbb{R}^3$ has full rank, and the POD requires the use of three variables for the description of the data. However, the helix is a one-dimensional manifold and can be smoothly parameterized with a single variable.

The POD can thus determine an appropriate embedding space for given data, but it cannot provide the most efficient description of a data set where nonlinear dependencies are present. To address this issue, researchers in the field of statistics and neural networks have developed alternatives to the POD that can take nonlinear correlations between the variables into account. NLPCA [44] and VQPCA [45] are two interesting alternatives that are briefly described in the following subsections.

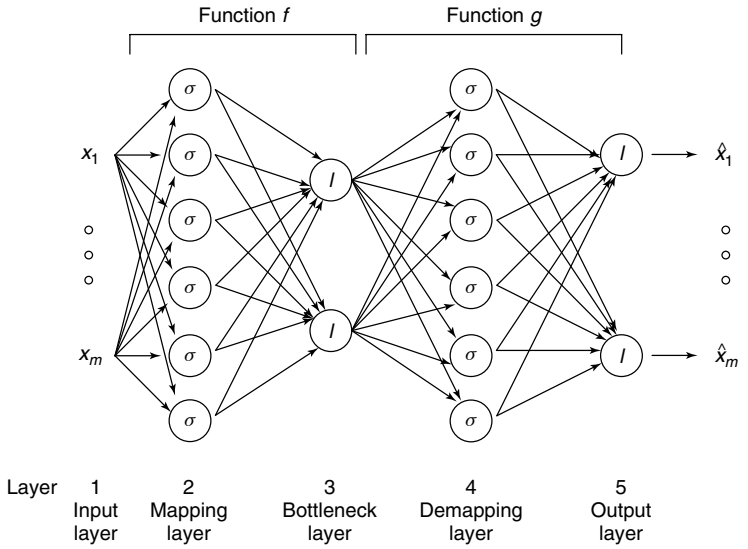


Figure 2. Network architecture for implementation of NLPCA (σ = sigmoidal transfer function, l = linear transfer function).

3.1 A global approach

Generally speaking, a dimension reduction technique provides an approximation to a data set, which is the composition of two functions f and g:

$$\mathbf{x} = \hat{\mathbf{x}} + \boldsymbol{\rho} = g(f(\mathbf{x}(t))) + \boldsymbol{\rho} \tag{18}$$

The projection function $f: \mathbb{R}^m \rightarrow \mathbb{R}^r$ projects the original m-dimensional data onto an r-dimensional subspace, whereas the expansion function $g: \mathbb{R}^r \rightarrow \mathbb{R}^m$ defines a mapping from an r-dimensional space back into the original m-dimensional space with $\boldsymbol{\rho}$ as the residue. Dimensionality reduction involves the determination of adequate functions f and g.

As the POD can be computed using AANNs with linear activation functions (Section 2.3.4), the most obvious generalization of the POD would be to consider nonlinear transfer functions in the hidden layer. However, a much better compression can be obtained by exploiting a theorem due to Cybenko [46]. It states that a three-layer neural network with m input neurons, nonlinear transfer functions in the second layer, and linear transfer functions in the third layer of r neurons can approximate to arbitrary accuracy any continuous functions from $\mathbb{R}^m$ to $\mathbb{R}^r$. This is true, provided that the number of neurons in the second layer is sufficiently large. This theorem suggests the use of a three-layer network to recover projection function f and the use of another three-layer network to recover expansion function g. The resulting five-layer network, which implements NLPCA, is shown in Figure 2.

A crucial step when using a neural network is the training phase. The neural network parameters (i.e., the weights and biases) are updated iteratively until the output of the network approximates the input as closely as possible. Eventually, this results in minimum information loss in the same sense as in the POD.

For further details about NLPCA, the reader can refer to [44]. In structural dynamics, NLPCA has been exploited for separating the structural changes from the changes caused by the environment in a damage detection context [47], for finite element model updating [48] and for sensor validation and reconstruction [49].

3.2 A local approach

The POD and NLPCA try to describe all the data using the same global features. An alternative paradigm is to capture data complexity by a combination of local linear POD projections. A local

model implementation of the POD involves a two-step procedure, i.e.,

1. a clustering of the data space into distinct regions by vector quantization;
2. the construction of separate low-dimensional coordinate systems in each local region using the POD.

The VQPCA algorithm is an extension of a standard vector quantizer [45]. VQPCA partitions the input space into a set of regions and approximates each region by a hyperplane defined by the POD, while a standard vector quantizer approximates each region by a codebook vector. The VQPCA algorithm is as follows:

1. Partition $\mathbb{R}^m$ into q disjoint regions $S_1, \ldots, S_q$ with a nearest neighbor approach using the Euclidean distance as the distortion function (other distortion functions can also be used [50]).
2. For each Voronoi cell S_j and its corresponding centroid $\boldsymbol{\mu}_j$, estimate the local covariance matrix

$$\boldsymbol{\Sigma}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in S_j} (\mathbf{x} - \boldsymbol{\mu}_j)(\mathbf{x} - \boldsymbol{\mu}_j)^{\mathrm{T}} \qquad (19)$$

   where N_j is the number of vectors mapped to S_j. Next, compute the eigenvectors $(\mathbf{p}_{j1}, \ldots, \mathbf{p}_{jr})$ of each matrix $\boldsymbol{\Sigma}_j$.
3. To reduce the dimension of any m-dimensional vector $\mathbf{x}_i$, determine the cell S_j, which contains the vector and project $\mathbf{x}_i$ onto the r leading eigenvectors to obtain the local linear coordinates

$$\begin{aligned} \mathbf{z}_i &= f_j(\mathbf{x}_i) = [\mathbf{p}_{j1} \ldots \mathbf{p}_{jr}]^{\mathrm{T}} (\mathbf{x}_i - \boldsymbol{\mu}_j) \\ &= \begin{bmatrix} \mathbf{p}_{j1}^{\mathrm{T}} (\mathbf{x}_i - \boldsymbol{\mu}_j) \\ \vdots \\ \mathbf{p}_{jr}^{\mathrm{T}} (\mathbf{x}_i - \boldsymbol{\mu}_j) \end{bmatrix} \qquad \text{if } \mathbf{x}_i \in S_j \end{aligned} \qquad (20)$$

   The compressed representation of $\mathbf{x}_i$ consists of the index j of the Voronoi cell in which $\mathbf{x}_i$ lies and the r-dimensional vector $\mathbf{z}_i$. The data are reconstructed from this representation according to

$$\begin{aligned} \hat{\mathbf{x}}_i &= g_j(f_j(\mathbf{x}_i)) = g_j(\mathbf{z}_i) \\ &= \boldsymbol{\mu}_j + [\mathbf{p}_{j1} \ldots \mathbf{p}_{jr}] \mathbf{z}_i \end{aligned} \qquad (21)$$

   The accuracy of the compressed representation is assessed using the normalized mean square error (MSE):

$$MSE = \frac{E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]}{E[\|\mathbf{x} - E[\mathbf{x}]\|^2]} \qquad (22)$$

For further details about VQPCA, the reader can refer to [45]. In structural dynamics, VQPCA has been exploited for model reduction [51] and for separating the structural changes from the changes caused by the environment (e.g., temperature) for damage detection purposes [52].

4 APPLICATION EXAMPLES

4.1 Reduced-order modeling

A number of studies in the technical literature have shown that the POD provides a suitable basis for reduced-order modeling in structural dynamics (e.g., [14, 23]). We also aim to demonstrate herein the efficiency of a nonlinear dimensionality reduction technique such as VQPCA with respect to the classical POD. To this end, the reconstruction of the dynamical response of a nonlinear cantilever beam is taken as an example. As shown in Figure 3, the clamped beam is modeled with seven beam elements, and the local nonlinearity k_{nl} is a spring that exhibits a cubic stiffness. The free vibration of the beam is simulated with an initial displacement given by a static force F_0 applied at the end of the beam.

The data set consists of seven vertical accelerations measured along the beam. The POD, which is equivalent to VQPCA with a single Voronoi cell, is first applied to the data. In a second step, the data are reconstructed using VQPCA with a number of

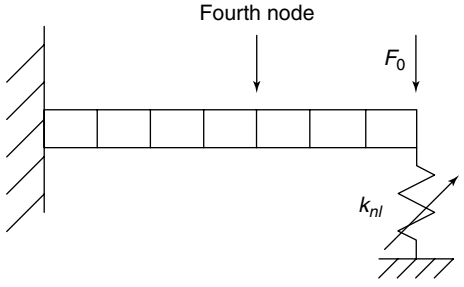


Figure 3. Model of the nonlinear beam.

Table 1. POD and VQPCA applied to the nonlinear beam example

Number of regions	MSE (%) 1 modes	MSE (%) 2 modes	MSE (%) 3 modes	MSE (%) 4 modes
1 (POD)	31.75	9.15	3.53	0.010
5	17.45	7.30	1.52	0.009
10	12.34	5.09	1.18	0.008
20	6.72	2.85	0.83	0.006
30	5.69	2.08	0.60	0.004

regions varying from 5 to 30. Table 1 summarizes the respective performance of the POD and VQPCA in terms of compression accuracy measured by the MSE. It is clear that VQPCA provides a more accurate representation of the system response than POD. For instance, for a unimodal representation, VQPCA with 20 cells attains about 80% lower error than POD and is still a better approximation than a bimodal POD representation.

4.2 Damage detection under varying environmental conditions

In addition to their dimensionality reduction capabilities, the POD and its nonlinear generalizations may also be used to eliminate the influence of environmental conditions in a damage detection context, and this without measuring the environmental variables. The basic idea of the methodology is to collect data on the healthy structure during a sufficiently long period of time, so that a reduced-order model capturing the environmental effects can be built. Eventually, this model can be used for a robust monitoring of the structural health under varying environmental conditions.

To illustrate the method, a geometric interpretation is presented in a two-dimensional case, i.e., when two features (e.g., the first two natural frequencies y_1, y_2 of a structure) are considered. As shown in Figure 4, the features collected on the healthy structure (represented by circles) are distributed around their geometric center (point O'). It is assumed that environmental variations are mainly responsible for the dispersion of the features. The application of the POD to this data set gives two POMs, namely, POM-I and POM-II. We note that the mean is not subtracted from the original data, for reasons explained in [31]. The first POM is associated with the highest singular value and is responsible for the largest variation of the features; it corresponds to the main environmental factor (or a combined effect of several factors). The second POM represents the effect of secondary factors. Considering point Y as an example, we first project this 2-D data onto the 1-D space spanned by the first POM. It results in a scalar equal to the length of segment $\overrightarrow{OX_1}$. Remapping this data point into the original 2-D space results in point Y_1, and

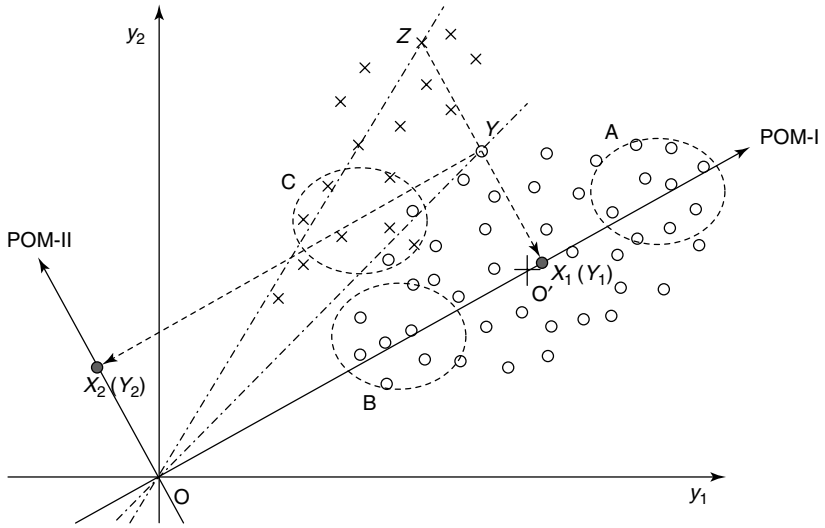


Figure 4. Geometric interpretation of the damage detection procedure under varying environmental conditions.

the corresponding residual error is measured by the length of segment $\overrightarrow{YY_1}$.

Let us now examine another set of identified features, indicated by the symbol X in Figure 4 and obtained from the damaged structure, still under varying environmental conditions. One inherent assumption of the proposed methodology is that the functional relationship between the features and damage is different from that between the features and the environmental factors. Considering point Z as an example, by using a similar projection process as for Y, the residual error $\overrightarrow{ZY_1}$ increases significantly with respect to $\overrightarrow{YY_1}$. In such a comparison between healthy and damaged states, the effect of the environmental factors has therefore been taken into account.

A damage indicator, the novelty index (NI), can be defined as the Euclidean norm of the computed residual error [53]. Defining $\overline{NI}$ and σ as the mean value and standard deviation of the NI for the prediction in the reference state, respectively, an X-bar control chart is constructed by drawing a centerline at $\overline{NI}$ and one additional horizontal line corresponding to the upper control limit, $UCL = \overline{NI} + 3\sigma$, which corresponds to a confidence interval of 99.7% in the case of a normal distribution. The presence of data above this control limit is potentially an indication of damage.

The POD and VQPCA algorithms were applied in [52] to experimental data resulting from 305 days of continuous monitoring of the Z24 bridge. During the monitoring period, accelerations were measured every hour on the structure, and modal parameters were identified automatically. From the first day of monitoring to the 268th day, the structure could be

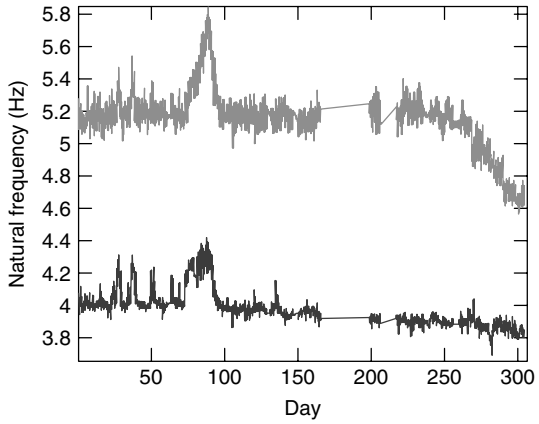


Figure 5. First two natural frequencies of the Z24 bridge during the monitoring period.

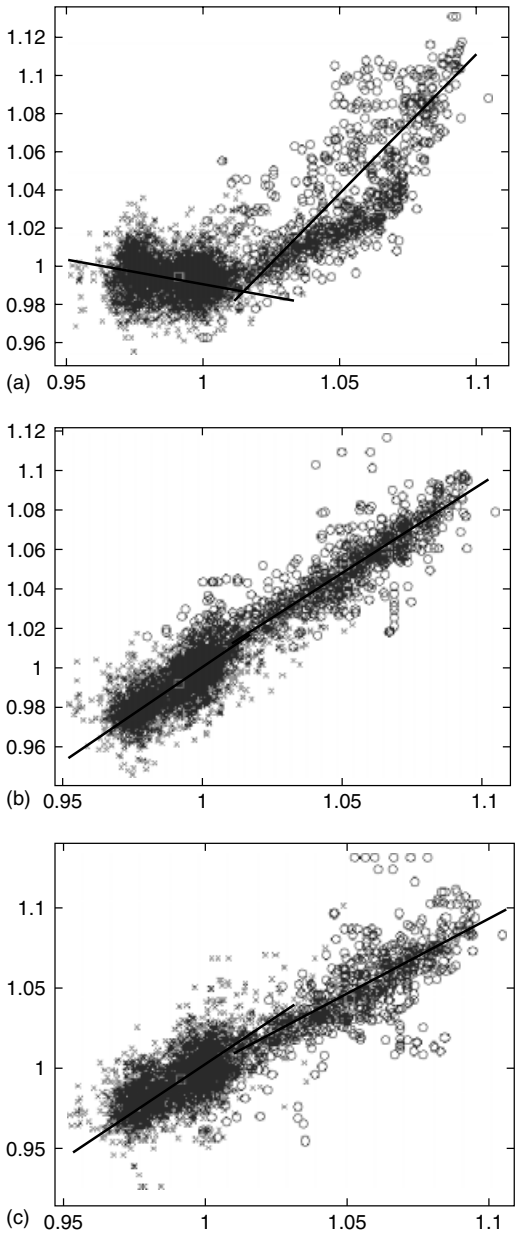


Figure 6. Two-dimensional projections of the space spanned by the first four normalized natural frequencies. (a–c) First frequency vs. second frequency; first frequency vs. third frequency and first frequency vs. fourth frequency.

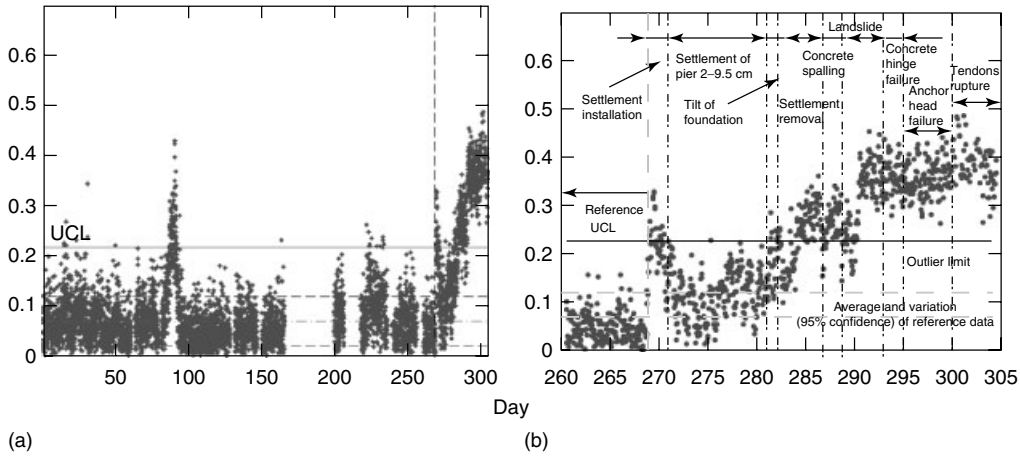


Figure 7. Damage detection results under varying environmental conditions. (a) POD; (b) close-up.

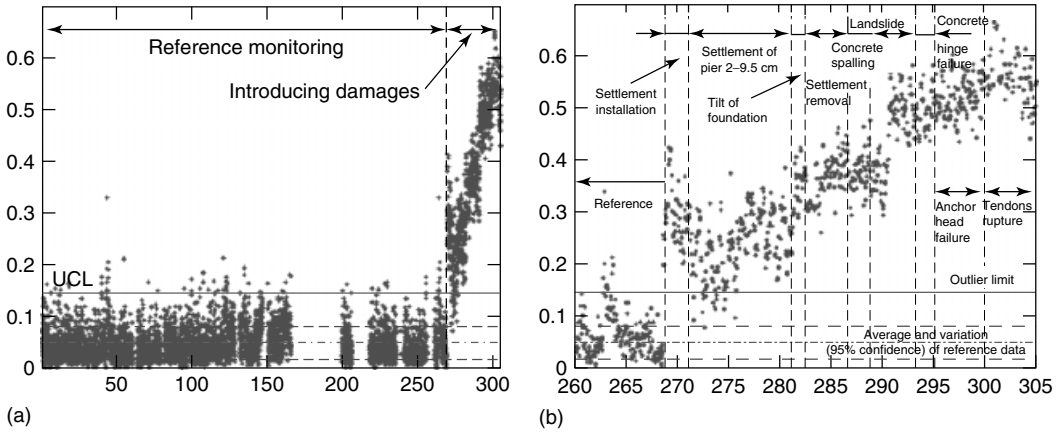


Figure 8. Damage detection results under varying environmental conditions. (a) VQPCA; (b) close-up.

assumed healthy; after the 268th day, the bridge was subjected to a series of realistic damage cases including pier settlement, concrete spalling, and tendons rupture.

One particularly challenging aspect of this application example is that the monitoring was performed under varying environmental conditions. A close look at the identified vibration features reveals that the natural frequencies of the first modes vary significantly with temperature (Figure 5). It was found that the asphalt layer on the bridge surface played a different role during warm and cold periods, which, in turn, caused the observed behavior. More precisely, the first four natural frequencies were shown to have a bilinear relation in function of the temperature, with a turning point at $0\,^{\circ}$C (Figure 6 for two-dimensional projections of this four-dimensional space).

The POD was first applied to the Z24 bridge data. The control chart together with a close-up is shown in Figure 7. A small incursion above the *UCL* is observed around the 270th day. The NI crosses the *UCL* for good after the 280th day, and the presence of a damage is detected. The results are therefore fairly satisfactory. However, outliers also appear between the 80th and 90th day, during which very cold weather caused freezing of the asphalt layer on the bridge surface.

To improve the results, VQPCA was then applied to the same data. Because of the bilinear relation between the first four natural frequencies, the space

was partitioned into two subregions, as indicated by the black solid lines in Figure 6. The comparison of the results obtained using VQPCA (Figure 8) and the POD (Figure 7) reveals two significant improvements:

- The abnormality between the 80th and 90th day disappears, because the nonlinear relation between the natural frequencies is effectively taken into account by the VQPCA method.
- The NI crosses the UCL after the 268th day, i.e., directly after the application of the damage and (aside from a few points) remains above this limit afterward. The VQPCA-based method is therefore more sensitive to damage.

5 CONCLUSION

With a clear trend toward very large data sets to analyze, dimensionality reduction has become an important step of structural health monitoring. In this context, the POD method is simple, fast, and efficient, which makes it a meaningful addition to the structural dynamicist's toolbox. This article has shown that when the linear nature of the POD is an important restriction, the VQPCA and NLPCA algorithms represent an interesting alternative to the POD for reduced-order modeling and feature extraction.

REFERENCES

[1] Lei Y, Kiremidjian AS, Nair KK, Lynch JP, Law KH. Algorithms for time synchronization of wireless structural monitoring sensors. *Earthquake Engineering and Structural Dynamics* 2005 **34**:555–573.

[2] Gao Y, Spencer BF, Ruiz-Sandoval M. Distributed computing strategy for structural health monitoring. *Structural Control and Health Monitoring* 2006 **13**:488–507.

[3] Lynch JP. An overview of wireless structural health monitoring for civil structures. *Philosophical Transactions of the Royal Society A: Mathematical Physical And Engineering Sciences* 2007 **365**:345–372.

[4] Staszewski WJ, Boller C, Tomlinson GR. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons: Chichester, 2004.

[5] Berkooz G, Holmes P, Lumley JL. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics* 1993 **25**:539–575.

[6] Jolliffe IT. *Principal Component Analysis*. Springer-Verlag: New York, 1986.

[7] Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 1901 **2**:559–572.

[8] Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 1933 **24**:417–441, 498–520.

[9] Mees AI, Rapp PE, Jennings LS. Singular value decomposition and embedding dimension. *Physical Review A* 1987 **36**(1):340–346.

[10] Wu CG, Liang YC, Lin WZ, Lee HP, Lim SP. A note on equivalence of proper orthogonal decomposition methods. *Journal of Sound and Vibration* 2003 **265**:1103–1110.

[11] Fitzsimons PM, Rui C. Determining low dimensional models of distributed systems. *Advances in Robust and Nonlinear Control Systems*. ASME DSC, 1993, Vol. 53.

[12] Cusumano JP, Bai BY. Period-infinity periodic motions, chaos and spatial coherence in a 10-degree-of-freedom impact oscillator. *Chaos, Solitons and Fractals* 1993 **3**:515–535.

[13] Al-Dmour AS, Mohammad KS. Active control of flexible structures using principal component analysis in the time domain. *Journal of Sound and Vibration* 2002 **253**:545–569.

[14] Azeez MFA, Vakakis AF. Proper orthogonal decomposition of a class of vibroimpact oscillations. *Journal of Sound and Vibration* 2001 **240**:859–889.

[15] Georgiou IT, Schwartz IB. Dynamics of large scale coupled structural-mechanical systems: a singular perturbation proper orthogonal decomposition approach. *SIAM Journal of Applied Mathematics* 1999 **59**:1178–1207.

[16] Kappagantu R, Feeny BF. Part 1: dynamical characterization of a frictionally excited beam. *Nonlinear Dynamics* 2000 **22**:317–333.

[17] Hemez FM, Doebling SW. Review and assessment of model updating for non-linear, transient dynamics. *Mechanical Systems and Signal Processing* 2001 **15**:45–73.

[18] Lenaerts V, Kerschen G, Golinval JC. Identification of a continuous structure with a geometrical nonlinearity, part II: proper orthogonal decomposition. *Journal of Sound and Vibration* 2003 **262**:907–919.

[19] Feeny BF. On proper orthogonal co-ordinates as indicators of modal activity. *Journal of Sound and Vibration* 2002 **255**:805–817.

[20] Iemma U. Digital holography and Karhunen-Loève decomposition for the modal analysis of two-dimensional vibrating structures. *Journal of Sound and Vibration* 2006 **291**:107–131.

[21] Kerschen G, Golinval JC, Vakakis AF, Bergman LA. The method of POD for dynamical characterization and order reduction of mechanical systems: an overview. *Nonlinear Dynamics* 2005 **41**:147–170.

[22] Azeez MFA, Vakakis AF. Numerical and experimental analysis of a continuous overhang rotor undergoing vibro-impacts. *International Journal of Non-linear Mechanics* 1999 **34**:415–435.

[23] Kerschen G, Feeny BF, Golinval JC. On the exploitation of chaos to build reduced-order models. *Computer Methods in Applied Mechanics and Engineering* 2003 **192**:1785–1795.

[24] Liang YC, Lin WZ, Lee HP, Lim SP, Lee KH, Sun H. Proper Orthogonal decomposition and its applications, part II: model reduction for MEMS dynamical analysis. *Journal of Sound and Vibration* 2002 **256**:515–532.

[25] Ghanem R, Spanos P. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag: Heildelberg, 1991.

[26] Feldmann U, Kreuzer E, Pinto F. Dynamic diagnosis of railway tracks by means of the Karhunen-Loève transformation. *Nonlinear Dynamics* 2000 **22**:183–193.

[27] Tumer IY, Wood KL, Busch-Vishniac IJ. Monitoring of signals from manufacturing processes using K-L transform. *Mechanical Systems and Signal Processing* 2000 **14**:1011–1026.

[28] Joyner ML, Banks HT, Wincheski B, Winfree WP. Reduced order computational methodology for damage detection in structures. *Proceedings SPIE*. Newport Beach, Vol. 3994, 2000; 10–17.

[29] Manson G. Identifying damage sensitive, environment insensitive features for damage detection. *Proceedings of the 3rd International Conference on Identification in Engineering Systems*, Swansea, 2002.

[30] De Boe P, Golinval JC. Principal component analysis of a piezo-sensor array for damage localization. *Structural Health Monitoring* 2003 **2**:137–152.

[31] Yan AM, Kerschen G, De Boe P, Golinval JC. Structural damage diagnosis under changing environmental conditions, part I: a linear analysis. *Mechanical Systems and Signal Processing* 2005 **19**:847–864.

[32] Mustapha F, Manson G, Pierce SG, Worden K. Structural health monitoring of an annular component using a statistical approach. *Strain* 2005 **41**:117–127.

[33] Lanata F, Del Grosso A. Damage detection and localization for continuous static monitoring of structures using a proper orthogonal decomposition of signals. *Smart Materials and Structures* 2006 **15**:1811–1829.

[34] Galvanetto U, Violaris G. Numerical investigation of a new damage detection method based on proper orthogonal decomposition. *Mechanical Systems and Signal Processing* 2007 **21**:1346–1361.

[35] Mujica LE, Vehi J, Ruiz M, Verleysen M, Staszewski W, Worden K. Multivariate statistics process control for dimensionality reduction in structural assessment. *Mechanical Systems and Signal Processing* 2008 **22**(1):155–171.

[36] Kerschen G, De Boe P, Golinval JC, Worden K. Sensor validation using principal component analysis. *Smart Materials and Structures* 2005 **14**:36–42.

[37] Holmes P, Lumley JL, Berkooz G. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press: New York, 1996.

[38] Feeny BF, Kappagantu R. On the physical interpretation of proper orthogonal modes in vibrations. *Journal of Sound and Vibration* 1998 **211**:607–616.

[39] Kerschen G, Golinval JC. Physical interpretation of the proper orthogonal modes using the singular value decomposition. *Journal of Sound and Vibration* 2002 **249**:849–865.

[40] Sirovich L. Turbulence and the dynamics of coherent structures, part I: coherent structures. *Quarterly of Applied Mathematics* 1987 **45**:561–571.

[41] Oja E. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* 1982 **15**:267–273.

[42] Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press: Oxford, 1995.

[43] Baldi P, Hornik K. Neural networks and principal component analysis: learning from examples without local minima. *Neural Networks* 1989 **2**:53–58.

[44] Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal* 1991 **37**:233–243.

[45] Kambhatla N. *Local Models and Gaussian Mixture Models for Statistical Data Processing*, Ph.D. Thesis.

Oregon Graduate Institute of Science & Technology, 1996.

[46] Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems* 1989 **2**:303–314.

[47] Sohn H, Worden K, Farrar CR. Statistical damage classification under changing environmental and operational conditions. *Journal of Intelligent Material Systems and Structures* 2002 **13**:561–574.

[48] Kerschen G, Golinval JC. A model updating strategy of non-linear vibrating structures. *International Journal for Numerical Methods in Engineering* 2004 **60**:2147–2164.

[49] Kerschen G, Golinval JC. Auto-associative neural networks, Part I: Replacement of missing sensor values. *Proceedings of the 8th International Conference on Recent Advances in Structural Dynamics*. Southampton, 2003.

[50] Kerschen G, Yan AM, Golinval JC. Distortion function and clustering for local linear models. *Journal of Sound and Vibration* 2005 **280**:443–448.

[51] Kerschen G, Golinval JC. Non-linear generalisation of principal component analysis: from a global to a local approach. *Journal of Sound and Vibration* 2002 **254**:867–876.

[52] Yan AM, Kerschen G, De Boe P, Golinval JC. Structural damage diagnosis under changing environmental conditions, part II: local PCA. *Mechanical Systems and Signal Processing* 2005 **19**:865–880.

[53] Worden K, Manson G, Fieller NRJ. Damage detection using outlier analysis. *Journal of Sound and Vibration* 2000 **229**:647–667.

Chapter 29
Higher Order Statistical Signal Processing

Paul White
Institute of Sound and Vibration Research, University of Southampton, Southampton, UK

1 INTRODUCTION

Spectral analysis is one of the most powerful and widely used tools in engineering (*see also* **Statistical Time Series Methods for SHM**). Thus, it is not surprising that it is a key tool in structural health monitoring (SHM). For example, single-channel spectral analysis of vibration data can allow one to measure modal frequencies and estimate damping parameters under the assumption that the excitation is broadband. Monitoring such response data over a period of time can allow one to identify changes within a structure. Identification of such changes does not usually lead to an effective SHM methodology, since it can be difficult to ensure that changes observed are the consequence of damage and not the result of benign processes that do not affect a structure's integrity.

This article considers the extension of classical spectral analysis into the so-called higher orders. Such methods are referred to as HOS, which is an acronym taken to interchangeably mean either *higher order statistics* or *higher order spectra*. Similar to spectral analysis, HOS is concerned with the analysis of stationary random processes. Extensions to both spectral analysis (*see also* **Time–frequency Analysis**; **Wavelet Analysis**) and HOS [1] for nonstationary signals exist, but are not considered here.

Classical spectral (and correlation) analysis is based on an analysis of a signal's power; the power is equivalent to the signal's variance. For a zero mean signal,[a] the variance is the average of the square of the signal; see equation (1). It is because of this squaring operation that we refer to classical spectral analysis as being based on the signal's second-order statistics. HOS extends the principles of spectral and correlation analyses to orders higher than 2.

In dealing with second-order statistics, spectral analysis is intimately linked to stationary Gaussian random processes. A Gaussian random process is a signal whose amplitude is distributed according to a Gaussian, or normal, distribution. A jointly Gaussian process is a signal in which the joint probability density function for pairs of the signal points separated by a fixed interval is also Gaussian. One can construct examples of stationary processes

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

that are Gaussian, but not jointly Gaussian; however, such examples are pathological. We shall assume that Gaussianity implies joint Gaussianity.

Second-order statistics are important to stationary, zero mean, Gaussian random processes since such processes are completely described through knowledge of their second-order statistics. The implication of this is that if a process is Gaussian then the spectrum, evaluated with sufficient resolution, completely characterizes that signal. Hence, considering statistics of order greater than 2 is only valuable when applied to non-Gaussian processes. As an illustration of this: it is always possible to construct a Gaussian process with the same spectrum as a given non-Gaussian process. Such processes would be indistinguishable from each other using second-order statistics, but, in general, their HOS will be different.

Gaussian signals are examples of stable distributions [2], meaning that the output of any linear system excited by a Gaussian input is itself Gaussian. Thus, when analyzing response signals that are Gaussian, it is common to assume that the input was Gaussian and that the system is linear. If the response signal is non-Gaussian, it is common to model the signal in one of two ways. Firstly, the signal can be modeled as the output of a linear system whose input is non-Gaussian. Secondly, one might assume that the input is Gaussian and the system is nonlinear. Evidently, there is no physical reason why the input cannot be non-Gaussian and the system nonlinear, but, in practice, this combination leads to system problems that are ill-posed, which can only be solved if additional information is available.

One use of HOS in SHM is based on determining whether or not a signal is Gaussian. The distinction between Gaussian and non-Gaussian processes can become important in SHM because in many cases the presence of a fault within a structure may result in the response of the structure being nonlinear. Such nonlinearities produce responses that are, in almost all cases, non-Gaussian. By detecting and characterizing the non-Gaussianity, it may be possible to provide information regarding the presence/absence of a fault and some information regarding its nature. It may be possible to exploit such an approach when only naturally occurring excitations are available.

In addition, HOS also offer opportunities to solve signal-processing problems that are intractable using only second-order statistics. This advantage arises because phase information is discarded when using second-order statistics, which is retained in the HOS. To succeed, such methods require that the underlying signals be non-Gaussian, so that one needs to carefully consider whether such an assumption is reasonable for the particular application of interest.

2 SECOND-ORDER STATISTICS

We begin by recapping the important elements of second-order spectral analysis. The power, P, of a signal is defined as

$$P = E[x(t)^2] \tag{1}$$

where $x(t)$ is the signal and $E[]$ is the expectation, or averaging, operator. Accordingly, the correlation function, r, is defined as

$$\begin{aligned} r(\tau) &= E[x(t)x(t-\tau)] \\ &= E[x(t)x(t+\tau)] = r(-\tau) \end{aligned} \tag{2}$$

The variable τ is referred to as the *lag variable* and clearly $r(0) = P$. Alternatively, one definition of the power spectrum is

$$\begin{aligned} S(f) &= E[|X(f)|^2] = E[X(f)X(f)^*] \\ &= E[X(f)X(-f)] \end{aligned} \tag{3}$$

where * denotes complex conjugation and $X(f)$ denotes the Fourier transform of $x(t)$. This definition overlooks some theoretical issues associated with defining the Fourier transform of stationary random processes, which requires one to formally modify equation (3), but without changing its essence. The form in equation (3) conveys all of the salient aspects, so we choose to sacrifice rigor to enhance clarity; a fully rigorous treatment can be found in many standard texts including [3, 4].

The spectrum and the correlation function are related via the Wiener–Khintchine theorem

$$\begin{aligned} S(f) &= \int_{-\infty}^{\infty} r(\tau)\mathrm{e}^{-2\pi i f\tau}\,\mathrm{d}\tau \\ \Rightarrow r(\tau) &= \int_{-\infty}^{\infty} S(f)\mathrm{e}^{2\pi i f\tau}\,\mathrm{d}f \end{aligned} \tag{4}$$

so that the correlation function and the power spectrum form a Fourier transform pair. Further, the spectrum is a decomposition of the signal's power as a function of frequency, in the sense that the area under $S(f)$ is equal to the power of the signal P:

$$P = \int_{-\infty}^{\infty} S(f)\, \mathrm{d}f \tag{5}$$

In the following sections, we discuss how the above-mentioned familiar principles can be extended to higher orders.

3 HIGHER ORDER STATISTICS

3.1 Higher order moments and cumulants

According to equation (1), the power, P, is defined as the variance of the signal, which is the second-order moment[b] of the signal. The definition of the general nth order moment, M_n, is

$$M_n = E[x(t)^n] \tag{6}$$

Evidently, $P = M_2$ and the case $n = 1$ corresponds to the mean, which we assume to be zero.

If the probability density function for $x(t)$ is symmetric, then $M_n = 0$ for all odd values of n. A particular example of this is the Gaussian distribution. It is key to what follows to realize that, for a Gaussian process, these higher order moments can be expressed in terms of the second-order moments, reflecting the fact that for a Gaussian process all information is contained in the second-order statistics, for example, for a Gaussian process $M_4 = 3M_2^2 = 3P^2$ and $M_6 = 15M_2^3 = 15P^3$.

It is rather unsatisfying that the even numbered higher order moments of a Gaussian process are nonzero, although they contain no information other than that available at second order. To avoid this, one can work with a modified form of the moments, which enjoy further useful properties. These quantities are referred to as the *cumulants*. They are defined in a manner that ensures that for a Gaussian process the cumulants are zero for all $n > 2$, i.e., all higher orders. The cumulants C_k of a particular order are defined in terms of the moments of that and lower orders. The first four cumulants (for four is the highest order commonly considered in HOS analysis) are

$$\begin{aligned} C_k &= \mathrm{cum}\{x(t)^k\} = E[x(t)^k] = M_k \quad k = 1, 2, 3 \\ C_4 &= \mathrm{cum}\{x(t)^4\} = M_k - 3{M_2}^2 \end{aligned} \tag{7}$$

Note that, for the first three orders, the moments and cumulants are identical; it is not until the fourth order that one needs to distinguish between moments and cumulants.

The definition of a cumulant used in equation (7) is an example of a more general form. Specifically in the above definition, the fourth-order cumulant considers a product of $x(t)$ with itself four times. This generalizes to the case where we consider the so-called cross cumulants between four variables, say, y_k, $k = 1, \ldots, 4$, in which case

$$\begin{aligned} C_4 &= \mathrm{cum}\{y_1, y_2, y_3, y_4\} \\ &= E[y_1y_2y_3y_4] - E[y_1y_2]E[y_3y_4] \\ &\quad - E[y_1y_3]E[y_2y_4] - E[y_1y_4]E[y_2y_3] \end{aligned} \tag{8}$$

The above is equivalent to (7) when $y_k = x(t)$.

Higher order moments have been used for condition monitoring purposes for a significant period of time. In particular, the kurtosis, defined as a normalized fourth-order moment, is a well-known measure of the impulsiveness of a signal and can be used to indicate various forms of fault [5].

3.2 Higher order moments and cumulant functions

The concepts in the preceding section can be extended to generate higher order extensions of the correlation function, $r(\tau)$. The correlation function is simply the cross moment/cumulant between $x(t)$ and $x(t - \tau)$; see equation (2). Consequently, the nth order moment function can be defined as

$$\begin{aligned} M_n(\tau_1, \ldots, \tau_{n-1}) = E[&x(t)x(t - \tau_1) \\ &\times x(t - \tau_2) \ldots x(t - \tau_{n-1})] \end{aligned} \tag{9}$$

This is a function of the $n - 1$ lag variables, τ_k. For a Gaussian process, the moment function is only nonzero for even n. For even values of n, the higher

order moment function can be expressed in terms of the correlation function $r(\tau)$, for example, at fourth order

$$M_4(\tau_1, \tau_2, \tau_3) = r(\tau_1)r(\tau_2 - \tau_3) + r(\tau_2)r(\tau_1 - \tau_3) + r(\tau_3)r(\tau_1 - \tau_2) \quad (10)$$

If the lags are set to zero, $\tau_k = 0$, then equation (10) simplifies to the earlier result $M_4 = 3M_2^2$.

Cumulant functions of a given order can also be defined. As before, in the cases where $n < 4$, the cumulant and moment functions are identical. For $n = 4$, the definition given in equation (8) is used, leading to the fourth-order cumulant function being defined as

$$\begin{aligned} C_4(\tau_1, \tau_2, \tau_3) &= \text{cum}\{x(t), x(t - \tau_1), \\ &\quad x(t - \tau_2), x(t - \tau_3)\} \\ &= M_4(\tau_1, \tau_2, \tau_3) - r(\tau_1)r(\tau_2 - \tau_3) \\ &\quad - r(\tau_2)r(\tau_1 - \tau_3) - r(\tau_3)r(\tau_1 - \tau_2) \end{aligned} \quad (11)$$

3.3 Higher order moments and cumulant functions

Spectra can be defined on the basis of the moment and cumulant functions discussed in the previous section. This is realized by taking the multidimensional Fourier transforms of the functions $M_k()$ or $C_k()$ to yield the higher order moment or cumulant spectra respectively. Specifically, the higher order moment spectrum, $\overline{S}()$ is defined as

$$\overline{S}_n(f_1, f_2, \ldots) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} M_n(\tau_1, \tau_2, \ldots, \tau_{n-1}) \times \mathrm{e}^{-2\pi i(f_1\tau_1 + f_2\tau_2 + \cdots + f_{n-1}\tau_{n-1})} \mathrm{d}\tau_1 \mathrm{d}\tau_2 \ldots \mathrm{d}\tau_{n-1} \quad (12)$$

and, for the higher order cumulant spectrum, $S()$, one has[c]

$$S_n(f_1, f_2, \ldots) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} C_n(\tau_1, \tau_2, \ldots, \tau_{n-1}) \times \mathrm{e}^{-2\pi i(f_1\tau_1 + f_2\tau_2 + \cdots + f_{n-1}\tau_{n-1})} \mathrm{d}\tau_1 \mathrm{d}\tau_2 \ldots \mathrm{d}\tau_{n-1} \quad (13)$$

By extension of the principles discussed earlier, the higher order moment spectrum is zero for Gaussian processes only for odd n, whilst the cumulant HOS for Gaussian process is zero for all $n > 2$, for Gaussian signals. The moment spectra can also be expressed in a form equivalent to equation (3) as follows:

$$\overline{S}(f_1, f_2, \ldots, f_{n-1}) = E[X(f_1)X(f_2)\ldots X(f_{n-1}) \times X(f_1 + f_2 + \cdots + f_{n-1})^*] \quad (14)$$

The cumulant spectrum of third order (and all odd orders) is identical to the moment spectrum, whereas the fourth-order cumulant spectrum is as follows:

$$\begin{aligned} S(f_1, f_2, f_3) = \overline{S}(f_1, f_2, f_3) &- S(f_1)S(f_2)\delta(f_2 + f_3) \\ &- S(f_2)S(f_3)\delta(f_1 + f_3) \\ &- S(f_3)S(f_1)\delta(f_1 + f_2) \end{aligned} \quad (15)$$

where $\delta(f)$ is a Kronecker delta function, wherein $\delta(0) = 1$ and $\delta(f) = 0$ for all $f \neq 0$[d].

The higher order spectra of orders 3 and 4 are referred to using their own terminology. Specifically, the third-order spectrum is referred to as the *bispectrum*[e] [6, 7] and the fourth-order spectrum is the *trispectrum* [8, 9]. Note that there are cumulant and moment forms for the trispectrum, whereas the definition of the bispectrum remains unaffected whether one chooses to consider it as a moment or a cumulant spectrum.

Cumulant spectra or cumulant functions have two distinct advantages over their moment counterparts. The first advantage, which we have used to motivate the introduction of the cumulants, is the fact that Gaussian processes have cumulant spectra that are identically zero for all orders greater than 2 (all higher orders). The second is that cumulants are additive, in the sense that if two processes are statistically independent then the cumulant of a process formed when these two processes are added is the sum of the constituents' cumulants. These two facts combine to provide one of the key motivations for employing HOS, specifically, the observation that adding Gaussian noise to a signal does not alter its cumulant spectra or cumulant function. Since in many applications noise is regarded as obeying a Gaussian

distribution, then the use of HOS offers the potential to conduct processing, which is unaffected by the noise. This theoretical advantage can be undermined by some of the more practical issues associated with HOS analyses. Probably the most troublesome aspect comes from the fact that to construct robust estimates of HOS quantities one needs to use long data lengths. This requirement becomes more stringent as the order of the HOS increases or if one requires high resolution in the spectra. Such problems are exacerbated by the requisite increase in computer resources imposed by the use of HOS of higher order and/or high resolution.

3.4 The bispectrum

The bispectrum is a measure of the third-order statistics of a signal and as such is only nonzero if the signal has a skewed distribution. Being the first of the higher order spectra, it is easiest to deal with, requiring least memory and imposing the lowest computational burden. The bispectrum can be expressed in the particular form of equation (14) as follows:

$$\begin{aligned} S(f_1, f_2) &= E[X(f_1)X(f_2)X(f_1 + f_2)^*] \\ &= E[X(f_1)X(f_2)X(-f_1 - f_2)] \quad (16) \end{aligned}$$

Consider a low-pass signal, so that $X(f) = 0$ for $f > f_{\max}$. Then evidently $S(f_1, f_2) = 0$ if $f_1 + f_2 > f_{\max}$, so that the bispectrum of a low-pass signal can only be nonzero in a triangular region, as shown in Figure 1.

There are symmetries inherent within the bispectrum. Specifically, it is apparent from the definition in equation (16) that

$$\begin{aligned} S(f_2, f_1) &= S(f_1, f_2) \quad \text{and} \\ S(-f_1, -f_2) &= S(f_1, f_2)^* \end{aligned} \quad (17)$$

The full set of symmetries is depicted in Figure 2 for a low-pass signal. The bispectrum is completely specified only through its values in the triangular region, $f_1 \geq 0$, $f_1 \leq f_2$, called the *principal domain* and is shown as a chequered region in Figure 2.

From its definition in equation (16), it is apparent that the bispectrum will, in general, be complex valued. In many applications, it is adequate to only consider its magnitude.

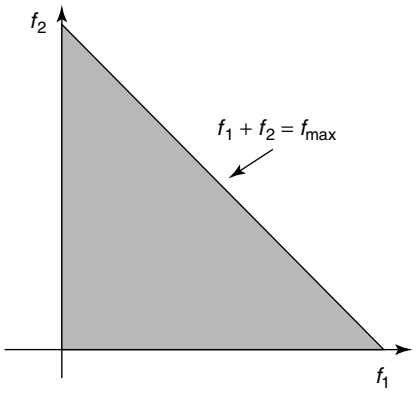


Figure 1. Possible nonzero values of the bispectrum of a low-pass signal.

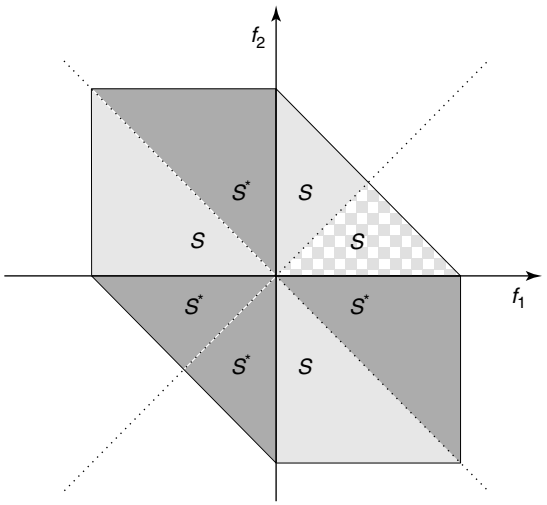


Figure 2. Symmetries in the bispectrum.

3.5 Computation of the bispectrum

The implementation of the bispectrum on a digital computer requires a formulation in terms of a digital signal. On the basis of equation (16), and assuming ergodicity, one method estimating the bispectrum is based on dividing the data into overlapping segments and applying a discrete Fourier transform (DFT) to

each segment [10]; specifically, one computes

$$\hat{S}(k_1, k_2) = \frac{1}{M} \sum_{m=1}^{M} X_m(k_1) X_m(k_2) X_m(k_1 + k_2)^* \tag{18}$$

where $X_m(k)$ is the DFT of the data in the mth segment. Each segment of data contains N consecutive samples; the blocks can be overlapped and maybe windowed. Such an estimation scheme is a natural extension of the segment averaging approach extensively employed in classical spectral estimation. The estimate in equation (18) can be scaled in various manners, as is the case with classical spectral estimation, but for simplicity we consider the unscaled form in equation (18).

The general form of bispectrum of a digital signal has additional properties that are not present in the analog formulation (16). Consider the principal domain as defined in the previous section, i.e., $k_1 \geq 0$, $k_1 \leq k_2$. Then this domain can itself be divided into two regions: the inner triangle (IT) and the outer triangle (OT), as shown in Figure 3 [11].

Inside IT, the arguments of the digital bispectrum k_1 and k_2 are such that $k_1 + k_2 < N/2$ and so all the terms in equation (18) are evaluated for frequencies between 0 and $f_s/2$, where f_s is the sampling frequency. Within the IT, for a properly sampled signal, the digital bispectrum can provide an unbiased estimate of bispectrum of the underlying analog process. In the OT, the digital bispectrum is zero assuming (i) that the signal is not aliased and (ii) that the signal is stationary [11, 12]. It has been proposed that examining the value of the bispectrum within this region provides a test for aliasing and/or stationarity [11, 13], whereas tests based in the IT have been used to test whether a signal is Gaussian or not [7, 14].

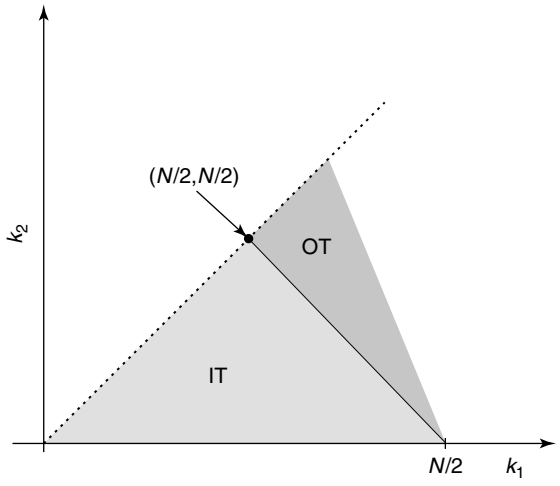


Figure 3. Regions in the digital bispectrum: inner triangle (IT) and outer triangle (OT).

Estimates of the bispectrum based on equation (18) are unbiased as long as an adequate block size, N, is used. The variance of this estimator has the following form [7, 15]:

$$\mathrm{var}\{\hat{S}(k_1, k_2)\} \propto S(k_1) S(k_2) S(k_1 + k_2) \tag{19}$$

Thus, the variance of the bispectrum estimate is dependent upon the spectrum of the signal. This can produce misleading results. Regions in the bispectrum can appear large simply because the variance of the estimate is large in that region, only as a consequence of the triple product of the spectrum, equation (19), being large. This effect can be mitigated by normalizing the bispectrum, $\tilde{S}(k_1, k_2)$, using

$$\tilde{S}(k_1, k_2) = \frac{\frac{1}{M} \sum_{m=1}^{M} X_m(k_1) X_m(k_2) X_m(k_1 + k_2)^*}{\sqrt{\hat{S}(k_1) \hat{S}(k_2) \hat{S}(k_1 + k_2)}},$$

$$\hat{S}(k) = \frac{1}{M} \sum_{m=1}^{M} |X_m(k)|^2 \tag{20}$$

where $\hat{S}(k)$ is an estimate of the power spectrum based on segment averaging. The quantity $\tilde{S}(k_1, k_2)$ is referred to as the *bicoherence* [6] and is a normalized form of the bispectrum. Other normalizations have been proposed for the bispectrum [10].

3.5.1 Example of the bispectrum for a simulated cracked beam

By way of example, consider data generated by randomly exciting a piecewise linear model of a cracked beam. This model is defined as [16, 17]

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + c \frac{\mathrm{d}y}{\mathrm{d}t} + \kappa(y) y = x(t) \tag{21}$$

where c is the damping coefficient and $\kappa(y)$ is the displacement-dependent stiffness; note that equation (21) assumes a unit mass. This stiffness is defined such that $\kappa(y) = \kappa_1$ for $y < 0$ and $\kappa(y) = \kappa_2$ for $y > 0$, where κ_1 and κ_2 are constants. We assume that $\kappa_1 < \kappa_2$, i.e., the beam is stiffer when the crack is compressed (closed), and for $y < 0$, the crack opens and the beam is less stiff. The proportionate depth of the crack is approximately equal to the ratio of the stiffnesses κ_1/κ_2 for small crack depths [18, 19]. In this simulation, we assume a stiffness ratio of 30% and a damping coefficient of 0.1. The excitation is band-limited Gaussian noise. White Gaussian measurement noise is also added to the response data. For comparison purposes, we generate a surrogate signal based on this cracked beam response data. The surrogate signal consists of Gaussian noise, which is filtered so that its spectrum matches that of the response generated using equation (21). Both signals are constructed so that they are 100 000 samples in duration. This surrogate data, being Gaussian, has a bispectrum, which is theoretically zero for all frequencies.

Figure 4 shows the power spectra of the response and the surrogate data. The data consists of a resonance at a normalized frequency of 0.19. A harmonic component at double the resonance frequency can be seen in the noise-free data, shown as a dotted line in Figure 4(b), and this harmonic is a consequence of the nonlinear nature of the dynamics described in equation (21). The addition of measurement noise almost completely obscures the presence of this second harmonic in the simulated measured data set, as shown by the solid line in Figure 4(b). A key difference between the response and surrogate data is that, for the response data, the phase close to the harmonic frequency is related to the phase near the resonance [6], whereas in the surrogate data no such correlations exist. These interrelationships between different frequency bands are not evident in the power spectrum because second-order spectra fail to represent phase information.

Figure 5 depicts normalized and unnormalized bispectra computed for the surrogate and response data. These bispectra are computed using an fast Fourier transform (FFT) size (N) of 256 samples. For both the unnormalized bispectral estimates, shown in Figure 5(a) and (b), there are horizontal and vertical lines at a normalized frequency of 0.19. These features are present in the bispectra as consequence of the increase in the power spectrum at the resonance. It is important to recognize that these peaks are presented in both the surrogate data and the response data. Such structure can lead to one drawing misleading conclusions from the bispectrum. By normalizing the bispectrum, the spurious structure is eliminated. The normalized bispectrum of the surrogate signal is close to zero everywhere, whereas for the response data there is a strong peak, at the bifrequency (0.19, 0.19), indicating that, at frequency 0.38, there is a significant contribution from the normalized frequency 0.19 interacting with itself.

3.6 The trispectrum

The fact is that the bispectrum analyses of a signal in terms of its skewness limits its applicability. In most of the applications, the nonlinearities encountered are symmetric and give rise to response data, that is, they have a symmetric distribution. In such cases, the bispectrum is zero. In general, the trispectrum,

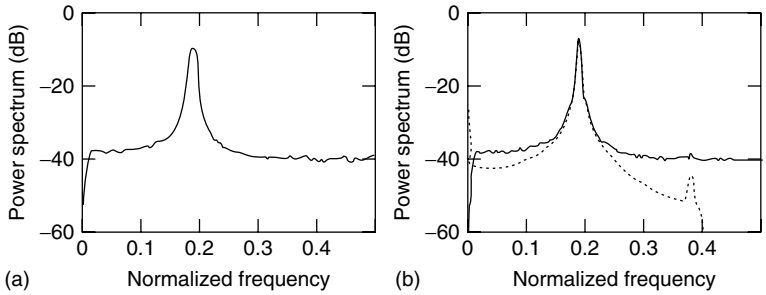


Figure 4. Spectral estimates: (a) spectrum of the surrogate Gaussian signal; (b) solid line: spectrum of the response signal with added noise and dotted line: spectrum of response signal in the absence of noise.

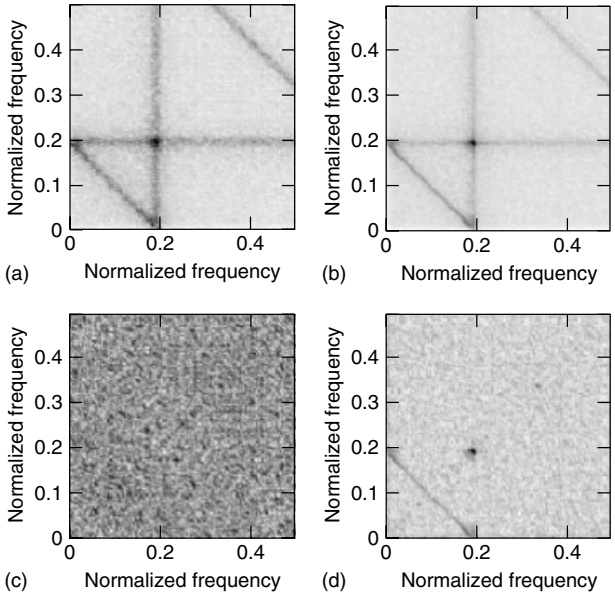


Figure 5. Bispectrum estimates, displayed on a logarithmic color axis with dark shades representing large values (each plot is individually scaled): (a) unnormalized bispectrum of the surrogate Gaussian signal, (b) unnormalized bispectrum of the response signal, (c) normalized bispectrum of the surrogate Gaussian signal, and (d) normalized bispectrum of the response signal.

the fourth-order HOS, is nonzero. It should be noted that the estimation of the trispectrum requires more data and imposes a greater computational burden than estimation of the bispectrum. As already discussed, there are definite theoretical advantages associated with the use of the cumulant, as opposed to the moment, form of the trispectrum and so it is only the cumulant form of the trispectrum that we consider in the following.

In general, the trispectrum exhibits a set of symmetries that parallel those of the bispectrum. The details of these symmetries can be found elsewhere [9]. The computation of the trispectrum can be performed in a manner that is a natural extension of the segmenting averaging approach described by equation (18).

The trispectrum is a function of three frequency variables, which makes displaying it problematic. There are several approaches that can be taken for representing the trispectrum; here we shall take the option of considering a specific subset of values, sometimes referred to as *a slice of the trispectrum*. Specifically, we consider the specific case of equation (15), where $f_1 = -f_2$, so that the cumulant trispectral slice, $T(f_1, f_2)^{\mathrm{f}}$, discussed here can be written as

$$T(f_1, f_2) = E[|X(f_1)|^2|X(f_2)|^2] - S(f_1)S(f_2) - S(f_1)^2\delta(f_1 - f_2) \tag{22}$$

This trispectral slice describes the correlations of the spectral amplitudes, $|X(f)|^2$, at the two frequencies f_1 and f_2. Since the slice is a subset of a cumulant HOS, by construction, it is zero for stationary Gaussian processes.

3.6.1 Example of the trispectrum for a simulated nonlinear spring

To illustrate the use of the trispectrum, we consider a simulated data set from a system with a nonlinear stiffness. This system can be regarded as a softening spring, i.e., a single degree of freedom system whose stiffness reduces as displacement away from equilibrium increases. The specific system modeled has the form:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + c\frac{\mathrm{d}y}{\mathrm{d}t} + k\sigma \tanh(y/\sigma) = x(t) \tag{23}$$

The hyperbolic tangent function ensures that for small displacements the spring stiffness is k. The scaling constant σ is included to control the linearity of the system; for large values of σ, the system behaves in a linear fashion, whilst for small values the system is highly nonlinear. The particular form of equation (23) ensures that the stiffness at small amplitudes remains unaffected by the value of σ.

The nonlinearity in the system means that, as the displacement is increased, the stiffness reduces. The consequence is that the response of this system has a larger fourth-order moment as the value of σ is reduced.

The excitation used for this simulation is band-limited Gaussian noise. The simulated response lasts for 100 000 samples. Results are shown for the system with the same stiffness parameter, $k = 160$, and damping, $c = 0.1$, but the value of nonlinearity parameter, σ, is varied. For large values of σ, the system has a resonance at a normalized frequency of 0.2. Four different values of σ are considered: 0.6, 0.3, 0.15, and 0.1. The largest value corresponds to a case where the system is very close to being linear, whereas for the smallest value there is a high degree of nonlinearity in the response. Figure 6 shows the spectra of the four response signals. Without explicit knowledge of the underlying system, these spectra provide no direct evidence of the nonlinearity. As the nonlinearity is increased, the peak at resonance broadens toward the lower frequencies, but this is not necessarily indicative of nonlinearity and could be explained by a change in the linear dynamics of the system. However, the slices of the normalized trispectrum, shown in Figure 7, betray the increasing levels of nonlinearity in the system. There is a peak in the slice of the trispectrum, indicating coupling between the spectral amplitudes at the different frequency pairs close to resonance.

4 OTHER USES OF HOS

In this discussion, we have considered the use of HOS for the limited problem of identifying the presence of a nonlinearity from response-only data. Such an application can be extremely useful if one only has output measurements under natural excitations. However, this is only one of the possible potential uses of HOS. There are several possible further uses of HOS. The most direct application is in situations where one has access to both input and response data. Then one can use the HOS spectra to construct estimates of the Wiener or Volterra kernels [20]

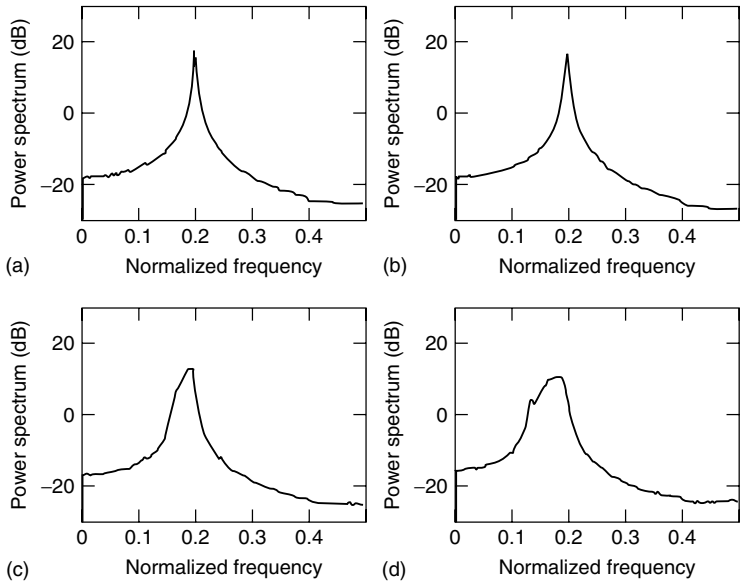


Figure 6. Power spectra of the simulated response form softening spring: (a) $\sigma = 0.6$, (b) $\sigma = 0.3$, (c) $\sigma = 0.15$, and (d) $\sigma = 0.1$.

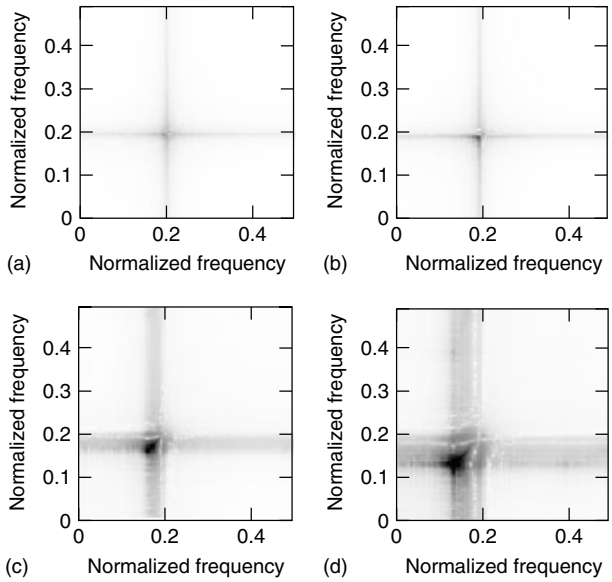


Figure 7. Estimates of the normalized slices of the trispectrum, displayed on a logarithmic gray scale with dark shades representing large values: (a) $\sigma = 0.6$, (b) $\sigma = 0.3$, (c) $\sigma = 0.15$, and (d) $\sigma = 0.1$.

that characterize the nonlinear system [21]. This is achieved using cross higher order spectra, in a manner analogous to the way that cross spectra are used to estimate transfer functions in linear systems.

Further to this, HOS can be used to solve at least two problems associated with linear systems excited by non-Gaussian processes. The first of these problems is the estimation of the parameters of a moving average (MA) system model (or by extension of the MA component of an autoregressive moving average (ARMA) system model) [22]. The second problem that HOS can be used to solve is blind source separation [23]. In both these problems, the fact is that using the spectrum to represent a signal discards phase information means that general solutions cannot be obtained, whereas the use of HOS, in which phase information is preserved, means that solutions can be formulated.

5 CONCLUSIONS

HOS can provide useful information that is not accessible through the spectrum (or correlation function). This information allows one a more complete description of non-Gaussian processes. These non-Gaussian processes are normally modeled as a result of non-Gaussian excitation of a linear system or Gaussian excitation of nonlinear system. In SHM, this can be of considerable benefit if the structure under investigation is such that nonlinearities are indicative of the presence of a fault.

END NOTES

[a.] Throughout we shall assume that signals of interest are zero mean.

[b.] Moments are commonly discussed in two forms: central and noncentral moments. The distinction being that the central moments are computed after the mean of the signal has been subtracted from the data. Since we assume the data to be zero mean, such a distinction is meaningless and we use the generic term "moment".

[c.] Note that the notation for spectrum and a higher order spectrum are both $S()$ and they are distinguished by the number of arguments in the function. This emphasizes the commonality between the concepts.

[d.] On a technical point: deriving (15) by Fourier transforming (11) yields a solution containing Dirac delta functions, rather than the Kronecker delta function appearing here. The reason for this difference is a consequence of relaxation of rigor in adopting (3), if

that rigor is applied throughout then no inconsistency occurs.

[e.] The prefix "bi" in the term bispectrum refers to the fact it is a function of two frequency variables and is not supposed to indicate that it is some measure of second-order properties.

[f.] It is necessary to adopt a notation for the trispectral slice, which is distinct from the bispectrum, so one cannot continue the convention of using S with different numbers of arguments.

REFERENCES

[1] Fonollosa JR, Nikias CL. Wigner higher-order moment spectra – definition, properties, computation and application to transient signal analysis. *IEEE Transactions on Signal Processing* 1993 **41**(1):245–266.

[2] Shao M, Nikias CL. Signal-processing with fractional lower order moments – stable processes and their applications. *Proceedings of the IEEE* 1993 **81**(7):986–1010.

[3] Priestly MB. *Spectral Analysis and Time Series*. Academic Press: London, 1982.

[4] Bendat JS, Piersol AG. *Random Data: Analysis and Measurement Procedures*, *Second Edition*. John Wiley & Sons: New York, 1986.

[5] Dyer D, Stewart RM. Detection of rolling element bearing damage by statistical vibration analysis. *Journal of Mechanical Design-Transactions of the Asme* 1978 **100**(2):229–235.

[6] Kim YC, Powers EJ. Digital bispectral analysis and its applications to non-linear wave interactions. *IEEE Transactions on Plasma Science* 1979 **7**(2):120–131.

[7] Subba Roa T, Gabr MM. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer-Verlag, 1984.

[8] Dallemolle JW, Hinich MJ. Trispectral analysis of stationary random time-series. *Journal of the Acoustical Society of America* 1995 **97**(5):2963–2978.

[9] Collis WB, White PR, Hammond JK. Higher-order spectra: the bispectrum and trispectrum. *Mechanical Systems and Signal Processing* 1998 **12**(3):375–394.

[10] Fackrell JWA, White PR, Hammond JK, Pinnington RJ. The interpretation of the bispectra of vibrational signals. 1. Theory. *Mechanical Systems and Signal Processing* 1995 **9**(3):257–266.

[11] Hinich MJ, Wolinsky MA. A test for aliasing using bispectral analysis. *Journal of the American Statistical Association* 1988 **83**(402):499–502.

[12] Hinich MJ, Messer H. On the principal domain of the discrete bispectrum of a stationary signal. *IEEE Transactions on Signal Processing* 1995 **43**(9):2130–2134.

[13] Hinich MJ. Detecting a transient signal by bispectral analysis. *IEEE Transactions on Acoustics Speech and Signal Processing* 1990 **38**(7):1277–1283.

[14] Brockett PL, Hinich MJ, Patterson D. Bispectral-based tests for the detection of Gaussianity and linearity in time-series. *Journal of the American Statistical Association* 1988 **83**(403):657–664.

[15] Brillinger DR, Rosenblatt M. Asymptotic theory of estimates of the k-th order spectra. In *Spectral Analysis of Time Series*, Harris B (ed). John Wiley & Sons: New York, 1967, pp. 153–188.

[16] Qian GL, Gu SN, Jiang JS. The dynamic behavior and crack detection of a beam with a crack. *Journal of Sound and Vibration* 1990 **138**(2):233–243.

[17] Rivola A, White PR. Bispectral analysis of the bilinear oscillator with application to the detection of fatigue cracks. *Journal of Sound and Vibration* 1998 **216**(5):889–910.

[18] Bouraou NI, Gelman LM. Theoretical bases of free oscillation method for acoustical non-destructive testing. *Noise Con-97*, Pennsylvania State University, 1997; pp. 519–524.

[19] Ryue J, White PR. The detection of cracks in beams using chaotic excitations. *Journal of Sound and Vibration* 2007 **307**(3–5):627–638.

[20] Schetzen M. *The Volterra and Wiener Theories of Nonlinear Systems*. Krieger, 2006.

[21] Collis WB, White PR, Hammond JK. Higher order spectra and Volterra analysis of mechanical systems. In *IEE Colloquium on Higher Order Statistics in Signal Processing: Are They Any Use?* IEE: London, 1995, pp. 2/1–2/6.

[22] Mendel JM. Tutorial on higher-order statistics (spectra) in signal-processing and system-theory – theoretical results and some applications. *Proceedings of the IEEE* 1991 **79**(3):278–305.

[23] Cardoso JF. Blind signal separation: statistical principles. *Proceedings of the IEEE* 1998 **86**(10): 2009–2025.

Chapter 28

Damage Detection Using the Hilbert–Huang Transform

Darryll J. Pines[1] and Liming W. Salvino[2]

[1] *Department of Aerospace Engineering, University of Maryland, College Park, MD, USA*
[2] *Structures and Composites, Carderock Division, Naval Surface Warfare Center, West Bethesda, MD, USA*

1 INTRODUCTION

Huang *et al.* [1] introduced a new adaptive method for nonlinear and nonstationary data analysis in 1998. This method consists of the combination of the empirical mode decomposition (EMD) and the Hilbert spectral analysis. Lin *et al.* [2] showed that the Hilbert–Huang transform (HHT) can be used to identify the structural parameters of a benchmark building and is quite accurate in detecting structural damage locations and severities. Quek *et al.* [3] illustrated the suitability of the HHT for damage detection problems, such as an aluminum beam with a crack, a sandwiched aluminum beam with an internal delamination, and a reinforced concrete slab with different degrees of damage. Crack and delamination in homogeneous beams can be located accurately and damage in the reinforced concrete slab can be identified if it has been previously loaded beyond first crack. Tua *et al.* [4] used the energy peaks in the Hilbert spectrum corresponding to crack-reflected waves to determine accurate flight times and also to estimate the orientation of the crack. Yang *et al.* [5] proposed two damage detection techniques based on the HHT. The first method based on the EMD extracted damage spikes due to a sudden change, whereas the second one based on the EMD and the Hilbert transform was capable of determining the damage time instants and determining the natural frequencies and damping ratios of the structure before and after damage. Pines and Salvino [6] showed the ability of the EMD to extract phase information from transient signals and used the results to infer damage in a structure. Salvino *et al.* [7–10] also compared the instantaneous phase detection approach with other data-driven damage detection methods, such as autoregressive modeling and nonlinear prediction approach using 2-D frame structure with multiple damages. Jha *et al.* [11] investigated the HHT to a multilevel structure. They showed that discontinuity in the intrinsic mode functions (IMFs) indicated the presence and location of

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

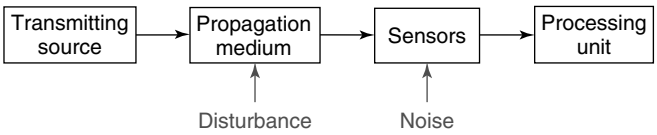


Figure 1. A common signal-processing system.

a damaging event in a structure monitored continuously. Bernal and Gunes [12] examined the instantaneous frequency of the IMFs provided by the EMD as a damage detection tool. Yu *et al.* [13] proposed a method for the fault diagnosis of roller bearings based on the HHT. The local Hilbert marginal spectrum is used to diagnose the faults in a roller bearing and to identify fault patterns. Quek *et al.* [14] compared the results obtained from the wavelet analysis with results from an EMD. The EMD was shown to provide a more direct method of extracting information needed for damage detection purposes.

This article demonstrates that the use of EMD coupled with Hilbert spectral analysis is an effective tool for locating and tracking progressive damage accumulation in a two-dimensional thin composite plate structure [15]. The next section introduces the concept of EMD as a new nonstationary time-series signal-processing approach (Figure 1).

2 THE HILBERT–HUANG TRANSFORM

The EMD and HHT method [12] addresses the limitations of the traditional Hilbert transform for signal analysis. The fundamental concepts of instantaneous frequency and time–frequency analysis are first described. The EMD and HHT method can be divided into two parts. In the first part, the data are decomposed into a collection of IMFs. This decomposition is viewed as an expansion of the data in terms of the IMFs. In other words, these IMFs are regarded as the basis of that expansion, which can be linear or nonlinear as dictated by the data. The IMFs, by definition, have well-behaved Hilbert transforms. In the second part, the corresponding instantaneous frequencies are calculated. The local energy and the instantaneous frequency derived from the Hilbert transforms of IMFs give us a full-energy–frequency–time distribution of the data, and such a representation is designated as the Hilbert–Huang spectrum. In this manner, the EMD and HHT method affords time–frequency analysis of a large class of signals.

2.1 Fundamental concepts

The Hilbert transform $H[x(t)]$ of a real-valued function $x(t)$ extended from $-\infty$ to $+\infty$ is a real-valued function defined by

$$\begin{aligned} H[x(t)] &= y(t) \\ &= \lim_{\varepsilon \to 0} \left[\int_{-\infty}^{0-\varepsilon} \frac{x(u)}{\pi(t-u)}\,\mathrm{d}u \right. \\ &\quad \left. + \int_{0-\varepsilon}^{\infty} \frac{x(u)}{\pi(t-u)}\,\mathrm{d}u \right] \end{aligned} \tag{1}$$

Assuming that $\int_{-\infty}^{\infty}[x(t)]^2\,\mathrm{d}t < \infty$, equation (1) can be rewritten as

$$H[x(t)] = y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(u)}{t-u}\,\mathrm{d}u \tag{2}$$

where P indicates the Cauchy principal value. This transform exists for all functions of class L^p, i.e., the space of p-power integrable functions. With this definition, $y(t)$ forms the complex conjugate of $x(t)$ and we can define the analytical signal, $z(t)$, as

$$z(t) = x(t) + iy(t) \tag{3}$$

The advantage of this representation lies in the fact that the possibility arises of uniquely determining genuine time-variant variables. These are the instantaneous parameters with amplitude as $a(t) = \sqrt{x(t)^2 + y(t)^2}$ and phase $\theta(t) = \arctan(y(t)/x(t))$.

Finally, equation (3) can be rewritten as

$$z(t) = a(t)\,\mathrm{e}^{i\theta(t)} \tag{4}$$

In this equation, the polar coordinate expression further clarifies the local nature of this representation.

It is the best local fit of an amplitude- and phase-varying trigonometric function to $x(t)$. In this respect, the Hilbert transform expresses a signal as a harmonic with time-dependent amplitude and frequency. The instantaneous frequency is defined as the rate of phase change

$$\omega = \frac{\mathrm{d}\theta}{\mathrm{d}t} \quad \text{and} \quad f = \frac{1}{2\pi}\frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{5}$$

The definition of the instantaneous frequency presented here is the most common form, but it is by no means unique. Ville [16] defined another estimator for the instantaneous frequency as the first moment of the distribution with respect to frequency. Cohen [17] defined the instantaneous frequency to be the average of the frequencies that exist in the time–frequency plane at a given time. A comprehensive discussion on the various proposed formulations may be found in Boashash [18]. Some limitations on the data are necessary, since the instantaneous frequency given in equation (5) is a single value function of time. The definition implies that at any given time, there is only one frequency value. This leads Cohen [19] to introduce the term *monocomponent function*, which has been loosely defined as narrow band. Finally, one reasonable and meaningful definition of the instantaneous frequency could be the representation of the frequency of the signal at one time, without any information about the signal at other times. For most multicomponent signals, the assumption of a single-frequency component at each time instant is no longer valid, leading to a meaningless instantaneous frequency. In order to obtain meaningful instantaneous frequency, restrictive conditions have to be imposed on the data as discussed by Gabor [20] and Cohen [19]: the frequency of the signal must be positive and the function must be symmetric locally with the respect to the zero mean level. These restrictions lead to the definition of a class of functions for which the instantaneous frequency can be defined everywhere. A new method needs therefore to be introduced in order to decompose any signal into a superposition of components with well-defined instantaneous frequency. The idea behind this decomposition method is to decompose a multicomponent signal into thesum of single-component functions. This new method has been introduced by Huang in 1998 [1] and is called the *empirical mode decomposition*.

2.2 The empirical mode decomposition

2.2.1 Intrinsic mode function

According to Huang, an intrinsic mode is a function that satisfies two conditions:

1. In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one;
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

An example of an IMF is given in Figure 2. IMFs represent oscillatory modes embedded within data where each IMF involves only one mode of oscillation. An IMF can be nonstationary and either be amplitude and/or frequency modulated. It has to be mentioned that, by definition, IMFs always have positive frequencies, because the oscillations in IMFs are symmetric with respect to the local mean.

2.2.2 The sifting process

The sifting process aims to decompose any data into a set of independent IMF components since most of the data are not naturally IMFs.

The decomposition is based on three assumptions:

1. the signal has at least two extrema, one maximum and one minimum;
2. the characteristic timescale is defined by the time lapse between the extrema; and
3. if the data were totally devoid of extrema but contained only inflection points, then it can be differentiated once or more times to reveal the extrema.

The sifting process to find the IMFs of a signal consists of several steps. We describe these steps using an arbitrary signal denoted $x(t)$.

1. Find the positions and amplitudes of all local maxima and minima in the input signal. These are marked by dots in Figure 3.
2. Create the upper envelope by spline interpolation of the local maxima and the lower envelope by spline interpolation of the local minima, denoted by $e_{\max}(t)$ and $e_{\min}(t)$. These are shown as curves in Figure 3.

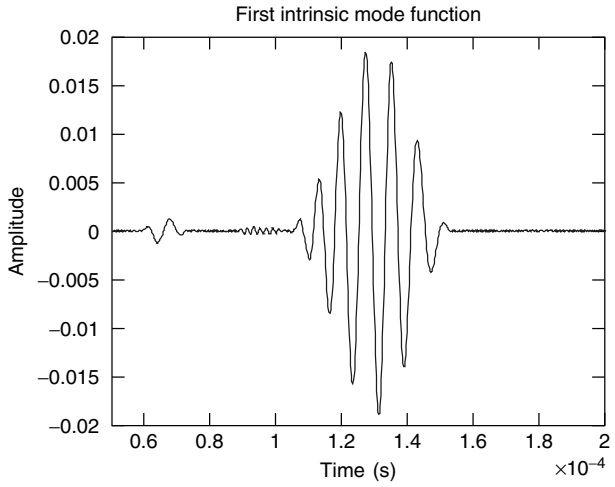


Figure 2. Intrinsic mode function (IMF) example.

3. For each time instant t, calculate the mean of the upper envelope and the lower envelope.

$$m_1 = \frac{e_{\max}(t) + e_{\min}(t)}{2} \tag{6}$$

This signal is referred to as the *envelope mean* and is shown as the black line in Figure 3.

4. Subtract the envelope mean signal from the input signal, yielding the results illustrated by dark grey in Figure 3.

$$h_1(t) = x(t) - m_1 \tag{7}$$

This is one iteration of the sifting process. The next step is to check if the signal $h_1(t)$ is an IMF or not. In the original work of Huang, the sifting process stops when the difference between two consecutive siftings is smaller than a selected threshold SD, defined by

$$SD = \sum_{t=0}^{T} \left[\frac{|(h_{1(k-1)}(t) - h_{1k}(t))|^2}{h_{1(k-1)}^2(t)} \right] \tag{8}$$

The choice of the stop criterion is crucial and is discussed further at the end of this section.

5. If $h_1(t)$ is not an IMF, iterate by repeating the process from step 1 with the resulting signal from step 4. If in the second sifting process, $h_1(t)$ is treated as the data, then

$$h_{11} = h_1 - m_{11} \tag{9}$$

We can repeat this sifting procedure k times, until h_{1k} is an IMF, that is,

$$h_{1k} = h_{1(k-1)} - m_{1k} \tag{10}$$

When the stop criterion is met, the IMF is defined as

$$c_1 = h_{1k} \tag{11}$$

After the IMF c_1 is found (illustrated by the dark grey curve in Figure 3), we define the residue r_1 as the result of subtracting this IMF from the input signal:

$$r_1 = x(t) - c_1 \tag{12}$$

The residue is illustrated by the dark grey curve in Figure 3.

6. The next IMF is found by starting over from step 1, now with the residue as the input signal.

Steps 1–6 can be repeated for all the subsequent r_j and the result is

$$r_1 - c_2 = r_2, \ldots, r_{n-1} - c_n = r_n \tag{13}$$

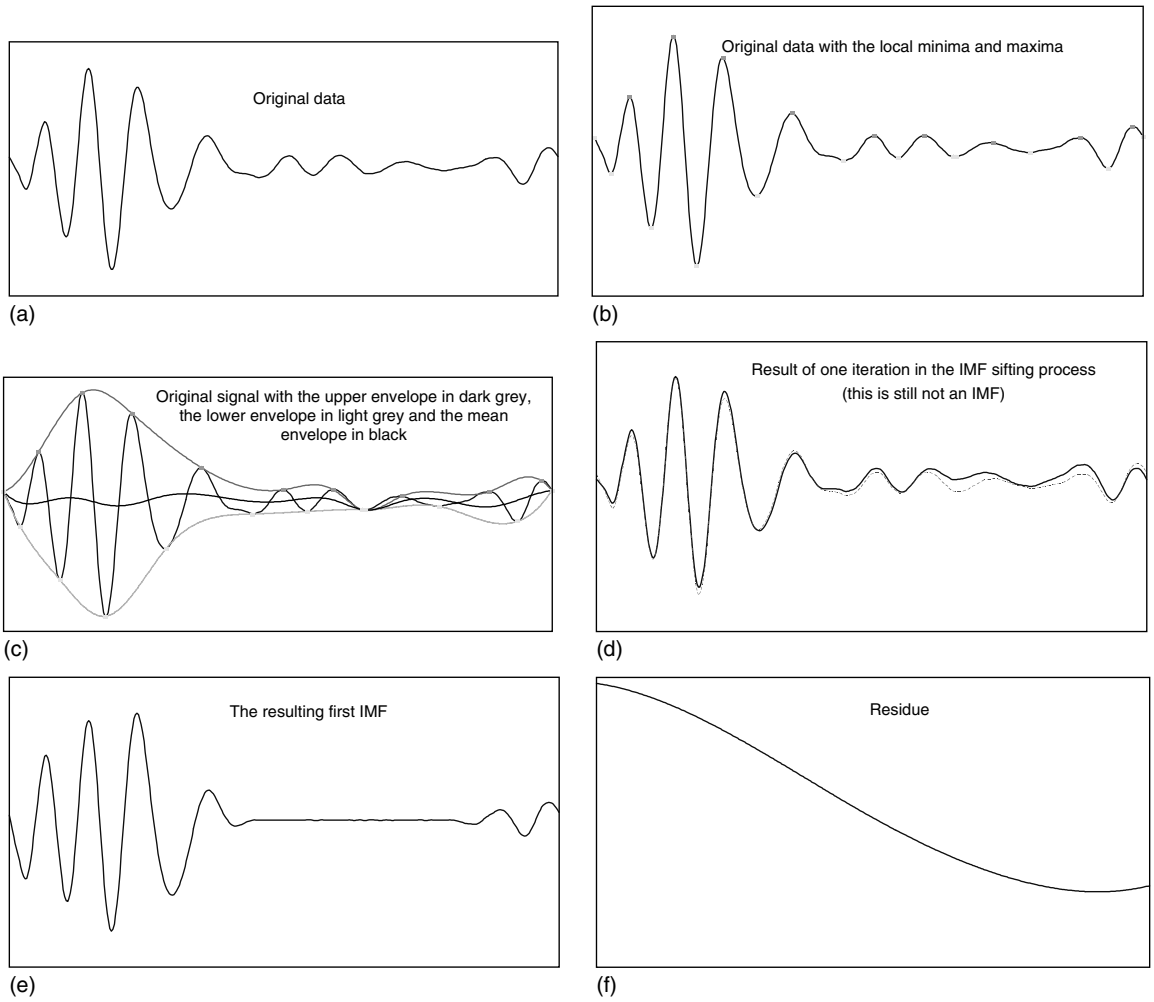


Figure 3. Sifting process illustration: (a) original signal; (b) localization of the local extrema; (c) spline fitting; (d) result of one iteration of the sifting process; (e) resulting first IMF; and (f) residue function.

The EMD is completed when the residue, ideally, does not contain any extrema points. This means that it is either a constant or a monotonic function. The signal can be expressed as the sum of IMFs and the last residue

$$x(t) = \sum_{i=1}^{n} c_i + r_n \tag{14}$$

The extracted IMFs are symmetric, have a unique local frequency, and different IMFs do not exhibit the same frequency at the same time. Another way of explaining how the EMD works is that it picks out the highest frequency possible that remains in the signal. The sifting process allows one to decompose the data into n-empirical modes and a residue. A brief summary of the EMD procedure is given in Figure 4.

It is important to note that certain features of IMFs are algorithm dependent. When more than one instantaneous frequency at a specified time exists for a complicated signal, its value can be uniquely assigned to an IMF component i. However, i can be different for a given time t, depending on the choice of computational parameters. In other words, there are different ways to decompose a given signal

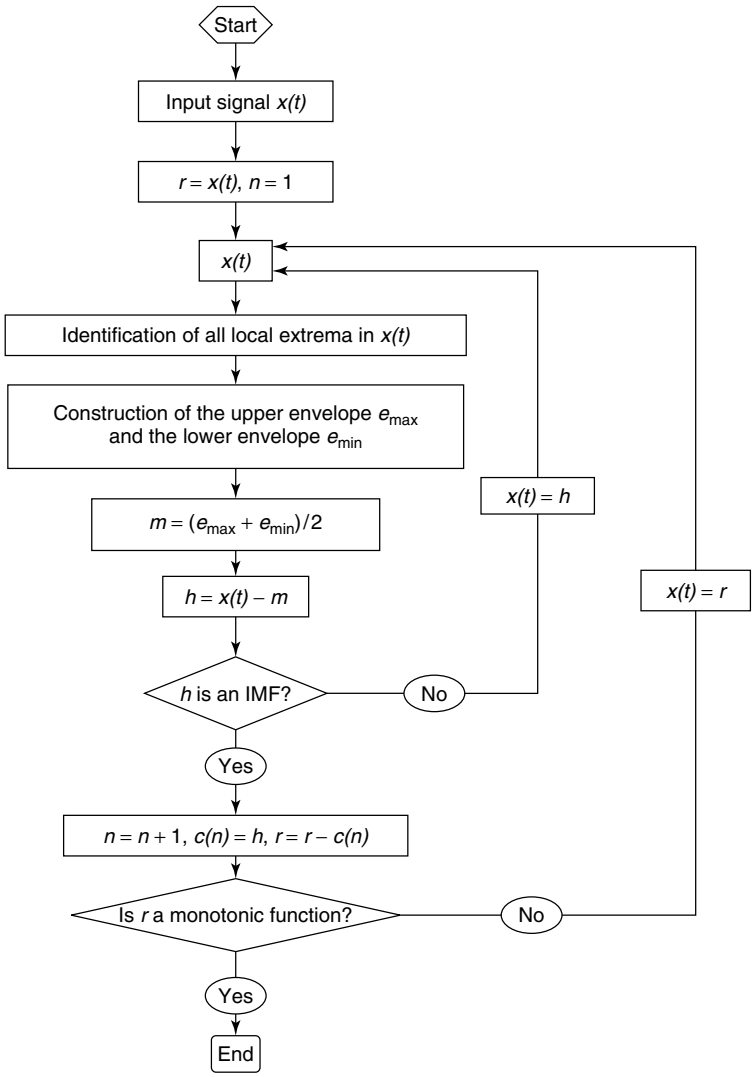


Figure 4. Flow chart of the EMD.

into IMFs. Different IMF sets can be generated using different sifting parameters such as stoppage criterion, maximum number of siftings, end-point boundary conditions, etc. The obvious, yet important question is which set of the many possible choices of shifting parameters gives a meaningful result and how to choose the sifting parameters to obtain the "best IMF set". A follow-on work of Huang *et al.* defines a confidence limit to find a stable range of stopping criteria for the sifting operation [9]. Within this stable parameter range, any IMF sets are good representations of the signal. This feature of EMD method can be used in a constructive way to examine data to achieve drastically improved robustness of the analysis results [10]. Further discussion about this is given in Section 2.5.

2.3 The Hilbert spectral analysis

The EMD gives us a set of independent IMF components with meaningful instantaneous frequencies.

Each of the IMF components can be therefore treated as a signal where the Hilbert transform can be applied.

$$c(t) = a_j(t) \exp\left(i \int \omega_j(t)\,\mathrm{d}t\right) \tag{15}$$

After performing the Hilbert transform on each IMF component, the original data $x(t)$ can be expressed as a real part (RE) of the complex expansion:

$$x(t) = RE\left[\sum_{j=1}^{n} a_j(t) \exp\left(i \int \omega_j(t)\,\mathrm{d}t\right)\right] \tag{16}$$

The residue r_n is omitted on purpose because it is either a monotonic function or a constant. Here, both amplitude $a_j(t)$ and instantaneous frequency $\omega_j(t)$ are functions of time t in contrast with the constant amplitude and frequency in the Fourier expansion.

$$x(t) = RE\left[\sum_{j=1}^{\infty} a_j \mathrm{e}^{i\omega_j t}\right] \tag{17}$$

The IMF represents a generalized Fourier expansion.

Equation (16) enables us to represent the amplitude and the instantaneous frequency as functions of time in a three-dimensional plot, in which the amplitude can be contoured on the frequency–time plane. This frequency–time distribution of the amplitude is designated as the Hilbert amplitude spectrum, $H(\omega, t)$, or simply Hilbert spectrum. Various forms of Hilbert spectra presentations can be made. The first one, the skeleton form, will emphasize frequency variations of each IMF and will be used if more quantitative results are desired. A second form, commonly used as a first look, is the smoothed Hilbert spectrum. If we apply a weighted spatial filter with large enough spatial averaging, we obtain a smoothed spectrum similar to what the wavelet analysis would give. Even if such a smoothing degrades both frequency and time resolutions, it could give a better physical interpretation. Hence, if more qualitative results are desired, the smoothed presentation is more appropriate. Both forms of the Hilbert spectrum are shown in Figure 5 for the example treated in the sifting process section.

2.4 Example of time–frequency response

To show the efficiency of the EMD and the Hilbert spectrum, a simple example has been studied and compared with the most common time–frequency methods. We consider the case of a simple sine wave with one frequency suddenly switching to another frequency and a Dirac-type impulse occurring at a certain time. This signal $s(t)$ is defined by equation (18)

$$\begin{cases} s(t) = \cos(2\pi f_1 t) & 0 < t \leq 0.125s \\ s(t) = \cos(2\pi f_2 t) + \delta(t = 0.1875s) & \\ & 0.125s < t \leq 0.25s \end{cases} \tag{18}$$

This signal has been chosen in order to demonstrate the ability of the EMD to separate the different

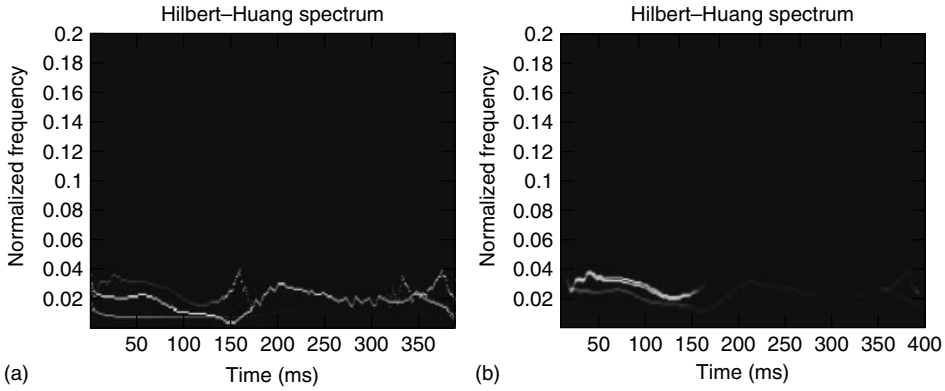


Figure 5. Hilbert–Huang spectrum: (a) skeleton form and (b) smoothed form.

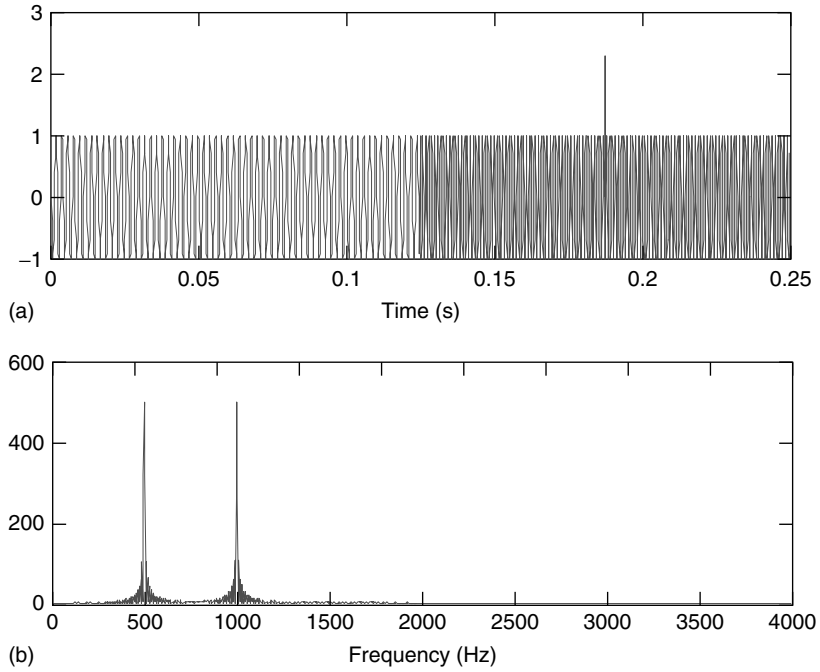


Figure 6. Example signal $s(t)$ (a) time-domain signal and (b) frequency spectrum.

frequency components as well as to identify irregularities in a signal. The time-domain representation of this signal along with its Fourier spectrum is given in Figure 6. As expected, the two frequencies present in the signal show up in the Fourier spectrum without any indication of when it is occurring and for how long these two frequencies are present. The spike in the signal is undetectable in Fourier spectrum.

The EMD with the Hilbert–Huang spectrum gives a high-resolution time–frequency representation. The 2-D plot provides fine details about the nature of the two sine waves as well as the impulse. The two frequencies can be accurately determined, as well as their time duration. Both frequency and time resolutions are excellent. Moreover, there is little, if any, leakage at the frequency switch point. The impulse appears as a discontinuity in the frequency band. This can be compared with a discontinuity for a mathematical function. This discontinuity is well localized in time and is spread out over the whole frequency range. This makes sense as an impulse in the time domain theoretically contains all frequencies. All these details can be seen in the three-dimensional plot in Figure 7.

After analyzing the signal $s(t)$ with the HHT, it is instructive to process the same signal through other time–frequency methods. A comparison with these methods will reveal the strengths and weaknesses of other time–frequency methods. The spectrogram of $s(t)$ corresponding to the short-time Fourier transform (STFT) and the scalogram corresponding to the wavelet transform are plotted in Figure 8. A random window size has been chosen for the STFT, that is, one-fourth of the signal length. A large time window implies a good frequency resolution and a poor time resolution. Indeed, we can notice that the two frequency components are overlapping in time around the switch point. Hence, we could believe that both frequencies are present in the signal during a short time period, which is not true. Furthermore, the impulse does not appear in the spectrogram. The main problem of the STFT is hence underlined. There is no guarantee that the window size adopted coincides with the stationary timescales. The impulse event being much localized in time, a narrow window must therefore be applied. However, this choice implies a poor frequency resolution, as shown in Figure 9(a). The impulse can now be seen but the spectrogram

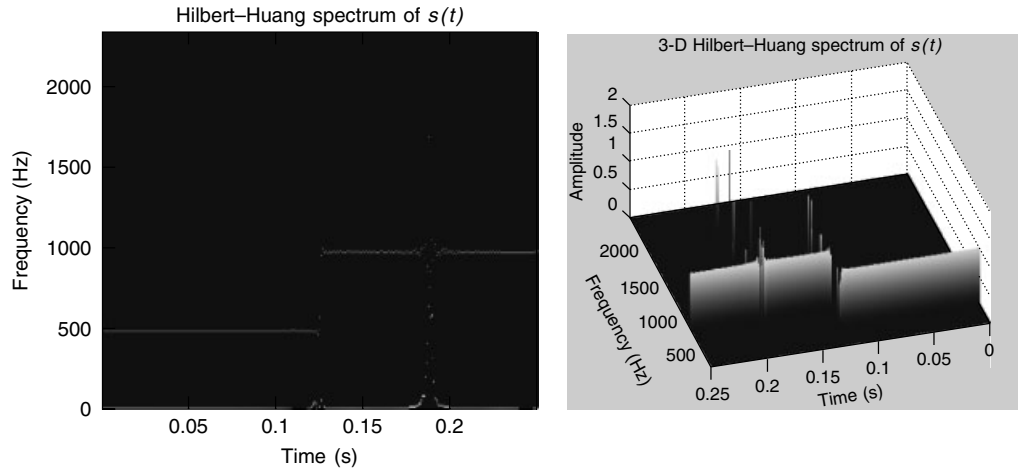


Figure 7. Hilbert–Huang spectrum of $s(t)$.

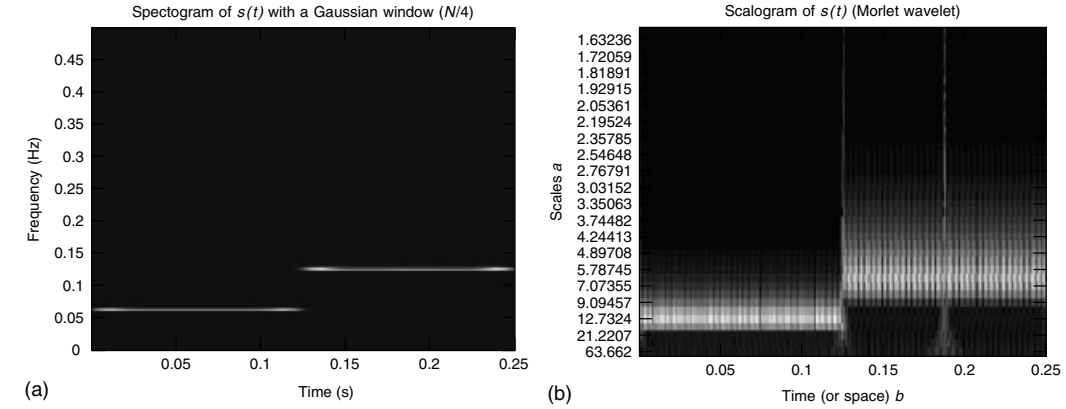


Figure 8. Spectrogram and scalogram of $s(t)$: (a) spectrogram ($N/4$) and (b) scalogram.

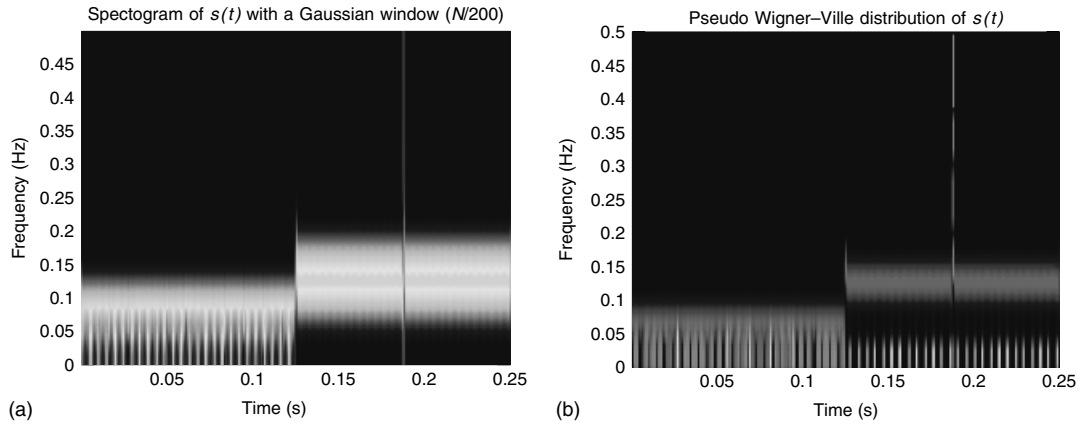


Figure 9. Similarity between the STFT and the PWVD: (a) spectrogram ($N/200$) and (b) pseudo Wigner–Ville distribution.

gives a smeared average frequency range over which the main wave energies reside. For the analysis of nonstationary data, the STFT turns out to be very limited compared to the HHT. Moreover, the pseudo Wigner–Ville distribution is plotted in Figure 9(b). In order to get rid of the cross terms and interferences, the time–frequency representation has been modified by smoothing in time and frequency. The result is, then, basically that of the STFT analysis. The Wigner–Ville distribution therefore suffers all the limitations of the STFT. The wavelet analysis is the best available nonstationary data analysis method prior to the development of HHT. The standard Morlet wavelet analysis identifies the local frequency before and after the frequency switch as well as the location of the frequency switch, which is shown in Figure 8(b). At the same time, the result also shows the leakage of energy to the neighboring modes. In the Hilbert–Huang spectrum (Figure 7), we can see much sharper frequency definitions and the time location of the frequency switch than those shown in the scalogram (Morlet wavelet). The range of variation is insignificant compared with the leakage in the wavelet analysis result. This suggests that the HHT is capable of providing a more crisp indication of the frequencies in the signal.

Another interesting observation is that the scalogram shows the smearing of the precise time location of the frequency switch event in the lower frequency range. If we only look in the low-frequency range, we cannot tell the exact time of the impulse, whereas it is well localized in the high-frequency range. More generally, to look for definition of a low-frequency event, we have to look at the high-frequency range in the wavelet spectrum. On the contrary, the energy of the impulse is well localized in both time and frequency domains in the Hilbert–Huang spectrum. This illustrates the unique property of the Hilbert spectrum in elimination of the spurious harmonic components to represent the nonstationary data. Another strength of the HHT is its ability to resolve the intrawave frequency modulation, while the wavelet can only describe the interwave frequency modulation. The analyzed signal $s(t)$ does not illustrate this property. Huang demonstrates it clearly through some examples [1].

This relatively simple example has been used to try to underline the strength as well as weaknesses of the basic wavelet transform, the STFT and the HHT. A comparative table (Table 1) summarizes the advantages and shortcomings of these different time–frequency methods based on the above simple observations. The STFT seems to be limited compared to the wavelet and the HHT in the analysis of nonstationary signals. The wavelet transform has been the best available nonstationary data analysis before the introduction of the HHT. However, the analysis of the signal $s(t)$ has proven that the HHT offers a better time and frequency resolutions than the basic wavelet transform used in this example and allows a better physical interpretation of the signal content. Section 3 is dedicated to the investigation of the HHT features that could be used to obtain robust and efficient metrics for damage detection in structures.

2.5 Summary, discussion, and recent development of EMD

EMD method has been proposed as an adaptive time–frequency data analysis method by decomposing a signal into a series of components known as *IMFs*. Each IMF contains only a single-frequency component at any instant in time. Instantaneous frequencies can be calculated using Hilbert transform for each IMF function. Taken collectively, the Hilbert spectra of the IMF set yield complete time–frequency information about the original signal. This approach, which has been termed the *Hilbert–Huang transform* (*HHT*), makes it possible to apply the Hilbert transform to an extremely general class of functions and signals. The HHT method has received considerable attention in a broad range of applications; many of these applications can be found in [21]. There are also many recent advancements in the EMD and HHT approach. For example, Flandrin *et al.* [22], and Wu and Huang [23] showed that the EMD algorithm is effectively an adaptive dyadic filter bank when applied to white noise.

One of the major drawbacks of the EMD algorithm is the frequent appearance of "mode mixing", which is defined as an IMF function either consisting of signals of different oscillation scales or a signal of a similar scale residing in different IMF components. Mode mixing is mainly due to time-series intermittency, which is a true physical phenomenon of measured systems. To overcome the scale mixed

Table 1. Comparison of time–frequency methods

	STFT	Wavelet	Hilbert–Huang
Strength	Easy to implement	Basis functions obtained by shifting and scaling a particular function	High time–frequency resolution
		Uniform resolution	Generalized Fourier analysis with variable amplitudes and frequencies
		Analytic form for the result	First local and adaptive method in frequency–time analysis
		Nonstationary data analysis	Can clearly define both inter- and intrawave frequency modulations
		Feature extraction	Robust nonlinear and nonstationary data analysis
			Feature extraction
Weakness	Piecewise stationarity of the data assumption not always justified	Uniformly poor resolution	End effects due to spline fitting and the Hilbert transform
	Time–frequency resolution limited by the Heisenberg principle	Leakage generated by the limited length of the basic wavelet function	Cannot separate signals with very close frequencies
	Nonadaptive nature	Nonadaptive nature	No physical meaning of some IMFs
	Feature extraction impossible	Cannot resolve intrawave frequency modulation	No mathematical formulation
		High-frequency range observation to define local events	

up between IMF components, a new approach is developed, termed *ensemble empirical mode decomposition* (*EEMD*) [24]. The EEMD utilizes important statistical characteristics of underlying scale separation principle of the EMD technique. It enables the EMD method to be a truly dyadic filter bank for general signal. By adding finite noise, the EEMD eliminates mode mixing in all cases automatically. This improvement is significant for real-world signal analysis.

3 EMD METRICS FOR DAMAGE DETECTION

The EMD with its associated Hilbert spectral analysis has shown promising results in the analysis of time-series data. This great potential could be applied to help the diagnosis and the prognosis of structures. The structural health monitoring aims at developing a damage identification method that provides complete damage information. Rytter [25] proposed a system of classification for damage identification techniques, which defined four levels of damage identification. The presence of damage, its location, its size, and the prediction of the remaining service life of the structure are the different hierarchical goals to reach for any damage detection scheme. An appropriate metric is an essential mechanism to validate the presence of a structural damage and the key tool to infer the size of the damage. From the measurement of a specific physical property of the structure (e.g., damping, stiffness, and energy), a metric will give the necessary information to quantify the severity of the damage. On the basis of the EMD method, we propose several metrics that can

be applied for damage detection and diagnostics of a structure.

3.1 The Hilbert instantaneous phase

Some researchers have recently suggested exploiting instantaneous phase features of the vibration signals combined with wave mechanics-based concepts for damage detection [26]. Compared to the other time–frequency methods, the Hilbert instantaneous phase is a unique feature that describes the traveling structural-wave propagation. Pines and Salvino [6] have indeed proved the dependency of the phase on the structural parameters as stiffness, mass, and damping. The Hilbert instantaneous phase has been shown to have an interesting feature for damage detection metrics.

3.1.1 Description

The Hilbert transform of a real-valued time-domain signal $x(t)$ is another real-valued time-domain signal, denoted by $H[x(t)]$, such that $z(t) = x(t) + iH[x(t)]$ is an analytic signal, where

$$H[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(u)}{t-u}\, \mathrm{d}u \tag{19}$$

We can define an envelope function $a(t)$ describing the instantaneous amplitudes of the original signal $x(t)$ and a phase function $\theta(t)$ describing the instantaneous phase of $x(t)$ versus time using $z(t) = x(t) + iH[x(t)] = a(t)\,\mathrm{e}^{i\theta(t)}$. These instantaneous parameters are hence defined as

$$a(t) = [x(t)^2 + H[x(t)]^2]^{1/2}$$
$$\text{and } \theta(t) = \arctan\left(\frac{H[x(t)]}{x(t)}\right) \tag{20}$$

The instantaneous Hilbert phase is therefore defined for the real-valued time-domain signal $x(t)$ in equation (20). Using superposition, the signal $x(t)$ can be decomposed into n-empirical modes $c_i(t)$ and can be expressed as

$$x(t) = \sum_{i=1}^{n} c_i(t) \tag{21}$$

The residue, which is a mean trend, has been left out on purpose. Then, the Hilbert transform is applied to each IMF and produces the instantaneous phase as functions of time,

$$\theta_i(t) = \arctan\left(\frac{H[c_i(t)]}{c_i(t)}\right) \tag{22}$$

The total instantaneous phase is the sum of the instantaneous phases corresponding to each IMF and is defined as

$$\theta(t) = \sum_{i=1}^{n} \arctan\left(\frac{H[c_i(t)]}{c_i(t)}\right) \tag{23}$$

Because the intrinsic modes have been restricted to be symmetrically local with respect to the mean zero level, the phase can be considered to be local and to increase monotonically as a function of time. The continuous phase function is represented using unwrapped radian phases with changing absolute jumps greater than π and their 2π complements. This unwrapped instantaneous Hilbert phase now needs to be investigated as a potential damage detection tool.

3.1.2 1-D finite element simulation

Before computing the phase for real undamaged and damaged data, a finite element model has been simulated in order to understand how the instantaneous phase can be used to detect damage in structures. A dynamic finite element model of a clamped-free rod was created in order to simulate the 1-D wave propagation. Damage was introduced by a loss of stiffness in an element and the excitation input was a sine burst signal. The simulation results are then processed through the EMD, and the phases for the undamaged case and the damaged case are then computed and compared.

First, we wish to analyze the response of an undamaged rod to a transient load. The problem will be that of a rod, fixed at $x = 0$ and subjected to a transient load $f(t)$ at $x = L = 30\,\mathrm{m}$. The goal of the dynamic finite element model (FEM) simulation is to plot the variation of the displacement $u(x, t)$ with time for any x along the rod. For the present study case, the transient load will be a sine burst at 10 kHz frequency applied at the tip, and the displacement at

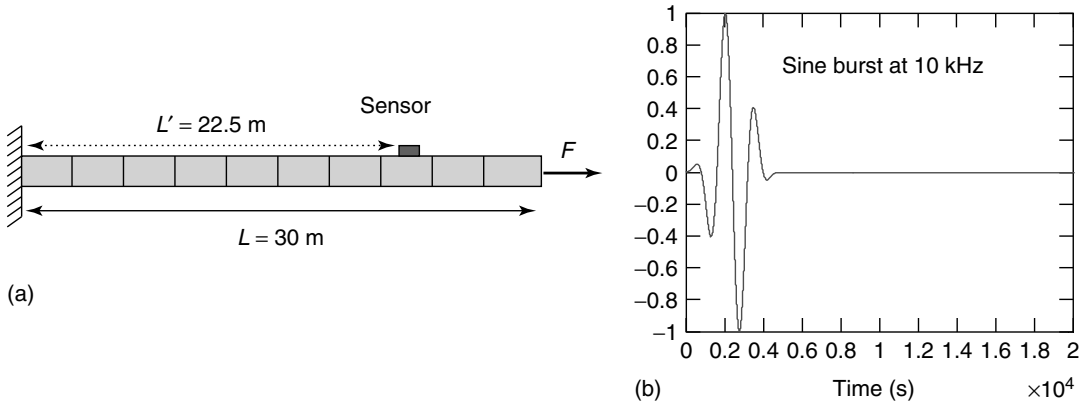


Figure 10. Forced vibrations model of a rod: (a) finite element model and (b) transient load.

$x = 22.5$ m is evaluated as shown in Figure 10. The wave-propagation solution of this kind of problem is very well known and can be found in Graff's book [27]. Waves in a rod are nondispersive. This means that their speed of propagation is constant, independent of frequency. The mathematical resolution of the wave propagation is therefore facilitated and the reflection from boundaries can be mathematically predicted. As shown in Graff's book [27], we will have a displacement reversal at the fixed boundary. This statement can be verified by the result of the simulation as seen in Figure 11. At first, the sensor detects the traveling wave generated by the tip excitation. The reflection from the clamped boundary is identified later and is reversed as predicted by the wave-propagation theory in a rod. The wave propagation in the undamaged rod is depicted in Figure 11.

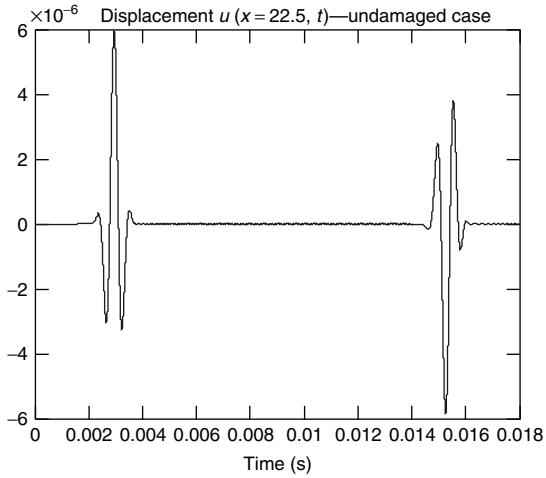


Figure 11. Response of an undamaged clamped-free rod to a transient load.

Reflections of waves may occur at discontinuities other than a boundary condition. One way to simulate damage in a computational model is to create a discontinuity in cross section, or material properties, or both, which is commonly referred to as an *impedance change*. A loss of stiffness in an element is hence introduced in the model, resulting in a damaged rod. The situation is shown in Figure 12.

Two damaged models have been created to introduce a loss of stiffness in the element located at $x = 7.5$ m such that (i) $E_{\text{dam}} = E_{\text{und}}/2$ for the first case, and (ii) $E_{\text{dam}} = E_{\text{und}}/4$ for the second case. Both wave-propagation simulation results are plotted in Figure 13. The reflections from the damage and the

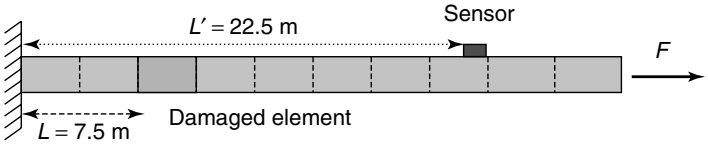


Figure 12. Damaged rod finite element model.

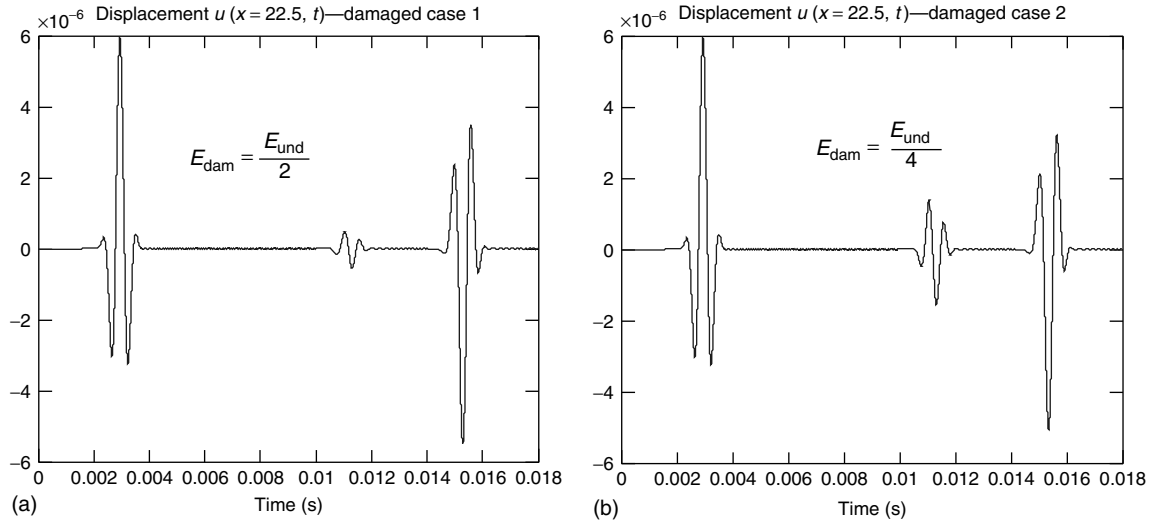


Figure 13. Response of a damaged clamped-free rod to a transient load: (a) damaged case 1 and (b) damaged case 2.

edge are clearly visible. The amplitude of these reflections increases with an increase in loss of stiffness in the element. A larger damage releases more energy in the structure, implying a greater magnitude in the waveform reflection.

A hamming window is applied in order to smooth the ends of the data to eliminate some aberrations and to allow a better spline interpolation in the EMD algorithm. The EMD along with the Hilbert transform can now be applied to these signals and the phase computed. The result for the Hilbert phase of the undamaged case and the damaged case is represented in Figure 14. From this plot, we can conclude that the reflection from the damage is interpreted by a slope change in the Hilbert phase. Physically, any damage in a structure alters the speed at which the energy traverses the structure. Once the wave passed through the damage, the energy speed is no more affected and the Hilbert phase behaves in the same manner for both undamaged and damaged cases as depicted in Figure 14. Furthermore, the slope change appears to be dependent on the size of the damage as seen in Figure 14. The energy speed propagation would be therefore altered in a different way depending on the

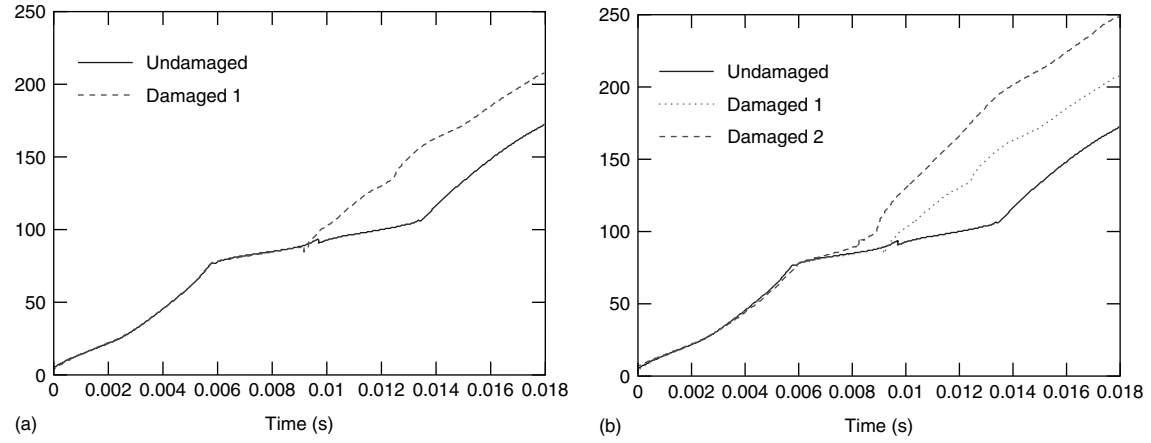


Figure 14. Hilbert phase: (a) damaged case 1 and (b) damaged case 2.

size of the damage. This implies that one can track increasing amount of damage as a function of phase. Thus, the Hilbert phase allows the size and location of damage to be determined. Even if the simulation results are not perfect and could be improved by augmenting the meshing and discretizing the rod by more elements, the damages are clearly localized by the slope change of the Hilbert phase.

3.2 The energy metric

One of the main advantages of the HHT is to provide a high-resolution energy–time–frequency representation. The Hilbert spectrum is capable of describing, with high precision, the frequency content of any nonstationary and nonlinear signals. The different features embedded in the time-domain signal can therefore be highlighted and better understood. For these reasons, the energy–time–frequency plot provided by the HHT could be used to identify and assess structural damage. A Hilbert energy spectrum describing the wave energy density can be created and the severity of the damage inferred from the reflected energy.

3.2.1 The Hilbert–Huang spectrum and the reflected energy

As seen earlier, after applying the EMD and the Hilbert transform on the IMFs, a real signal $x(t)$ can be expressed as $x(t) = \sum_{j=1}^{n} a_j(t) \exp\left(i \int \omega_j(t)\, \mathrm{d}t\right)$, where $a_j(t)$ are the instantaneous amplitudes and $\omega_j(t)$ are the corresponding instantaneous frequencies. Because both amplitude and frequency of each IMF are a function of time, they can be used to define a three-dimensional space or ordered triplet $[t, \omega(t), a(t)]$. This space is generalized by means of a function of two variables $H(\omega, t)$ to $[t, \omega(t), H(\omega, t)]$, where $a(t) = H(\omega(t), t)$. Therefore, we obtain a three-dimensional plot, in which the amplitude can be contoured on the frequency–time plane. The amount of energy carried by a wave is related to its amplitude. This transported energy is directly proportional to the square of the amplitude of the wave. Thus, the squared values of amplitude can be substituted in the Hilbert spectrum to represent the energy density and to produce the Hilbert energy spectrum. The idea behind this is that any reflection from damage can be quantitatively estimated through the energy transported by the reflected wave. The amplitude of the reflection increases with the size of the damage. This increase in amplitude should be revealed locally in the energy–time–frequency representation. We can, therefore, expect to infer the size of the damage through the Hilbert energy spectrum.

3.2.2 The Hilbert–Huang spectrum as a damage detection parameter

A damage index can be introduced in order to quantify the severity of the damage as a function of reflected energy:

$$E = \sum_{f} \sum_{t} |H_{\mathrm{e}}(t, f)| = \sum_{f} \sum_{t} |a(t)^2| \quad (24)$$

where $H_{\mathrm{e}}(t, f)$ is the energy density spectrum and $a(t)$ is the instantaneous amplitude.

This damage index is applied to the reflected waveform from the damage, and the result can be displayed next to each spectrum. As expected, this damage index increases with the size of the damage and a trend could be interpolated to infer the increasing damage.

In general, it will be difficult to identify the reflections from boundaries and other discontinuities. The EMD serves this purpose. The ability of the EMD to decompose any complicated data into a set of simple oscillatory functions would allow extracting embedded reflections.

To summarize, the objective is first to extract signal structures embedded in the data. The EMD is therefore the fundamental key to reveal the damage signatures that may otherwise remain hidden. The next step is to create the Hilbert energy spectrum for both undamaged and damaged cases. At this point, the energy density representation accurately describes the wave propagation in the structure. The presence of damage can hence be inferred and an estimation of the size predicted through the amount of reflected energy.

3.3 The phase shift metric

The high frequency and energy definitions provided by the HHT are used to locate and infer the size of

damage. Another interesting parameter to investigate is the high time resolution of the Hilbert–Huang spectrum. The time of flight is an important feature that is often used in damage detection schemes. The time of flight between the actuator and sensor in a structure is directly dependent on the wave-propagation properties in the structure.

The time resolution is therefore explored to obtain an accurate time of flight and, consequently, a precise position of the structural discontinuity. The track of increasing damage would be also possible as long as the initial size of the damage is known. A wave propagates at a certain speed in a structure. The time difference between the initial sensing actuation and the reflection from a discontinuity gives the time needed for the wave to travel. The basic formula length = speed × time locates the damage. When a wave encounters damage, part of the incident wave front is reflected back while the rest is transmitted through the damaged region. This reflection occurs at the beginning of the damage, which implies that a bigger damage would engender a sooner reflection than a smaller damage located at the same position. Hence, the times of flight would be different for damages with different sizes at the same location. Thus, if the time resolution of the HHT is powerful enough to perceive the small fluctuations of the different times of flight, an estimation of the severity of the damage would be possible.

Assuming that the size and the shape of the initial damage are known, simple geometric calculations would allow one to estimate the growth of the damage from the time difference between the initial and the new reflections. The severity of the damage can be thus quantified through the arrival time of the reflection as long as the initial damage is identified. Hence, the importance of an accurate signal processing for obtaining the flight times cannot be overemphasized. Accurate identification of localized events is required to determine the location of the damage, especially when complications due to the dispersive nature of waves in plates exist.

3.4 Summary

The EMD and the Hilbert spectrum offer a lot of possibilities for damage detection. First, the Hilbert phase represents a unique feature describing the energy propagation in a structure. The Hilbert phase as a damage detection parameter seems to work quite well for one-dimensional models with nondispersive wave propagation. The presence of damage is interpreted by a slope change in the Hilbert phase plot, which corresponds to a modification of the speed at which the energy traverses the structure. The finite element simulation as well as the civil building model analysis by Pines and Salvino [6] has shown promising results in the detection and evaluation of structural damage. Secondly, the energy metric relies on the high energy–time–frequency resolution of the Hilbert spectrum. This metric listens to the energy release in the structure upon defects growth. The amount of reflected energy is described and quantified with the Hilbert energy spectrum and an estimation of the severity of the damage can be deduced. Finally, a phase shift metric is developed on the basis of the arrival times of reflected waves from damage. Depending on the size of the damage, the wave will need a certain time to travel. This time of flight can be exploited along with geometric considerations to determine the defect growth from the knowledge of the initial damage properties.

Hence, the EMD along with the Hilbert transform seems to be the appropriate time–frequency analysis method for health monitoring of structures. The ability of the EMD to extract embedded oscillation, to reveal hidden reflections in the data, and to accurately describe the frequency content of a signal through the high-definition Hilbert spectrum make it an ideal tool for damage detection. The different damage detection schemes discussed in this article must now be investigated for two-dimensional structures where the waves are dispersive and thus more complicated to describe. The next section describes a low-velocity impact experiment on a cross-ply composite plate and an analysis of the results with the HHT.

4 LOW-VELOCITY IMPACT DAMAGE

A real-world application is now investigated to confirm the real potential of the HHT for a structural health monitoring system. The most commonly encountered type of damage is caused by impact due to the low interlaminar strength of composites. The mechanical properties can be severely degraded as

a result of a low-velocity impact even for barely visible damage. An impact on the structure can induce different type of damages such as matrix cracking, delaminations, and broken fibers. Such impact can occur in reality if a worker accidentally drops a tool or if a bird crashes into the aircraft during takeoff. A good way to experimentally simulate a low-velocity impact is to build a swinging pendulum mechanism. The energy of the impact could therefore be quantified through the potential energy and related to the damage growth.

4.1 Experimental setup

The experiment was conducted with an eight-ply composite plate [0/90]s, the setup of which is shown in Figure 15. The plate was instrumented with a piezoceramic actuator and a PVDF sensor array. A 0.635 cm × 0.635 cm PZT-5H patch actuator was bonded to the surface of the laminate at the left edge and used to excite the structure with interrogating A0 waves. A sensor array was bonded at a distance of 30.48 cm. from the actuation element. The array contained 19 circular sensor elements with a diameter of 0.3175 cm. and a spacing of 0.635 cm. The plate was 88.9 cm. long and 60.96 cm. wide. A tone burst signal of a 40 kHz frequency and five cycles was used to interrogate the plate and the resulting transient responses were observed and saved for each sensor of the array. The pendulum mechanism is shown in Figure 16. A hammer type of impactor was conceived and it was able to carry circular masses to increase the energy impact. The impactor was free to swing around a circular shaft introduced through the two columns. Instead of a bearing, shaft collars were used to avoid unwanted oscillations of the impactor. Thus, some frictional effects were neglected in this experiment and should be taken into account for a most accurate estimation of the energy released by the system. The experimental setup is pictured in Figure 17. The horizontal position of the plate was chosen in order to avoid any bending after the impact.

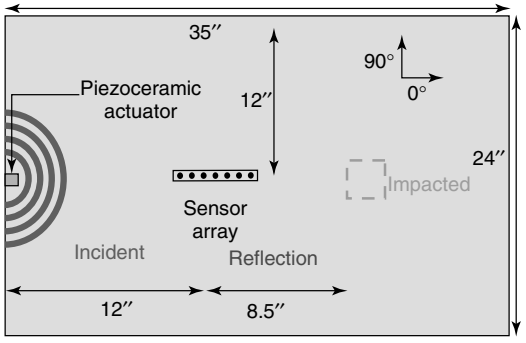


Figure 15. Experimental setup.

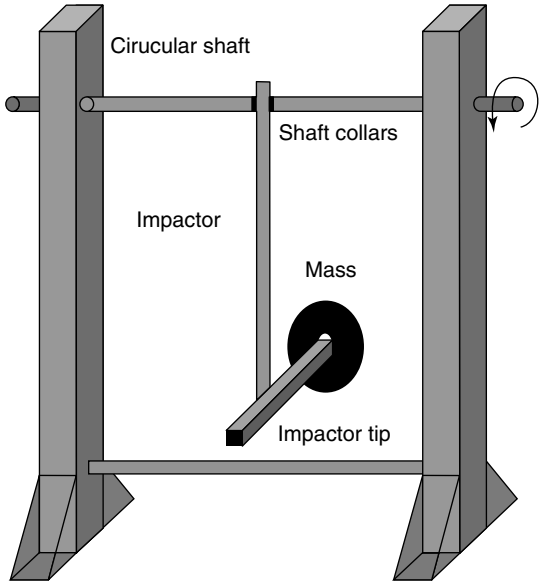


Figure 16. Swinging pendulum mechanism.

This experiment setup allows systematic impact as desired. For each impact, the response signals are gathered without removing the plate. The boundary conditions remain identical for every impact and thus will not influence the results. A picture of the impacted region is also taken after each impact. In this way, a correlation between the visual inspection (or local nondestructive evaluation (NDE)) of the plate and the signal-processing results will be possible. Each impact starts at zero initial velocity and at a certain initial angle. The impact energy comes from the gravitational potential energy. The experimental setup therefore makes pendulum impacts repeatable with only one varying parameter, the impact energy.

4.2 Impact energy quantification

Let us take the zero of potential energy to be at 90° angle of the swing, as shown in Figure 18. When the

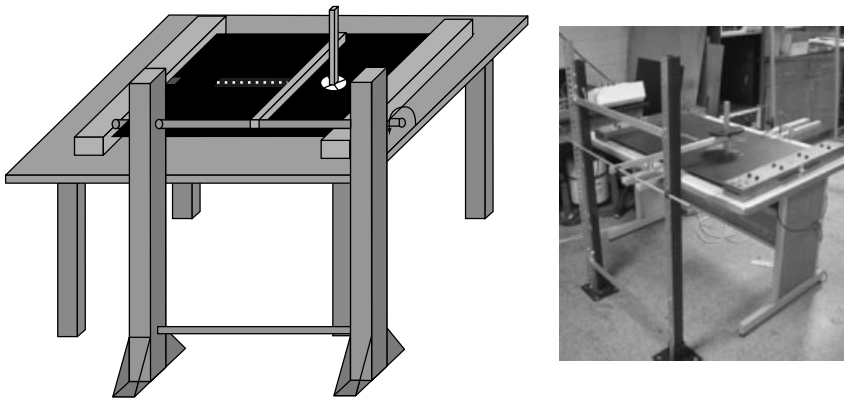


Figure 17. Low-velocity impact experimental setup.

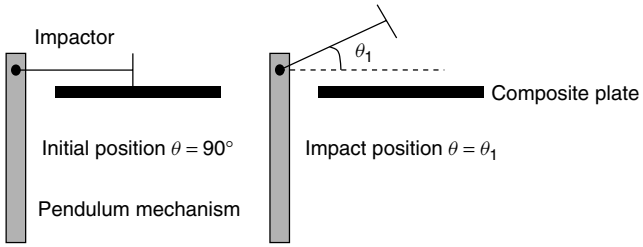


Figure 18. Potential energy calculation.

pendulum is brought to a certain height before being released, it is higher than at the initial position of the swing, and the potential energy is equal to the weight of the pendulum multiplied by the change in height. If the pendulum has length L, the change in height is $L \sin\theta$ and the potential energy becomes

$$E = mgL \sin\theta \tag{25}$$

In the study case, the pendulum is not limited at a single mass but at a hammer impactor with a significant weight. Thus, the impactor system will be replaced for the calculation by a single mass defined by its center of mass. The center of mass is defined as the average of the constituting element positions weighted by their masses. Each time a mass is added to the system, the center of mass changes.

The damage detection methodology follows these steps:

1. The plate is impacted. Each impact has more energy than the previous one. The initial angle is increased or a mass is added for this purpose.
2. A picture of the impacted region is taken.
3. The plate is interrogated by a tone burst at 40-kHz frequency with five cycles. The response signals are observed and saved on a disk.
4. This set of data is processed by using HHT.
5. Repeat the process.

The different impact cases with the corresponding potential energy are given in Table 2.

4.3 Transient signal analysis

A first undamaged set of data is taken as a baseline. The plate is then impacted and the middle

Table 2. Summary of the different impact cases

Case #	Mass m (kg)	Length L (m)	Initial angle θ (°)	Energy (J)
1	0.787	0.422	90	3.27
2	3.05	0.601	45	12.7
3	3.05	0.601	90	17.9
4	5.32	0.625	45	23.1

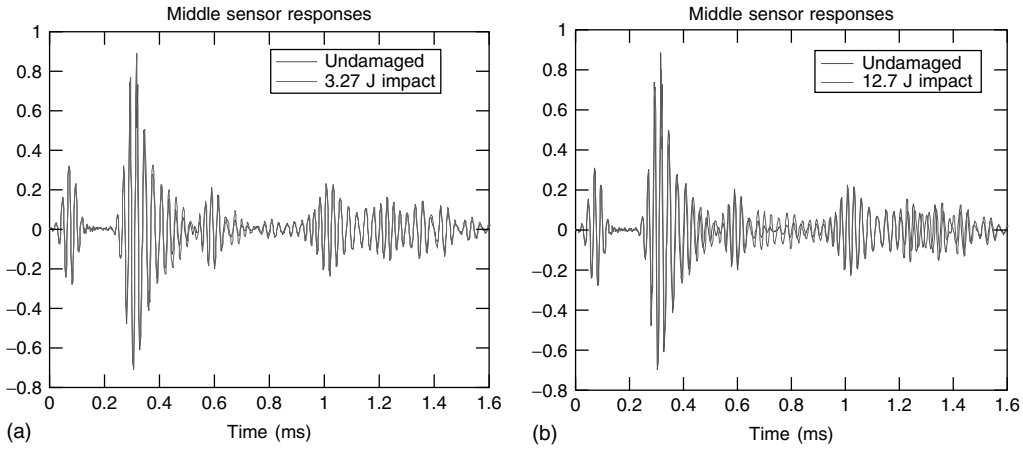


Figure 19. Middle sensor responses for the first two impact cases: (a) impact 1 and; (b) impact 2.

sensor responses are examined to infer the presence of damage in the plate. A first impact of 3.27 J is applied and the middle sensor signals for the undamaged and impacted plates are compared in Figure 19(a). The location of the impacted region was chosen in order for the reflected waveform to be visible in the transient signals. The pendulum mechanism was therefore impacted by the plate at a distance of 21.59 cm. from the middle sensor. If damage is created due to the impact, the reflected waveform should appear between the top/bottom edges and the right edge reflections. The observation of the transient signals in Figure 19(a) does not give any information about the presence of damage due to the first low-velocity impact. The impact energy was then increased by adding a mass to the impactor and the middle sensor responses are plotted in Figure 19(b). In this case, a difference can be seen between the undamaged and the impacted signals at ∼0.7 ms. A reflected waveform, though weak, shows up as expected in case of damage. The Lamb waves encounter a discontinuity through the thickness of the plate and part of the incident wavefront is reflected back. The middle sensor response is able to exhibit this reflection. An impact of 12.7 J damaged the composite plate. A C-scan examination would precisely determine the type of damage engendered by the impact. This NDE technique being unavailable, a visual inspection of the impacted region may provide some understanding of the impact consequence on the structure. A picture of both impacted region surfaces were taken and are shown in Figure 20. The observation of these pictures

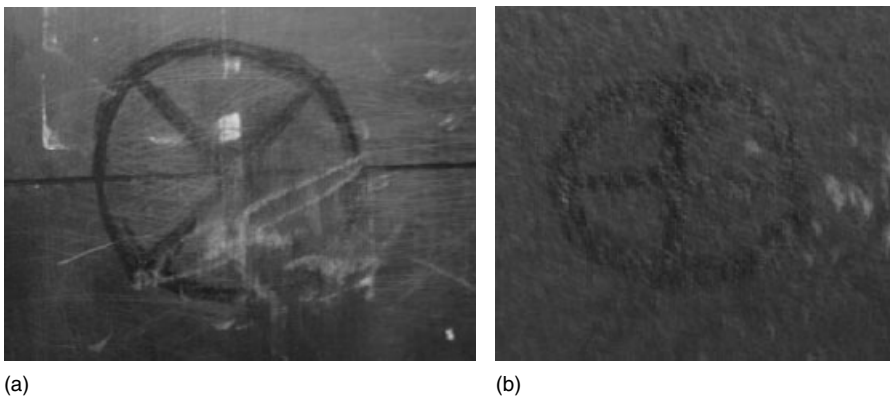


Figure 20. Impacted region pictures: (a) front surface and (b) back surface.

and a touch examination of the impacted region do not show the presence of any defects in the plate. The common visual inspection technique used for the structural health monitoring of composite structures would not detect the presence of damage in the plate. However, the transient response detects a discontinuity in the structure at the impact location. The reflected waveform localized by the sensor signal is, nevertheless, weak and needs to be effectively identified as a damage reflection. To accomplish this goal, the wavenumber filtering technique is applied to extract the leftward propagating waves. The damage reflection would be better characterized and quantified with this representation. The leftward propagating waves for the undamaged and the first two impacted cases are plotted in Figure 21. The first impact does not affect the health of the composite plate. However, the leftward propagating waves for the impact of 12.7 J of energy confirm the presence of damage with a new waveform detected between 0.6 and 0.8 ms. Although invisible on both surfaces of the impacted region, damage is present in the structure. The source of impact damage may include delamination between plies, matrix cracking, or broken fibers. The use of Lamb waves to interrogate composite structures enables subsurface damage that is invisible to the naked eye to be detected. A third impact is used to determine whether the damage region grows. The middle sensor response corresponding to this impact case 3 is plotted in Figure 22 along with the picture of the impacted region. Again, a weak reflection is visible for the impacted case around 0.7 ms that is not present for the undamaged case. In order to have a better idea of the magnitude of this reflection, the leftward propagation is also extracted and plotted in Figure 23. The picture of the front surface of the plate clearly shows that the composite structure is damaged. Broken fibers as well as matrix cracking can be seen. Compared to the previous impact, the damage is evident by visual inspection of the plate. However, the magnitude of the reflected waveform shown in Figure 24 slightly increases from the 12.7-J impact to the 17.9-J impact, whereas the difference between the shapes of the impacted region is obvious. From this observation, two theories can be assumed. The first theory is that the Lamb wave is not able to track the increasing damage. The second is that the structural properties of the plate have been equally affected by the impacts even though the visual inspection leads to different interpretations. A stronger impact analysis will confirm either of these assumptions. A mass of 10 lb is added to the hammer system and the swinging impactor is released at a 45° angle, producing a 23.1-J impact (case 4). The middle sensor response, along with the leftward propagation, is plotted in Figure 24. Both sensor signal and leftward propagating waves show that the magnitude of the damage reflection increases with the applied

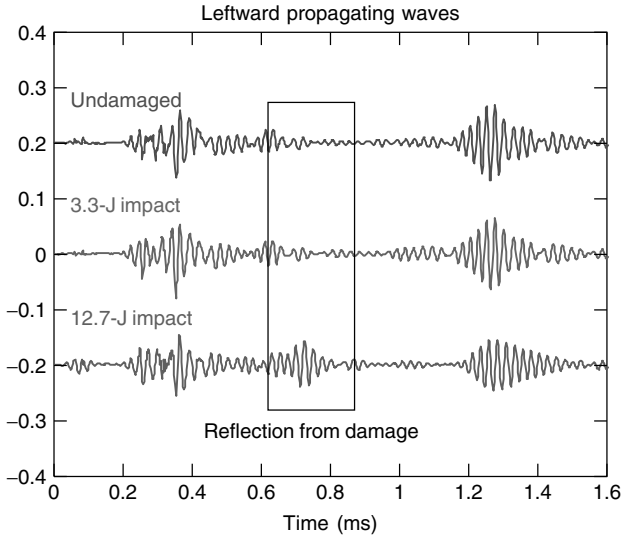


Figure 21. Leftward propagating waves—impacted cases 1 and 2.

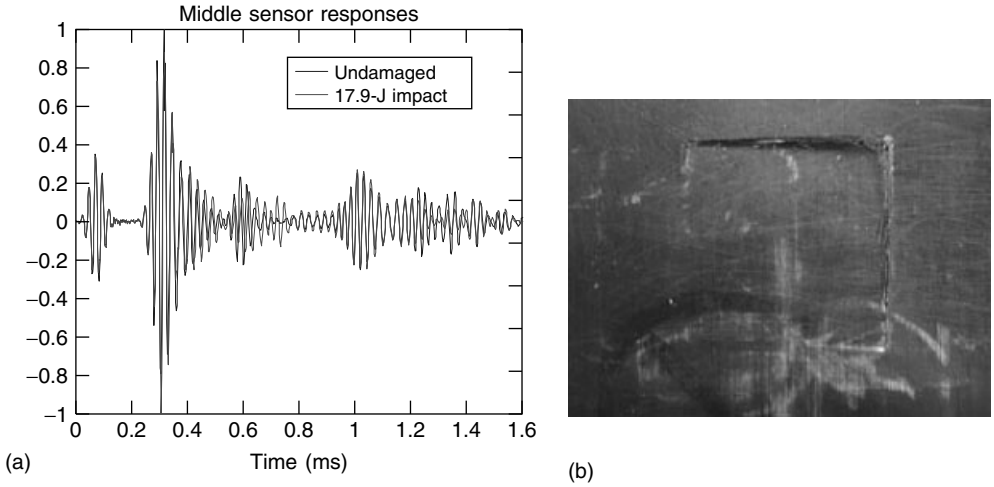


Figure 22. 17.9-J impact case: (a) middle sensor responses and (b) front surface picture.

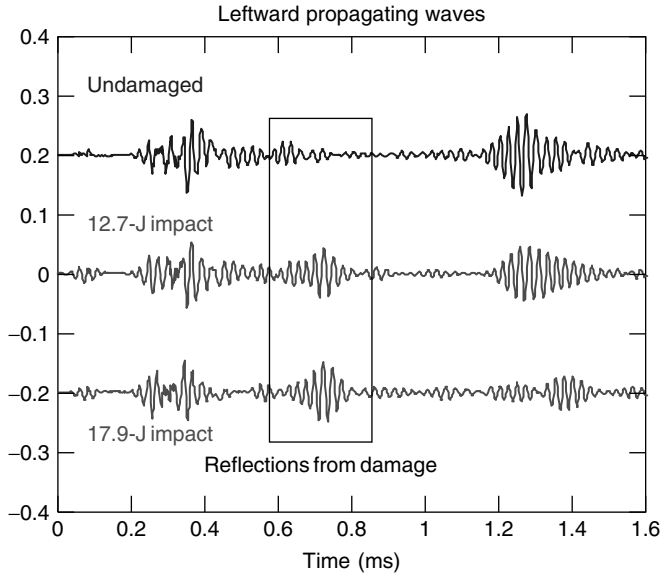


Figure 23. Leftward propagating waves—impacted cases 2 and 3.

impact energy and thus with the size of the damage. The second assumption is therefore the right one. The reflections from the right edge also decay with the increasing energy impact. Before applying the HHT to these signals, several conclusions can be drawn from the transient analysis:

- The Lamb waves are able to detect damage in a composite plate that is invisible to the naked eye.
- A composite structure can be seriously damaged without any physical signs being shown on the surfaces.
- The physical marks of damage and the real loss in mechanical properties are not related.
- The amplitude of the reflection from damage increases with the impact energy and therefore with the size of the damage.

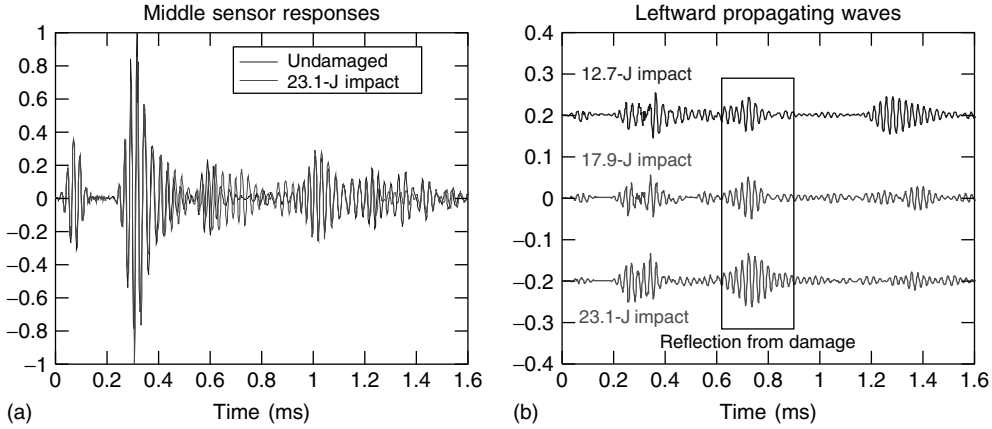


Figure 24. Case 4 analysis (a) middle sensor responses; (b) leftward propagation.

4.4 Hilbert–Huang spectrum and energy metric

The EMD is then used to apply the Hilbert transform to the IMFs extracted and to plot the time–frequency representation of the data. Once the Hilbert–Huang spectrum is obtained, the energy released by the damage can be quantified and linked to the impact energy and the size of the damage. The energy metric is applied to the unfiltered set of data. The Hilbert–Huang spectra for the unfiltered undamaged and damaged (12.7-J impact) are given in Figure 25. The results of the Hilbert energy measurement of the reflected frequency band are given in Table 3. Assuming that each impact weakens the plate and contributes to the creation of damage in the next impact, a cumulative energy impact approach seems to be more representative of the effective energy involved in damaging the plate. The energy metric results can therefore be plotted to describe the growth of damage with respect to the energy impact, as seen in Figure 26. The energy released upon the damage growth perfectly fits a parabolic interpolation curve. The track of increasing damage is therefore possible for this experiment through the energy released by the damage and using only one sensor.

Table 3. Energy metric results for the unfiltered data

Potential energy	0	3.3	12.7	17.9	23.1
Released energy	0.0182	0.1172	0.4946	1.1773	2.8956

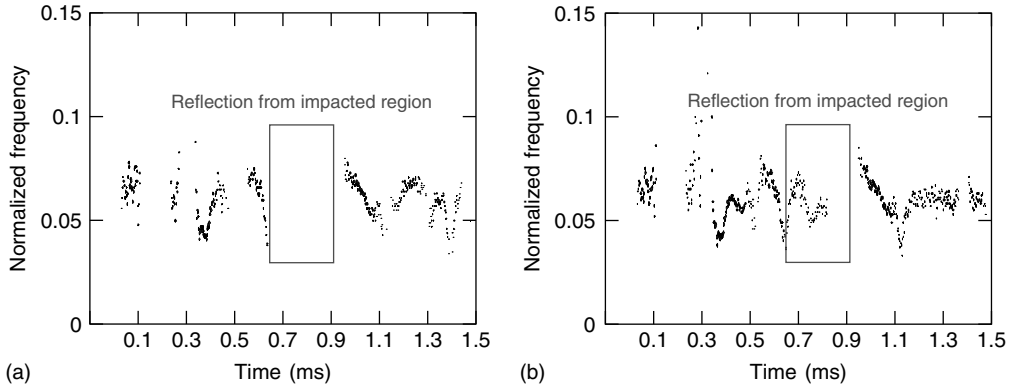


Figure 25. Hilbert–Huang spectra—unfiltered data: (a) undamaged and (b) 12.7-J impact.

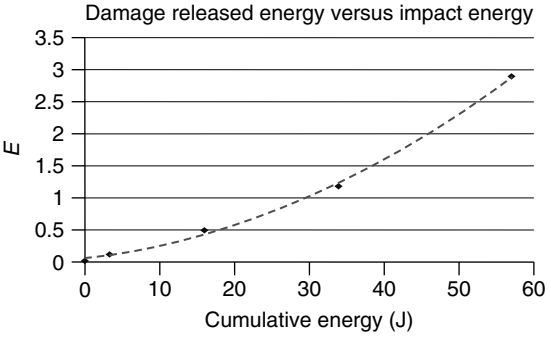


Figure 26. Growth of damage with impact energy.

4.5 Energy–time spectrum and phase shift metric

The energy–time spectrum aims to localize the damage through a precise determination of the time of flight and to quantify the size of the damage through the phase shift between the reflected waveforms. The energy–time spectrum for the low-velocity impact experiment described in the previous sections is given in Figure 27. For a better visibility and readability, the results have been separated in two distinct plots. As expected, the localization in time of the reflected waveform through the energy spectrum turns out to be very precise. A phase shift between the reflections from increasing damage can even be temporally quantified and related to the size of the damage. The amplitude of the reflected waveforms also increases with the energy impact. However, for case 4 corresponding to the 23.1-J impact energy, the reflection is spread out in time, has a weaker magnitude, and does not exhibit a clear peak. These characteristics can be explained by the fact that the addition of the 10-lb mass to the system along with a large initial angle deviated the point of impact and created a second damage in the plate. The reflected waveform represented in the energy spectrum therefore encloses two adjacent reflections. This assumption is confirmed by the picture of the impacted region in Figure 28 that shows two damages in the plate. Using the group wave velocity and the time of flight given by the energy spectrum, the location of the damage can be inferred.

In order to confirm that the plate has effectively been damaged by the impact, a cross section cut along the damage is realized. As we can see in Figure 29, the different impacts created matrix crackings, broken fibers, and a delamination.

4.6 The Hilbert phase

So far, the Hilbert phase turned out to be not only the most promising feature for delamination detection in composite structures using the HHT but also the most sensitive and difficult to apply. This real-world type of experiment aspires to confirm the capability of the Hilbert phase to describe the changes in the wave-propagation speed energy. The Hilbert phase results are plotted in Figure 30. For case 1, the transient analysis and the wavenumber filtering technique did not reveal the presence of damage even though the energy measurement showed a slight increase. The

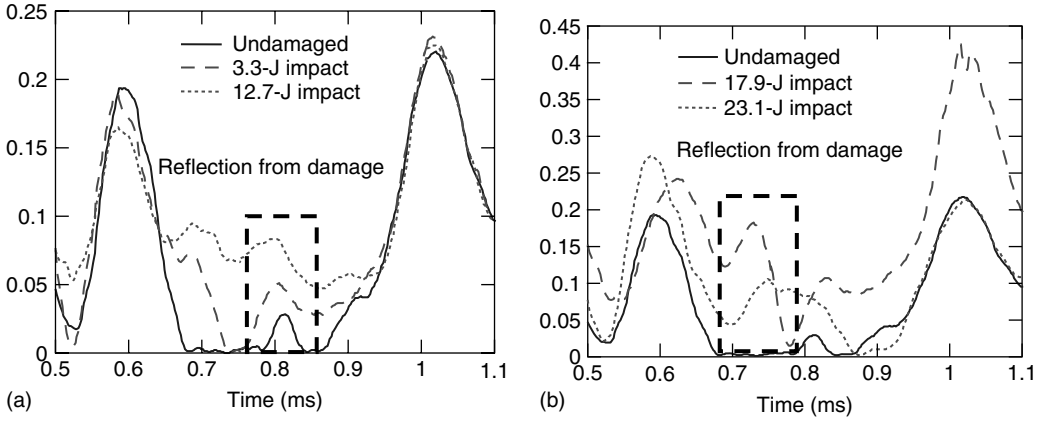


Figure 27. Energy–time spectra: (a) cases 1 and 2 and (b) cases 3 and 4.

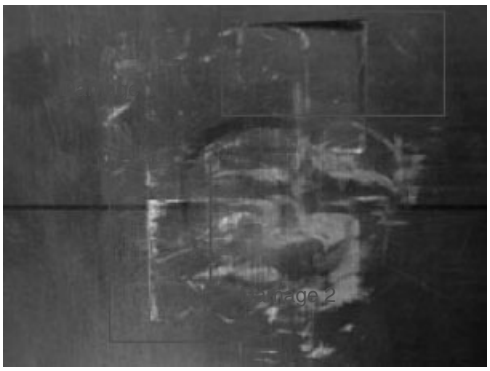


Figure 28. Front surface picture—case 4.

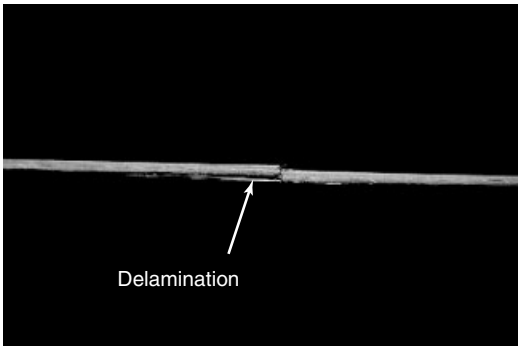


Figure 29. Cross-section picture of the impacted region.

Hilbert phase confirms these results. The phases for the undamaged case and the 3.3-J impact behave in the same way, indicating that there is no discontinuity encountered by the wave. In the case of damage, the wave propagation is altered, resulting in a slope change in the Hilbert phase. The Hilbert phase of cases 2 and 3 show these deviations. At around 0.7 ms, the undamaged and damaged phases start to diverge denoting the presence of damage in the composite plate as seen in Figure 30. The location of damage can be inferred from the time divergence and the amount of damage can be quantified by the mean phase error metric. The behaviors of both impacted phases are pretty close even if the 17.9-J impact appears to damage the plate more than the 12.7-J impact. The Hilbert phase is therefore consistent with all the previous conclusions and can perfectly describe the wave propagation and interaction with damage in the composite plate.

5 SUMMARY

This article has illustrated how the EMD method can be used to track damage in simple one-dimensional and thin two-dimensional composite structures. For thin plates, wave propagation is described in terms of Lamb waves. The antisymmetric fundamental transverse wave mode A_0 was able to describe the damage interactions in composite plates. The study has focused on developing and validating damage detection schemes based on the features of the EMD and the associated Hilbert spectral analysis. Three damage detection techniques were developed on the basis of the energy density, the time of flight, and the energy speed propagation. The instantaneous phase can be used as a damage detection tool. A finite

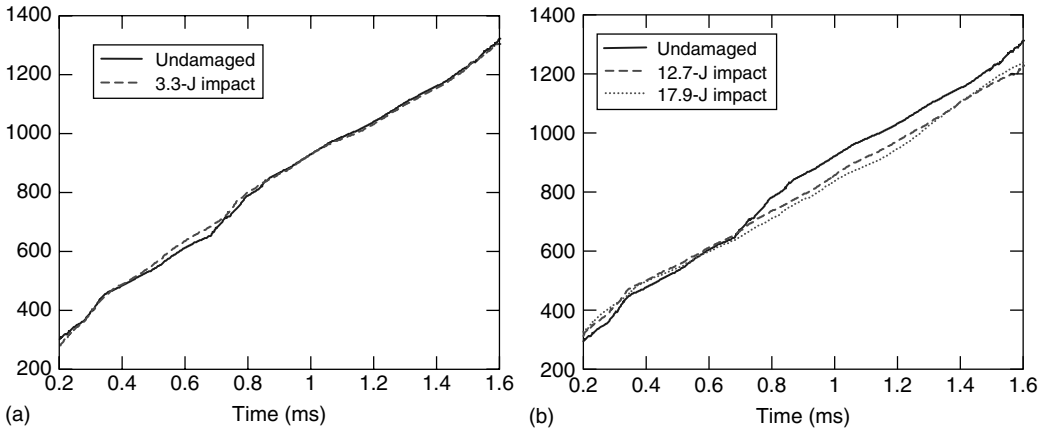


Figure 30. Hilbert phase plots: (a) case 1 and (b) cases 2 and 3.

element model of a rod was used to show the effect of a stiffness loss on the phase behavior. The principle is that any damage in a structure alters the speed at which the energy traverses the structure. The phase reflects this speed change by a slope deviation. As a result, the Hilbert phase can not only detect the presence of the defect in the structure but also quantify the extent of the damage. The Hilbert–Huang spectrum described the energy propagation in the structure. Any reflection is measurable by its energy density. The energy released upon defect growth can be therefore quantified and used to track the increasing damage. A larger reflection due to a larger damage will result in an increase of the energy in the Hilbert energy spectrum. The time of flight between the actuation waveform and a damage reflection can be accurately extracted from the energy–time spectrum provided by the HHT. The peaks on the energy–time spectrum for the IMF component, containing the highest energy, give the wave arrival times of interest. The wave group velocity combined with the time of flight locates the damage. If the initial shape and size of the damage is known, basic geometric considerations associated with the time of flight and wave-propagation characteristics provide a means to estimate the severity of the damage. A low-velocity impact was successfully experimented and tested. The repetitive impact system allowed a correlation of the visual inspection of the impacted region with the potential energy and the HHT results. Subsurface damage is detectable by examining the leftward propagating wave dynamics from the middle sensor. The energy–time spectrum could locate the damage, whereas the energy density spectrum was used to quantify the damage. The increasing damage was shown to follow a parabolic trend. The Hilbert phase also gave consistent results with the other damage detection techniques and was able to detect and quantify the amount of damage created by the pendulum impactor mechanism. This real-world application showed all the HHT potential as a damage detection tool.

RELATED ARTICLES

Signal Processing for Damage Detection
Data Preprocessing for Damage Detection
Statistical Time Series Methods for SHM
Cepstral Methods of Operational Modal Analysis
Hilbert Transform, Envelope, Instantaneous Phase, and Frequency
Time–frequency Analysis
Wavelet Analysis

REFERENCES

[1] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proceedings of the Royal Society of London* 1998 **A454**:903–995.

[2] Lin S, Yang J, Zhou L. Damage identification of a benchmark building for structural health monitoring. *Smart Materials and Structures* 2005 **14**:162–169.

[3] Quek S, Tua P, Wang Q. Detecting anomalies in beams and plate based on the Hilbert-Huang transform of real signals. *Smart Materials and Structures* 2003 **12**:447–460.

[4] Tua P, Quek S, Wang Q. Detection of cracks in plates using piezo-actuated Lamb waves. *Smart Materials and Structures* 2004 **13**:643–660.

[5] Yang L, Lei Y, Lin S, Huang N. Hilbert-Huang based approach for structural damage detection. *Journal of Engineering Mechanics* 2004 **130**(1):85–95.

[6] Pines DJ, Salvino LW. Structural damage detection using empirical mode decomposition and HHT. *Journal of Sound and Vibration* 2006 **294**:97–124.

[7] Salvino LW, Pines DJ, Todd M, Nichols J. EMD and instantaneous phase detection of structural damage. In *Hilbert-Huang Transform: Introduction and Applications*, Huang N, Chen S (eds). World Scientific Publishing, 2005; pp. 227–262.

[8] Salvino LW, Pines DJ, Todd M, Nichols J. Signal processing and damage detection in a frame structure excited by chaotic input force. *Proceedings of the 10th Annual SPIE Smart Material and Structures Conference*, San Diego, CA, March 2003.

[9] Huang NE, Wu ML, Long SR, Shen SS, Qu WD, Gloersen P, Fan KL. A confidence limit for the empirical mode decomposition and the Hilbert spectral analysis. *Proceedings of the Royal Society of London* 2003 **A459**:2317–2345.

[10] Wu Z, Huang NE. Statistical significant test of intrinsic mode functions. In *Hilbert-Huang Transform: Introduction and Applications*, Huang NE,

Shen SSP (eds). World Scientific Publishing: Singapore, 2005; pp. 125–148, 360.

[11] Jha R, Xu S, Ahmadi G. Health monitoring of a multi-level structure based on empirical mode decomposition and Hilbert spectral analysis. In *Fifth International Workshop on Structural Health Monitoring*, Chang FK (ed). Stanford University Press: Stanford, CA, 2005.

[12] Bernal D, Gunes B. An examination of instantaneous frequency as a damage detection tool. *14th Engineering Mechanics Conference*. Austin, TX.

[13] Yu D, Cheng J, Yang Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mechanical Systems and Signal Processing* 2005 **19**:259–270.

[14] Quek S, Tua P, Wang Q. Comparison of Hilbert-Huang, wavelet and Fourier transforms for selected applications. *Mini-Symposium on Hilbert-Huang Transform in Engineering Applications*, Newark, November 2003.

[15] Purekar AS. *Piezoelectric Phased Array Acousto-Ultrasonic Interrogation of Damage in Thin Plates*, Ph.D. Thesis. University of Maryland, Department of Aerospace Engineering, 2006.

[16] Ville J. Theorie et applications de la notion de signal analytique. *Cables et Transmissions*. 1948; Vol. 2A, pp. 61–74.

[17] Cohen L. Time-frequency distributions—a review. *Proceedings of the IEEE* 1989 **77**(7):941–981.

[18] Boashash B. Estimating and interpreting the instantaneous frequency of a signal—part 1: fundamentals. *Proceedings of the IEEE* 1992 **80**(4): 520–537.

[19] Cohen L. *Time-Frequency Analysis*. Prentice-Hall: Englewood Cliffs, NJ, 1995.

[20] Gabor D. Theory of communication. *Proceedings of the IEEE* 1946 **93**:429–457.

[21] Huang N, Chen S. *Hilbert-Huang Transform: Introduction and Applications*. World Scientific Publishing, 2005.

[22] Flandrin P, Rilling G, Goncalves P. Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters* 2004 **11**:112–114.

[23] Wu Z, Huang NE. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London* 2004 **A460**:1597–1611.

[24] Wu Z, Huang NE. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Proceedings of the Royal Society of London Series A*, 2005, to appear in.

[25] Rytter A. *Vibration Based Inspection of Civil Engineering Structure*, Ph.D. Dissertation. Department of Building Technology and Structural Engineering, Aalborg University: Aalborg, 1993.

[26] Jha R, Cross K, Janoyan K, Sazonov E, Fuchs M, Krishnamurthy V. Experimental evaluation of instantaneous phase based index for structural health monitoring. *Proceedings of SPIE, Smart Structures and Integrated Systems*. SPIE, 2006; Vol. 6173.

[27] Graff KF. *Wave Motion in Elastic Solids*. Dover Publications: New York, 1975.

Chapter 36

Model-based Statistical Signal Processing for Change and Damage Detection

Michèle Basseville
IRISA, Campus de Beaulieu, Rennes, France

1 INTRODUCTION

Detecting changes in signals and dynamical systems has been the topic of a number of both theoretical and practical investigations for about 30 years, as can be seen e.g., from the survey papers [1–3] and books [4–6] among many other references. Early investigations of statistical change detection took place in the area of quality control and trace back to the 1930s. More recent applications to industrial monitoring problems [7–11] have highlighted the relevance of statistical inference approaches to the problems of fault detection, isolation, and diagnosis. As for vibration-based structural health monitoring (SHM), a number of simulations, laboratory test beds, and real test cases [12–17] have also exhibited some interesting properties of algorithms based on a statistical change detection approach to damage detection and localization.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1.1 Motivations for change detection

Many monitoring problems can be stated as the problem of detecting and isolating a change in the parameters of a static or dynamic stochastic system. The use of a model of the monitored system is reasonable, since many industrial processes rely on physical principles, which are written in terms of (differential) equations, providing us with (dynamical) models. Moreover, the use of (physical) parameters is mandatory when diagnostics is sought. In the sequel, we equally use the words deviation, change, fault, and damage, considering that all these events are reflected

by a change in the parameter vector of a model of the monitored structure or system.

The change detection framework and methodology is one way to approach the analysis of nonstationary phenomena. Statistical decision tools for detecting and estimating changes are useful for different purposes:

1. automatic segmentation of signals as a first step in recognition-oriented signal processing;
2. gain updating in adaptive identification algorithms for improving their tracking ability;
3. quality control;
4. monitoring complex structures and industrial processes (damage and fault detection and diagnosis) for fatigue prevention, aided control, and condition-based maintenance.

Even though this article focuses on the use of change detection for damage detection and isolation, the same methodology and tools apply to the other problems as well.

1.2 Motivations for statistical methods

It has been widely acknowledged that the fault or damage detection and isolation problem can be split into two steps: *generation of residuals*, which are ideally close to zero under no-fault conditions, minimally sensitive to noises and disturbances, and maximally sensitive to faults; and *residual evaluation*, namely, design of decision rules based on these residuals. The basic statistical approach to residual generation consists in deriving sufficient statistics, namely, transformations of the measurements, which capture the entire information about the fault contained in the original data. Residual evaluation is typically answerable to statistical methods, which are basically aimed at deciding if a residual discrepancy from zero is significant.

The main advantage of the statistical approach is its ability to perform the early detection of deviations with respect to a reference undamaged state of the monitored system (possibly *before* a deviation develops into a damage). Another major advantage is its ability to assess the level of significance of those deviations with respect to noises and uncertainties. Whereas the accuracy of parameter estimates provides us with the relative size of the estimation error with respect to the noises on the system measurements, the statistical tests described below can similarly tell us if the relative size of the parameter discrepancy in the monitored system with respect to the accuracy of the reference parameter value is significant or not. These are crucial issues when addressing real SHM problems.

1.3 Two types of changes

Let θ be the model parameter, and $y_1, y_2, \ldots, y_n$ a sequence of measurements. In most cases of interest for SHM, both θ and the y_i's are multidimensional. For both intuitive and mathematical reasons, it is useful to classify the change detection problems into two categories: *additive* changes, such as changes in the mean value of the measurements, and *nonadditive* (or spectral) changes, which basically affect the correlations in the data or the dynamics of the system and associated transfer functions.

The main issue is which function of the data should be handled for making a decision about the presence of a change. The main lesson is that, whereas the innovation is convenient for monitoring additive changes, the innovation is *not sufficient* for monitoring changes in the system dynamics.

1.4 Three types of change detection problems

From now on, we assume that we are given a reference value θ_0 of the model parameter. Generally, such a reference parameter is identified with data from the safe system. The monitored system is often subject to other types of nonstationarities than the parameter deviations of interest, typically changes in the functioning modes of a machine, nonstationarities in the environment of a structure, etc. In such cases, the reference value θ_0 should be identified on long data samples containing as many of these nuisance changes as possible. The detection problem may be solved on the basis of data samples of smaller size. Depending on the relative time constants of the system to be monitored, on the sampling of the data, and on the size, speed and rate of the deviations to be detected, three types of detection problems may

occur in practice, when processing real data, onboard or not.

- **Model validation**

Given, on the one hand, a reference value θ_0 of the model parameter and, on the other hand, a new data sample, the problem is to decide whether or not the new data are still well described by this parameter value. Of course, this problem may be stated either off-line (fixed sample size) or on-line (varying sample size). A fixed-size sliding window may be useful. Model validation may be a relevant issue for onboard processing: for example, batch processing is appropriate for onboard monitoring of aging.

- **Off-line change detection**

Given a data sample of size n, the problem is to decide whether a change in the parameter has occurred, from θ_0 to θ_1, at an unknown time υ in the sample.

- **On-line change detection**

At every time instant n, the problem is to decide whether a change in the parameter has occurred, from θ_0 to θ_1, at an unknown time instant $\upsilon < n$.

Of course, the most difficult problem is the third one, because in this problem the amount of information in the data about the new parameter value θ_1 is the least. Also, the criteria for designing the detection algorithms and analyzing their performances depend on the detection problem. These are mean time between false alarms, probability of wrong detection, mean delay to detection, probability of nondetection, and accuracy of the estimates of the change onset time and of the change magnitude. However, even though the decision functions for solving these three problems are not the same, they all can be viewed as different implementations of the same primary residual.

This article is organized as follows. First, in Section 2, the key detection tools, likelihood ratio and CUSUM (cumulative sum) tests, are introduced for the model validation and on-line change detection problems, respectively. Then, the on-line detection problem is addressed for changes in the mean of scalar signals, additive changes in multidimensional independent or dependent time series, and changes in the spectrum of scalar signals, in Sections 3, 4, and 5, respectively. Other tools useful for handling small changes are introduced in Section 6. Back to the model validation point of view, Section 7 addresses the problem of detecting changes in the dynamics of a linear state-space system, of which vibration-based monitoring is an important instance. Finally, in Section 8 some additional issues are addressed, such as damage or fault isolation and diagnostics, robustness in the decision, damage detectability, and sensor positioning.

2 LIKELIHOOD RATIO AND CUSUM TESTS

The key detection tools are first introduced for hypotheses testing, and then for on-line change detection, distinguishing between independent and dependent observed data.

2.1 Hypotheses testing

The measured data $y_1, y_2, \ldots, y_n$ are assumed to be random variables with probability distributions p_θ. Hypotheses about the parameter vector θ are stated, reflecting the *a priori* knowledge on the changes to be monitored. Typically, the null (no-change) hypothesis is $\mathbf{H}_0 : \theta \in \Theta_0$ and the alternative (change) hypothesis is $\mathbf{H}_1 : \theta \in \Theta_1$. Two types of hypotheses are distinguished. A *simple* hypothesis is reduced to one point in the parameter space; typically, the no-change hypothesis reduces to the reference value of the parameter vector: $\Theta_0 = \{\theta_0\}$, and the change hypothesis to a different value: $\Theta_1 = \{\theta_1\}$. A *composite* hypothesis is defined by a subset of the parameter space; typically the no-change hypothesis is defined by Θ_0, and the change hypothesis by Θ_1.

2.1.1 Independent observations

Assuming that the parameter is θ, the probability of observing the measurement y_i is $p_\theta(y_i)$ and is called the *likelihood*. First, consider the case of simple hypotheses. For deciding whether the parameter vector is more likely to be θ_1 than θ_0, it is natural to compute the *likelihood ratio* $p_{\theta_1}(y_i)/p_{\theta_0}(y_i)$. For computational purposes, it is also useful to handle the

log-likelihood function: $l_\theta(y_i) \stackrel{\text{def}}{=} \ln p_\theta(y_i)$ and the log-likelihood ratio:

$$s_i \stackrel{\text{def}}{=} \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} = l_{\theta_1}(y_i) - l_{\theta_0}(y_i) \tag{1}$$

The basic feature of s_i, which makes it a good candidate for making the decision, is that

$$\mathbf{E}_{\theta_0}(s_i) < 0, \ \mathbf{E}_{\theta_1}(s_i) > 0 \tag{2}$$

In other words, s_i reflects the change in θ by a change in the sign of its expected value.

The likelihood of n independent observations $y_1, \ldots, y_n$, stacked into a vector $\mathcal{Y}_1^n$, is written as a product: $p_\theta(\mathcal{Y}_1^n) = \prod_{i=1}^n p_\theta(y_i)$. For making the decision between the two simple hypotheses, relevant functions of the data to deal with are the likelihood ratio [18]:

$$\Lambda_n \stackrel{\text{def}}{=} \frac{p_{\theta_1}(\mathcal{Y}_1^n)}{p_{\theta_0}(\mathcal{Y}_1^n)} = \frac{\prod_{i=1}^n p_{\theta_1}(y_i)}{\prod_{i=1}^n p_{\theta_0}(y_i)} \tag{3}$$

and the log-likelihood ratio, which is written as a CUSUM:

$$S_n \stackrel{\text{def}}{=} \ln \Lambda_n = \sum_{i=1}^n s_i \tag{4}$$

A typical behavior of S_n, reflecting the property in equation (2), is displayed in Figure 2 for the simple case of a change in the mean of a Gaussian sequence shown in Figure 1.

When the hypotheses are simple, the test comparing the likelihood ratio to a threshold,

$$\text{If } \Lambda_n < \varrho : \mathbf{H}_0 \text{ is chosen; If } \Lambda_n \geq \varrho : \mathbf{H}_1 \text{ is chosen}$$

or equivalently the log-likelihood ratio,

$$\text{If } S_n < h : \mathbf{H}_0 \text{ is chosen; If } S_n \geq h : \mathbf{H}_1 \text{ is chosen}$$

is optimal (Neyman–Pearson lemma). This means that for a given probability of wrong rejection of $\mathbf{H}_0$ (false alarm), this test minimizes the probability

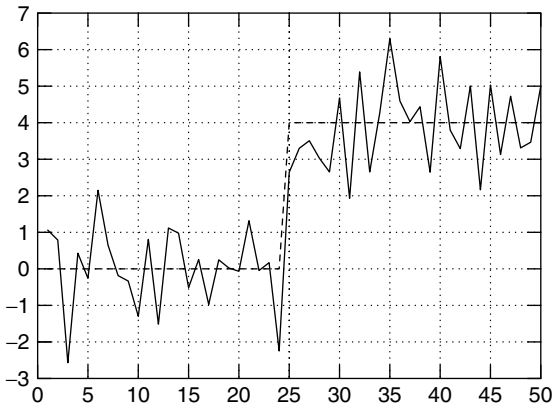


Figure 1. A change in the mean of a Gaussian sequence with constant variance.

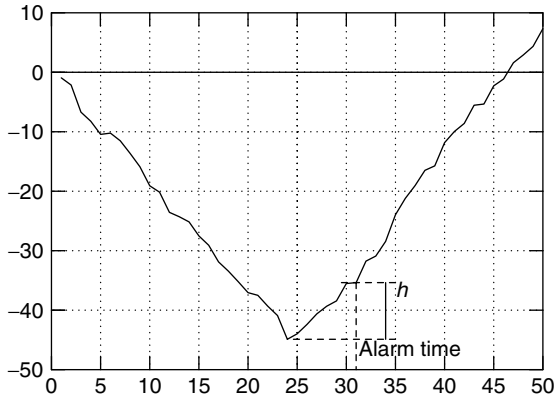


Figure 2. Typical behavior of the log-likelihood ratio S_n: negative drift before and positive drift after the change.

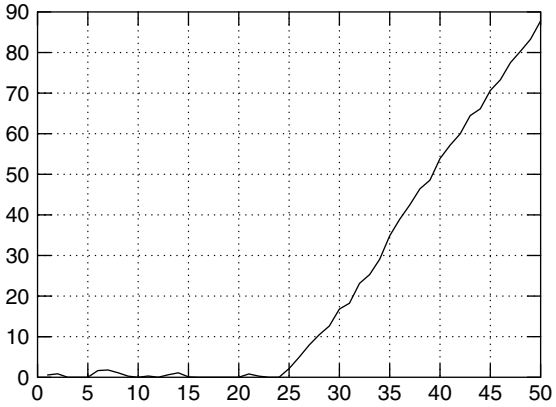


Figure 3. Typical behavior of the CUSUM decision function g_k.

of making a wrong acceptance of $\mathbf{H}_1$. The threshold (ϱ, or h) is determined from a fixed probability of false alarm.

When the hypotheses are composite, the decision is based on the generalized likelihood ratio (GLR), which consists in maximizing the likelihood functions with respect to the unknown parameter:

$$\widehat{\Lambda}_n = \frac{\sup_{\theta_1 \in \Theta_1} p_{\theta_1}(\mathcal{Y}_1^n)}{\sup_{\theta_0 \in \Theta_0} p_{\theta_0}(\mathcal{Y}_1^n)} \tag{5}$$

It is more tricky to precisely state in what ways this test is optimal.

2.1.2 Dependent observations

Assuming that the parameter is θ, one should use the *conditional likelihood* of y_i given the previous data $y_1, \ldots, y_{i-1}$, denoted by $p_\theta(y_i|\mathcal{Y}_1^{i-1})$ with the convention: for $i = 1$, $p_\theta(y_i|\mathcal{Y}_1^{i-1}) \stackrel{\text{def}}{=} p_\theta(y_1)$. The log-likelihood ratio is written as

$$s_i \stackrel{\text{def}}{=} \ln \frac{p_{\theta_1}(y_i|\mathcal{Y}_1^{i-1})}{p_{\theta_0}(y_i|\mathcal{Y}_1^{i-1})} \tag{6}$$

and has the same basic property as in equation (2). If n observations are recorded, the decision between two simple hypotheses is to be made with the aid of the likelihood ratio:

$$\Lambda_n \stackrel{\text{def}}{=} \frac{p_{\theta_1}(\mathcal{Y}_1^n)}{p_{\theta_0}(\mathcal{Y}_1^n)} = \frac{\prod_{i=1}^{n} p_{\theta_1}(y_i|\mathcal{Y}_1^{i-1})}{\prod_{i=1}^{n} p_{\theta_0}(y_i|\mathcal{Y}_1^{i-1})} \tag{7}$$

since, because of the Bayes rule, the joint likelihood is the product of conditional likelihoods:

$$p_{\theta_\ell}(\mathcal{Y}_1^n) = \prod_{i=1}^{n} p_{\theta_\ell}(y_i|\mathcal{Y}_1^{i-1}) \quad (\ell = 0, 1) \tag{8}$$

When using the log-likelihood ratio, the CUSUM property still holds true:

$$S_n \stackrel{\text{def}}{=} \ln \Lambda_n = \sum_{i=1}^{n} s_i \tag{9}$$

In case of composite hypotheses, the GLR test should involve the conditional probabilities as well.

2.2 On-line change detection

At the current time instant k, it is now assumed that there exists an unknown change time ν at which the parameter θ of the distribution p_θ changes.

The problem is stated as that of deciding between the two hypotheses:

$$\begin{aligned} \mathbf{H}_0 &: \forall\, i, 1 \le i \le k, \quad \theta_i = \theta_0 \\ \mathbf{H}_1 &: \exists\, \nu, 1 \le \nu \le k, \theta_i = \theta_0 \ (i < \nu) \\ &\text{and } \theta_i = \theta_1 \ (i \ge \nu) \end{aligned} \tag{10}$$

At time k, the null hypothesis $\mathbf{H}_0$ enforces observations $y_1, \ldots, y_k$ to have the same distribution p_{θ_0}, and the alternative (change) hypothesis $\mathbf{H}_1$ assumes that there exists an unknown change time ν such that $y_1, \ldots, y_{\nu-1}$ have the distribution p_{θ_0} and $y_\nu, \ldots, y_k$ have the distribution p_{θ_1}. The decision on the presence of a change is made with the aid of a stopping (alarm) time:

$$t_a = \min\{k \ge 1 : g_k \ge h\} \tag{11}$$

The parameter θ_0 is assumed to be known, up to an (on-line) estimation. The problem is to design the decision function g_k, namely, the function of the observations with which the decision will be made, to tune the threshold h, and to design a change time estimate $\widehat{\nu}_k$.

2.2.1 Independent observations

In the case of simple hypotheses, namely when both θ_0 and θ_1 are known, the key detection tool is, again, the likelihood ratio between the two hypotheses in equation (10):

$$\frac{\prod_{i=1}^{\nu-1} p_{\theta_0}(y_i) \cdot \prod_{i=\nu}^{k} p_{\theta_1}(y_i)}{\prod_{i=1}^{k} p_{\theta_0}(y_i)} = \frac{\prod_{i=\nu}^{k} p_{\theta_1}(y_i)}{\prod_{i=\nu}^{k} p_{\theta_0}(y_i)} \stackrel{\text{def}}{=} \Lambda_\nu^k \tag{12}$$

The change onset time is estimated by maximizing the likelihood under $\mathbf{H}_1$, that is,

$$\begin{aligned} \widehat{\nu}_k &\stackrel{\text{def}}{=} \arg \max_{1 \le j \le k} \prod_{i=1}^{j-1} p_{\theta_0}(y_i) \cdot \prod_{i=j}^{k} p_{\theta_1}(y_i) \\ &= \arg \max_{1 \le j \le k} \Lambda_j^k \end{aligned} \tag{13}$$

where

$$\Lambda_j^k = \frac{\prod_{i=j}^{k} p_{\theta_1}(y_i)}{\prod_{i=j}^{k} p_{\theta_0}(y_i)} \tag{14}$$

Equation (13) holds because the denominator in the left-hand side of equation (12) does not depend on ν. Let

$$S_j^k \stackrel{\text{def}}{=} \ln \Lambda_j^k \tag{15}$$

Then, because ln(.) is an increasing function, the estimated change time can be written as

$$\widehat{\nu}_k = \arg \max_{1 \le j \le k} S_j^k \tag{16}$$

Again, because of the products in equation (14), S_j^k defined in equation (15) and can be written as a *CUSUM*:

$$S_j^k = \sum_{i=j}^{k} \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} \tag{17}$$

As a decision function g_k in equation (11), consider the following one, based on S_j^k

$$g_k \stackrel{\text{def}}{=} \max_{1 \le j \le k} S_j^k \tag{18}$$

Thanks to equation (15) and equation (16), this decision function can be also written as

$$g_k = \ln \Lambda_{\widehat{\nu}_k}^k \tag{19}$$

This is known as the *CUSUM algorithm* [19], first investigated by Page in the 1950s [20]. A typical behavior of g_k is displayed in Figure 3 for the simple case of a change in the mean of a Gaussian sequence shown in Figure 1.

The CUSUM algorithm can be interpreted as a repeated sequential probability ratio test (SPRT). It is optimal in the sense that it minimizes the mean delay to detection for a given probability between false alarms. Different definitions of the delay have been considered in the literature, and different optimality results have been proven, where some are asymptotic (infinite number of observations) [21] and some are not [22]. See also [2, 23, 24].

In the case of a composite alternative hypothesis $\mathbf{H}_1 : \theta_1 \in \Theta_1$, modified CUSUM algorithms can be designed, based either on a notion of minimum magnitude of change, or on a weighting of the likelihood with respect to the unknown values of θ_1. The GLR methodology can be used as well, resulting in a double maximization, with respect to the change time as above, and with respect to the change magnitude. This is further investigated in Section 3 for a simple case.

2.2.2 Dependent observations

Similar approaches can be used, replacing likelihoods $p_{\theta_\ell}(y_i)$ by conditional likelihoods $p_{\theta_\ell}(y_i|\mathcal{Y}_1^{i-1})$ for $\ell = 0, 1$. Again, the CUSUM property (equation 17) holds:

$$S_j^k = \sum_{i=j}^{k} \ln \frac{p_{\theta_1}(y_i|\mathcal{Y}_1^{i-1})}{p_{\theta_0}(y_i|\mathcal{Y}_1^{i-1})} \tag{20}$$

This is investigated in Section 5 for detecting changes in the spectrum of a signal.

A major issue is the computational complexity of the resulting algorithms in more sophisticated cases. Another key difficulty is the performance analysis of those algorithms. These are the main motivations for the approach described in Section 6.

2.2.3 Adaptive thresholds and windows

The decision function g_k in equation (18) can be easily written as the deviation of the CUSUM S_1^k with respect to its minimum value: $g_k = S_1^k - m_k$, with $m_k \stackrel{\text{def}}{=} \min_{1 \le j \le k} S_1^j$. The stopping rule, when written accordingly, namely,

$$t_a = \min \{k \ge 1 : S_1^k \ge m_k + h\} \tag{21}$$

appears to apply an *adaptive threshold* to the sum S_1^k. Note that the adaptation of the threshold involves m_k, and is thus based on the *entire past* of the sequence of observations.

The CUSUM algorithm also involves a *random size moving window* since g_k can also be written as

$$g_k = \left(S_{k-N_k+1}^k\right)^+ \tag{22}$$

with

$$N_k \stackrel{\text{def}}{=} N_{k-1} \cdot \mathbf{1}_{\{g_{k-1}>0\}} + 1 \tag{23}$$

3 ON-LINE DETECTION OF CHANGES IN THE MEAN

Consider now the problem addressed in Section 2 in the simple case where θ is the mean value of the distribution $p_\theta(y)$ assumed to be Gaussian, namely,

$$p_\theta(y) = 1/(\sigma\sqrt{2\pi})\mathrm{e}^{(y-\theta)^2/2\sigma^2} \tag{24}$$

Assuming that θ_1 is known after the change, the CUSUM S_j^k reduces to a data *integrator*:

$$S_j^k = \frac{\upsilon}{\sigma^2} \sum_{i=j}^{k} \left(y_i - \theta_0 - \frac{\upsilon}{2} \right) \stackrel{\text{def}}{=} S_j^k(\upsilon) \tag{25}$$

In equation (25), $\upsilon \stackrel{\text{def}}{=} \theta_1 - \theta_0$ is the *change magnitude* and $b \stackrel{\text{def}}{=} \upsilon/\sigma$ is the *change-to-noise ratio*, which is the relevant counterpart of the usual signal-to-noise ratio (SNR).

Because of equations (22) and 23), the CUSUM algorithm (11), (18), and (25) in this case can be seen as an *integration* of the observations *over a sliding window with adaptive size*. This algorithm is especially efficient for *small* change-to-noise ratio b [4].

When θ_1 is unknown and assumed to belong to a known set, $\theta_1 \in \Theta_1 \subseteq \mathbf{R}$, several approaches can be used. A weighting of the likelihood with respect to the unknown θ_1 may be introduced when some *a priori* information about the distribution of θ_1 is available [24, 25]. Another approach consists in introducing a minimum change magnitude υ_m in equation (25), and in running two CUSUM tests in parallel, since the change direction, namely increase or decrease in the mean, is not known either [4, Chapter 2]. This is summarized as

Decreasing mean

$$T_1^k \stackrel{\text{def}}{=} \sum_{i=1}^{k} \left(y_i - \theta_0 + \frac{\upsilon_m}{2} \right)$$
$$M_k \stackrel{\text{def}}{=} \max_{1\le j\le k} T_1^j$$
$$t_a = \min\{k \ge 1 : M_k - T_1^k \ge h\}$$

Increasing mean

$$S_1^k \stackrel{\text{def}}{=} \sum_{i=1}^{k} \left(y_i - \theta_0 - \frac{\upsilon_m}{2} \right)$$
$$m_k \stackrel{\text{def}}{=} \min_{1\le j\le k} S_1^j$$
$$t_a = \min\{k \ge 1 : S_1^k - m_k \ge h\}$$

Tuning two parameters υ_m and h might seem more tricky than tuning a single threshold. However, the two choices turn out to be easy since both parameters are well decoupled in practice. Because of equation (25), the threshold h should be of the form $\tilde{h}\sigma^2$. In case of unknown σ, involving a good estimate $\hat{\sigma}^2$ in the threshold may contribute to the performances of the algorithm. This approach has proven useful in many application cases.

The GLR approach maximizes the likelihood ratio with respect to the two unknown parameters (change time *and* magnitude):

$$g_k \stackrel{\text{def}}{=} \max_{1\le j\le k} \sup_{\upsilon} S_j^k(\upsilon) \tag{26}$$

The second maximization is explicit in the present Gaussian case, which results in a *quadratic* test:

$$g_k = \max_{1\le j\le k} \frac{1}{2(k-j+1)} \left(\frac{\sum_{i=j}^{k} (y_i - \theta_0)}{\sigma} \right)^2 \tag{27}$$

4 EXTENSION TO ADDITIVE JUMPS

These algorithms extend to the multidimensional case.

4.1 Changes in a vector mean

First consider the problem of detecting an additive change in a sequence of independent *vector* Gaussian variables $Y_1, \ldots, Y_k$, with mean $M\Upsilon$ and covariance Σ. The hypotheses can now be written as follows:

$$\begin{aligned} &\mathbf{H}_0 \ : \ \forall\, i, 1 \le i \le k, \Upsilon_i = 0 \\ &\mathbf{H}_1 \ : \ \exists\, \nu, 1 \le \nu \le k, \Upsilon_i = 0 \ (i < \nu) \\ &\quad \text{and } \Upsilon_i = \Upsilon \ (i \ge \nu) \end{aligned} \tag{28}$$

The GLR test for detecting such a deviation in Υ is written as

$$\begin{aligned} g_k &\stackrel{\text{def}}{=} \max_{1\le j\le k} \sup_{\Upsilon} S_j^k(\Upsilon) \\ &= \max_{1\le j\le k} \frac{k-j+1}{2} \end{aligned} \tag{29}$$

$$\overline{Y}_j^{k\,\mathrm{T}}\ \Sigma^{-1}\ M(M^{\mathrm{T}}\ \Sigma^{-1}\ M)^{-1}M^{\mathrm{T}}\ \Sigma^{-1}\ \overline{Y}_j^{k} \quad (30)$$

where $\overline{Y}_j^{k} \stackrel{\mathrm{def}}{=} \sum_{i=j}^{k} Y_i/(k-j+1)$. The test, equation (30), is the multidimensional counterpart of equation (27).

4.2 Additive jumps in state-space models

More generally, consider the case of additive jumps on linear Gaussian state-space systems:

$$\begin{cases} X_{k+1} = F\ X_k + G\ U_k + W_k + \Gamma\ \Upsilon_x(k,\nu) \\ Y_k = H\ X_k + J\ U_k + V_k + \Xi\ \Upsilon_y(k,\nu) \end{cases} \quad (31)$$

where X is the unknown state; U, Y are the measured input and output vectors, with dimensions s, m, r, respectively; $(W_k)_k$ and $(V_k)_k$ are independent Gaussian white noise sequences, with covariance matrices Q and R, respectively; Υ_x, Υ_y are the fault vectors, with compatible dimensions; and ν is the jump onset time. Sensor and actuator faults are important instances of equation (31).

As shown in [4, Chapter 7] and [26], this problem reduces to detecting a change in the mean vector of the *innovation* ε_k of system (31), computed with a Kalman filter, namely,

$$\begin{cases} \hat{X}_{k+1|k} = F\hat{X}_{k|k} + G\ U_k \\ \hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k\ \varepsilon_k \\ \varepsilon_k = Y_k - H\ \hat{X}_{k|k-1} - J\ U_k \end{cases} \quad (32)$$

where the Kalman gain K_k is computed through

$$\begin{aligned} K_k &= P_{k|k-1}\ H^{\mathrm{T}}\ \Sigma_k^{-1} \\ P_{k+1|k} &= F\ P_{k|k}\ F^{\mathrm{T}} + Q \\ P_{k|k} &= (I_s - K_k\ H)\ P_{k|k-1} \end{aligned} \quad (33)$$

Under the hypothesis $\mathbf{H}_0$ $(k < \nu)$, the innovation ε_k is Gaussian with mean zero and covariance matrix $\Sigma_k = HP_{k|k-1}H^{\mathrm{T}} + R$. In the presence of a change, namely under $\mathbf{H}_1$, ε_k is Gaussian with mean vector $\rho(k,\nu)$ and covariance Σ_k. Actually, because of the assumption of additive changes in equation (31), the signature of the change on the innovation is also additive, and $\rho(k,\nu)$ can be computed recursively. Assuming a steady-state Kalman filter, a closed-form expression of $\rho(k,\nu)$ can be derived. Using the transfer function notation, $\rho(k,t_0) = \mathcal{K}_x(z)\Upsilon_x(k,t_0) + \mathcal{K}_y(z)\Upsilon_y(k,t_0)$, straightforward computations lead, asymptotically for large k, to

$$\begin{aligned} \mathcal{K}_x(z) &= H(zI_s - \bar{F})^{-1}\Gamma \\ \mathcal{K}_y(z) &= \left[I_r - H(zI_s - \bar{F})^{-1}FK\right]\Xi \\ \bar{F} &\stackrel{\mathrm{def}}{=} F(I_s - KH) \end{aligned} \quad (34)$$

When using the GLR methodology for detecting the change in the mean vector of ε_k, the double maximization with respect to the change time *and* "magnitude" in equation (29) reduces to a single maximization, since the maximization over the fault vector Υ is again explicit:

$$\begin{aligned} \sup_{\Upsilon} S_j^k = &\left(\sum_{i=j}^{k} \tilde{\rho}^{\mathrm{T}}(i,j)\Sigma_i^{-1}\varepsilon_i\right)^{\mathrm{T}} \\ &\times \left(\sum_{i=j}^{k} \tilde{\rho}^{\mathrm{T}}(i,j)\Sigma_i^{-1}\tilde{\rho}(i,j)\right)^{-1} \\ &\times \left(\sum_{i=j}^{k} \tilde{\rho}^{\mathrm{T}}(i,j)\Sigma_i^{-1}\varepsilon_i\right) \end{aligned} \quad (35)$$

(compare with equation 30). An estimate of the change onset time is provided by $\hat{\nu}_k = \arg\max_{k-\ell+1\le j\le k} S_j^k$. The information contained in the change time and magnitude estimates can be used for updating the Kalman filter after detection [4, Chapter 7] and [26].

5 CHANGES IN THE SPECTRUM

For detecting changes in the spectrum of a scalar signal, an intuitive approach is discussed and the application of the CUSUM and GLR methods is presented.

5.1 Monitoring one innovation

Assuming an AR (autoregressive) model, $y_k = \sum_{l=1}^{p} a_l\ y_{k-l} + \mathrm{e}_k$, the problem is to detect changes in the

parameter vector $\theta^{\mathrm{T}} \stackrel{\mathrm{def}}{=} (a_1 \ldots a_p)$. A natural idea is to compute the innovation:

$$\varepsilon_k(\theta) \stackrel{\mathrm{def}}{=} y_k - \sum_{l=1}^{p} \hat{a}_l \; y_{k-l} \tag{36}$$

and compare it with a threshold. From a statistical inference point of view, however, this type of residual suffers from two limitations. First, whenever $p > 1$, the dimension of ε in equation (36) is *smaller* than the dimension of the parameter vector.[a] This means that the innovation ε lives in a smaller dimensional space than the parameter θ, and thus might be unable to discriminate between, if not insensitive to, different changes in that vector. Stated otherwise, a statistics (function of the parameter vector and of the raw data) should have at least the same dimension as the parameter to be inferred and monitored.

Second, ε is a *linear* combination of the measured data. Stated otherwise, residual (equation 36) is a *first-order* statistic. When the changes of interest affect the dynamics (here the AR coefficients a_i's), using linear combinations of the data is *not* sufficient in the statistical sense: a nonnegligible set of spectral changes result in *no* change in either the mean or the variance of the innovation [27]. A more complex expression in ε, such as

$$\sum_i \left(\frac{\varepsilon_i^2(\theta)}{\sigma^2} - 1 \right) \tag{37}$$

shares the same type of limitations: the statistics in equation (37) is known to be poorly efficient in a number of cases [4, Chapter 8] and [27].

This does not mean, however, that the innovation should not play any role in detecting changes in spectra. Indeed, how to use the prediction error for parameter estimation is clearly stated in the system identification literature [28]: a parameter estimate should be updated with the aid of the *gradient* of the squared prediction error with respect to the parameter:

$$-\frac{1}{2} \; \partial / \partial \theta \left(\varepsilon_k^{\mathrm{T}}(\theta) \; \varepsilon_k(\theta) \right) \tag{38}$$

Note that the dimension of that gradient satisfies the above requirement. If the changes affect the dynamics of the system, a residual built on the vector in equation (38) is relevant. Note also that in the AR case it involves *second-order* statistics, namely covariances, which is mandatory.

More generally, a residual for detecting (parametric) changes in the dynamics of the system should build on a parameter *estimating function*. This is further described in Sections 6 and 7.

5.2 Monitoring two innovations

For detecting changes in the coefficients a_i's of an AR model, the CUSUM and GLR approaches can be used. Again, since the data are *dependent observations*, the computations should involve the conditional likelihoods $p_{\theta_\ell}(y_i | y_{i-1}, \ldots, y_{i-p})(\ell = 0, 1)$.

In the present scalar AR case, the CUSUM algorithm (11), (18), (20) runs with

$$S_j^k = \sum_{i=j}^{k} \frac{\varepsilon_i^2(\theta_0) - \varepsilon_i^2(\theta_1)}{2\sigma^2} \tag{39}$$

which involves *two innovations*, and not only one. The implementation in the (actual) case of unknown θ_0 and θ_1 consists in plugging in equation (39) estimates of $\widehat{\theta_0}$ and $\widehat{\theta_1}$ computed within a growing window and a fixed-size sliding window, respectively. Variations on this theme have resulted in a CUSUM test based on a function of two innovations different from equation (39) (but equivalent for small deviations) [27]. This algorithm has been shown to be useful for detecting small spectral changes.

This approach extends to scalar autoregressive moving average (ARMA) processes, although the likelihood and innovation computations are more involved in that case because of the MA part [4, Chapter 8].

5.3 Parameterized spectral distance measures

Comparing with a threshold the spectral distance between two models estimated on different data windows is a natural approach to detecting spectral changes. Some care should be taken for selecting a relevant spectral distance measure [29–31].

Of common use in speech processing, the family of log-spectral deviations between two spectral densities

s_0 and s_1 can be written as follows:

$$d_q(s_0, s_1) \stackrel{\text{def}}{=} \| \ln s_0 - \ln s_1 \|_q = \left\| \ln \frac{s_0}{s_1} \right\|_q \quad (40)$$

where, for a spectral density s, the q-norm is defined by

$$\|s\|_q \stackrel{\text{def}}{=} \left(\int_{-\pi}^{\pi} \frac{1}{2\pi} |s(\omega)|^q \, \mathrm{d}\omega \right)^{1/q} \quad (41)$$

Back to the linear prediction framework of the AR process above, namely, for s of the form

$$\frac{1}{s(\omega)} \stackrel{\text{def}}{=} \frac{|A(\mathrm{e}^{j\omega})|^2}{\sigma^2} \quad (42)$$

where $A(z) \stackrel{\text{def}}{=} \sum_{k=0}^{p} a_k \, z^{-k}$, the *cepstral coefficients* $(c_k)_k$ are defined as $\ln A(z) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} c_k \, z^{-k}$ and satisfy

$$\ln \frac{\sigma^2}{|A(\mathrm{e}^{j\omega})|^2} = \sum_{k=-\infty}^{\infty} c_k \, \mathrm{e}^{-jk\omega},$$
$$c_0 \stackrel{\text{def}}{=} \ln \sigma^2, \qquad c_{-k} \stackrel{\text{def}}{=} c_k \quad (43)$$

The *cepstral distance*, defined as *the mean quadratic distance* (d_q in equation (40) with $q = 2$), can be written as an Euclidean distance between the two collections of cepstral coefficients:

$$d_2^2(s_0, s_1) = \sum_{k=-\infty}^{\infty} \left(c_k^{(0)} - c_k^{(1)} \right)^2 \quad (44)$$

of which the finite sum

$$d^2(L) \stackrel{\text{def}}{=} \sum_{k=-L}^{L} \left(c_k^{(0)} - c_k^{(1)} \right)^2 \quad (45)$$

is a good approximation. It is important to note that equation (45) is *not* an Euclidean distance between the two collections of AR coefficients a_k. When written in terms of the poles z_i, equation (45) becomes

$$d^2(L) = \sum_{k=-L}^{L} \frac{1}{k^2} \left(\sum_{i=1}^{p} \left(z_i^{(1)\,n} - z_i^{(0)\,n} \right) \right)^2 \quad (46)$$

It is very important to realize that equation (45) is thus *not* an Euclidean distance between the two collections of poles z_i. Similarly it is *not* an Euclidean distance between the two collections of Fourier spectrum lines.

6 DETECTING SMALL CHANGES

Consider now more complex situations involving nonadditive changes in multidimensional dependent data (signals). For overcoming both computational and performance analysis issues associated with the likelihood ratio approach, a possible approach consists in assuming *small changes* and using basically a sensitivity analysis relative to various noise and uncertainty levels. This approach is known as the *statistical local approach*.

The capability of performing early detection and isolation of small deviations of a structure or a process, with respect to a reference behavior considered as normal, is most useful in condition-based and aging monitoring problems. This is why this approach is of interest for SHM.

The local approach is first introduced in the framework of the likelihood function, and then for other estimating functions. Indeed, some important monitoring problems cannot be addressed with the aid of the likelihood. This is the case of vibration-based SHM addressed in Section 7.

6.1 Likelihood: efficient score

A key concept turns out to be very useful for designing and analyzing hypotheses testing and change detection algorithms, under the assumption of a small change. This is the sensitivity of the log-likelihood function with respect to the parameter vector:

$$z_i^* \stackrel{\text{def}}{=} \left. \frac{\partial l_\theta(y_i | \mathcal{Y}_1^{i-1})}{\partial \theta} \right|_{\theta=\theta^*} \quad (47)$$

also called *efficient score*. It enjoys a basic sign change property:

$$\mathbf{E}_{\theta_0}(z_i^*) < 0 \quad \text{for } \theta_0 = \theta^* - \Upsilon,$$
$$\mathbf{E}_{\theta_1}(z_i^*) > 0 \quad \text{for } \theta_1 = \theta^* + \Upsilon \quad (48)$$

which again makes it a good candidate for making the decision about θ.

If n observations are recorded, the efficient score can be written as follows:

$$\mathcal{Z}_n(\theta^*) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \left.\frac{\partial \ln p_\theta(\mathcal{Y}_1^n)}{\partial \theta}\right|_{\theta=\theta^*} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} z_i^* \tag{49}$$

Again, the CUSUM property holds. The efficient score is characterized by the key identity:

$$\mathbf{E}_\theta(\mathcal{Z}_n(\theta^*)) = 0 \quad \Longleftrightarrow \quad \theta = \theta^* \tag{50}$$

The expectation of the score is zero if, and only if, the expectation is computed at the very point where the derivative is computed. The efficient score is thus the ML (maximum likelihood) parameter estimating function. *The efficient score is generically different from the innovation*.

The efficient score enjoys the following asymptotic property. Consider the hypotheses:

$$\begin{aligned} &\text{(Safe) } \mathbf{H}_0 : \theta = \theta_0 \\ &\text{(Damaged) } \mathbf{H}_1 : \theta = \theta_0 + \Upsilon/\sqrt{n} \end{aligned} \tag{51}$$

where vector Υ is unknown, but fixed. For large n, hypothesis $\mathbf{H}_1$ corresponds to *small* deviations.

Thanks to that assumption, the efficient score $\mathcal{Z}_n(\theta_0)$ behaves asymptotically as a Gaussian vector with the same covariance matrix under both hypotheses:

$$\mathcal{Z}_n(\theta_0) \rightarrow \begin{cases} \mathcal{N}(0, \mathbf{I}(\theta_0)) & \text{under} \quad \mathbf{P}_{\theta_0} \\ \mathcal{N}(\mathbf{I}(\theta_0)\ \Upsilon\theta, \mathbf{I}(\theta_0)) & \text{under} \quad \mathbf{P}_{\theta_0+\frac{\Upsilon}{\sqrt{n}}} \end{cases} \tag{52}$$

where $\mathbf{I}(\theta_0)$ is the covariance matrix of the efficient score:

$$\begin{aligned} \mathbf{I}(\theta_0) &\stackrel{\text{def}}{=} \lim_{n\to\infty} \mathbf{I}_n(\theta_0), \\ \mathbf{I}_n(\theta_0) &\stackrel{\text{def}}{=} \mathbf{E}_{\theta_0}\ \mathcal{Z}_n(\theta_0)\ \mathcal{Z}_n^{\mathrm{T}}(\theta_0) \\ &= -\frac{1}{n}\ \mathbf{E}_{\theta_0}\ \left.\frac{\partial^2 \ln p_\theta(\mathcal{Y}_1^n)}{\partial \theta^2}\right|_{\theta=\theta_0} \end{aligned} \tag{53}$$

and is known as *the Fisher information matrix*. The efficient score reflects the small change in θ as a change in its own mean value: $\mathcal{Z}_n$ has zero mean under $\mathbf{H}_0$, and mean $\mathbf{I}(\theta_0)\Upsilon$ under $\mathbf{H}_1$. Note that matrices $\mathbf{I}(\theta_0)$ and $\Sigma(\theta_0)$ depend on neither the sample size n nor the change vector Υ in the hypothesis $\mathbf{H}_1$. Thus the detection of a change in θ reduces to the detection of a change in the mean of the (asymptotically) Gaussian vector $\mathcal{Z}_n$. Exactly as in equation (30), the test is *quadratic* in $\mathcal{Z}_n$:

$$\mathcal{Z}_n^{\mathrm{T}}(\theta_0)\ \mathbf{I}^{-1}(\theta_0)\ \mathcal{Z}_n(\theta_0) \geq \varrho \tag{54}$$

This test is distributed as a χ^2-variable, with $\mathrm{rank}(\mathbf{I}(\theta_0))$ degrees of freedom (dof). From this, a threshold can be deduced, for a given false alarm probability. The noncentrality parameter under $\mathbf{H}_1$ is $\Upsilon^{\mathrm{T}}\ \mathbf{I}(\theta_0)\ \Upsilon$. This test is asymptotically optimum for testing the hypothesis of a small change [4].

6.2 Other estimating functions

The efficient score may not be the relevant function to base the decision on. Actually, there are many identification algorithms that do not build on the likelihood function, but on other parameter estimating functions instead [28, 32]. For such functions, a similar asymptotic Gaussianity result holds. Let $K(\theta, Y_k)$ be a function of the parameter vector and of the data, enjoying the property:

$$\mathbf{E}_{\theta_0}\ K(\theta, Y_k) = 0 \quad \Longleftrightarrow \quad \theta = \theta_0 \tag{55}$$

which makes it a good candidate for being the basis of a parameter estimation algorithm. Define

$$\zeta_n(\theta_0) = \frac{1}{\sqrt{n}}\ \sum_{k=1}^{n} K(\theta_0, Y_k) \tag{56}$$

which can be thought of as a generalized score. Also define the sensitivity and covariance matrices:

$$\begin{aligned} M(\theta_0) &\stackrel{\text{def}}{=} -\mathbf{E}_{\theta_0}\ \left.\frac{\partial K(\theta, Y_k)}{\partial \theta}\right|_{\theta=\theta_0}, \\ \Sigma(\theta_0) &\stackrel{\text{def}}{=} \lim_{n\to\infty} \mathbf{E}_{\theta_0}\ \zeta_n(\theta_0)\ \zeta_n^{\mathrm{T}}(\theta_0) \end{aligned} \tag{57}$$

Then, the asymptotic Gaussianity result holds for this residual [4, 7, 33–35]:

$$\zeta_n(\theta_0) \rightarrow \begin{cases} \mathcal{N}(0, \Sigma(\theta_0)) & \text{under} \quad \mathbf{P}_{\theta_0} \\ \mathcal{N}(M(\theta_0)\ \delta\theta, \Sigma(\theta_0)) & \text{under} \quad \mathbf{P}_{\theta_0+\frac{\Upsilon}{\sqrt{n}}} \end{cases} \tag{58}$$

This asymptotic property holds for a large class of (not necessarily Gaussian) random processes. Thus,

as the efficient score does, the residual reflects the small change in the parameter vector as a change in its own mean value. Hence the χ^2 test:

$$\chi_n^2 \stackrel{\text{def}}{=} \zeta_n^{\mathrm{T}}\, \Sigma^{-1}\, M\, \left(M^{\mathrm{T}}\, \Sigma^{-1}\, M\right)^{-1}\, M^{\mathrm{T}}\, \Sigma^{-1}\, \zeta_n \;\geq \varrho \tag{59}$$

where the dependence on θ_0 has been removed for simplicity.

This test is distributed as a χ^2 variable, with rank(M) dof and noncentrality parameter $\Upsilon^{\mathrm{T}} M^{\mathrm{T}} \Sigma^{-1} M\, \Upsilon$. Let $\mathbf{F}(\theta_0)$ be the asymptotic Fisher information on θ_0 contained in $\zeta_n(\theta_0)$:

$$\mathbf{F}(\theta_0) \stackrel{\text{def}}{=} M^{\mathrm{T}}(\theta_0)\; \Sigma^{-1}(\theta_0)\; M(\theta_0) \tag{60}$$

Note that if the estimating function K in equations (55) and (56) is the efficient score in equation (49), namely when using a likelihood-based approach, both matrices M and Σ in equation (57) equal the Fisher matrix $\mathbf{I}$ in equation (53), and the χ^2 test (equation 59) reduces to the test in equation (54).

7 CHANGES IN THE SYSTEM DYNAMICS AND VIBRATION-BASED SHM

Consider the multidimensional counterpart of the problem of detecting changes in a spectrum addressed in Section 5. This problem can be stated as that of detecting changes in the AR part of a multidimensional ARMA model, and is highly related to vibration-based SHM. More than the computational complexity of the likelihood function for vector ARMA processes, a tight coupling between the AR part (basically the structure) and the MA part (basically the excitation) prevents the use of a likelihood-based approach when the unknown excitation is nonstationary.

Another solution resorts to the equivalent representation of any vector ARMA model as a linear dynamical system [36]. The problem is then to detect changes in the pair (H, F) of

$$\begin{cases} X_{k+1} = F\; X_k + W_{k+1} \\ Y_k = H\; X_k \end{cases} \tag{61}$$

An important instance of this problem is the detection of changes in the observed eigenstructure of state-transition matrix F, namely, the $(\lambda, \varphi_\lambda)$'s, where $\varphi_\lambda \stackrel{\text{def}}{=} H\phi_\lambda$, and ϕ_λ is the eigenvector of F associated with the eigenvalue λ. It is well known that this is one relevant formulation of the vibration-based SHM problem [37, 38]. By stacking the modes and mode shapes $(\lambda, \varphi_\lambda)$'s into a vector parameter θ, damage detection reduces to the problem of detecting changes in θ. Assuming that a reference value θ_0 is available, identified on data recorded on the undamaged system, and given a new data sample, the problem is to decide whether or not this sample is still well described by θ_0.

A possible solution consists in using an estimating function, different from the likelihood, within the framework of small changes described in Section 6. The estimating function associated with the covariance-driven subspace identification algorithm is a good candidate, which exploits the following property. Let $R_i \stackrel{\text{def}}{=} \mathbf{E}(Y_k\; Y_{k-i}^{\mathrm{T}})$ be the theoretical covariance matrix for lag i, and

$$\mathcal{H}_{p+1,q} \stackrel{\text{def}}{=} \begin{pmatrix} R_0 & R_1 & \vdots & R_{q-1} \\ R_1 & R_2 & \vdots & R_q \\ \vdots & \vdots & \vdots & \vdots \\ R_p & R_{p+1} & \vdots & R_{p+q-1} \end{pmatrix} \tag{62}$$

be the theoretical output covariance and Hankel matrices, respectively. Introducing the cross-covariance between the state and the observed outputs, $G \stackrel{\text{def}}{=} \mathbf{E}(X_k Y_k^{\mathrm{T}})$, direct computations of the R_i's from (61) lead to $R_i = H F^i G$ and to the well-known factorization [39]:

$$\mathcal{H}_{p+1,q} = \mathcal{O}_{p+1}(H, F)\; \mathcal{C}_q(F, G) \tag{63}$$

where

$$\mathcal{O}_{p+1}(H, F) \stackrel{\text{def}}{=} \begin{pmatrix} H \\ HF \\ \vdots \\ HF^p \end{pmatrix}$$

$$\text{and} \qquad \mathcal{C}_q(F, G) \stackrel{\text{def}}{=} (G\; FG\; \cdots\; F^{q-1}G) \tag{64}$$

are the observability and controllability matrices, respectively. Note that the observed dynamics (H, F)

is found in the left-hand side term only. For estimating the reference θ_0, the covariance-driven subspace identification algorithm performs the singular value decomposition (SVD) of the empirical Hankel matrix $\widehat{\mathcal{H}}^0_{p+1,m}$ computed with data recorded on the undamaged system. Let θ_0 be the parameter vector resulting from this identification.

The subspace interpretation of the SVD provides the following characterization for θ_0:

$$N(\theta_0)^{\mathrm{T}}\ \widehat{\mathcal{H}}^0_{p+1,m} = 0 \tag{65}$$

where the orthonormal matrix N is subject to

$$N(\theta_0)^{\mathrm{T}} N(\theta_0) = I,\ \ N(\theta_0)^{\mathrm{T}} \mathcal{O}_{p+1}(\theta_0) = 0 \tag{66}$$

and matrix $\mathcal{O}_{p+1}(\theta_0)$ can be computed from θ_0. Matrix N depends implicitly on parameter θ and is not unique. However, it can be treated as a function of parameter θ, denoted by $N(\theta)$ [40].

Let $\widehat{\mathcal{H}}^1_{p+1,m}$ be the empirical Hankel matrix filled with the covariances of the *new measurements* collected on the (possibly damaged) system, and let n be the length of this new data set. Inspired by the left-hand side of equation (65), it is proposed to base the decision on the following residual in [40, 41]:

$$\zeta_n(\theta_0) \stackrel{\text{def}}{=} \mathrm{vec}\big(N(\theta_0)^{\mathrm{T}}\ \widehat{\mathcal{H}}^1_{p+1,m}\big) \tag{67}$$

where vec is the column stacking operator.

Thanks to the local approach in Section 6.2, vector $\zeta_n(\theta_0)$ in equation (67) has the asymptotic Gaussian behavior summarized in (58) [40]. A change in θ, namely in the system dynamics, is reflected as a change in its mean, which switches from zero under $\mathbf{H}_0$, to $M(\theta_0)\Upsilon$ under $\mathbf{H}_1$. Since the matrices $M(\theta_0)$ and $\Sigma(\theta_0)$ depend on neither the sample size n nor the fault vector Υ in the hypothesis $\mathbf{H}_1$, they can be estimated prior to testing, using data on the safe system (as the reference θ_0).

The global test χ^2_n in equation (59) to be used performs a *sensitivity* analysis of the residual to the damages, *relative to* uncertainties in the modal estimates and noises on the available data [41]. It is often necessary to select the threshold for χ^2_n experimentally. How to achieve this from histograms of empirical values obtained on data for undamaged cases is explained in [15]. This damage detection algorithm has been run successfully on a number of test cases [42]. A nonparametric version is described in [43].

In this section, we have considered a batch-wise computation of the residual and an off-line detection problem statement (equation 58). A sample-wise computation of ζ_n, together with a sample-wise CUSUM-like detection algorithm for solving an on-line detection problem as in (equation 28), is described in [8, 13]. Recent and ongoing research includes model-based statistical handling of the environmental effects in vibration-based monitoring of civil engineering structures [43, 44], and of the complex aeroservoelastic effects underlying the flutter phenomenon [8, 45].

8 FURTHER ISSUES

Several other issues relating to the above-mentioned statistical framework can be addressed.

8.1 Damage isolation and diagnostics

In a statistical framework, the fault or damage isolation problem can be addressed from different points of view. The first issue is to discriminate between the components of a Gaussian vector, using nuisance elimination methods. This problem is of wide interest, thanks to the asymptotic Gaussianity result underlying the design of change detection methods presented in Section 6. In practice, focusing the monitoring on some subsets of components of the parameter vector θ is of interest for many purposes. For example, in the vibration-based SHM problem in Section 7, it may be desirable to monitor a specific mode λ and the associated mode-shape φ_λ.

A more involved issue is multiple hypotheses testing, for which the definition of optimality criteria is tricky. A still more involved problem is online isolation: the design of both optimality criteria and optimal algorithms raises several questions and these are still not completely resolved.

Finally, in most SHM applications, a complex physical system, characterized by a generally nonidentifiable parameter vector Φ, has to be monitored using a simple model characterized by an identifiable parameter vector θ. A possible direct solution is sketched for the corresponding diagnostics problem expressed in terms of Φ.

8.1.1 Isolation as nuisance elimination

The fault or damage vector is now partitioned as follows: $\Upsilon = \begin{pmatrix} \Upsilon_a \\ \Upsilon_b \end{pmatrix}$, with Υ_a, Υ_b of known dimensions l_a, l_b, respectively ($l_a + l_b = l$). The matrices M in equation (58) and $\mathbf{F}$ in (60) are partitioned accordingly:

$$M = (M_a M_b), \mathbf{F} = \begin{pmatrix} \mathbf{F}_{aa} & \mathbf{F}_{ab} \\ \mathbf{F}_{ba} & \mathbf{F}_{bb} \end{pmatrix} = \begin{pmatrix} M_a^T \Sigma^{-1} M_a & M_a^{\mathrm{T}} \Sigma^{-1} M_b \\ M_b^T \Sigma^{-1} M_a & M_b^{\mathrm{T}} \Sigma^{-1} M_b \end{pmatrix} \quad (68)$$

Matrix M is assumed to be full column rank (fcr). Let notation $\mathbf{F}_a^{*-1}$ stand for the upper-left term of $\mathbf{F}^{-1}$, where $\mathbf{F}_a^* = \mathbf{F}_{aa} - \mathbf{F}_{ab}\, \mathbf{F}_{bb}^{-1}\, \mathbf{F}_{ba}$, and notation $p_{\Upsilon_a, \Upsilon_b}(\zeta)$ stands for the density of a Gaussian vector $\zeta \sim \mathcal{N}(M_a \Upsilon_a + M_b \Upsilon_b, \Sigma)$. The problem is to isolate changes either in Υ_a or in Υ_b, with the other component being unknown and considered as a nuisance parameter.

Isolation through projection A rather intuitive statistical solution to the isolation problem, called *sensitivity approach*, assumes Υ_b and consists in projecting the deviations in Υ onto the subspace generated by the components Υ_a to be isolated. The sensitivity test $\widetilde{t}_a$ for monitoring Υ_a is GLR test between $\Upsilon = (0, 0)$ and $\Upsilon = (\Upsilon_a, 0)$, where $\Upsilon_a \neq 0$, namely

$$\widetilde{t}_a \stackrel{\text{def}}{=} 2\, \ln \frac{\max_{\Upsilon_a}\ p_{\Upsilon_a, 0}(\zeta)}{p_{0,0}(\zeta)} = \widetilde{\zeta}_a^{\mathrm{T}}\, \mathbf{F}_{aa}^{-1}\, \widetilde{\zeta}_a \quad (69)$$

where the partial residual

$$\widetilde{\zeta}_a \stackrel{\text{def}}{=} M_a^{\mathrm{T}} \Sigma^{-1} \zeta \quad (70)$$

is nothing but the efficient score with respect to Υ_a. Under both hypotheses, test statistics $\widetilde{t}_a$ is distributed as a χ^2-variable with l_a dof, and noncentrality parameter $\Upsilon_a^{\mathrm{T}} \mathbf{F}_{aa} \Upsilon_a$ under $\Upsilon_a \neq 0$ *if* $\Upsilon_b = 0$ *actually holds*. Note that, for $a \neq b$, $\widetilde{\zeta}_a$ and $\widetilde{\zeta}_b$ are correlated, and $\widetilde{t}_a$ might be sensitive to a change in Υ_b [46]. The decision about the faulty or damaged component Υ_a or Υ_b is then taken according to the largest sensitivity test $\widetilde{t}_a$ or $\widetilde{t}_b$.

Isolation through rejection Another statistical solution to the problem of isolating Υ_a consists in viewing parameter Υ_b as a nuisance, and using an existing method for inferring part of the parameters while ignoring and being robust to the complementary part. This method is called *min–max approach* and consists in replacing the nuisance parameter component Υ_b by its least favorable value, namely the value that minimizes the power of test t, or equivalently the noncentrality parameter. This is equivalent to the GLR method:

$$t_a^* \stackrel{\text{def}}{=} 2\, \ln \frac{\max_{\Upsilon_a, \Upsilon_b}\ p_{\Upsilon_a, \Upsilon_b}(\zeta)}{\max_{\Upsilon_b}\ p_{0, \Upsilon_b}(\zeta)} \quad (71)$$

which consists in replacing the nuisance component Υ_b by its ML estimate. This can be written as follows:

$$t_a^* = \zeta_a^{*\mathrm{T}}\, \mathbf{F}_a^{*-1}\, \zeta_a^* \quad (72)$$

where $\zeta_a^* \stackrel{\text{def}}{=} \widetilde{\zeta}_a - \mathbf{F}_{ab} \mathbf{F}_{bb}^{-1} \widetilde{\zeta}_b$ is called *the effective score*. Note that ζ_a^* is the residual from regression of the partial residual in Υ_a on the (nuisance) partial residual in Υ_b. This is a fairly intuitive statistical approach, which actually traces back to Neyman's work in the 1950s [47]. Under both hypotheses, test statistics t_a^* is distributed as a χ^2-variable with l_a dof and, under $\Upsilon_a \neq 0$, noncentrality parameter $\Upsilon_a^T \mathbf{F}_a^* \Upsilon_a$ *for all* Υ_b. The isolation of the damaged component Υ_a or Υ_b is then performed according to the largest min–max test t_a^* or t_b^*.

The sensitivity (projection) and min–max (rejection) tests enjoy a nice relationship. Global test can be written as follows: $t = t_a^* + \widetilde{t}_b$, where the decomposition is orthogonal in the sense that the two terms are uncorrelated, namely, $\mathrm{cov}_0(t_a^*, \widetilde{t}_b) = \mathrm{cov}_1(t_a^*, \widetilde{t}_b) = 0$ under both safe and faulty hypotheses. This additive decomposition holds for general distributions and more than two faults, whereas the orthogonal decomposition holds in the Gaussian case for only two faults.

The min–max test can be used for designing damage detection tests insensitive to environmental effects. This has been done in [44], where a simplified model of the temperature effect on the structural dynamics has been plugged within the min–max rejection test. This approach has been tested on a simulated bridge deck and a laboratory test case within a climatic chamber.

8.1.2 Isolation as multiple hypotheses testing

The isolation of multiple (more than two) faults raises several issues. The first issue is to distinguish between embedded multiple faults with causality constraints and independent multiple faults, and to state the corresponding simultaneous multiple hypotheses testing problem. The second issue concerns optimality criteria, which are the relevant counterpart of the trade-off between false alarms and detection probabilities in the ordinary hypotheses testing (detection) problem. The third issue is to exhibit tests performing the isolation in an optimal manner. It turns out that the above sensitivity and rejection approaches are relevant, from both theoretical and experimental points of view. For example, the set of all the sensitivity tests monitoring each fault, or the set of all the rejection tests, solve different multiple hypotheses testing problem, optimally with respect to different optimality criteria [48–50].

8.1.3 On-line isolation

The on-line isolation problem requires both on-line detection of a change and on-line discrimination between multiple hypotheses. Several types of performance indexes can be defined: probabilities of wrong isolations, mean delay to detection/isolation, etc. For designing the on-line detection/isolation algorithms, different optimization problems can be defined, based on those criteria. A major issue, for which research investigations are still needed, is to define criteria and optimization problems, which admit an explicit, and not approximate, optimal solution [2, 51].

8.1.4 Damage diagnostics

Considering again the vibration-based SHM problem of Section 7, damage localization and diagnostics can be stated as a detection problem, and not an (usually ill-posed) inverse estimation problem. This problem is addressed by plugging aggregated sensitivities of the modes and mode shapes with respect to FEM structural parameters in the above setting of subspace-based residuals in equation (67) and partial residuals such as equation (70). This results in directional tests as in equation (69), which perform the same type of damage-to-noise sensitivity analysis of the residual as for damage detection. The computation, the analytical-to-experimental matching, and the aggregation of the sensitivities, are performed off-line at a design stage, whereas the directional tests may be computed onboard. See [41] for more details and [52] for an example.

8.2 Increasing the robustness of the test

The main idea for increasing the robustness of the tests with respect to operating conditions, for example, consists in using confidence ellipsoids for plugging the uncertainty in the reference parameter θ_0 and the *a priori* information on the changes in the shape of the hypotheses to be tested, instead of using a confidence interval for the decision function itself [7]. The hypotheses are then defined by

$$\begin{aligned}\Theta_0 &\stackrel{\text{def}}{=} \{\Upsilon | \Upsilon^{\mathrm{T}}\ M^{\mathrm{T}}\ \Sigma^{-1}\ M\ \Upsilon \leq \upsilon_0^2\} \\ \Theta_1 &\stackrel{\text{def}}{=} \{\Upsilon | \Upsilon^{\mathrm{T}}\ M^{\mathrm{T}}\ \Sigma^{-1}\ M\ \Upsilon \geq \upsilon_1^2\}\end{aligned} \tag{73}$$

In equation (73), $\Upsilon = \sqrt{n}(\theta - \theta_0)$ is the change vector, θ_0 is the reference parameter value, and the sizes υ_0 and υ_1 have to be chosen. Typically, υ_0 captures uncertainties on the reference value θ_0, and may be chosen by learning on the data while identifying θ_0. The other radius υ_1 reflects a minimum change magnitude to be detected and should be learned from the data based on the monitoring problem specifications. The GLR test for testing between the hypotheses in equation (73) consists in computing

$$\overline{\chi}_n^2 \stackrel{\text{def}}{=} \min_{\Upsilon \in \Theta_0} l_\Upsilon(\zeta_n) - \min_{\Upsilon \in \Theta_1} l_\Upsilon(\zeta_n) \tag{74}$$

where $l_\Upsilon(\zeta_n) \stackrel{\text{def}}{=} (\zeta_n - M\ \Upsilon)^{\mathrm{T}}\ \Sigma^{-1}\ (\zeta_n - M\ \Upsilon)$. This can be easily written as [4, 7]:

$$\begin{aligned}&\frac{1}{n}\ \overline{\chi}_n^2 \\ &= \begin{cases} -(\chi - \upsilon_1)^2 & \text{for} \quad \chi \leq \upsilon_0 \\ -(\chi - \upsilon_1)^2 + (\chi - \upsilon_0)^2 & \text{for} \quad \upsilon_0 \leq \chi \leq \upsilon_1 \\ +(\chi - \upsilon_0)^2 & \text{for} \quad \chi \geq \upsilon_1 \end{cases}\end{aligned} \tag{75}$$

where $\chi = \chi_n/\sqrt{n}$ and χ_n^2 is in equation (59).

Experiments on real applications have shown that this simple idea turns out to be a useful tool for

obtaining robustness against both changes in the functioning modes and undermodeling of the system to be monitored, which are often reported as being a crucial issue for model-based onboard fault or damage detection algorithms [9, 11].

8.3 Damage detectability and sensor positioning

There are two by-products of the statistical approach: information-based damage detectability and isolability definitions, and criteria for optimizing the sensor positioning for SHM.

Possible detectability definitions are of two types [53]. Intrinsic definitions capture the signature of the fault on the system. The signature may be defined in terms of the amount of information about the fault contained in the observed data. Performance-based definitions may capture, among other performance indexes, the signature of the fault on residuals. Again, the signature may be defined in terms of the amount of information about the fault contained in the residual.

The definition of a criterion for assessing the quality of a given sensor location for monitoring purposes can be based upon the power (probability of correct detection) of the statistical tests [54]. Such a criterion can be used in two ways. For a given sensor pool location (it is not always possible to choose), find the detectable and diagnosable faults; alternatively, for one given (set of) fault(s), find the sensor pools and locations with which their detection and diagnosis are possible.

9 CONCLUSION

In this article, the key elements in the design of change detection algorithms have been described. Basic concepts (likelihood ratio, estimation, innovation) have been introduced. They involve a number of familiar operations (integration, averaging, sensitivity, adaptive thresholds, and windows). It has been suggested that a statistical framework enlightens the meaning and increases the power of those operations.

One of the approaches in that framework is dedicated to the detection of small changes that is of interest for condition-based and aging monitoring problems, and thus for SHM. When applied to the case of vibration-based SHM, the small changes detection approach handles a residual that exploits the left null space of a particular matrix capturing the nominal (safe) state of the structure, and results in a damage detection algorithm proven useful on a number of test cases.

Handling nonlinear systems remains a challenging issue. Computing likelihood or other estimating functions may raise numerical and computational problems. Monte Carlo sequential simulation methods, of which a well-known instance is particle filtering, have been developed [55–57]. These types of methods can be used for e.g., efficient numerical approximation of the efficient score. Application of these methods to fault and damage detection problems is the topic of recent and ongoing works [58–60].

Finally, the methods described above have a limitation in terms of the number of sensors that can be handled at a time. This is especially true with the recent advances in sensors and wireless networking, which allow for a huge amount of synchronized distributed measurements. New SHM concepts, methods, and algorithms are necessary under such measurement setups.

END NOTES

[a.] The same situation occurs when the number of measured outputs is smaller than the order of the dynamics.

RELATED ARTICLES

Statistical Time Series Methods for SHM

Statistical Pattern Recognition

Novelty Detection

Uncertainty Analysis

Damage Measures

Principles of Structural Degradation Monitoring

Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance

Ambient Vibration Monitoring

Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge

Gas Turbine Engines

REFERENCES

[1] Laï TL. Sequential change-point detection in quality control and dynamical systems (with discussion). *Journal of the Royal Statistical Society, Series B* 1995 **57**(4):613–658.

[2] Laï TL. Sequential analysis: some classical problems and new challenges (with discussion). *Statistica Sinica* 2001 **11**(2):303–408.

[3] Willsky AS. A survey of design methods for failure detection in dynamic systems. *Automatica* 1976 **12**(6):601–611.

[4] Basseville M, Nikiforov IV. *Detection of Abrupt Changes—Theory and Applications*. Prentice Hall: Englewood Cliffs, NJ, 1993, http://www.irisa.fr/sisthem/kniga.

[5] Carlstein E, Müller H-G, Siegmund D (eds). *Change-Point Problems*, *IMS Monograph Series*. California Institute of Mathematical Statistics: Hayward, CA, 1994; Vol. 23.

[6] Gustafsson F. *Adaptive Filtering and Change Detection*. John Wiley & Sons, 2000.

[7] Basseville M. On-board component fault detection and isolation using the statistical local approach. *Automatica* 1998 **34**(11):1391–1416.

[8] Basseville M, Benveniste A, Goursat M, Mevel L. In-flight monitoring of aeronautic structures: vibration-based on-line automated identification versus detection. *IEEE Control Systems Magazine* 2007 **27**(5):27–42.

[9] Cussenot C, Basseville M, Aimard F. Monitoring the vehicle emission system components. *Proceedings of the 13th IFAC World Conference*. San Francisco, CA, July 1996; Vol. Q, pp. 111–116.

[10] Nikiforov IV. New optimal approach to global positioning system/differential global positioning system integrity monitoring. *AIAA Journal of Guidance, Control, and Dynamics* 1996 **19**(5):1023–1033.

[11] Zhang Q, Basseville M, Benveniste A. Early warning of slight changes in systems. *Automatica* 1994 **30**(1):95–113.

[12] Basseville M, Mevel L, Vecchio A, Peeters B, van der Auweraer H. Output-only subspace-based damage detection—application to a reticular structure. *Structural Health Monitoring: An International Journal* 2003 **2**(2):161–168.

[13] Mevel L, Basseville M, Benveniste A. Fast in-flight detection of flutter onset: a statistical approach. *AIAA Journal of Guidance, Control, and Dynamics* 2005 **28**(3):431–438.

[14] Mevel L, Goursat M, Basseville M. Stochastic subspace-based structural identification and damage detection—application to the steel-quake benchmark. *Mechanical Systems and Signal Processing* 2003 **17**(1):91–101.

[15] Mevel L, Goursat M, Basseville M. Stochastic subspace-based structural identification and damage detection and localization—application to the Z24 bridge benchmark. *Mechanical Systems and Signal Processing* 2003 **17**(1):143–151.

[16] Mevel L, Hermans L, van der Auweraer H. Application of a subspace-based fault detection method to industrial structures. *Mechanical Systems and Signal Processing* 1999 **13**(6):823–838.

[17] Peeters B, Mevel L, Vanlanduit S, Guillaume P, Goursat M, Vecchio A, van der Auweraer H. Online vibration based crack detection during fatigue testing. *Key Engineering Materials* 2003 **245**(2): 571–578.

[18] Lehmann EL. *Testing Statistical Hypotheses*. John Wiley & Sons: New York, 1986.

[19] Hinkley DV. Inference about the change point from cumulative sum-tests. *Biometrika* 1971 **58**(3): 509–523.

[20] Page ES. Continuous inspection schemes. *Biometrika* 1954 **41**(1):100–115.

[21] Moustakides GV. Optimal procedures for detecting changes in distributions. *Annals of Statistics* 1986 **14**(4):1379–1387.

[22] Ritov Y. Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics* 1990 **18**(3):1464–1469.

[23] Shiryaev AN. *Optimal Stopping Rules*. Springer-Verlag: New York, 1978.

[24] Siegmund D. *Sequential Analysis—Tests and Confidence Intervals*. Springer-Verlag: New York, 1985.

[25] Wald A. *Sequential Analysis*. John Wiley & Sons: New York, 1947.

[26] Willsky AS, Jones HL. A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control* 1976 **21**(1):108–112.

[27] Basseville M, Benveniste A. Sequential detection of abrupt changes in spectral characteristics of digital signals. *IEEE Transactions on Information Theory* 1983 **29**(5):709–724.

[28] Ljung L. *System Identification—Theory for the User, Second Edition*. PTR Prentice Hall: Upper Saddle River, NJ, 1999.

[29] Basseville M. Distance measures for signal processing and pattern recognition. *Signal Processing* 1989 **18**(4):349–369.

[30] Gray A, Markel J. Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1976 **24**(5):380–391.

[31] Gray R, Buzo A, Gray A, Matsuyama Y. Distortion measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1980 **28**(4):367–376.

[32] Heyde CC. *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation, Springer Series in Statistics*. Springer-Verlag: New York, 1997.

[33] Basawa IV. Neyman-Le Cam tests based on estimating functions. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Le Cam L, Olshen RA (eds). Wadsworth, 1985, pp. 811–825.

[34] Basawa IV. Generalized score tests for composite hypotheses. In *Estimating Functions*, Godambe VP (ed). Clarendon Press: Oxford, 1991, pp. 121–132.

[35] Hall WJ, Mathiason DJ. On large sample estimation and testing in parametric models. *International Statistical Review* 1990 **58**(1):77–97.

[36] Akaïke H. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics* 1974 **26**:363–387.

[37] Ewins DJ. *Modal Testing: Theory, Practice and Applications, Second Edition*. Research Studies Press: Letchworth, 2000.

[38] Fritzen C-P. Recent developments in vibration-based structural health monitoring. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford, CA, September 2005.

[39] Stoïca P, Moses RL. *Introduction to Spectral Analysis*. Prentice Hall: Upper Saddle River, NJ, 1997.

[40] Basseville M, Abdelghani M, Benveniste A. Subspace-based fault detection algorithms for vibration monitoring. *Automatica* 2000 **36**(1):101–109.

[41] Basseville M, Mevel L, Goursat M. Statistical model-based damage detection and localization: subspace-based residuals and damage-to-noise sensitivity ratios. *Journal of Sound and Vibration* 2004 **275**(3–5):769–794.

[42] Basseville M, Benveniste A, Goursat M, Mevel L. Subspace-based algorithms for structural identification, damage detection, and sensor data fusion. *Journal of Applied Signal Processing* 2007 **2007**:13; Article ID 69136. Special Issue on *Advances in Subspace-Based Techniques for Signal Processing and Communications*, 2007.

[43] Balmès É, Basseville M, Bourquin F, Mevel L, Nasser H, Treyssède F. Merging sensor data from multiple temperature scenarios for vibration-based monitoring of civil structures. *Structural Health Monitoring* 2008 **7**:DOI:10.1177/1475921708089823 (Published online).

[44] Basseville M, Bourquin F, Mevel L, Nasser H, Treyssède F. A statistical nuisance rejection approach to handling the temperature effect for monitoring civil structures. *Proceedings of the 4th World Conference on Structural Control and Monitoring*, Paper 123. San Diego, CA, July 2006.

[45] Zouari R, Mevel L, Basseville M. Mode-shapes correlation and CUSUM test for on-line flutter monitoring. *Proceedings of the 17th IFAC Symposium on Automatic Control in Aerospace (ACA)*. Toulouse, June 2007.

[46] Basseville M. Information criteria for residual generation and fault detection and isolation. *Automatica* 1997 **33**(5):783–803.

[47] Neyman J. Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Harold Cramér Volume*, Grenander U (ed). John Wiley & Sons: New York, 1959, pp. 213–234.

[48] Basseville M, Nikiforov IV. Fault isolation for diagnosis : nuisance rejection and multiple hypotheses testing. *Annual Reviews in Control* 2002 **26**(2): 189–202.

[49] Baum CW, Veeravalli VV. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory* 1994 **40**(6):1994–2007.

[50] Spjøtvoll E. On the optimality of some multiple comparison procedures. *Annals of Statistics* 1972 **21**(3):1486–1521.

[51] Nikiforov IV. A simple recursive algorithm for diagnosis of abrupt changes in random signals. *IEEE Transactions on Information Theory* 2000 **46**(7):2740–2745.

[52] Balmès É, Basseville M, Mevel L, Nasser H, Zhou W. Statistical model-based damage localization: a combined subspace-based and substructuring approach. *Structural Control and Health Monitoring*, 2007 **15** (Published online).

[53] Basseville M. On fault detectability and isolability. *European Journal of Control* 2001 **7**(6): 625–638.

[54] Basseville M. On sensor positioning for structural health monitoring. *Proceedings of the 2nd European Workshop on Structural Health Monitoring*. Munich, July 2004; pp. 583–590.

[55] Andrieu C, Doucet A, Singh SS, Tadić VB. Particle methods for change detection, identification and control. *Proceedings of the IEEE* 2004 **92**: 423–438.

[56] Doucet A, de Freitas N, Gordon NJ (eds). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[57] Pitt MK, Shephard N. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association* 1999 **94**(446):590–599.

[58] Das S, Kyriakides I, Chattopadhyay A, Papandreou-Suppappola A. Particle filter based matching pursuit decomposition for damage quantification in composite structures. *Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. Newport, RI, May 2006.

[59] Ghanem R, Ferro G. Health monitoring for strongly non-linear systems using the Ensemble Kalman filter. *Structural Control and Health Monitoring* 2006 **13**(1):245–259.

[60] Li P, Kadirkamanathan V. Fault detection and isolation in nonlinear stochastic systems—a combined adaptive Monte Carlo filtering and likelihood ratio approach. *International Journal of Control* 2004 **77**(2):1101–1114.

Chapter 30
Statistical Pattern Recognition

Hoon Sohn[1] and Chang Kook Oh[2]

[1] *Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea*
[2] *Department of Civil Engineering, California Institute of Technology, Pasadena, CA, USA*

1 INTRODUCTION

The problem of searching for patterns in data has a long history. Statistical pattern recognition (SPR) is concerned with the discovery of regularities in data and subsequent actions such as classification [1]. Ample examples of pattern recognition are available. For instance, recognition of handwritten characters employs a machine-learning technique called *neural network* to take digital pixel images as inputs and to assign them into different characters. Although this is a nontrivial problem owing to the wide variations in handwriting, pattern-recognition techniques currently available can address such a classification problem. Thanks to their flexibility and diverse potential capacities, these pattern-recognition techniques have found a variety of applications.

One specific application of interest is structural health monitoring (SHM). SHM is a process of evaluating and assessing the integrity and safety of a structural system based on data autonomously and continuously collected from sensors [2]. The goals of SHM are (i) to identify the presence of damage (level I), (ii) to locate and quantify the damage (levels II and III), and (iii) finally, to predict the remaining useful life of the system (level IV) [2]. These objectives of SHM can be readily cast in the context of SPR. For instance, level I damage identification can be viewed as a classification problem where a binary output ($0 =$ undamaged and $1 =$ damaged) is assigned to the input data.

To achieve these goals, an SHM process involves the definition of potential damage scenarios for the system, observation of the system over a period

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

of time using periodically spaced measurements, extraction of features from these measurements, and statistical analysis of these features to determine the current state of the system's health [3]. For long-term SHM, the output of this process periodically updates information regarding the ability of the structure to perform its intended function in light of the inevitable aging and degradation resulting from operational environments. After extreme events, such as earthquakes or blast loading, SHM is used for a rapid condition screening and aims to provide, in near real time, reliable information regarding the current integrity of the structure.

Operational evaluation is the first step required before deploying an SHM system. Operational evaluation attempts to provide justifications for implementing an SHM system. In this stage, one should be able to answer the following questions: (i) What are the life-safety and economic motives for deploying an SHM system on the structure of interest? (ii) What are the definitions and types of damage that are of most concern for the system being monitored? (iii) Under what operational and environmental conditions does the system function? (iv) What are the costs and limitations on acquiring data in such operational environments? Once these questions are answered and the deployment of the SHM system is justified, we proceed to the second stage to design the data acquisition system.

The data acquisition process involves selecting the types of sensors to be used, choosing the location where the sensors should be placed, determining the number of sensors to be used, and defining the data acquisition/storage/transmittal hardware. This process is application specific, and economic considerations play a major role in these decisions. In this stage, additional questions such as how to improve the quality of data (data cleansing), how to compress the necessary data (data condensation), and how to alleviate undesirable variations of the data due to varying operational and environmental conditions (data normalization) should be addressed.

Feature extraction is a process of mining features that are sensitive to target damage from measured raw signals. Inherent in the feature selection process is the condensation of the data. The sensors employed to perform SHM typically produce a large amount of data. Therefore, data reduction is not only advantageous but also necessary, particularly if collection and comparison of data sets over the lifetime of the structure are envisioned (*see* **Dimensionality Reduction Using Linear and Nonlinear Transformation**). In addition, because data may be acquired from a structure over an extended period of time and in an operational environment, robust data-reduction techniques must retain the sensitivity of the chosen features to the structural damage of interest in the presence of environmental variability. Damage features that are used to distinguish damaged structures from undamaged ones have received the most attention in the technical literature.

Statistical model development is concerned with the implementation of the algorithms that operate on the extracted features to quantify the damage state of the structure. This statistical modeling can be done in either supervised or unsupervised learning mode. Unsupervised learning can be applied to data obtained from the undamaged structure, but this approach is often limited to level I or II damage classification, which identifies only the presence and locations of damage. When data are available from both the undamaged and damaged structure, the supervised learning approach can be taken to move forward to higher-level damage identification to classify and quantify the damage. Often, numerical simulations using an analytical model of a structure are employed to augment the scarce training data associated with the damaged structure.

Out of these four steps comprising the SHM process, the SPR approach is most relevant to feature extraction and statistical modeling processes. Therefore, the discussion in this article is mainly limited to these two steps of the SHM process. This article is organized as follows. First, it is discussed how various SPR techniques can be used in the feature-extraction process. Next, two distinctive statistical modeling approaches—unsupervised and supervised learning—are introduced. Finally, there is a separate section on data normalization because of its significance in the field deployment of an SHM system. The article concludes with a summary and a discussion on the scope for future work.

2 FEATURE EXTRACTION

For most SPR applications, it is advantageous to transform the raw data into some new form before

proceeding with statistical modeling; the feature-extraction process is one of the most important factors in determining the successful performance of the statistical modeling described in the next step (*see* **Data Preprocessing for Damage Detection**). Note that, feature extraction often involves signal preprocessing, data condensation, and data normalization. In the simplest form, feature extraction may take the form of a linear transform of the raw data as shown in Figure 1.

For the sake of discussion, let us assume that features 1 and 2 in Figure 1 denote the first two fundamental frequencies obtained from undamaged and damaged states of the system. In Figure 1, the data points in a lighter color are associated with the natural frequencies obtained from the undamaged state, and the darker data points correspond to the damaged state. These two-dimensional input data can be projected onto a one-dimensional feature space using various linear transforms. In particular, it is shown in Figure 1 that the separation (or damage sensitivity) of the data can be maximized by projecting the data using a projection technique called Fisher's linear discriminant. The description of Fisher's discriminant is provided later. In this process, the extraction of a damage-sensitive feature (the data after projection) and data reduction (from two-dimensional input data space to one-dimensional feature space) are accomplished simultaneously.

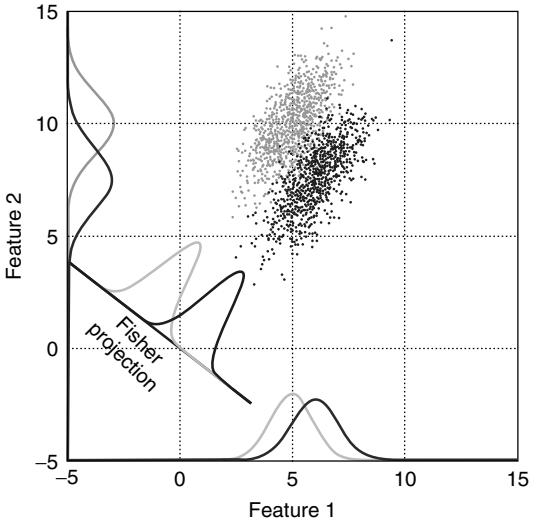


Figure 1. Simultaneous feature extraction and data condensation performed on the raw data.

The best features for damage detection are typically application specific. Numerous features are often identified for a structure and assembled into a feature vector. From a statistical discrimination perspective, a low-dimensional feature vector is desirable, and it is also desirable to obtain many samples of the feature vectors. There are no restrictions on the types or combinations of data contained in the feature vector. A variety of methods are employed to identify features for damage detection. Past experience with measured data from a system, particularly if damaging events have been previously observed for that system, is often the basis for feature selection. Numerical simulation of the damaged system's response to simulated inputs is another means of identifying features. The application of engineered flaws, similar to the ones expected in actual operating conditions, to laboratory specimens can identify parameters that are sensitive to the expected damage. Damage accumulation testing, during which significant structural components of the system under study are subjected to a realistic accumulation of damage, can also be used to identify appropriate features. A large number of features used for SHM are reported in [3, 4].

2.1 Curse of dimensionality

The need for low dimensionality in the feature vectors is referred to as curse of dimensionality and is discussed in general texts on density function estimation [5]. As discussed later in this article, statistical modeling requires statistical characterization of data, before any statistical inference can be reached. However, the difficulty of approximating local behavior of the data in a high-dimensional space is often reported. The problem is that to properly approximate a distribution of the data, the required sample size should grow exponentially with respect to the dimension of the sample space. This is called *curse of dimensionality*. This term was first coined by Bellman [6], when he observed the astronomical growth of computational effort in solving combinatorial optimization over many dimensions.

For one-dimensional feature space, 99.74% of the data will be inside $m \pm 3\sigma$ if the data have a normal distribution. Then, for two-dimensional feature spaces, only 66% of the data will be within

a circle of $m \pm 3\sigma$. Therefore, as the feature space expands to a higher dimension, more data are required to describe the distribution of the data, and it may impose serious limitations on how the unsupervised techniques can be used for SHM applications. Sometimes, it seems to be counterintuitive because we often feel that extraction of a higher-dimensional feature is better.

The true dimension of multivariate data in a high-dimensional space is almost always of dimension less than what it seems. Therefore, a parsimonious representation of the underlying structure can be obtained by eliminating a significant number of dimensions. The distinction between feature extraction and data reduction is not always clear cut. Feature extraction is the process of identifying damage-sensitive properties from the measured response signals, and this process often results in some form of data reduction. Data compression into feature vectors of small dimension is necessary if accurate estimates of the statistical distribution of the feature are to be obtained.

2.2 Multivariate analyses for dimensionality reduction

To avoid the curse of dimensionality, data are often projected onto a lower-dimensional feature space using specially designed mapping functions. This process is called *data reduction* or *data condensation*. In terms of data reduction, the main reason for using multivariate analysis is to analyze the simultaneous relationship among variables, to minimize the redundancy among them, and to obtain a parsimonious description of the structure underlying a set of multivariate data. Furthermore, data condensation is often done simultaneously as a part of feature extraction. For instance, the previously described illustrative example in Figure 1 can be seen as a feature-extraction process, but it is also conducting data reduction as well. Generally, this can be accomplished by constructing a linear/nonlinear combination of the original variables [7]. The choice of a multivariate technique for the analysis of interdependent variables is dependent on the nature of the data. Here, a few examples of multivariate techniques are introduced.

2.2.1 Principal component analysis

The primary goal of principal component analysis (PCA) is to construct a linear combination of the original variables so that the total variance (entropy) of the original variables can be kept as large as possible [8]. The variable projected onto the first principal component has the largest variance, and the variables projected using the subsequent projection components have successively smaller contributions to the total variables. These principal components are found by solving a singular-value problem of the covariance matrix among the original variables, and they are extracted in such a way that they are uncorrelated with each other.

Additional techniques often used for data reduction include canonical correlation analysis, multidimensional scaling, project pursuit, clustering analysis, and independent component analysis [1, 7, 9]. In particular, factor analysis and independent component analysis are similar to PCA in a sense that all of them explore the covariance structure for dimensionality reduction. For instance, factor analysis focuses on the common part of the total variance that a specific variable shares with the remaining variables.

2.2.2 Nonlinear principal component analysis

Multilayer neural networks such as autoassociative neural network (AANN) can be used to perform nonlinear dimensionality reduction, overcoming some of the limitation of linear techniques [10]

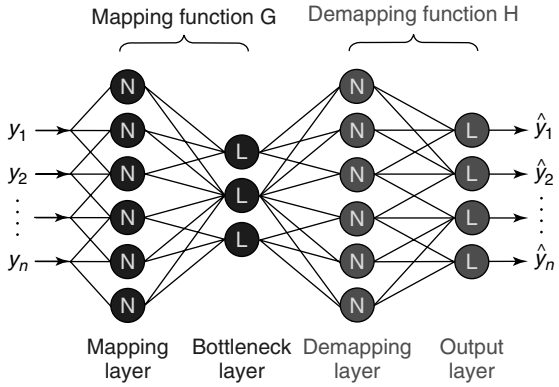


Figure 2. A schematic of an autoassociative neural network that performs nonlinear principal component analysis. [Reproduced with permission from Ref. 10. © Sage Publications, 2002.]

(*see* **Nonlinear Features for SHM Applications**). A typical AANN consists of three layers; mapping, bottleneck, and demapping layers as shown in Figure 2. The network is trained so that the outputs are simply the input vectors themselves. Because the network attempts to map each input vector onto itself, it is referred to as autoassociative network. In AANN, the dimension of the mapping and demapping layers (the number of neurons in each layer) are identical (m), but the dimension of the bottleneck layer (d) is designed to be less than those of the mapping and demapping layers ($m > d$). The first mapping layer projects the m-dimensional input space onto a d-dimensional subspace. Similarly, the demapping layer defines a mapping from the reduced subspace to the original input space. Owing to the reduced dimensionality of the bottleneck layer, AANN performs nonlinear PCA [11]. AANN has the advantage of not being limited to linear transforms, but nonlinear optimization techniques are needed to find the parameters of the network. Similarly, self-organizing maps can be also viewed as a nonlinear generalization of PCA.

2.3 Data preprocessing—standardization versus whitening

Typically, data from sensors are measured in different units and have significantly different variability. These different scaling and variability in data can cause the subsequent pattern-recognition algorithms to weigh input variables differently. To place equal importance for each variable, data can be standardized so that each variable has zero mean and unit variance, i.e., $z = (x - \overline{x})/\sigma$, where $\overline{x}$ and σ are the mean and the standard deviation of the data, respectively. Using PCA, the data can be further processed so that different variables become uncorrelated. This operation is called *whitening*.

$$z = \Lambda^{-\frac{1}{2}} U^T (x - \overline{x}) \tag{1}$$

where Λ and U represent diagonal matrices with elements of eigenvalues of the data covariance matrix and the orthogonal matrix with columns of eigenvectors, respectively.

3 UNSUPERVISED LEARNING

Unsupervised learning can be seen as one-class supervised learning, meaning that training is done using data only from one class. For SHM applications, this one class is most likely to be the pristine condition of the structure. On the basis of the training data sets obtained from the healthy condition of the structure, a statistical model of the healthy condition can be built. Once the statistical model of the healthy condition is constructed and a new set of data becomes available, typical unsupervised learning can be viewed as two-class hypothesis testing.

$$\begin{aligned} &\mathrm{H}_0 : \text{the system is not damaged} \\ &\mathrm{H}_1 : \text{the system is damaged} \end{aligned} \tag{2}$$

where H_0 is the null hypothesis stating that there is no defect in the system. If the new data set supports the null hypothesis, the null hypothesis is accepted. Otherwise, the null hypothesis is rejected. As can be seen from this example, hypothesis testing is a process of inferring whether a given null hypothesis is true or not based on the given data set. This dependency on the data can be explicitly expressed by employing the following conditional probabilities:

$$\begin{aligned} &\text{If } P(X|H_0) > P(X|H_1), \text{ accept } H_0 \\ &\text{otherwise reject } H_0 \end{aligned} \tag{3}$$

where $P(X|H_0)$ is a conditional probability of observing the data set X given that the null hypothesis is correct. $P(X|H_1)$ is defined in a similar manner. Most of unsupervised learning techniques can be viewed as variations of this hypothesis testing.

3.1 Type I error versus type II error

During hypothesis testing, there are two types of errors that need to be minimized: type I and II errors. Type I error is also referred to as a *false-positive error*, and this is the error of rejecting the null hypothesis when it is, in fact, correct. Type II error is the case where the null hypothesis is accepted even if it is incorrect and often called a *false-negative error*. In hypothesis testing, one needs to compromise between these two different types of

errors since it is not possible to simultaneously reduce both errors. Therefore, the users need to make conscious decisions as to which type of error is more critical for their applications. For instance, it might be more critical to minimize the false-negative error than to reduce the false-positive error for nuclear power plants because an unchecked defect (false-negative error) could lead to a catastrophic disaster. On the other hand, frequent false-positive alarms of a smoke detector in an ordinary house could lead to the deactivation of the device.

3.2 Fixed-sample test versus sequential test

On the basis of availability of data or how data are used, hypothesis testing can be classified into fixed-sample hypothesis testing or sequential hypothesis testing [12]. In fixed-sample hypothesis testing, that is, the conventional hypothesis testing, statistical inference is conducted after all necessary data are collected or the sample size is fixed. On the other hand, hypothesis testing is continuously (or sequentially) repeated in sequential hypothesis testing whenever a new set of data becomes available. For instance, an automobile company may want to examine the effectiveness of its newly designed air-bag system through actual car-crash tests. In this case, the null hypothesis could be that the newly designed air-bag system works properly with a certain confidence level. To verify this null hypothesis, the automobile company could have crashed 10 cars and performed fixed-sample hypothesis testing afterward. However, it would be more advantageous to repeatedly conduct hypothesis testing whenever another crash test is conducted, until one can arrive at a conclusion. For instance, if the air-bag system works properly five times in a row, probably there might be no reason to continue to crash the remaining cars. For continuous SHM, sequential testing may be more advantageous than fixed-sample testing.

3.3 Parametric approach versus nonparametric approach

As shown in equation (3), hypothesis testing requires statistical characterization of the data or extracted features, before the testing can be performed. This statistical modeling can be done in two different ways: parametric estimation versus nonparametric estimation [13]. In parametric estimation, the statistical distribution of the data is assumed in advance, and the corresponding parameters of the assumed distribution are estimated from the given data. For instance, the data can be often assumed to have a Gaussian distribution, and the parameters associated with the Gaussian distribution, the mean and standard deviation of the data, can be estimated from the given data sets. This type of parametric approach is advantageous when the available data are scarce because there are only a few parameters that need to be estimated. However, there is no guarantee that the selected distribution type properly captures the characteristics of the data, and it may impose unnecessary constraints on the data. For instance, a unimodal distribution such as a Gaussian distribution can be used, although the data, in fact, have a bimodal distribution. Therefore, it is important to verify the goodness of fit of the data.

On the other hand, nonparametric estimation techniques do not impose unnecessary constraints on the data and let the data define their own distribution. This nonparametric approach is more desirable when a plethora of training data is available. However, this nonparametric approach may fail to properly present the distribution of data when only a limited amount of data is available. Particularly, the nonparametric approach is more susceptible to the curse of dimensionality than the parametric approach. Examples of nonparametric approaches include but are not limited to: histogram, naive estimator, k-nearest neighbor method, and kernel estimator.

3.4 Normality test

As mentioned previously, the parametric approach assumes a known distribution type for an unknown data set. This can yield potential problems for later damage diagnosis. On the basis of available literature in SHM, most parametric techniques applied to SHM assume that the data have a Gaussian distribution without proper validation of such an assumption. There are techniques available for determining how close the data are to a normal distribution, such as a normal probability plot, skewness and

kurtosis tests, and the Kolmogorov–Smirnov test. It is recommended that users examine whether the normality assumption of the data is valid before using a normal distribution for their parametric estimation. When the data fail these normality tests, there are other types of distributions available.

3.5 Establishment of decision boundaries

The success of the hypothesis test not only depends on the proper characterization of the data's distribution but it also relies on the appropriate establishment of the decision boundary. A decision boundary or a threshold for damage diagnosis is often determined on the basis of the data distribution type and the given confidence level. For instance, when the training data have a normal distribution, a threshold for outlier analysis can be set to be at $m \pm 3\sigma$ with a 99.74% confidence. When a new data sample goes beyond this threshold value, it is likely that the null hypothesis is rejected with a 99.74% confidence.

3.6 Extreme value statistics

Often the decision boundary for SHM applications resides near the tail of the estimated distribution. Therefore, the errors of misclassification heavily depend on how well the tail of the distribution is modeled rather than the central population of the data. In fact, there is a large body of statistical theory that is explicitly concerned with modeling the tails of distributions, and these statistical procedures can be applied to damage-detection problems. The relevant field is referred to as extreme value statistics (EVS) [14], a branch of order statistics. EVS focuses on modeling those extreme points near the tails without prior knowledge of the parent distributions. It is shown that the extreme values can asymptotically follow one of only three possible extreme value distributions: Gumbel, Weibull, or Fréchet. Furthermore, a generalized extreme value distribution (GEV) can be employed to unify these three distributions, and a process to automatically estimate the associated parameters is available [15]. It has been demonstrated that false alarms resulting from unwarranted assumption of data distributions can be reduced by employing EVS [16].

3.7 Examples of unsupervised learning techniques

3.7.1 Outlier/novelty analysis

The idea of outlier/novelty analysis is to extract features from the data that characterize only the normal condition and these are used as a reference. During monitoring, data are measured, and the appropriate features are extracted and compared (in some sense) with the reference. Any significant deviation from the reference is considered to signal an outlier. Here, an outlier in a data set is an observation that is surprisingly different from the rest of the data in some sense, and therefore is believed to be generated by a different mechanism to the other data. The discordancy of a candidate outlier is some measure that can be compared against the corresponding objective criterion, and allows the outlier to be judged as statistically likely or unlikely to have come from the assumed generating model. For SHM applications, the discordancy should be evaluated with respect to a model constructed from a normal condition of the system of interest [17].

The case of outlier detection in univariate data is relatively straightforward in the sense that outliers will stick out from one end or the other of the data set. There are numerous discordancy tests. One of the most common, and the one whose extension to multivariate data is employed later, is based on deviation statistics given by [18]:

$$z_\zeta = \frac{x_\zeta - \overline{x}}{s} \tag{4}$$

where x_ζ is the potential outlier and $\overline{x}$ and s are the mean and standard deviation of the sample, respectively. The latter two values may be calculated with or without the potential outlier in the sample depending upon whether *inclusive* or *exclusive* measures are preferred. This discordancy value is then compared with a threshold value and the observation is declared to be an outlier or not.

The value of the threshold is critical; unfortunately, it is usually necessary to make some assumptions about the data to establish it. Suppose the normal condition data are assumed as Gaussian. In this case, there is a 95% probability that discordancy, z_ζ computed using equation (4) will lie in the range $[-1.96, 1.96]$. That is, there is only a 5% chance that

the point is a sample from the same normal condition distribution, but it lies outside this range.

3.7.2 Statistical process control

Statistical process control (SPC) is essentially similar to outlier/novelty analysis. For instance, control chart analysis, which is the most commonly used SPC technique and very suitable for automated continuous system monitoring, is applied for continuous SHM. When the system of interest experiences abnormal conditions, the mean and/or variance of the extracted features are expected to change. In [19], an X-bar control chart is employed to monitor the changes in the means of the selected features and to identify samples that are inconsistent with the past data sets. Application of the S control chart, which measures the variability of the structure over time, to the current test structure is presented in [20]. Several variations of the control charts can be found in [21].

3.7.3 Clustering

Cluster analysis is the grouping of sample data in a population to discover a structure in the data. In this article, it is considered as one of the unsupervised learning techniques since it explores natural groupings present in the data. For example, cluster analysis may find out that the data available are from two distinctive groups (one from an undamaged state and the other from a damaged state) without training. One approach of particular interest is a sum-of-squares method. The sum-of-squares method partitions a set of data samples into a prespecified number of clusters. This method finds partitioning of the data that minimizes within-class distances and maximizes between-class distances. There are several variations of the sum-of-squares method depending on the choice of the distance criteria and the optimization procedure adopted [22].

4 SUPERVISED LEARNING

Now, we turn our attention to supervised learning. The goal of supervised learning is to approximate a functional relationship between input and output variables using a set of input and output examples. For instance, a function $\mathbf{y} = f(\mathbf{x})$, which maps each input vector, $\mathbf{x}$, to an output vector $\mathbf{y}$, can be estimated using given training input and output data. Once the function is trained, it can be used for classification with discrete outputs or for regression when the outputs are continuous variables. Similarly, a conditional probability density function $p(\mathbf{y}|\mathbf{x})$ can be modeled using the training data set. Often this conditional probability is estimated using Bayes' theorem in the following form.

$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{x})} \tag{5}$$

Note that the posterior probability $p(\mathbf{y}|\mathbf{x})$ can be directly determined and used for the subsequent decision stages. Approaches that model the posterior probabilities directly are called *discriminative models*. In generative models, the conditional probability $p(\mathbf{x}|\mathbf{y})$ and the prior probability $p(\mathbf{y})$ are computed first. Then, the posterior probability $p(\mathbf{y}|\mathbf{x})$ is estimated using the above form of Bayes' theorem. The relative merits of these alternatives are further discussed in [1].

4.1 Linear models for regression: polynomial curve fitting

Here, the basic concepts of the supervised learning are illustrated using a polynomial equation since it is one of the simplest forms of linear regression models used in supervised learning [1]. Note that this polynomial equation is linear in terms of the unknown parameters w_i. Here, the polynomial equation is used as a running example through this article to highlight a number of key concepts. Suppose that a training data set is composed of N observations of the input variable x together with corresponding observations of the target variable t. The goal is to exploit the training data set to make a prediction, y, of the target variable, t, for a new input variable, x_{new}, using the polynomial function of the form.

$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_d x^d = \sum_{i=0}^{d} w_i x^i \tag{6}$$

where d is the order of the polynomial and x^i denotes x raised to the ith power. The coefficients of the polynomial are collectively represented by

the coefficient vector $\mathbf{w}$. The coefficient values can be determined by minimizing an error function that measures the sum of the squares of the errors between the target outputs and the corresponding predictions:

$$J(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2 \tag{7}$$

x_n and t_n denote the nth observations of x and t, and y_n represents the prediction corresponding to x_n. Because the error function is a quadratic function of $\mathbf{w}$, there exits a unique solution. This solution can be obtained by taking derivatives of the error function with respect to $\mathbf{w}$.

$$\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{T} \tag{8}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^d \\ 1 & x_2 & \cdots & x_2^d \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^d \end{bmatrix} \quad \text{and}$$

$$\mathbf{T} = [\, t_1 \quad t_2 \quad \cdots \quad t_N \,]^{\mathrm{T}} \tag{9}$$

Using this linear prediction model in training, several issues need to be addressed. These issues are discussed below.

4.1.1 Model selection

The first issue is how to determine the order d of the polynomial. More generally, this issue is called *model selection*, and it is an issue common to most of the supervised learning techniques. In general, a lower-order polynomial gives poor fits to the data and does not fully capture the underlying structure of the data. As the order increases, the fitness of the model to the data improves. At some point, the polynomial passes exactly through all the data points and the error function becomes zero. However, as the model order keeps increasing, the fluctuation of the fitted model also increases and the model starts fitting the noises rather than the underlying function. That is, the prediction capability of the fitted model deteriorates as the model order increases. This behavior is known as overfitting. The behavior of the polynomial function also depends on the number of the available training data. For a fixed model order, the overfitting becomes less severe as data size increases. A rule of thumb for model selection is that the number of data points should be at least 5 or 10 times (or some multiple) larger than that of the free parameters in the prediction model.

4.1.2 Regularization

One approach often used to avoid overfitting is regularization [1, 23, 24]. Regularization introduces a penalty term in the error function to discourage a model with unnecessary complexity. A simple example of such a penalty term takes the form of a sum of squares of the coefficients, leading to a modified error function.

$$J(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} \tag{10}$$

where $\mathbf{w}^{\mathrm{T}}\mathbf{w} = w_0^2 + w_1^2 + \cdots + w_d^2$, and the Lagrange parameter $\lambda(\geq 0)$ controls the relative importance of the regularization term in comparison with the sum-of-squares error term. Note that the sum of squares of the coefficients typically increases as the model becomes more complex (or as the order of the polynomial model increases).

Another approach to avoid overfitting is to divide available data into training and validation data sets. The weight values of a range of models are estimated using the training data, and the prediction performances of the trained models are evaluated using the validation data. In theory, when the training is done properly, the prediction error level for the validation date should be roughly equivalent to that of the training data. This process can be repeated for an increasing order of the model, and the one having the best predictive performance is selected. Note that the use of validation data set is also commonly adopted when an iterative training of the prediction model is necessary, such as for neural networks. The prediction performance for the training and validation data sets can be evaluated and compared at each iteration step, and the training can be stopped when the prediction errors for the validation data start increasing during the iteration. Optionally, a third test set can be reserved so that the performance of the selected

model can be finally evaluated after completion of training. This process is called *early stopping*.

In many applications, the availability of data is scarce and the accessible data should be taken advantage of as much as possible for training. If the size of the validation data set is limited, it gives a poor estimate of predictive performance. One solution to this problem is to use cross validation. In cross validation, the data are partitioned into D groups. Then, data only from $D-1$ groups are used for training and the remaining data from the remaining group are used to evaluate the predictive capability. This process is repeated D times for all possible choices of the divided training and validation data sets, and the final predictive performance is averaged from the D runs.

4.2 Examples of supervised learning techniques

4.2.1 Linear discriminant function

The method using linear discriminant functions divides the given training data sets into the prescribed number of classes of different properties by determining linear separating boundaries. In case of binary classification where the prescribed number of classes is two, i.e., 0 or 1 for undamaged or damaged state of the structure, respectively, the linear boundary is defined as $f_1(\mathbf{x}) = 0$ using equation (11). This linear boundary is called a *hyperplane* whose unknown parameters such as weight vector $\mathbf{w}$ and bias w_0 are estimated through minimizing the training error, for example, the sum-of-squares error. The prediction is made for a new data $\mathbf{x}_{\text{new}}$ by assigning class C_1 if $f_1(\mathbf{x}_{\text{new}}) > 0$ or C_2 if $f_1(\mathbf{x}_{\text{new}}) < 0$. This can be easily extended to multiclass cases [23].

$$f_1(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + w_0 \tag{11}$$

Instead of using a linear combination, any monotonic function denoted by $g(\cdot)$ in equation (12) can be adopted. This function $g(\cdot)$ is called an *activation function*, and a specific choice is the logistic sigmoid function $g(x) = 1/(1+\exp(x))$.

$$f_2(\mathbf{x}) = g(\mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + w_0) \tag{12}$$

The separating boundary obtained by equation (12) is also linear, but it provides the probabilistic interpretation on outputs. There are, however, many SHM problems for which linear discriminant functions are insufficient, which leads to nonlinear classification methods presented next.

4.2.2 Neural networks

Neural networks, or artificial neural networks (ANN), were developed on the basis of biological neural networks and classified as two distinct groups, namely, *recurrent* and *feedforward* networks, depending on whether the form of networks is circled or not, respectively (*see* **Artificial Neural Networks**). The most commonly used feedforward network is the multilayer perceptron, which consists of the input, hidden, and output layers with weights that connect neurons in the layers.

A neural network is trained by using a given training data set to minimize a predefined cost function. This cost function is decided depending on the task; a popular choice is the mean-squared error between the given outputs in the training data set and the outputs generated from the network. Typically, there is one output node corresponding to each class, except in a binary classification where there is only one output node. The training procedure with a cost function of mean-squared error consists of the following steps: (i) to assign initial random weights and biases to nodes in the hidden and output layers; (ii) to generate outputs through the designed networks; (iii) to compare the outputs with the given values in the training data set; (iv) to go back to layers to modify the weights and biases if the discrepancy is larger than threshold (back-propagation algorithm); and (v) to iterate procedures (ii)–(iv) until the network outputs converge.

When designing neural networks, it is necessary to decide how many layers and neurons are used. Excessive number of layers or neurons causes an overfitting problem, while insufficient number of them causes an underfitting problem. In [25], Bayesian model class selection was adopted to overcome these problems by determining the most probable model class based on the given data set.

4.2.3 Radial basis function

The Radial basis function (RBF) is a real-valued function taking the form $\phi(\|\mathbf{x} - \mathbf{x}^n\|)$, where $\phi(\cdot)$ is a nonlinear function and $\|\cdot\|$ is a norm (usually

Euclidean norm). The most commonly used nonlinear function is the Gaussian, $\phi(\mathbf{x}) = \exp(-\mathbf{x}^2/(2\sigma^2))$, where σ controls the smoothness. The linear sum of M RBFs as shown in equation (13) forms the so-called RBF network.

$$f(\mathbf{x}) = \sum_{i=1}^{M} w_i \phi(\|\mathbf{x} - \mathbf{c}_i\|) + w_0 \qquad (13)$$

where $\mathbf{w}$, w_0, and $\mathbf{c}_i$ represent weight vector, bias, and center of the basis functions, respectively, and M is smaller than the number of training data.

In classification, RBF networks provide the distributions of separate classes instead of estimating separating boundaries. This distribution is approximated by RBF networks as a posterior probability for each class through Bayes' theorem in equation (5) [23].

4.2.4 Support vector machine

The support vector machine (SVM) is a recently developed state-of-art pattern-recognition method [24]. 0SVM determines the separating boundaries between data of different classes by maximizing the margin as well as minimizing the misclassification errors: margin is defined as the distance of the closest point to the hyperplane as shown in Figure 3. The SVM inherently considers regularization via this margin term to alleviate ill-conditioning, and the solution is computed by solving a convex optimization problem. The convex optimization has the advantage that if a local minimum is obtained, it is guaranteed to be a global minimum, and usually, there exists only one such minimum.

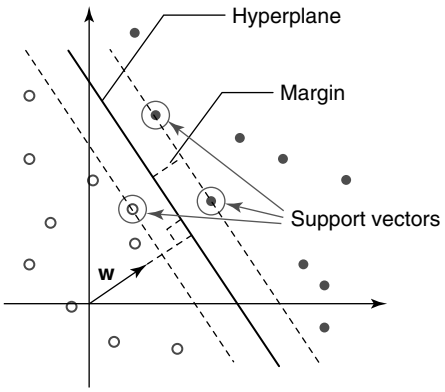


Figure 3. Illustration of SVM as a maximum margin classifier: SVM algorithm estimates a hyperplane to produce a maximum margin.

Nonlinear separating boundaries can also be computed by applying the kernel method. This method enables the same algorithm for the hyperplane evaluation to be applied for the nonlinear classification problems simply by replacing every dot product with any continuous, symmetric, and positive definite kernels, $k(\mathbf{x}, \mathbf{x}_i)$ [26]. The Gaussian kernel $k(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2/(2\sigma^2))$ is one of the frequent choices and the corresponding boundary has the form

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i) + \alpha_0 = 0 \qquad (14)$$

where α_i, y_i, and α_0 are Lagrangian multiplier, class label, and bias, respectively.

An illustrative example is presented in Figure 4(a). Classification is performed for 200 randomly generated data points from two mixtures of two Gaussian probability density functions (one shown as crosses and another as dots) using SVM. The circled data specify support vectors. Only these data play an important role in deciding the separating boundaries. The misclassification rate for SVM classification is 5.5% using 66 support vectors.

4.2.5 Relevance vector machine

The relevance vector machine (RVM) is a Bayesian classification methodology using an identical function form for separating boundary as the SVM described in equation (15) [27].

$$f(\mathbf{x}) = \sum_{i=1}^{N} w_i k(\mathbf{x}, \mathbf{x}_i) + w_0 \qquad (15)$$

RVM is known to overcome some of disadvantages of SVM, and a specific prior known as the *automatic relevance determination prior* gives a unifying framework to control model complexity, to overcome the overfitting problem, and to consider regularization as well [28].

The same dataset previously utilized for SVM classification is selected to perform RVM classification and relevance vectors to control a separating boundary are circled. As shown in Figure 4(b), RVM

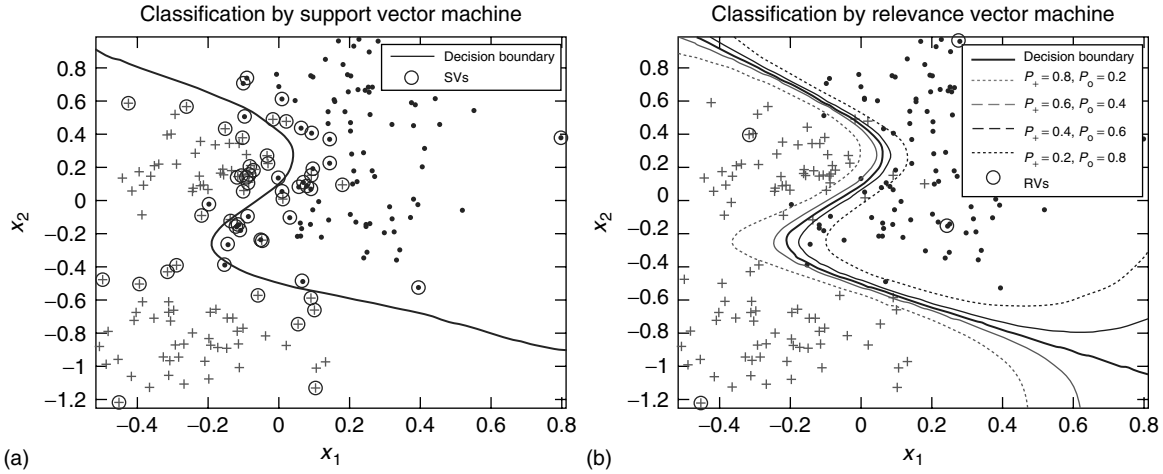


Figure 4. Comparison of classification results by SVM (a) and RVM (b) (RVM provides several separating boundaries according to the corresponding probabilities while SVM yields only one boundary. In Figure 4(b), P_+ and P_o denote the probabilities of crosses and dots, respectively. The misclassification rates for SVM and RVM are 5.5% identically but the number of support vectors and relevant vectors to control the separating boundary is 66 and 4, respectively.)

can provide a probabilistic description on results so that different separating boundaries can be obtained with the corresponding probabilities. Suppose that the crosses and dots in Figure 4 are data extracted from measurements representing undamaged and damaged states, respectively. If one wants to decide a separating boundary to reduce more critical false-negative alarm by imposing higher probabilities to an undamaged state, for example, probability for undamaged state $= 0.6$, the separating boundary with $P_+ = 0.6$ and $P_o = 0.4$ in Figure 4(b) can be estimated by RVM. The misclassification rate for RVM classification using the same data set is 5.5% using only four relevance vectors. Additional applications in SHM and comparison results with SVM are presented in [29].

5 DATA NORMALIZATION

Stated in its most basic form, the objective of SHM is to ascertain whether damage is present or not based on measured dynamic or static characteristics of a system to be monitored. Until now, damage diagnosis has been discussed by comparing the data collected from a healthy part of the structure with a new set of data obtained from a potentially damaged part of the system. The basic premise here is that any deviation from the healthy condition of the system is a result of damage. In reality, structures are subject to changing environmental and operational conditions that affect measured signals, and these ambient variations of the system can often mask subtle changes in the system's response signal caused by damage. It has been shown that these natural variations of the system could lead to false-positive indications of damage. This might be one of the biggest hurdles for deploying a reliable continuous SHM system to field applications. A procedure to normalize data sets so that signal changes caused by operational and environmental variations of the system can be separated from structural changes of interest, such as structural deterioration or degradation is called *data normalization* here. First, reviews of the effects of environmental and operational variations on real structures as reported in literature are presented. Then, research progresses that have been made in the area of data normalization are provided.

Many existing SHM techniques neglect the important effect of changing environmental and operational conditions on the underlying structure. For in-service structures, the variability in dynamic properties can be a result of time-varying environmental and operational conditions. Environmental conditions include wind, temperature, and humidity, while operational conditions include ambient loading conditions, operational speed, and mass loading. Note that data

normalization materials presented here are mainly taken from [30].

5.1 Temperature

The effects of temperature variability on the measured dynamic response of structures have been addressed in several studies. It is intuitive that temperature variation may change the material properties of a structure. Wood reported that the changes of bridge responses were closely related to the structure's temperature based on the vibration test of five bridges in the United Kingdom [31]. Analyses based on the data compiled suggested that the variability of the asphalt elastic modulus due to temperature effects was a major contributor to the changes in the structural stiffness. Temperature variation not only changes material stiffness but also alters the boundary conditions of a system. On the basis of a field test conducted on the Sutton Creek Bridge in Montana, USA, Moorty and Roeder reported that the movements obtained from both the analytical model and the measured values showed significant expansion of the bridge deck as temperature increased [32]. Rohrmann *et al.* noted that when a bridge structure was obstructed from expanding or contracting, the expansion joints could be closed, significantly altering the boundary conditions [33]. Finally, other studies on the influence of temperature variation on in-service structures have been reported in [30].

5.2 Boundary condition

Changes in the structure's surroundings or boundary conditions such as thermal expansion can produce more significant changes in dynamic responses than damage. Using an analytical model of a cantilever beam, Cawley compared the effect of crack formation on the resonant frequency to that of the beam's length [34]. In this study, the crack was introduced at the fixed end of the cantilever beam, and the length of the beam was varied. His results demonstrated that the resonance frequency change caused by a crack, which was 2% cut through the depth of the beam, was 40 times smaller than that caused by a 2% increase in the beam's length. Alampalli reported that, for a 6.76 m $\times$ 5.26 m bridge span that they tested, the natural frequency changes caused by freezing of the bridge supports were an order of magnitude larger than the variation introduced by an artificial saw cut across the bottom flanges of both girders [35]. In particular, the changing boundary conditions would impose difficulties on the implementation of SHM systems because it is often challenging to directly measure the boundary conditions of a structure.

5.3 Mass loading

Mass loading such as traffic loading would be another operational variable that is difficult to precisely measure, but might be important for data normalization. The researchers seem to agree that the mass loading effect of moving vehicles varies depending on the vehicles' mass relative to the magnitude of the bridge. Kim *et al.* reported that the measured natural frequencies of a 46-m-long simply supported plate girder bridge decreased by 5.4% because of heavy traffic [36]. However, for the medium- and long-span bridges, changes in the measured natural frequencies due to different types of vehicle loading (heavy vs light) were hardly detectable. Zhang *et al.* found the damping ratios are sensitive to the traffic mass, especially when the deck vibration exceeded a certain level [37]. It is believed that damping ratio increases because the energy dissipation capacity in the material and at the joint increases at higher traffic loading. In addition, it is speculated that the secondary structure–vehicle interaction effects could also influence the dynamic characteristics of bridge structures.

5.4 Wind-induced vibration

Wind-induced vibration plays an important role for long-span bridges. As a bridge vibrates in the wind, the energy input from the wind-induced vibration becomes larger than the energy dissipated by damping, causing flutter or buffeting. Fujino and Yoshida observed that the dynamic behavior of cable-stayed or suspension bridges is amplitude dependent [38]. For instance, the fundamental frequency of a suspension bridge reduced as the wind speed increased.

On the other hand, the modal damping increased when the wind velocity exceeded a certain level.

Mahmoud *et al.* utilized continuously measured vibration data from the Hakucho Suspension Bridge in Japan to study the dynamic behavior of a suspension bridge [39]. It was found that the vertical amplitude of the bridge response was almost a quadratic function of the wind speed, and the damping ratio was dependent on the vibration amplitude. In particular, the lower natural frequencies were significantly affected by wind speeds. However, the dependency of mode shapes on the vibration amplitude was mainly observed for higher modes and near the tower.

5.5 Data normalization techniques

Because data can be measured under varying conditions, the ability to normalize the data becomes very important to the SHM process. This section contains new technical developments that attempt to tackle the aforementioned environmental and operational issues of SHM. When the environmental or operating variability is an issue, there are three different situations for data normalization. First, when direct measurements of the varying environmental or operational parameters are available, various kinds of regression and interpolation analyses can be performed to relate measurements relevant to structural damage and those associated with environmental and operation variation of the system [17, 33, 40, 41]. For example, the dependency of two-dimensional features on some environmental variable T can be approximated by using regression analysis as shown in Figure 5(a) when a large set of extracted features and measured environmental variables are available. Note that there might be some damage cases that cannot be distinguished from the undamaged conditions unless the environmental variable is measured: for example, the damage case shown in Figure 5(a).

On the other hand, there are situations in which direct measurements of these operational and environmental parameters are impractical or difficult to achieve and damage produces changes in the extracted features, which are orthogonal to the

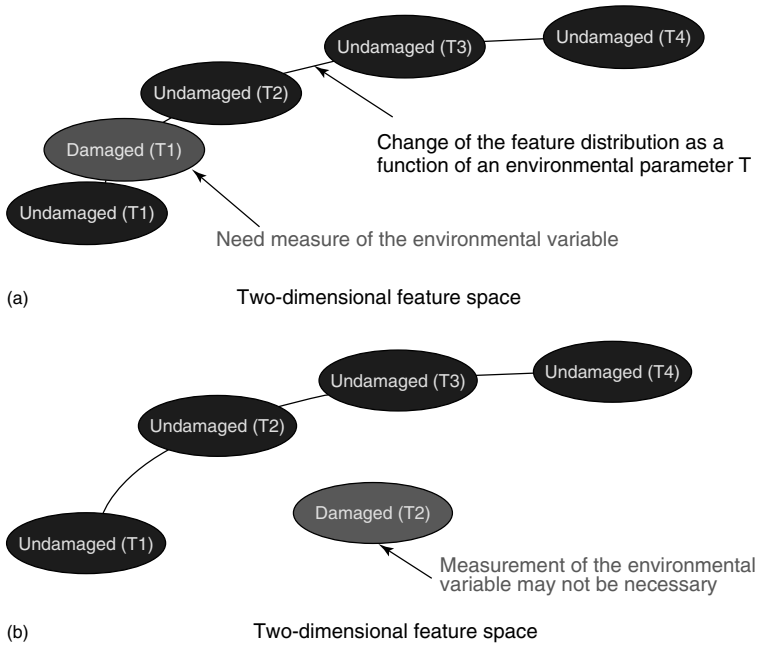


Figure 5. Three conceptional situations exist for data normalization: (a) direct measurements of the varying environmental or operational parameters are available, (b) direct measurements of these parameters are impractical or difficult to achieve, but damage produces changes orthogonal to changes caused by the operational and environmental variation, and (c) features that are mainly sensitive to damage but insensitive to operational and environmental variations can be extracted (not shown in the figure).

changes caused by the operational and environmental variation of the system as shown in Figure 5(b). For this situation, it may be possible to distinguish changes caused by damage from those caused by the operational and environmental variation of the system without measuring the operational and environmental parameters. Several researchers have tackled this situation by implicitly modeling the underlying relationship between the environmental variables and the damage-sensitive features [10, 42–44]. Others have attempted to divide baseline data sets into subsets, each corresponding to a different operational and environmental condition [45, 46].

Finally, there are other researchers who attempt to explicitly extract features that are mainly sensitive to damage but insensitive to operational and environmental variations. In particular, a suite of reference-free damage-detection techniques that do not require direct comparison with baseline data have been developed to address data normalization [47–49].

One of the reference-free approaches in [48] utilizes the polarization characteristics of the lead zirconate titanate (PZT) wafers attached on the both sides of the structure as shown in Figure 6. PZT materials develop an electrical charge or voltage when a mechanical pressure is applied. Conversely, they produce deformation (strain) when exposed to an applied electric field. PZTs are crystalline materials that are artificially polarized through a thermal poling process. The overall behavior of a piezoelectric material as well as its electrical characteristics is governed by the polarization direction of the material. That is, the "sign" of the output voltage depends on the bending of the PZT with respect to its poling direction.

PZTs A and D are collocated on either side of the plate, and PZTs B and C are placed in a similar manner. The arrows in Figure 6(a) show the positive poling direction of each PZT. When the plate is in a pristine condition and four identical PZTs are instrumented as shown in Figure 6(a), it is shown that signal AC becomes identical to signal BD as illustrated in Figure 6(b). Here, signal AC denotes the response signal measured at PZT C when a toneburst excitation is applied at PZT A, and signal BD is defined in a similar fashion. However, signal AC becomes no longer identical to signal BD when there is a crack between PZTs A and B (or PZTs C and D) as shown in Figure 6(c) and (d). Crack formation creates Lamb wave mode conversion due to a sudden thickness change in the structure. Note that, while the S_0 and A_0 modes in Figure 6(b) are in-phase, the S_0/A_0 and A_0/S_0 modes in signals AB and CD are fully out-of-phase as shown in Figure 6(d). (S_0, A_0, S_0/A_0, and A_0/S_0 are defined in Figure 6.) Therefore, the modes converted by the crack can be extracted simply by subtracting signal AC from signal BD.

Since this approach relies only on comparison of two signals instantaneously obtained at the current

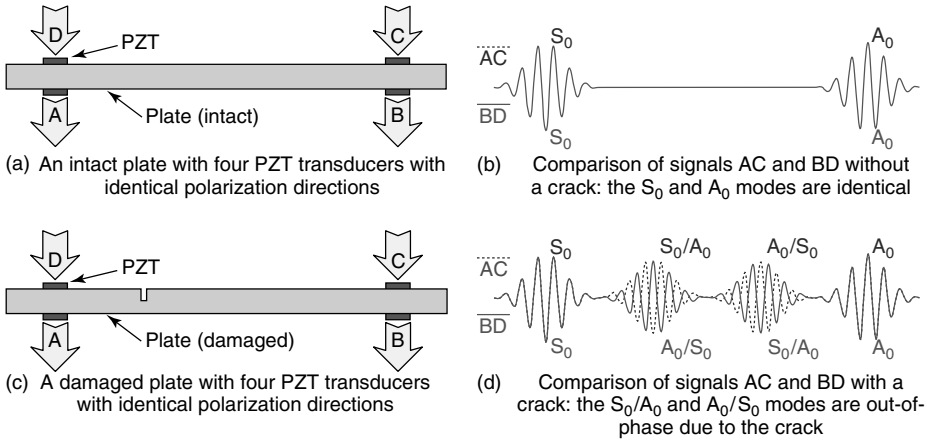


Figure 6. Extraction of the additional Lamb wave modes generated by a crack using the poling directions of the PZT transducers. (A_0 and S_0 denote the fundamental symmetric and antisymmetric modes. A_0/S_0 denotes an A_0 mode converted from an S_0 mode when it passes through a crack. S_0/A_0 is defined similarly. Signal AC denotes the response signal measured at PZT C when the excitation is applied at PZT A. Signal BD is defined similarly.)

state of the system, rather than on comparison with previously recorded reference data, it is experimentally and theoretically shown that this approach reduces false alarms of defect due to changing operational and environmental conditions of the system.

6 CONCLUSION

The problem of searching for patterns in data has a long history. SPR is concerned with the discovery of regularities in data and subsequent actions such as classification and regression. The introduction of this SPR paradigm to SHM has brought a new perspective to the SHM community, and together they have undergone substantial development over the past decade. In particular, the SPR approach of SHM problems has grown from being viewed as a special treatment of the SHM problems to becoming a mainstream. This article is intended to reflect these recent developments in SHM, while providing a brief introduction to the field of SPR. Many new concepts and novel SHM techniques based on SPR view points are constantly being developed and implemented in the SHM community. In addition, the practical applicability and benefits of the SPR approaches have been reported here through the illustrations of a range of numerical, experimental, and field examples.

ACKNOWLEDGMENTS

This research is supported by the Radiation Technology Program under Korea Science and Engineering Foundation (KOSEF) and the Ministry of Science and Technology (M20703000015-07N0300-01510) and Korea Research Foundation Grant funded by the Koran Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-331-D00462). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] Bishop CM. *Pattern Recognition and Machine Learning*. Springer, 2007.

[2] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Report LA-13976-MS. Los Alamos National Laboratory, 2004.

[3] Sohn H, Farrar CR, Hunter NF, Worden K. Structural health monitoring using statistical pattern recognition techniques. *Journal of Dynamic Systems Measurements and Control-Transactions of the ASME* 2001 **123**(4):706–711.

[4] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics: A Literature Review*, Report LA-13070-MS. Los Alamos National Laboratory, 1996.

[5] Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 1992.

[6] Bellman RE. *Adaptive Control Processes*. Princeton University Press, 1961.

[7] Dillon WR, Goldstein M. *Multivariate Analysis: Methods and Applications*. John Wiley & Sons, 1984.

[8] Fukunaga K. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[9] Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. Prentice Hall, 1998.

[10] Sohn H, Worden K, Farrar CR. Statistical damage classification under changing environmental and operational conditions. *Journal of Intelligent Material Systems and Structures* 2002 **13**(9): 561–574.

[11] Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. *AICHE Journal* 1991 **37**:233–243.

[12] Ghosh BK. *Sequential Tests of Statistical Hypotheses*. Addison-Wesley, 1970.

[13] Silverman BW. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1992.

[14] Castillo E, *Extreme Value Theory in Engineering*, *Series in Statistical Modeling and Decision Science*. Academic Press, 1988.

[15] Park HW, Sohn H. Parameter estimation of the generalized extreme value distribution for structural health monitoring. *Journal of Probabilistic Engineering Mechanics* 2006 **21**(4):366–376.

[16] Sohn H, Allen DW, Worden K, Farrar CR. Structural damage classification using extreme value

statistics. *Journal of Dynamic Systems Measurement and Control-Transactions of the ASME* 2005 **127**(1):125–132.

[17] Worden K, Sohn H, Farrar CR. Novelty detection in a changing environmental: regression and interpolation approaches. *Journal of Sound and Vibration* 2002 **258**(4):741–761.

[18] Barnett V, Lewis T. *Outlier in Statistical Data*. John Wiley & Sons, 1994.

[19] Sohn H, Czarnecki JJ, Farrar CR. Structural health monitoring using statistical process control. *Journal of Structural Engineering ASCE* 2000 **126**(11): 1356–1363.

[20] Fugate ML, Sohn H, Farrar CR. Vibration-based damage detection using statistical process control. *Mechanical Systems and Signal Processing* 2001 **15**(4):707–721.

[21] Montgomery DC. *Introduction to Statistical Quality Control*. John Wiley & Sons, 1997.

[22] Webb A. *Statistical Pattern Recognition*. John Wiley & Sons, 2002.

[23] Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press, 1998.

[24] Vapnik VN. *Statistical Learning Theory*. John Wiley & Sons, 1998.

[25] Lam HF, Yuen KV, Beck JL. Structural health monitoring via measured ritz vectors utilizing artificial neural networks. *Computer-Aided Civil and Infrastructure Engineering* 2006 **21**(4):232–241.

[26] Aizerman M, Braverman E, Rozonoer L. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control* 1964 **25**:821–837.

[27] Tipping ME. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 2001 **1**:211–244.

[28] Mackay DJC. Bayesian non-linear modelling for the prediction competition. *ASHRAE Transactions* 1994 **100**(2):1053–1062.

[29] Oh CK, Beck JL. Sparse Bayesian learning for structural health monitoring. *Proceedings of the Fourth World Conference on Structural Control and Monitoring*. San Diego, CA, 2006.

[30] Sohn H, Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society A on Structural Health Monitoring* 2007 **365**:539–560, a special issue.

[31] Wood MG. *Damage Analysis of Bridge Structures using Vibrational Techniques*, Ph.D. Thesis. Department of Mechanical and Electrical Engineering, University of Aston: Birmingham, 1992.

[32] Moorty S, Roeder CW. Temperature-dependent bridge movements. *Journal of Structural Engineering ASCE* 1992 **118**:1090–1105.

[33] Rohrmann RG, Baessler M, Said S, Schmid W, Ruecker WF. Structural causes of temperature affected modal data of civil structures obtained by long time monitoring. *Proceedings of the XVII International Modal Analysis Conference*. Kissimmee, FL, 1999; pp. 1–6.

[34] Cawley P. Long range inspection of structures using low frequency ultrasound. *Proceedings of DAMAS 97: Structural Damage Assessment using Advanced Signal Processing Procedures*. University of Sheffield, 1997; pp. 1–17.

[35] Alampalli S. Influence of in-service environment on modal parameters. *Proceedings of IMAC XVI*. Santa Barbara, CA, 1998; pp. 111–116.

[36] Kim CY, Jung DS, Kim NS, Yoon JG. Effect of vehicle mass on the measured dynamic characteristics of bridges from traffic-induced test. *Proceedings of the IMAC XIX*. Kissimmee, FL, 1999; pp. 1106–1110.

[37] Zhang QW, Fan LC, Yuan WC. Traffic-induced variability in dynamic properties of cable-stayed bridge. *Earthquake Engineering and Structural Dynamics* 2002 **31**:2015–2021.

[38] Fujino Y, Yoshida Y. Wind induced vibration and control of Trans-Tokyo bay crossing bridge. *Journal of Structural Engineering ASCE* 2002 **128**(8):1012–1025.

[39] Mahmoud M, Abe M, Fujino Y. Analysis of suspension bridge by ambient vibration measurement using the time domain method and its application to health monitoring. *Proceedings of IMAC XIX*. Kissimmee, FL, 2001; pp. 504–510.

[40] Peeters B, De Roeck G. One year monitoring of the Z24 bridge: environmental influences versus damage effects. *Proceedings of IMAC-XVIII*. San Antonio, TX, 2000; pp. 1570–1576.

[41] Fritzen CR, Mengelkamp G, Guemes A. Elimination of temperature effects on damage detection within a smart structure concept. *Proceedings of the 4th International Workshop on Structural Health Monitoring*. Stanford University, CA, 2003; pp. 1530–1538.

[42] Kullaa J. Is temperature measurement essential in structural health monitoring. *Proceedings of*

the 4th International Workshop on Structural Health Monitoring. Stanford University, CA, 2003; pp. 717–724.

[43] Ruotolo R, Surace C. Damage detection using singular value decomposition. *Proceedings of DAMAS 97: Structural Damage Assessment using Advanced Signal Processing Procedures*. University of Sheffield, 1997; pp. 87–96.

[44] Manson G. Identifying damage sensitive, environmental insensitive features for damage detection. *3rd International Conference on Identification in Engineering Systems*. University of Wales Swansea, 2002.

[45] Sohn H, Worden K, Farrar CR. Novelty detection using auto-associative neural network. *Symposium on Identification of Mechanical Systems: International Mechanical Engineering Congress and Exposition*. New York, 2001.

[46] Sohn H, Farrar CR. Damage diagnosis using time series analysis of vibration signals. *Journal of Smart Materials and Structures* 2001 **10**: 446–451.

[47] Sohn H, Park HW, Law KH, Farrar CR. Combination of a time reversal process and a consecutive outlier analysis for baseline-free damage diagnosis. *Journal of Intelligent Material Systems and Structures* 2007 **18**(4):335–346.

[48] Kim SB, Sohn H. Instantaneous reference-free crack detection based on polarization characteristics of piezoelectric materials. *Smart Materials and Structures* 2007 **16**:2375–2387.

[49] Park HW, Sohn H, Law KH, Farrar CR. Time reversal active sensing for health monitoring of a composite plate. *Journal of Sound and Vibration* 2007 **302**:50–66.

Chapter 32
Artificial Neural Networks

Steve Reed
QinetiQ, Farnborough, UK

1 INTRODUCTION

The aim of this article is to describe the development and the key elements of artificial neural networks (ANNs) within the context of structural health monitoring (SHM). Chang [1] describes the essence of SHM technology as developing autonomous systems for the continuous monitoring, inspection, and damage detection of structures with minimum labor involvement. The characteristics of ANNs in their ability to learn from experience, generalize from examples, and identify underlying information from within noisy data mean that they have the potential to play a leading role in such SHM systems.

Within this article, the development of ANNs and the basic principles behind their functionality, training, and deployment are described. Emphasis is placed upon the multilayer perceptron (MLP), both as the workhorse of ANN methods and as a vehicle for describing the principles behind these methods; however, other frequently used techniques, including radial basis functions (RBFs) and self-organizing feature maps (SOFMs), are also detailed. Additionally, the application of ANN techniques, including examples of parameter-based structural usage models and damage detection methods using ANNs, are presented. The use of ANN techniques in novelty detection (*see* **Novelty Detection**), pattern-recognition applications (*see* **Statistical Pattern Recognition**), operational loads measurement (*see* **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**), and fixed-wing fatigue monitoring (*see* **Aerospace Applications of SMART Layer Technology**) are discussed elsewhere. Reference has been made throughout this article to a wide range of texts. Readers with a desire for further knowledge of ANNs are recommended to investigate [2–6] as introductory texts and [7–9] as more seminal works.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 ANN DEVELOPMENT

2.1 ANN definition

There is no universally accepted definition of what constitutes an ANN. However, a useful definition is provided by Gurney [2]:

> "... an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the interunit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns."

2.2 Biological inspiration

The origins of ANNs lie primarily in the world of cognitive science. The human brain is an extremely complex system, capable of remembering, thinking, and problem solving. The neuron is the fundamental unit of the brain and receives and combines signals along input paths (dendrites) from other neurons. If the combined signal is sufficiently strong, the neuron "fires" and an output is transmitted via the axon to other neurons. Any signal entering a neuron passes through a synaptic junction. This is a minute gap in the dendrite, filled with neurotransmitter fluid that accelerates or retards the electrical flow [2]. Adjustment to the impedance of the synaptic gap is a critical process linked with memory and learning. Increasing the synaptic strengths allows the brain to learn or retain information (Figure 1).

2.3 McCulloch and Pitts neuron

While the inspiration may be biological, the development of ANNs has blossomed with the advances made in computing. The origins of ANNs are generally traced back to the work of McCulloch and Pitts [10]. They recognized that combining together many simple neurons was a potential source of increased computational power. The activation of their neuron was binary and it either fired (activation of 1) or did not fire (activation of 0) (Figure 2).

McCulloch–Pitts neurons were connected to other neurons by directed weighted paths. A connecting path excited the connection if the weight was positive;

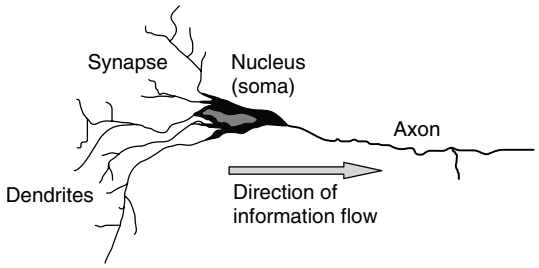


Figure 1. Illustration of a biological neuron.

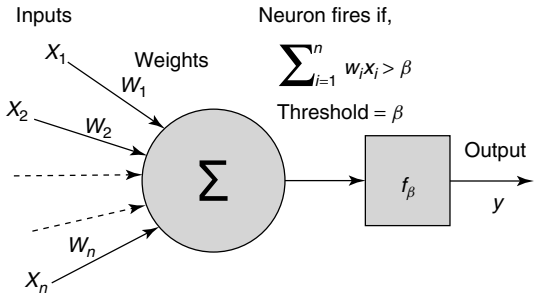


Figure 2. Representation of a simple McCulloch–Pitts Neuron.

otherwise it was inhibitory. All excitatory connections into a neuron had the same weight. Furthermore, each neuron had a fixed threshold, and if the net input was greater than the threshold, it fired. Also, it took one time step for a signal to pass over one connection link. As a single layer, the McCulloch–Pitts neuron was only able to solve linearly separable problems. However, McCulloch–Pitts neurons are still used widely in logic circuits today.

2.4 Hebb's learning rule

In the late 1940s, Hebb [11] developed the first learning rule for ANNs. Hebb stated that if two neurons were active simultaneously, then the strength of the connection between them should be increased. Hebb's rule was adapted by later researchers to allow it to be used in computer simulations.

2.5 Rosenblatt's perceptron

In the early 1960s, the McCulloch–Pitts neuron was developed further by Rosenblatt [12]. Rosenblatt's perceptron networks were composed of an input

layer (retina), a hidden layer (associative layer), and an output layer (decision layer). Rosenblatt's perceptron learning rule used a supervised iterative weight adjustment that used a learning coefficient, a more powerful form of the Hebb rule. Rosenblatt avoided the credit-assignment problem (i.e., how to update the weights between the input and hidden layers) by only allowing the connections between the decision nodes and the associative nodes to be adjusted. The connections between the inputs and hidden (associative) layer were fixed before learning commenced (Figure 3).

The early successes with perceptrons led to enthusiastic claims [6]. However, the problem was that the size of the network grew exponentially with the increased dimensions of the problem. Methods were developed to address this issue with restricted connections between nodes, referred to as *diameter-limited perceptrons*. Then, in the late 1960s, Minsky and Papert [13] published a rigorous investigation into the capabilities of perceptrons and concluded that they were of limited use. Possible solutions proposed included the use of several hidden layers within the perceptron but this approach failed over the credit-assignment problem that Rosenblatt had avoided. The result of Minsky and Papert's book was that nearly all ANN research was abandoned until Hopfield's paper [14] led to a rekindling of interest.

2.6 Hopfield's network

Hopfield [14] used a fully connected recurrent network of McCulloch–Pitts neurons to create a content-addressable memory system. Hopfield considered the output of the neurons as dynamical states and identified that a network could be used to store patterns that were the stable states of the network. A pattern can be retrieved from the network using only a subset of the pure pattern. Hopfield is also largely credited with reviving interest in neural network research.

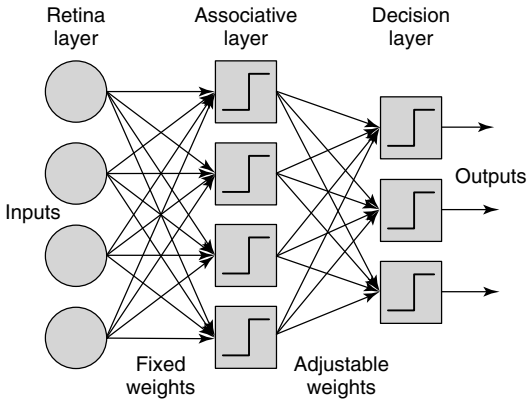


Figure 3. Rosenblatt's perceptron.

2.7 Development of the multilayer perceptron

The development of the MLP is generally attributed to Rumelhart and McClelland [7]. Interestingly, the backpropagation learning algorithm (or the generalized delta rule), the method of solving the credit-assignment problem, was outlined independently by Werbos in his 1974 doctoral thesis (published as [15]). Elsewhere in the literature, the independent development of the backpropagation method is also attributed to Parker [16] and LeCun [17].

3 SUPERVISED AND UNSUPERVISED LEARNING MECHANISMS

3.1 Supervised learning

ANNs can be subdivided by the methods they use to learn relationships. In supervised learning methods, the ANN is presented with inputs and examples of the required outputs, or targets. The inputs are multiplied by weights and biases, which are adjusted by iteration to reduce an error function. This error function is based upon the difference between the desired output and that predicted by the network. The concept is simple but the complexity is in the training algorithm that is required to converge the training to a minimum-error solution efficiently. Well-known examples of supervised learning ANNs include the MLP and RBF. Supervised learning is illustrated in Figure 4.

3.2 Unsupervised learning

Unsupervised learning allows the network to determine the characteristic relationships within the data

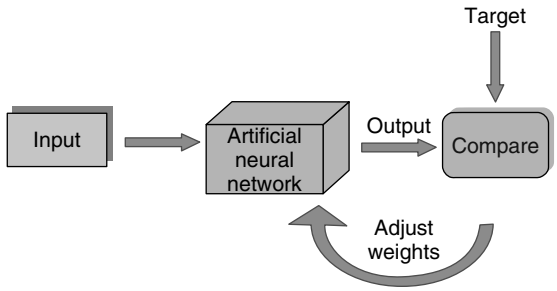


Figure 4. Supervised learning schematic for an ANN.

without reference to target examples. Information about the characteristic features of input data is created during the learning process and stored in the weights. Output signals describe relationships between the current input signals and the weight vectors. Well-known unsupervised techniques include SOFMs and learning vector quantization (LVQ) (incorporating both unsupervised and supervised learning).

4 MULTILAYER PERCEPTRON

4.1 Weights and biases

A simplified schematic of an MLP is illustrated in Figure 5. The normalized inputs are each linked to the hidden layer neurons by a weight vector ($\boldsymbol{w}$) and a bias vector ($\boldsymbol{b}$) is applied to each neuron. These weights and biases can be thought of as analogous to the equation used in linear regression ($y = mx + c$). The weight can be considered as the gradient (m) and the bias as the intercept (c). The network in Figure 5 has been illustrated with one output node; however, many output nodes can be used if required.

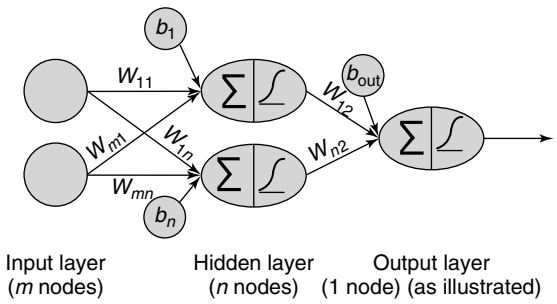


Figure 5. Schematic of simplified MLP–ANN.

4.2 Hidden layer activation functions

An MLP is a feed-forward network and the hidden layer contains a number of neurons that receive weighted data from each input node and pass it to the next layer node, via the activation function. The function of each of these nonlinear activation functions is akin to a mechanical nonlinear gearbox. The relationship between the input and output is nonlinear but constant. In the network illustrated in Figure 5, logistic sigmoid functions are used. Other shaped functions can be used as long as they are continuous, differentiable, monotonically nondecreasing functions. Being able to differentiate the activation function is the key element that separates an MLP from previous developments and the backpropagation of the error (solving the credit-assignment problem) is dependent upon this facet. Consequently, hard-limit functions cannot be used in an MLP.

4.3 Output layer

Data from each neuron in the final hidden layer are passed to the output layer. Here the output layer is a logistic sigmoid function (range 0–1), again with a bias. However, simple linear summation, again with a bias, is often used for an output without range restriction. The outputs of the two-layer network, illustrated in Figure 5, with both layers using logistic sigmoid activation functions would be calculated by using equation (1):

$$\mathbf{y} = \text{logistic}\left[\sum\left\{\boldsymbol{w}_2 \times \left[\text{logistic}\left\{\sum(\boldsymbol{w}_1 \times \mathbf{x} + \boldsymbol{b}_1)\right\}\right] + \boldsymbol{b}_2\right\}\right] \quad (1)$$

where, $\mathbf{y}$ is the output matrix, *logistic* describes the logistic sigmoid activation function (equation 2), $\mathbf{x}$ is the input matrix, $\boldsymbol{w}_1$ is the first-layer weight matrix, $\boldsymbol{w}_2$ is the second-layer weight matrix, $\boldsymbol{b}_1$ is the first-layer bias vector, and $\boldsymbol{b}_2$ is the second-layer bias vector.

4.4 Activation function derivatives

As already discussed, being able to differentiate the activation function is the key element of the

MLP and the importance of this becomes clear in the following sections. The activation functions and derivatives of the most widely used logistic sigmoid, tangent sigmoid, and linear functions are detailed in equations (2–4):

Logistic sigmoid

$$o_{\mathrm{pj}} = f\left(net_{\mathrm{pj}}\right) = \frac{1}{1+\mathrm{e}^{-k\cdot net}}$$
$$f'\left(net_{\mathrm{pj}}\right) = k\cdot o_{\mathrm{pj}}\left(1-o_{\mathrm{pj}}\right) \qquad (2)$$

Tangent sigmoid

$$o_{\mathrm{pj}} = f\left(net_{\mathrm{pj}}\right) = \frac{\mathrm{e}^{k\cdot net}-\mathrm{e}^{-k\cdot net}}{\mathrm{e}^{k\cdot net}+\mathrm{e}^{-k\cdot net}}$$
$$f'\left(net_{\mathrm{pj}}\right) = k\cdot\left(1-o_{\mathrm{pj}}^2\right) \qquad (3)$$

Linear

$$o_{\mathrm{pj}} = f\left(net_{\mathrm{pj}}\right) = net_{\mathrm{pj}}$$
$$f'\left(net_{\mathrm{pj}}\right) = 1 \qquad (4)$$

where w_{ij} is the output from pattern p at node j, f is the internal activation, k is the slope parameter, and $net_{\mathrm{pj}} = \Sigma w_{\mathrm{ij}}\cdot o_{\mathrm{pj}}$, where o_{pj} is the weight from node i to j. This formulation for the derivatives makes the computation of the gradient more efficient since the output has already been calculated in the forward pass. Plots of the logistic sigmoid function and its derivative, used in the MLP illustrated in Figure 5, are produced in Figure 6.

4.5 Backpropagation training algorithm

The aim of the training algorithm is to reduce the error function to a minimum. This error function is usually based upon the difference between the desired output (target) and that predicted by the ANN, sometimes with the addition of a regularization term to prevent weights from growing too large. Training the ANN can best be visualized by using a simplified two-dimensional error surface, illustrated in Figure 7. Here the aim is to reach the global minimum value in the error surface without getting stuck in local minima or saddle points.

There is insufficient space here to show the derivation of the backpropagation algorithm and hence the solution is merely stated using the notation reproduced by Beale and Jackson [3]. Firstly, the network output from the initialized weights and input values is calculated and compared with the desired output. Then the weights are adjusted starting from the output layer and working backward using equation (5):

$$w_{\mathrm{ij}}(t+1) = w_{\mathrm{ij}}(t) + \eta\delta_{\mathrm{pj}}o_{\mathrm{pj}} \qquad (5)$$

where $w_{\mathrm{ij}}(t)$ represents the weights from node i to node j at time t, η is the learning rate (0–1), which is

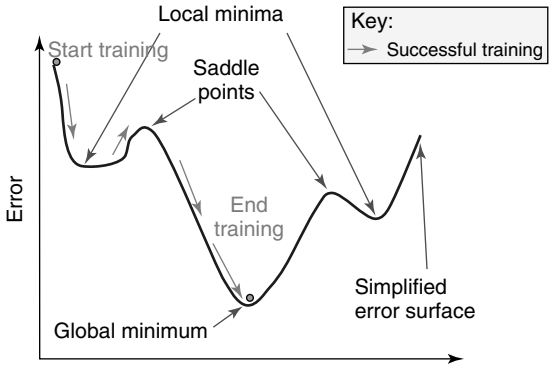


Figure 7. Simplified schematic of error surface.

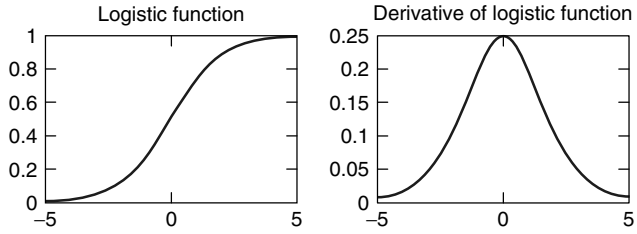


Figure 6. Logistic activation function and derivative.

effectively a gain term, δ_{pj} is an error term for pattern p on node j and o_{pj} is the output from pattern p at node j. For the output units the error term is expressed by equation (6):

$$\delta_{pj} = f'(net_{pj})(t_{pj} - o_{pj}) \tag{6}$$

For a logistic sigmoid activation function (equation 2), this would be represented by equation (7):

$$\delta_{pj} = k \cdot o_{pj}(1 - o_{pj}) \cdot (t_{pj} - o_{pj}) \tag{7}$$

For the hidden units, the error term is expressed by equation (8):

$$\delta_{pj} = f_j'(net_{pj}) \cdot \sum_k \delta_{pk} w_{jk} \tag{8}$$

where the sum is over the k nodes in the layer above node j. Again, for a logistic sigmoid activation function this would be represented by equation (9):

$$\delta_{pj} = k \cdot o_{pj}(1 - o_{pj}) \cdot \sum_k \delta_{pk} w_{jk} \tag{9}$$

The importance of being able to calculate the derivatives of the activation functions (equations 2–4) is now evident. Using these relationships, the network can either be trained in batch mode, where the weight updates are calculated after the entire training data set has passed through the network, or in sequential mode, where weight updates are carried out after the presentation of each training example.

4.6 Improvements to the backpropagation algorithm

There is a plethora of methods in the literature for improving the performance of the backpropagation algorithm. The addition of a momentum term is widely referenced; this allows previous updates to persist for a period of time. Hence, a large change of weights will occur if the changes are large and these will decrease as the changes become smaller. This means that the network is less likely to get stuck in local minima early in the training, as the momentum will push the changes over local increases in the error (or energy) function [8]. The inclusion of a momentum term modifies equation (5) to produce equation (10):

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj} + \alpha \Delta w_{ij}(t) \tag{10}$$

where α is the momentum coefficient and $0 \leq \alpha \leq 1$.

Other heuristic methods for improving the performance of MLPs with the backpropagation algorithm are described in the literature [6, 8, 9, 18, 19]; many of these are attributed to Jacobs [20] or seek to address the limitations identified by Jacobs and are based around adaptation of learning.

4.7 Second-order training algorithms

Although these developments, which are largely heuristic, have been shown to improve the speed of convergence, it is reported that the information provided by the first-order local error gradient can sometimes be an oversimplification of an often highly complex error surface [8]. Higher-order information may be required to produce significant improvements in the convergence performance of the MLP [9]. Here the problem becomes one of numerical optimization and the main methods used are based around conjugate gradient methods and quasi-Newtonian methods.

In the conjugate gradient algorithms, a search is performed along conjugate (i.e., not interfering with each other) directions. For most conjugate gradient algorithms, the length of weight update, or step size, is adjusted after each iteration of the algorithm. The initial step is taken in the steepest descent direction. Then a line search is performed to determine the optimum distance to move along the current search direction. The next search direction is then determined so that it is conjugate to previous search directions. The general procedure for determining the new search direction is to combine the new steepest descent direction with the previous search direction [8]. The various versions of the conjugate gradient method are distinguished by the way they calculate the step size and the step direction [18, 21].

The basis of Newtonian methods is that the change with the next step of the algorithm is equal to the product of the gradient of the error surface and the inverse Hessian matrix [9], where the Hessian matrix is the second derivative of the error

surface. However, calculation of the inverse Hessian matrix is extremely computationally expensive for large networks, and, in many cases, it may not be computable (as the Hessian must be nonsingular and positive definitive). Hence, in quasi-Newtonian (or secant) methods, an estimate of the Hessian matrix is calculated at each iteration of the algorithm, using the error surface gradient information. This estimation reduces the computational effort significantly and eliminates the risk of the inverse Hessian being incomputable [8].

Alternative approaches have also been suggested. These include the methods developed by MacKay [22]. In this method, a Bayesian framework is established and the weights and biases of the ANN are assumed to be random variables with specific distributions. The regularization parameters are related to the unknown variances associated with these distributions. Estimates of these parameters are produced using statistical techniques during training.

4.8 Limitations of MLP with backpropagation training

Although the network training may have arrived at a local minimum value, there is no guarantee that it is the absolute or global minimum (Figure 7). Consequently, the literature is brimming with detail on the risks of training halting in local minima and methods to avoid this, although, in practice, by training a reasonably large number of models and with the use of the large data sets typical of SHM applications, the risk of training halting in local minima is often found to be small [23].

Another curse of the MLP is overfitting. This occurs when the ANN no longer represents the underlying relationships in the data. The implication is that the ANN maps well the relationships within the training data, but fails to produce similar results on previously unseen data and fails to produce a generalized solution. This was illustrated by the now famous ANN tank detection photographic software that turned out to be a weather forecaster [24]. The literature is full of cautionary examples of the risks of overfitting and how to avoid it, such as early stopping (or test-error activation) techniques, regularization, or the addition of noise. In practice, the risk is very problem dependent, and where large training data sets and relatively small networks are used, the risk is often small [8].

4.9 Use of training, test, and validation data sets

Data are divided into training, test, and validation sets. Training data are used to optimize the weights and biases within the model. The statistically similar test data are used to ensure the solution is not overfitting to the training data (good generalization). Thereafter, the validation data are used for independent assessment of the performance of the model and to identify solution confidence limits.

5 RADIAL BASIS FUNCTIONS

Whereas the MLP applies a method of optimization by adjusting weights to vary the network response, the RBF is more of a curve-fitting approximation in high-dimensional space. The learning procedure involves finding a surface in multidimensional space that best approximates the training data. Considering the structure of a neural network in this case, the hidden units provide a set of functions that provide an arbitrary basis for the input patterns when they are expanded into the hidden-unit space. Hence these functions are termed *radial basis functions*.

In its most basic form, the RBF network consists of three entirely different layers—the input layer made up of source nodes (sensory units), the hidden layer (which serves a different purpose from that in an MLP), and the output layer (which supplies the response of the network). A schematic of an RBF network is shown in Figure 8. As for the MLP, the function contained within the RBF units can be varied. For example, these functions could be Gaussian, multiquadratic, inverse multiquadratic, thin plate spline, or r4 log r functions [9, 25].

The function of an RBF net can be understood directly by considering each of the nodes in the hidden layer. Each one contributes a "hump", possibly Gaussian, in n dimensions. This hump is weighted before being blended with others in the network at the output. Therefore, quite complex decision regions may be built up from relatively few nodes. An example is illustrated in Figure 9.

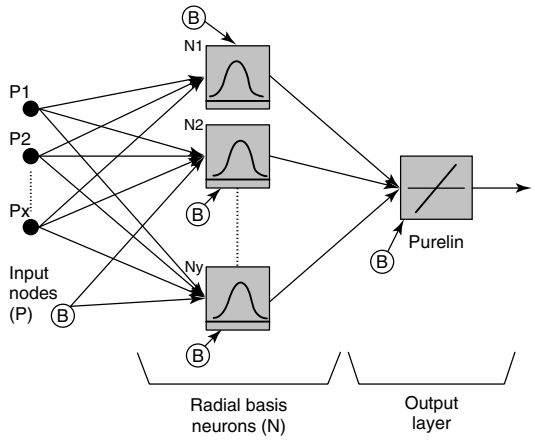


Figure 8. Example radial basis function network.

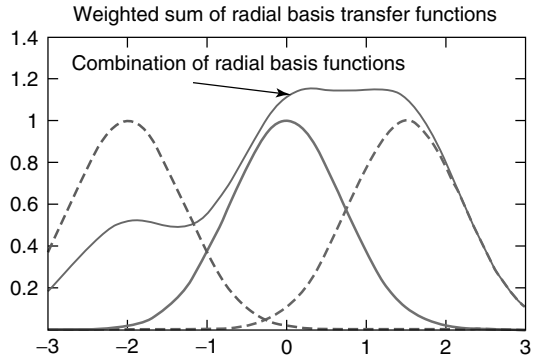


Figure 9. Creation of complex function from Gaussian distributions.

Training an RBF can be considerably faster than training MLP networks. The process usually consists of a two-stage training procedure. In the first stage, the parameters governing the basis functions are determined using relatively fast, unsupervised methods, i.e., without reference to target data. The second stage of the training then involves the determination of the final layer weights, which requires the solution of a linear problem and is therefore also fast. A set of vectors is applied to represent the input data, usually by minimizing the Euclidean distance. A set of weights is then simply adjusted to combine these vectors to represent the target. The Euclidean distance function is sometimes referred to as the *radial distance* and is the magnitude of the difference between the input vector and the locations of the centers of the radial functions. MLP and RBF networks play very similar roles in that they both provide techniques for approximating nonlinear relationships and both are feed-forward networks. They are also the most common forms of supervised networks in use.

6 AUTOASSOCIATIVE NETWORKS

An autoassociative network is a modified MLP whereby the network is trained to predict the inputs. However, the number of neurons in the hidden layer or layers is significantly less than the number of inputs (Figure 10—in this case, $\mathrm{N}y \ll \mathrm{P}x = \mathrm{M}x$). Hence, the network is forced to learn the key underlying relationships within the data.

When used as a data novelty detector [26] (*see* **Novelty Detection**), new input data are presented to this network and the inputs are predicted using the trained network. If the error in the prediction is significantly larger than the error seen in training, then this indicates that the new input data are outside of the experience of the network and are hence novel data.

7 SELF-ORGANIZING SYSTEMS

In self-organized or unsupervised learning, the purpose of the algorithm is to discover significant patterns in the input data without a target set. To achieve this, the algorithm applies rules of a local nature, meaning that changes applied to a neuron, in terms of weight variance, are confined to its immediate neighborhood.

A class of ANNs called *self-organizing feature maps* is based on this premise. These apply competitive learning, where the output neurons compete amongst themselves to be activated or fired. In an SOFM, neurons are placed at the nodes of a lattice that is one or two dimensional and these become tuned to various input patterns. These tuned neuron locations tend to become ordered in such a way that a meaningful coordinate system for different features is created over the lattice. Hence, an SOFM is characterized by the formation of a topographic map in which the spatial locations of the neurons correspond to intrinsic features of the input patterns.

One of the most well known SOFMs is the process developed by Kohonen [27]. The principal goal of

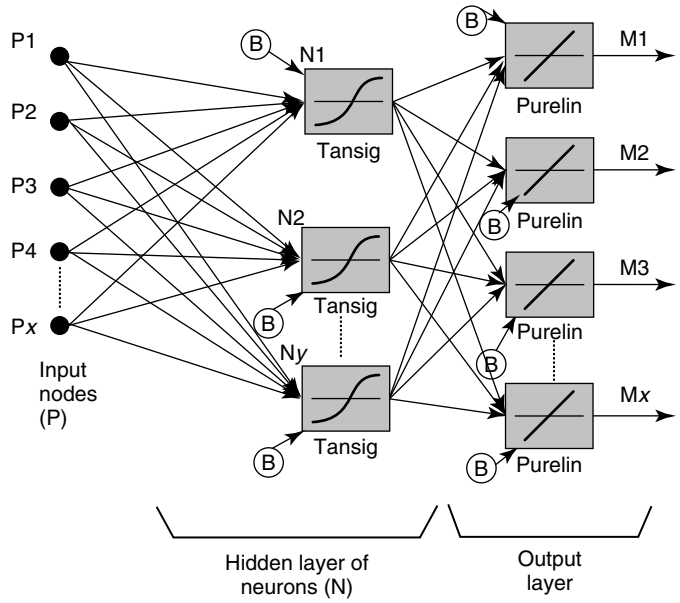


Figure 10. Schematic of an autoassociative network.

the Kohonen SOFM algorithm is to transform an incoming signal pattern of arbitrary dimension into a 1D or 2D discrete map.

8 LEARNING VECTOR QUANTIZATION

An LVQ combines supervised and unsupervised learning methods. The unsupervised competitive layer classifies the input data into a predetermined number of subclasses in the input multidimensional space. This activity is unsupervised, and the mapping is a result of the inherent clustering in the input data. Thereafter, these subclasses are mapped to the target classes, using a supervised learning algorithm. LVQ is a nearest-neighbor classification method that has been studied extensively in the field of pattern recognition [27]. Figure 11 is a simplified schematic of an LVQ.

9 PROBABILISTIC NEURAL NETWORKS

The probabilistic neural network (PNN) [28] is a Bayesian classifier contained within an ANN [5]. An input (feature) vector is used to identify a category and the network classifiers are trained by being shown data of known classifications. The PNN uses these training data to develop distributions that are used to estimate the likelihood of a feature vector being within several given categories. Ideally, this is combined with *a priori* knowledge such as the relative frequency of each category.

10 DYNAMIC NEURAL NETWORKS

Dynamic neural networks include an element of memory that allows the ANN to exhibit temporal behavior that is dependent not only on present inputs but also on prior inputs. There are two major classes of dynamic networks: recurrent neural networks (RNNs) and time delayed neural networks (TDNNs).

10.1 Recurrent neural networks

RNNs contain internal time delayed feedback connections. The two most common designs are the Elman network [29] and the Jordan network [30]. In an Elman network, the hidden-layer outputs are fed back

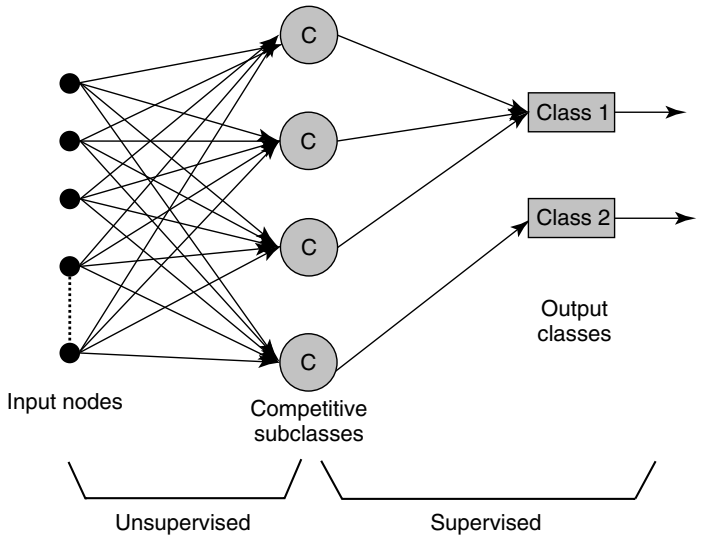


Figure 11. Simplified learning vector quantization example.

through a one-step delay to dummy input nodes. The Elman network can learn temporal patterns as well as spatial patterns because it can store information. The Jordan network is a recurrent architecture similar to the Elman network, but it feeds back the output layer rather than the hidden layer. RNN are more difficult to train than conventional networks because of the feedback connections.

10.2 Time delay neural networks

TDNNs can learn temporal behavior by using both present inputs and past inputs by simply delaying the input signal. The ANN is usually based upon a standard MLP, but it can also be an RBF, a PNN, or any other feed-forward network architecture. Since the TDNN has no feedback terms, it is easily trained with standard training algorithms [5].

11 ADAPTIVE AND NONADAPTIVE SYSTEMS

Adaptive prediction methods continue to train when it is in service, whereas for nonadaptive methods, the weight and bias coefficients are fixed after training. Without formal control over the training data set, currently, certification of such a mechanism for a safety-related application, such as an aircraft SHM system, would be most unlikely [31].

12 APPLICATIONS

ANNs can provide solutions to a variety of SHM-related problems including regression (mapping), classification, detection (abnormality), and autoassociation (reconstruction). This section contains some examples of the application of ANN technology within SHM systems.

12.1 Parametric-based aircraft structural loads and usage prediction

Several researchers in the aeronautical arena have published work using ANNs and similar mathematical techniques to predict stresses/strains or loads from flight parameters, to identify loading actions, or to supplement sparse flight test data. Smiths Aerospace are leading exponents of these techniques and Azzam *et al.* [32–34] and Wallace *et al.* [35] have described the use of a "mathematical network" to predict stresses/strains or loads in a cooperative work program involving Smiths Aerospace, the UK Ministry of Defence, and BAE SYSTEMS.

Successes reported included excellent correlation between measured strains and strains predicted from flight parameters. Cumulative fatigue damages from the predicted strains were reported as being within 3.6% of those calculated from measured strains for two wing locations, the fin, and taileron of the Tornado combat aircraft. These results were obtained from more than 1000 sorties, relating to data spread over 15 years of recording. Azzam [33] also described the development of methods to predict damage in helicopter components and the prediction of high-frequency events, or *rare events* as they termed them, such as buffet loading on the fin of a fixed-wing aircraft.

Reed and Cole [36] and Reed [23, 37, 38] reported the development of ANN-based fatigue monitors predicting strain from flight parameters for a wide range of fatigue critical locations for combat, transport, and trainer aircraft under both maneuver and buffet loading. An example plot comparing a measured strain time history with a predicted strain time history is presented in Figure 12. Fatigue damage calculated from predicted strains was within 6% of that calculated from the measured strains from unseen data.

Recent papers have illustrated the spread of this technology with the development of MLP–ANN parameter-based fatigue monitoring systems proposed for the Airbus A330-MRTT (Multi Role Tanker Transport) [39] and the Boeing F-18 combat aircraft, operated by the Finnish Air Force [40]. Other researchers have explored the use of this technology in discrete fields: the synthesis of helicopter loads was described by Hill *et al.* [41], in which a back-propagation ANN was used to synthesize strain and fatigue damage in several Lynx helicopter components. Manry and coauthors [42] also applied ANN to flight load synthesis, using Bell helicopter data. The prediction of strains in the empennage of a Cessna 172P General Aviation aircraft during discrete maneuvers, using a backpropagation ANN, was reported by Kim and Marciniak [43]. Additionally, Jacobs and Perez [44] used a hybrid cascade ANN to augment sparse buffet data by predicting the power spectral densities of surface pressures on the vertical tail of a fighter aircraft, from parametric data. Finally, Levinski [45] used ANN technology to synthesize loads on the vertical tail of an F/A-18 aircraft, using wind tunnel pressure data.

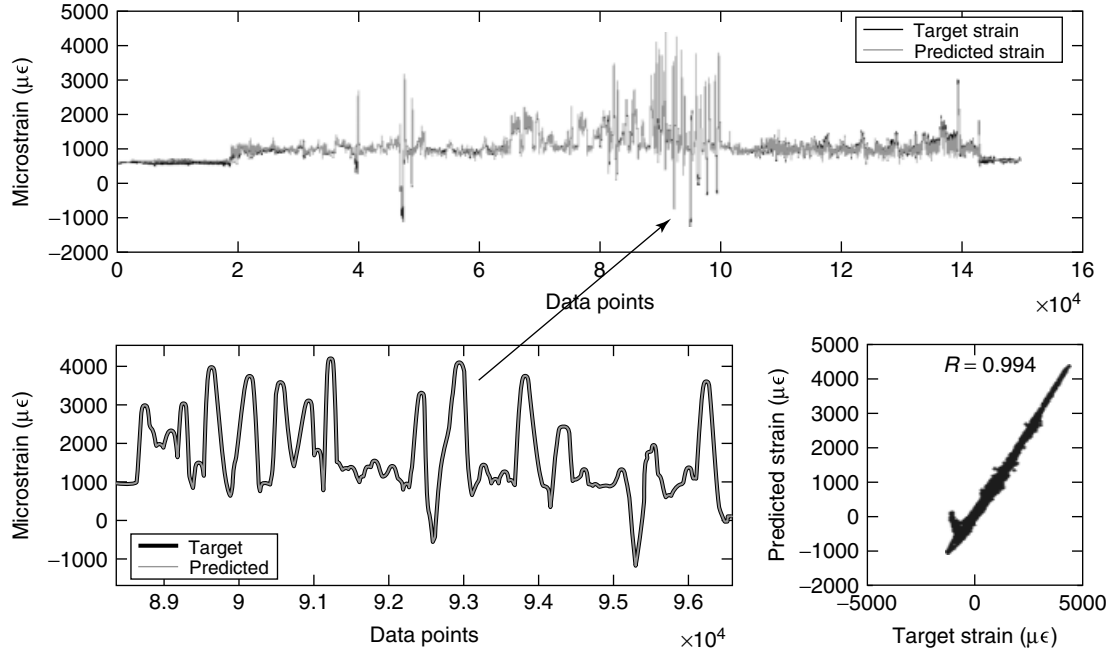


Figure 12. Example wing strain prediction for a trainer aircraft.

12.2 Damage detection in structures

There are a great many papers in the literature describing programs to develop or improve damage detection methods in both metallic and composite structures [46]. Advances in ultrasonic techniques, in particular, promise significant steps forward in this field [47]. ANN techniques are used within many of these techniques, primarily undertaking signal conditioning roles (*see* **Statistical Pattern Recognition**; **Novelty Detection**; **Data Fusion of Multiple Signals from the Sensor Network**).

For example, Lopes *et al.* [48] describe a nonmodel technique used to detect and locate structural damage using ANN. The method used piezoelectric sensors and actuators (*see* **Piezoelectricity Principles and Materials**; **Piezoelectric Wafer Active Sensors**; **Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection**) with high-frequency structural excitation (typically $>30\,\text{kHz}$) to detect changes in structural point impedance caused by the presence of damage. The impedance-based technique was very sensitive to minor changes in the near field of the piezoelectric sensor. However, owing to the difficulties in developing analytical models at such high frequencies, the technique could not correlate a change in electrical impedance with a specific change in structural properties and hence only provided limited information on the nature of the damage. Therefore, a method was developed to combine the impedance-based technique with ANN. Two MLP–ANNs were used to detect, locate, and characterize the structural damage, having been trained upon experimental data. The first ANN was used to identify the presence of damage and a second network was trained to identify particular types of damage, using experimental data. The authors used this technique to locate and identify the damage in a four-bay structure with bolted joints.

Chetwynd *et al.* [49] described a program of work involving the damage detection in a curved carbon-fiber reinforced panel with two omega stiffeners; this was also investigated using ultrasonic Lamb waves. The statistical technique of outlier analysis was used here as a way of preprocessing experimental data, prior to damage classification. MLP–ANNs were used successfully for both classification and regression problems of damage detection. The authors highlighted that the most intensive aspects of creating this SHM system were the collection of the training data and the generation of the optimum MLP network. Compared with these, the time spent in investigating the sensor network with Lamb waves and generating the MLP response was considered negligible. Hence, they concluded that such a system could be effectively used for real-time SHM applications.

REFERENCES

[1] Chang F-K. Structural health monitoring 2000. In *Proceedings of the 2nd International Workshop on Structural Health Monitoring*, ISBN 1-56676-881-0, Chang F-K (ed). Stanford University: Stanford, CA, 1999.

[2] Gurney K. *An Introduction to Neural Networks*, ISBN 1-85728-503-4, *First Edition*. Routledge: London, 1997.

[3] Beale R, Jackson T. *Neural Computing: An Introduction*, ISBN 0-85274-262-2. Institute of Physics Publishing: Bristol, 1990.

[4] Callan R. *The Essence of Neural Networks: From the Essence of Computing Series*, ISBN 0-13-908732X. Pearson Education: Essex, 1999.

[5] Tsoukalas LH, Uhrig RE. *Fuzzy and Neural Approaches in Engineering*, ISBN 0-471-16003-2. John Wiley & Sons, 1996.

[6] Fausett LV. *Fundamentals of Neural Networks, Architectures, Algorithms and Applications*, ISBN 0-13-334186-0. Prentice-Hall: Englewood Cliffs, NJ, 1994.

[7] Rumelhart DE, McClelland JL. *Parallel Distributed Processing, Exploration in the Microstructure of Cognition, Volume 1: Foundations and Volume 2: Psychological and Biological Models*, ISBN 0-262-18120-7 and 0-262-13218-4. MIT Press: Cambridge, MA, 1986.

[8] Haykin S. *Neural Networks: A Comprehensive Foundation*, ISBN 0-13-273350-1, *Second Edition*, Prentice Hall, 1999.

[9] Bishop CM. *Neural Networks for Pattern Recognition*, ISBN 0-19-853864-2. Oxford University Press, 1995.

[10] McCulloch WS, Pitts W. A logical calculus of the ideas imminent in nervous activity. *The Bulletin of Mathematical Biophysics* 1943 **5**:115–133, Reprinted in Anderson JA, Rosenfield E (eds). Neurocomputing: Foundations of Research, Cambridge, MA, 1988, pp. 18–28.

[11] Hebb DO. *The Organization of Behavior*. John Wiley & Sons: New York, 1949.

[12] Rosenblatt F. *Principles of Neurodynamics*. Spartan: New York, 1962.

[13] Minsky ML, Papert SA. *Perceptrons*. MIT Press: Cambridge, MA, 1969.

[14] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America* 1982 **79**:2554–2558.

[15] Werbos PJ. *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*, ISBN 0-471-59897-6. John Wiley & Sons: New York, 1994. (Originally presented as the Author's 1974 PhD Thesis—Beyond Regression).

[16] Parker D. *Learning Logic*, Technical Report TR-87. Centre for Computational Research in Economics and Management Science, MIT: Cambridge, MA, 1985.

[17] Le Cun Y. Learning processes in an asymmetric threshold network. In *Disordered Systems and Biological Organization*, *NATO ASI Series, F20*, Bienenstock E, Fogelman-Souli F, Weisbuch G (eds). Springer-Verlag: Berlin, 1986.

[18] Hagan MT, Demuth HB, Beale MH. *Neural Network Design*, ISBN 0534943322, *First Edition*. Brooks Cole, 1996.

[19] Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, 1993.

[20] Jacobs RA. Increased rates of convergence through learning rate adaptation. *Neural Networks* 1988 **1**:295–307.

[21] Möller MF. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* 1993 **6**:525–533.

[22] MacKay DJC. *Information Theory, Inference and Learning Algorithms*, ISBN 0521-64298-1. Cambridge University Press, 2003, pp. 527–534.

[23] Reed SC. Development of a parametric-based indirect aircraft structural usage monitoring system using artificial neural networks. *The Aeronautical Journal*, Paper No. 3078, 2007 **111**(1118):209–230.

[24] Swingler K. *Applying Neural Networks—A Practical Guide*, ISBN 0-12-679170-8. Morgan Kaufman Publishers: San Francisco, CA, 1996.

[25] Nabney IT. *Netlab: Algorithms for Pattern Recognition—Advances in Pattern Recognition*, ISBN 1-85233-440-1. Springer, 2002.

[26] Worden K. Structural fault detection using a novelty measure. *Journal of Sound and Vibration* 1997 **201**(1):85–101.

[27] Kohonen T. *Self Organization and Associative Memory*, ISBN 0-387-18314-0, *Second Edition*. Springer-Verlag: Berlin, 1987.

[28] Specht D. A general regression neural network. *IEEE Transactions on Neural Networks* 1991 **2**(5):568–576.

[29] Elman JL. Finding structure in time. *Cognitive Science* 1990 **14**:179–211.

[30] Jordan M. Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Annual Conference on Cognitive Science Society*. Amherst, 1986; pp. 531–546.

[31] UK MoD. *Structural Monitoring Systems Using Non-Adaptive Prediction Methods, Design and Airworthiness Requirements for Service Aircraft*, Defence Standard 00-970, Part 1, Issue 5, Section 3.2.29–3.2.55 and Leaflet 42, 2007.

[32] Azzam H. A practical approach for the indirect prediction of structural fatigue from measured flight parameters. *Proceedings of the Institution of Mechanical Engineers. Part G, Journal of Aerospace Engineering* 1997 **211**(G1):29–38.

[33] Azzam H. The use of mathematical models and artificial intelligence techniques to improve hums prediction capabilities. *Proceedings of the Royal Aeronautical Society, Innovation in Rotorcraft Technology Conference*. London, 1997.

[34] Azzam H, Hebden I, Gill L, Beavan F, Wallace M. Fusion and decision making techniques for structural prognostic health management. *IEEE Aerospace Conference*, Paper #1535. Montana, MT, 2005.

[35] Wallace M, Azzam H, Newman S. Indirect approaches to individual aircraft structural monitoring. *Proceedings of the Institution of Mechanical Engineers. Part G, Journal of Aerospace Engineering* 2004 **218**:329–346.

[36] Reed SC, Cole DG. Development of a parametric aircraft fatigue monitoring system using artificial neural networks. In *Proceedings of the 22nd Symposium of the International Committee on Aeronautical Fatigue*, Lucerne, ISBN 0954345428, Guillaume M (ed). EMAS Publishing: Sheffield, 2003, pp. 47–68.

[37] Reed SC. A parametric-based empennage fatigue monitoring system using artificial neural networks. In *Proceedings of the 23rd Symposium of the International Committee on Aeronautical Fatigue*, Hamburg, ISBN 3-932182-42-1, Dalle Donne C (ed). DGLR-Breicht, 2005, pp. 693–704.

[38] Reed SC. Indirect aircraft structural monitoring using artificial neural networks. *Proceedings of the Structural Health Monitoring and Damage Prognosis of Metallic and Non-Metallic Structures*. Institution of Mechanical Engineers: University of Sheffield, 18 October 2006.

[39] Gómez-Escalonilla J, García J, Cabrejas J, Armijo JI. A full scale parametric-based fatigue monitoring system using neural networks. Presented at *The 24th Symposium of the International Committee on Aeronautical Fatigue*, Naples, 2007.

[40] Tikka J, Salonen T. Parameter based fatigue life analysis for F-18 aircraft. Presented at *The 24th Symposium of the International Committee on Aeronautical Fatigue*. Naples, 2007.

[41] Hill K, Hudson RA, Irving PE, Vella AD. Loading spectra, usage monitoring and prediction of fatigue damage in helicopters. *Proceedings of the 18th Symposium of the International Committee on Aeronautical Fatigue*. Melbourne, 1995.

[42] Manry MT, Hsieh CH, Chandrasekaran H. Near-optimal flight load synthesis using neural networks, *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*. Madison, WI, 1999.

[43] Kim D, Marciniak M. *A Methodology to Predict the Empennage In-Flight Loads of a General Aviation Aircraft Using Backpropagation Neural Networks*, DOT/FAA/AR-00/50, Washington, DC, 2001.

[44] Jacobs JH, Perez PA. Combined approach to buffet response analysis and fatigue life prediction, NATO AGARD report 797—an assessment of fatigue damage and crack growth prediction techniques. Papers presented at *The 77th Meeting of the AGARD Structures and Materials Panel*. Bordeaux, 1993.

[45] Levinski O. Australian Defense Science and Technology Organization. *Prediction of Buffet Loads using Artificial Neural Networks*, Document DSTO-RR-0218, 2001.

[46] Staszewski WJ, Worden K, Signal processing for damage detection. In *Health Monitoring of Aerospace Structures—Smart Sensor Technologies and Signal Processing*, ISBN 0-470-84340-3. Staszewski WJ, Boller C, Tomlinson GR (eds). John Wiley & Sons: Chichester, 2004, pp. 163–206.

[47] Culshaw B, Pierce SG, Staszewski WJ. Condition monitoring in composite materials—an integrated systems approach. *Proceedings of the Institution of Mechanical Engineers, Journal of Systems and Control Engineering* 1998 **213**(3):189–202.

[48] Lopes V, Park Jr G, Cudney HH, Inman DJ. Smart structures health monitoring using artificial neural network. In *Structural Health Monitoring 2000*, ISBN 1-56676-881-0, Chang FK (ed). Technomic Publishing: Pennsylvania, PA, 2000, pp. 976–985.

[49] Chetwynd D, Mustapha F, Worden K, Rongong JA, Pierce SG, Dulieu-Barton JM. Damage localisation in a stiffened composite panel, Strain: *An International Journal for Experimental Mechanics* 2008 **41**:117–127.

Chapter 37

Data Fusion of Multiple Signals from the Sensor Network

Zhongqing Su[1], Xiaoming Wang[2] and Lin Ye[3]

[1] *Department of Mechanical Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong, China*
[2] *CSIRO Sustainable Ecosystems, Commonwealth Scientific and Industrial Research Organisation, Melbourne, VIC, Australia*
[3] *School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, Sydney, NSW, Australia*

1 INTRODUCTION

Structural health monitoring (SHM) is somewhat like detective work to catch the *criminal*—the damage to a structure. We capture in-field signals using a sensor or sensor network, canvas signals with appropriate processing tools, and propose a suitable model that enables us to infer damage parameters based on extracted signal features. Such an inferential procedure is *data fusion*. Literately, *data fusion* is a *multilevel* and *multifaceted* process of combining *multiple* features extracted from a *multitude* of spatially distributed *independent* sources, so as to provide capabilities in automatic detection, classification, and identification [1]. An appropriate data fusion can provide a pathway to quantitatively describe a damage event (appearance, location, and severity) and further evaluate the health status of the structure under inspection.

Implied by the definition, extracting appropriate signal features is the prerequisite for correct data fusion. Though application specific, the basic principle in extracting signal features for data fusion is to select those that are most sensitive to the occurrence and alteration of structural health status. There are other relevant articles on this topic in this encyclopedia (*see* **Signal Processing for Damage Detection**; **Nonlinear Features for SHM Applications**; **Novelty Detection**).

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 DATA FUSION ARCHITECTURE

A single sensor performing local data acquisition may easily fail to provide sufficient information concerning structural health status because of uncertainty, imprecision, and/or incompleteness, thereby eroding confidence in the evaluation results. A number of spatially distributed sensors form a sensor network by *communicating* with each other or through a centralized controller, to render the

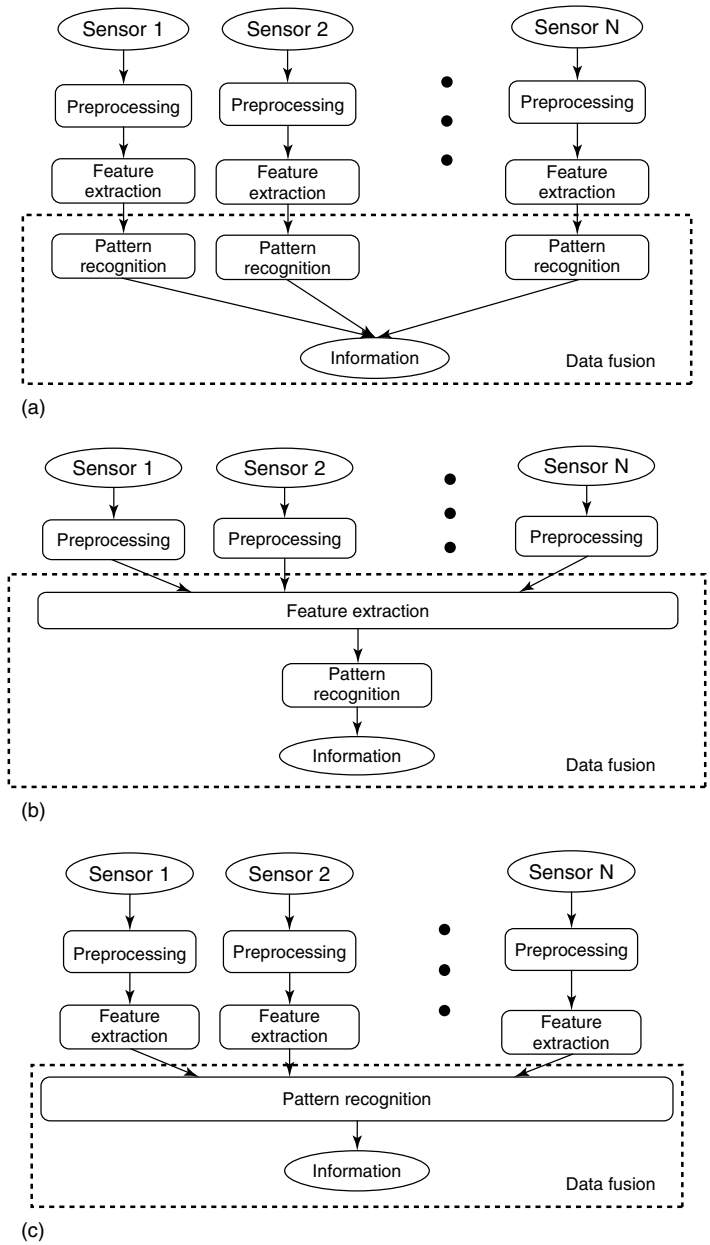


Figure 1. Major data fusion strategies: (a) independent sensor architecture; (b) centralized fusion architecture; and (c) decentralized fusion architecture [3].

possibility of aggregating and fusing all captured information. Basically, data fusion amalgamates redundant or/and complementary data from a multitude of sensors with the objective of making a robust and reliable decision. In general, the currently prevailing data fusion approaches for damage identification use one of three architectures, which are collated in Figure 1. The *independent sensor architecture* is the simplest scheme, where each sensor extracts signal features and carries out recognition independently; the *centralized fusion architecture* is used to fuse the information from data preprocessing and then extract features for future pattern recognition; and the *decentralized fusion architecture* executes feature extraction and selection for each sensor independently, and then the data are fused at the pattern recognition level to deliver a posterior recognition. In practice, various combinations of these three architectures such as the Waterfall and Omnibus strategies [2] are possible for different number of sensors.

3 DATA FUSION ALGORITHMS

Data fusion is implemented through a specific algorithm. A popular taxonomy for major data fusion algorithms currently available is displayed in Figure 2 [1], including mainly *physical models, feature-based inference techniques,* and *cognitive-based models*. In particular, the feature-based inference plays a dominating role and is widely adopted by the SHM community. It performs classification and identification by mapping data (such as statistical knowledge about an object or recognition of an object) into a declaration of identity, using either *parametric techniques* or *information theoretic techniques*. *Parametric techniques* directly map parametric data (such as signal features) into a declaration of identity, independent of a physical model. Some representative approaches under this category are *classical inference*, *Bayesian inference* (*BI*), *Dempster–Shafer inference*, and *generalized evidence processing*. On the other hand, *information theoretic*

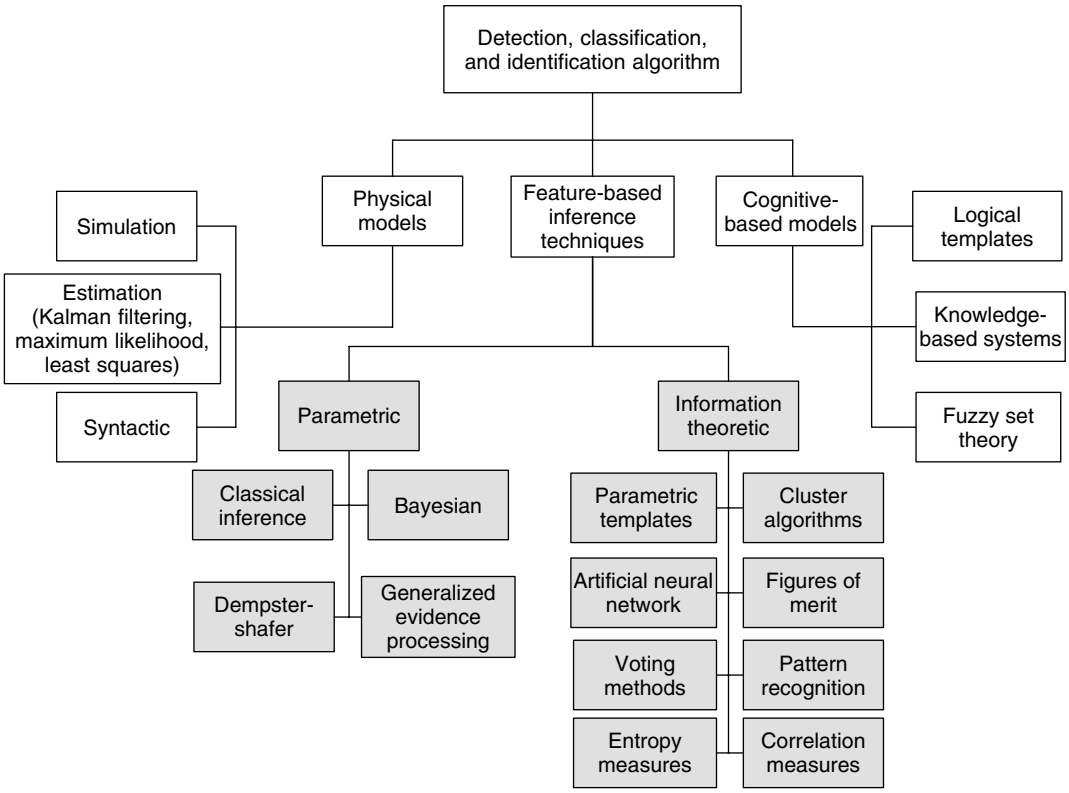


Figure 2. Taxonomy of detection, classification, and identification algorithms for data fusion [1].

Table 1. Comparison of major algorithms of detection, classification, and identification for data fusion

	Algorithm	Attributes
Parametric techniques	Classical inference (CI)	CI, based on a multitude of observations from different sources, gives the probability that an observation can be attributed to the presence of an object or event, given an assumed hypothesis. The major disadvantages of the algorithm include the difficulty in obtaining the probability density function that describes the observable used to classify the object, and complexities that arise when multivariate data are encountered. It has limitation in assessing multiple hypotheses and inability to take direct advantage of a prior likelihood probabilities. Examples can be referred to in [1].
	Bayesian inference (BI)	BI is a probability-based data fusion discipline. It uses prior knowledge to estimate and update the conditional posterior probability of a hypothesis or identify an event when given supporting evidence. As evidence accumulates, the degree of belief in the hypothesis becomes high or low, and it persistently modifies the perceptions when additional information becomes available. The hypothesis is accepted as true with a very high degree of belief or rejected as false when the degree of belief goes very low. Such a fusion algorithm can be used to assess more than two hypotheses. Its major disadvantages include the difficulty in defining prior probability functions, complexities when there are multiple potential hypotheses, mutual exclusivity required of competing hypotheses, and inability to account for general uncertainty. Examples can be referred to in [1, 4].
	Dempster–Shafer method (DSM)	DSM was developed under the *theory of evidence*, to deal with the degree of belief extracted from multiple evidence sources. Such a rule may represent both uncertainties and imprecision. The degree of belief on a hypothesis is described by a mass function. The fusion process takes a union of the mass of all similar beliefs among sensors through conjunction, also considering information conflict. The major disadvantages of the algorithm include the need to define the processes in each sensor that assign mass functions representing the degree of beliefs for a hypothesis. Examples can be referred to in [1, 3, 4].
Information theoretic techniques	Parametric templates (PT)	In this algorithm, multisensor data acquired over a period of time and multisource information are matched with preselected scenarios, models, or knowledge to determine if the observations contain evidence to identify an entity. The algorithm can be applied to event detection, situation assessment, and single object identification. Examples can be referred to in [1, 5].
	Artificial neural network (ANN)	ANN is a mathematical model or computational model based on biological neural networks, which is trained to map input data into selected output categories. The transformation of the input data into output classifications is performed by artificial neurons that attempt to emulate processes that occur in biological nervous systems. By fusing data from a multitude of sources, such an algorithm can be used to model complex relationships between inputs and outputs or to find patterns in data. Examples can be referred to in [3, 6, 7].
	Cluster algorithm (CA)	CA groups data into natural sets or clusters that are interpreted by an analyst to see if they represent a meaningful object category with similar characteristics or natures. All CAs need a similarity measure that described the closeness between any two feature vectors. A CA basically includes steps of sample selection, definition of features, computation of similarities among the data, use of a cluster analysis method to create groups of similar entities, and validation of resulting cluster solution. Examples can be referred to in [1].
	Voting method (VM)	VM fuses information from multiple sensors by weighting each sensor's belief as a vote to reach overall decision. The weight for each vote may be subjective. Examples can be referred to in [1, 4].

Table 1. (*continued*)

Algorithm	Attributes
Entropy measuring (EM)	EM measures the importance of an event by its probability of occurrence. Frequently occurring events or data are of low value while rare events or data are of higher values. By fusing all available information, EM can be used to measure the similarity of an event with regard to other similar events. Examples can be referred to in [1].
Figures of merit (FoM)	FoM is metrics derived from plausible or heuristic arguments that aid in establishing a degree of association between observations and object identity. It is often used to quantitatively characterize the probability of occurrence of an event or similarity with regard to other similar events, by fusing information concerning several variables. For damage identification, FoM is often used to indicate the probability that a damage event occurs, where, when FoM exceeds a critical value, damage is suspected. Examples can be referred to in [1, 8].
Pattern recognition (PR)	Pattern recognition is the identification of an individual object based on either *a priori* knowledge or on statistical information extracted from a series of patterns [7], and the term *pattern* is referred to as an entity to represent an abstract concept or a physical object. It is a field in the area of machine learning. Since diverse damage cases can be regarded as patterns that present various symptoms in a structure, identification of damage falls into the category of pattern recognition. Statistical and syntactic methods are two major implementations of pattern recognition. Statistical modeling is based on the statistical characterizations of patterns, generated by a probabilistic system using statistical density functions; syntactic recognition classifies data based on structural interrelationships of features and is not commonly used for damage identification. Examples can be referred to in [7].
Correlation measures (CMs)	In probability theory and statistics, correlation indicates the strength and direction of a linear relationship between two random variables. In general statistical usage, correlation reflects the independence of two variables. CMs can be used for detection, classification, and identification. Examples can be referred to in [3, 9].

techniques transform or map parametric data into an identity declaration based on the concept that similarity in identity is reflected by similarity in observable parameters, and no attempt is made to directly model the stochastic aspects of the observables [1]. Approaches under this category include *parametric templates*, *artificial neural network*s (*ANN*s) (*see* **Artificial Neural Networks**), *cluster algorithms, voting method, entropy measures, figures of merit, patter recognition*, (*see* **Statistical Pattern Recognition**) and *correlation measure*s (*CM*s). Attributes of the aforementioned major algorithms are compared in Table 1.

In addition, the cognitive-based models attempt to emulate the decision-making processes used by human analysts; representative approaches under this category include *logical templates, knowledge-based systems,* and *fuzzy set theory* [1].

4 EXAMPLES OF DATA FUSION FOR SHM

Various data fusion algorithms have been employed in different damage identification techniques. Four representative examples are presented here.

4.1 Fusion using independent sensor architecture

In this case study, each sensor independently extracts signal features (time-of-flight (ToF)) at the sensor level in the form of *independent sensor architecture* as illustrated in Figure 1(a). ToF is the time lag between two compositions in a signal, which is one of the most explicit and important features in a signal

for SHM. In terms of ToFs between the damage-scattered and incipient energy measured from a multitude of actuator–sensor pairs, the damage can be triangulated by referring to a benchmark. Pitch–catch and pulse–echo are two major implementations of ToF-based approaches, which measure forward and back scattering of incident energy from the damage, respectively [10]. Structural complication and noise often impairs the reliability and accuracy of the triangulation exercise. Data fusion by combining all available information from the sensor network provides an effective means of dealing with this problem.

Figure 3(a) [9] shows an example of a sensor network consisting of 12 lead zirconate titanate (PZT) elements and embedded in a composite laminate containing delamination and offering $12 \times 11 = 132$ sensing paths (*see* **Piezoceramic Materials—Phenomena and Modeling**; **Sensor Network Paradigms**; **Integrated Sensor Durability**

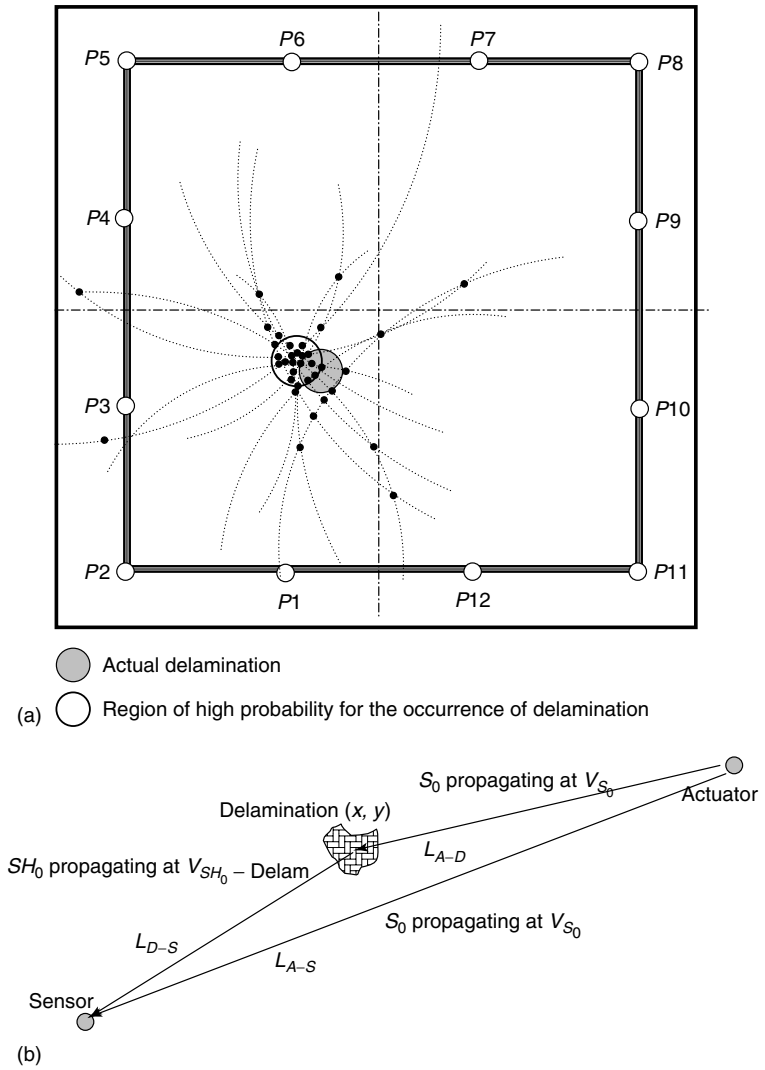


Figure 3. Damage triangulation in terms of ToF: (a) sensor arrangement and identification results and (b) spatial relationship of the actuator, sensor, and damage [9].

and Reliability). In the network, any one pair of PZT elements configures an actuator–sensor path, to generate and collect Lamb wave signals. A coordinate system is introduced for each path, where the actuator is at the origin. In the case that transducer $P1$ acts as the actuator and the damage center is presumed to be (x, y), it has

$$\frac{L_{A-D}}{V_{S_0}}+\frac{L_{D-S}}{V_{SH_{0\text{-Delam}}}}-\frac{L_{A-S}}{V_{S_0}}=T_{1-i},\ (i=2,3,\ldots,12)$$

$$L_{D-S}=\sqrt{(x-x_i)^2+(y-y_i)^2},\ \ L_{A-D}=\sqrt{x^2+y^2},$$

$$L_{A-S}=\sqrt{x_i^2+y_i^2}\ (i=2,3,\ldots,12) \qquad (1)$$

as elucidated in Figure 3(b), where L_{A-D}, L_{D-S}, and L_{A-S} represent the distances between $P1$ and the damage center (x, y), the damage center and the ith sensor, and $P1$ and the ith sensor, respectively. $V_{SH_0-\text{Delam}}$ and V_{S_0} are the velocities of the damage-induced fundamental shear-horizontal Lamb mode (SH_0) and the incipient fundamental symmetric Lamb mode (S_0), respectively. T_{1-i} denotes the ToF extracted from the signal acquired by sensing path $P1$-Pi; and (x_i, y_i) stands for the coordinates of the ith transducer.

Analogously, repeating the above analysis for all sensing paths in the sensor network, using equation (1), where each actuator, in turn, serves as the origin in its coordinate system, a nonlinear equation group is established. For each equation (quadratic) in the group, the solutions configure a circular root trajectory, indicating possible locations of the damage. Any two trajectories lead to an intersection and the region containing more interactions has a higher likelihood of occurrence of damage. Prediction results of the location of delamination in the composite laminate using several major actuator–sensor paths in the sensor network are shown in Figure 3(a).

In the above example, the procedure of two loci of roots leading to an intersection and indicating a possible damage location is considered as a low-level data fusion based on two pairs of actuator–sensor paths. As the number of sensors increases, a higher level data fusion is executed. In such a process, the prior beliefs or perceptions at low levels are fused into a posterior belief. As a consequence, the reliability of damage prediction is enhanced. Furthermore, the spatial distribution of the prior beliefs can become the basis for assessing the damage size and severity.

4.2 Fusion using correlation measures

In this case study, data fusion using CMs picks up the location of structural damage in terms of damage-induced singularity in the signals. The correlation coefficient, λ_{xy}, of two discretized signals in the same length, x_i and y_i $(i = 1, 2, \ldots, n)$, is defined as [9]

$$\lambda_{xy}=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{n\sum_{i=1}^{n}x_i{}^2-\left(\sum_{i=1}^{n}x_i\right)^2}\cdot\sqrt{n\sum_{i=1}^{n}y_i{}^2-\left(\sum_{i=1}^{n}y_i\right)^2}}\leq 1 \qquad (2)$$

When composition in signal x_i is correlated (similar) with any in signal y_i, the correlation coefficient reaches a local extremum; and the more the similarity is, the closer to 1 is the value of the coefficient. In the case that $x_i = y_i$, the correlation is autocorrection. The autocorrelation curves of two signals acquired from a carbon fibre/epoxy (CF/EP) composite panel before and after the occurrence of delamination are compared in Figure 4 [9]. The time lag between the moment that singularity takes place and the moment that incipient energy is generated by the actuator, ΔT, can be exactly determined from the ratio curve, Figure 4(c). An exceptional advantage of the approach is the weakening of boundary effect, since the same boundary reflection remains in all signals, regardless of the existence of damage.

In such a procedure, signals captured by individual sensors are correlated with corresponding signals from the benchmark, considered as the prior beliefs. By taking into account all the prior beliefs from sensors in the network, an overall belief or a posterior belief about the difference between the structure under inspection and benchmark is depicted.

From the above two examples, it can be seen that, upon extraction of interested signal features such as ToF or correlation coefficients, distributed

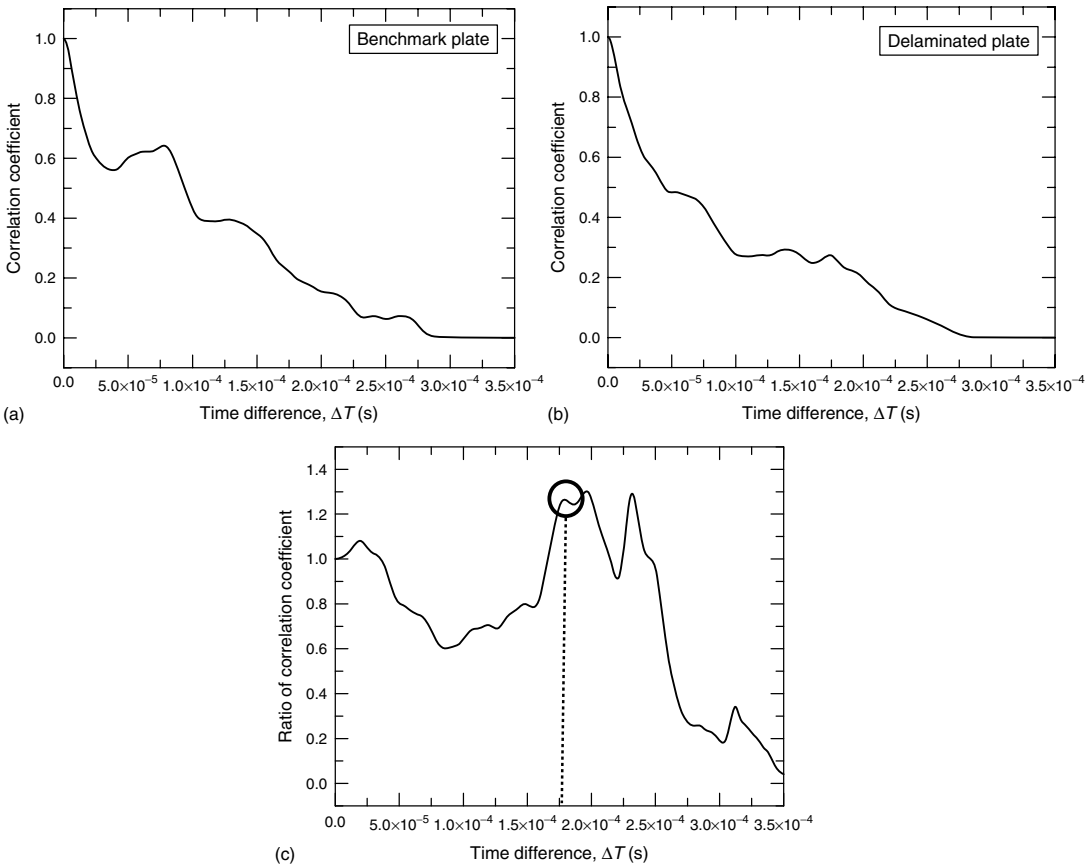


Figure 4. Correlation coefficient curves of Lamb wave signals acquired from (a) benchmark, (b) delaminated laminates, and (c) ratio of the curve in (b) to the curve in (a), where the delamination-induced singularity becomes pronounced [9].

sensors first figure out the individual beliefs that represent their own interpretation and perception on the damage event. A fusion process subsequently combines all these beliefs to form a consensus as to the overall view on the damage event and structural health status. During this procedure, fusion rules can be different for processing individual signals captured by distributed sensors. By this method, the fusion of multilevel decisions significantly reduces the risk caused by the inaccurate and even inappropriate judgment from an individual sensor or decision.

4.3 Fusion using artificial neural network (ANN)

The human nervous system consists of billions of small cellular units (*neurons*) connected by nerve fiber to form a neural net. ANN has been developed as a computational model to mimic such a way to process information based on a loose neural construction. ANN is able to explore the mathematical connection between a series of inputs (*conditions*) and outputs (*outcomes*) for a given system, whereby the consequence under an unknown stimulus can be predicted. A typical ANN is composed of one input layer, single or multiple processing (neural) layers, and one output layer, linked by transfer functions. Each neuron is weighted in parallel by an adjustable variable (*weight*) and offset by a constant (*bias*). One ANN developed for damage identification [6] is displayed in Figure 5(a).

In the ANN, the input, output, and two processing layers host m inputs, n outputs, J and K computing units (*neurons*), respectively; The ith neuron in the

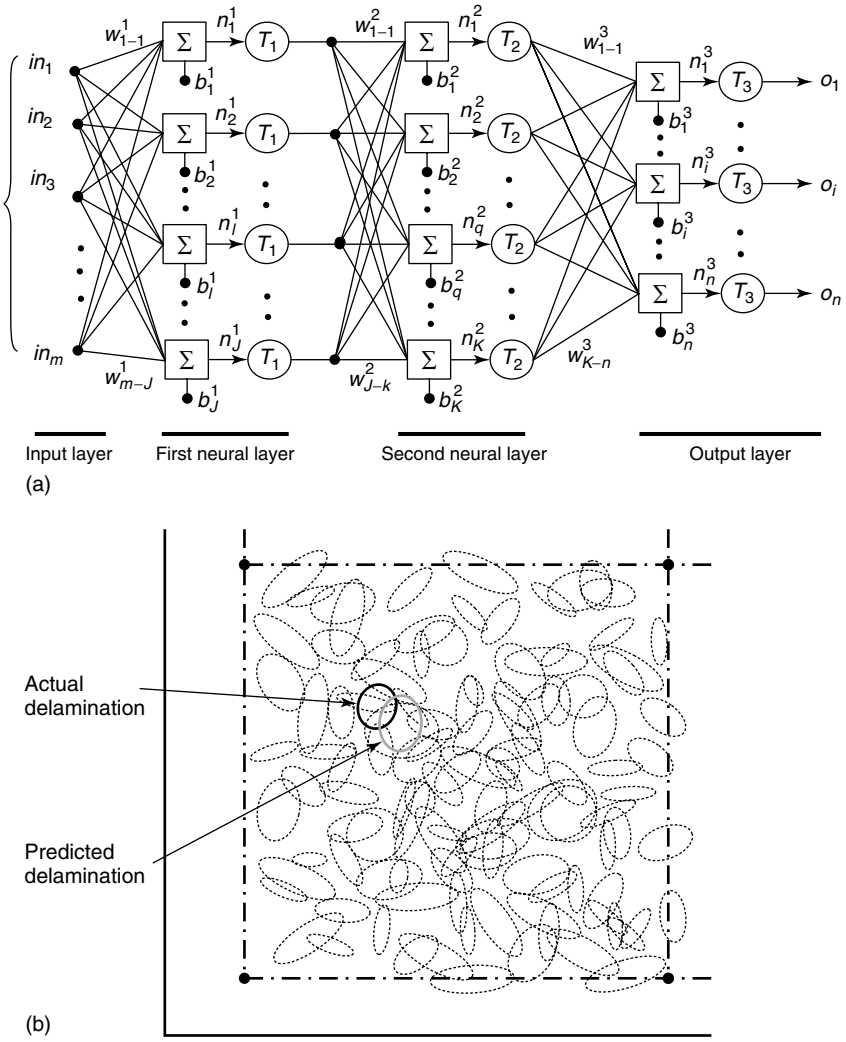


Figure 5. (a) An ANN designed for Lamb wave-based SHM and (b) prediction results (including location, shape, and size) of delamination in a CF/EP composite laminate using the well-trained ANN (dotted ellipses stand for the assumed pattern for ANN training) [6].

neural layer summates all the weighted inputs and biases to give the scalar output, which serves as the input for the next layer, in accordance with

$$o_i = T_3\left(\sum_{q=1}^{K} T_2\left(\left(\sum_{l=1}^{J} T_1\left(\left(\sum_{p=1}^{m} in_p \cdot w^1_{p-l} + b^1_l\right)\right.\right.\right.$$
$$\left.\left.\left.\cdot\, w^2_{l-q}\right) + b^2_q\right) \cdot w^3_{q-i}\right) + b^3_i\Bigg)$$
$$(i = 1, 2, \ldots, n) \qquad (3)$$

where o_i denotes the ith output element ($i = 1, 2, \ldots, n$). n^3_i is the output of the final output layer, and likewise, n^2_q and n^1_l are the outputs of two neural layers. in_p denotes the pth input element ($p = 1, 2, \ldots, m$); w^r_{p-q} represents the weight joining the pth input element (or neuron) in the rth layer with the qth neuron (or output) in the next layer; b^i_q is the bias for the qth element in the ith layer. T_i is the transfer function in the ith layer. A well-trained ANN is adept at inferring general rules or predicting

consequences from a specific example that has never been examined, making it possible to train an ANN using just a limited number of representative damage cases. Technically, an ANN network fuses all inputs to produce outputs as described above through a black box.

In one case study [6], up to 120 cases of delamination (with exclusive locations and geometric identities randomly selected in one-quarter of a quasi-isotropic CF/EP laminate (500 mm × 500 mm × 1.275 mm)) were simulated in terms of finite element method (FEM), respectively. For each damage case, signals were captured using a sensor network of nine piezoelectric elements surface mounted on the laminate, and signal features were extracted from the energy spectra. These extracted signal features were used to train the ANN shown in Figure 5(a). Without losing generality, the delamination was presumed elliptic, defined with six parameters: presence (0 or 1), location (ξ, ζ), semimajor/minor axes (α, β), and orientation of θ. The well-trained ANN was validated by quantitatively evaluating actual delamination in laminates of the same properties and dimension. Excellent prediction on all six damage parameters has been achieved (Figure 5b). The results predicted by the ANN do not superpose any damage pattern used for ANN training, implying that the results are ratiocinated rather than obtained by finding the fittest patterns from the training database.

In data fusion taking advantage of ANN, signals are acquired by different sensing paths for all the presumed structural health conditions and properly stacked to form a knowledge database. The training procedure of ANN in terms of the knowledge database is a kind of data fusion to achieve a belief on those concerned damage parameters (ANN outputs).

4.4 Fusion using Bayesian inference (BI)

BI is a probability-based data fusion discipline. It uses prior knowledge to estimate and update the conditional posterior probability of a hypothesis or to identify an event when given supporting evidence [1]. As evidence accumulates, the degree of belief in the hypothesis becomes high or low, and it persistently modifies the perceptions when additional information becomes available. The hypothesis is accepted as true with a very high degree of belief or rejected as false when the degree of belief goes very low, similar to the procedure of making judgments of human beings. The BI theorem is described by [4]

$$P(H_i|E_v)=\frac{P(E_v|H_i)P(H_i)}{P(E_v)}=\frac{P(E_v|H_i)P(H_i)}{\sum_i[P(E_v|H_i)P(H_i)]} \tag{4}$$

where $P(H_i|E_v)$ is the posterior probability that hypothesis H_i is true when given evidence E_v. $P(E_v|H_i)$ is the probability of evidence E_v given that H_i is true, defined as the likelihood function. $P(H_i)$ is the prior probability given that hypothesis H_i is true and satisfies $\sum_i P(H_i) = 1$. $\sum_i P(E_v|H_i)P(H_i)$ is the sum of the probability of evidence E_v when all mutually exclusive hypotheses H_i are true. The factor $P(E_v|H_i)/P(E_v)$ represents the impact of the new evidence on the degree of belief in the hypothesis; and if the hypothesis is true, this factor becomes large.

A two-level data fusion scheme using the BI algorithm was proposed for Lamb wave-based SHM [4]. In the approach, level-one decision fusion was based on signal features extracted from individual Lamb wave signals rendered by a sensor network to create perceptions of damage status and a knowledge database. H_i in equation (4) was the posterior probability for damage occurrence, and $P(E_v|H_i)$ estimated the confidence of individual sensor on the perception of damage existence. Subsequently, the perceptions from each sensor were integrated to represent decision fusion at the second level, to determine the detailed description for damage. The prediction results for through-hole damage in a composite laminate are displayed in Figure 6.

From the four examples aforementioned, it can be seen that whichever algorithms are used during data fusing, different criteria or methodologies can be adopted for processing individual signals captured by distributed sensors, so as to form an individual beliefs that represents their own interpretation and perception of damage status. All these beliefs or knowledge are further fused through one of the fusion strategies described in Figure 1 to deliver a comprehensive consensus on the entire

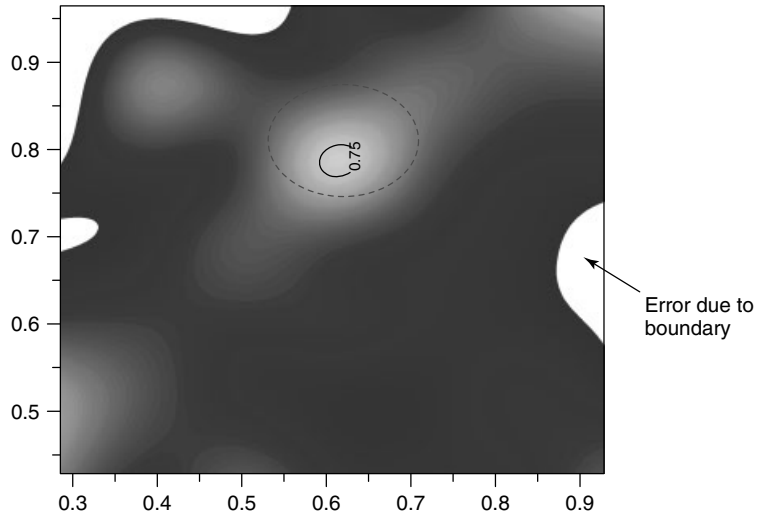


Figure 6. Damage prediction using BI (probability for the occurrence of damage is represented by grayscale—the lighter, the higher the possibility) [4].

damage event and health status of the structure under inspection.

5 CONCLUDING REMARKS

Data fusion is a *multilevel* and *multifaceted* process of combining *multiple* features extracted from a *multitude* of spatially distributed *independent* signal sources, so as to describe a damage event or health status of a structure under inspection. Triangulation of a damage site in simple structures can be fulfilled cost-effectively by using simple fusion schemes in terms of extracted signal features such as ToF or correlation. However, an SHM practice pertaining to quantitative evaluation of damage (including size assessment) is generally a highly complicated inverse procedure, which can only be implemented through appropriate inference-based data fusion. To this end, ANN and BI are two efficient solutions.

ACKNOWLEDGMENTS

Z. Su and L. Ye are grateful to the Hong Kong Polytechnic University for Grant A-PA8G and to the Australian Research Council for a Grant of Discovery Project.

REFERENCES

[1] Klein LA. *Sensor and Data Fusion Concepts and Applications*. SPIE Press: Washington, DC, 1999.

[2] Worden K, Dulieu-Barton JM. An overview of intelligent fault detection in systems and structures. *Structural Health Monitoring: An International Journal* 2004 **3**:85–98.

[3] Staszewski WJ, Boller C, Tomlinson GR. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons: New York, 2004.

[4] Wang X, Foliente G, Su Z, Ye L. Multilevel decision fusion in a distributed active sensor network for structural damage detection. *Structural Health Monitoring: An International Journal* 2006 **5**:45–58.

[5] Li CL, Hui KC. Feature recognition by template matching. *Computers and Graphics* 2000 **24**: 569–582.

[6] Su Z, Ye L. Lamb wave-based quantitative identification of delamination in CF/EP composite structures using artificial neural algorithm. *Composite Structures* 2004 **66**(1–4):627–637.

[7] Looney CG. *Pattern Recognition Using Neural Networks*. Oxford University Press: New York, 1997.

[8] Keilers Jr CH, Chang FK. Identifying delamination in composite beams using built-in piezoelectrics: part I—experiments and analysis. *Journal of Intelligent Material Systems and Structures* 1996 **6**:649–663.

[9] Su Z, Wang X, Chen Z, Ye L, Wang D. A *built-in* active sensor network for health monitoring of composite structures. *Smart Materials and Structures* 2006 **15**:1939–1949.

[10] Giurgiutiu V, Cuc A. Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention. *The Shock and Vibration Digest* 2005 **37**(2):83–105.

Chapter 38

Optimization Techniques for Damage Detection

Keith Worden[1], Wieslaw Staszewski[1], Graeme Manson[1], Aldo Ruotulo[2] and Cecilia Surace[2]

[1] *Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*
[2] *Department of Geotechnical and Structural Engineering, Politecnico di Torino, Torino, Italy*

1 INTRODUCTION

Finding the optimum solution for a given problem is an important area of research and application in many engineering fields. A variety of different optimization problems can be identified in structural health monitoring (SHM); a number of different possible mathematical tools can be offered to find a solution. This article summarizes some of the most commonly used optimization techniques. The focus is on combinatorial methods, which are computationally efficient and widely used in many engineering areas. These methods are illustrated by examples related to the problem of determining the optimal location of sensors in damage detection.

In mathematics, optimization is related to finding the maximum and/or minimum value of a function $f(x_1, x_2, \ldots, x_n)$ of n (i.e., $x_1, x_2, \ldots, x_n$) variables. This function, often called *the objective function*, describes the optimization problem defined in engineering. Often *a priori* limitations (or constraints) are imposed on the solution leading to constrained optimization problems. For example, if optimal locations for transducers are sought, certain locations are not possible owing to design constraints. Within this context, optimization is the task of finding values of *variables* for which the *objective function* reaches the extreme value satisfying the *constraints*. This definition points out three important elements of any optimization problem, i.e., *variables, objective function*, and *constraints*. Since the objective function is not always required (for example, in feasibility analysis), variables form the only essential element for any optimization problem.

There exist a variety of different optimization methods used in practice. The method selected for a

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

given optimization method often depends upon four major criteria: (i) Is the problem one- or multidimensional? (ii) Are functions describing the problem efficiently differentiable? (iii) How noisy is the function describing the problem? (iv) Is the proposed optimization algorithm computationally efficient and cheap? Early optimization techniques, often known as *ad hoc* methods, are based on rough and ready ideas that do not have any theoretical background.

Altogether, optimization algorithms can be classified into various groups. These are (i) local and global, (ii) one-dimensional and multidimensional, (iii) constrained and unconstrained, (iv) deterministic and nondeterministic, and (v) linear and nonlinear. The entire classification is not trivial as there is substantial overlap between the various groups of methods. Local methods (e.g., gradient descent, Newton's method) are used to find local optimal values in a finite neighborhood of the analyzed space, whereas global methods (e.g., simplex method, simulated annealing (SA), genetic algorithms (GAs)) search for truly global extreme values. Optimization problems with only one variable can be solved using one-dimensional methods such as simple minimization techniques (e.g., golden section search, parabolic interpolation). Finding a number of parameters that optimize the problem is a much harder task and requires multidimensional methods (e.g., downhill simplex method, GAs). Deterministic optimization (gradient descent, simplex method) methods deal with problems formulated with all known parameters. In contrast, nondeterministic methods (e.g., Tabu search, SA, GAs) are applied to problems with uncertainties. Classical deterministic optimization techniques can be classified into unconstrained (e.g., gradient descent, nonlinear conjugate gradient methods, Newton's methods) and constrained optimization (e.g., method of Lagrange multipliers, GAs) where the latter involves limitations imposed on solutions. Classical constrained optimization has a great degree of complexity. When the objective function and constraints are linear, the methods involved are often referred to as *linear optimization* (e.g., simplex algorithm) or linear programming. In contrast, nonlinear optimization (e.g., branch and bound technique) involves a nonlinear objective function or constraints. Nonlinear programming is always associated with substantial computation.

A good theoretical background of various optimization techniques can be found in [1–3]. In the text that follows, a brief introduction to the most widely used techniques together with application examples is given.

2 CLASSICAL OPTIMIZATION TECHNIQUES

2.1 Gradient descent

The gradient descent optimization technique finds extreme values by following the gradient of the objective function. If the optimization space is represented as a surface, finding the maximum value is associated with "going up" with the gradient. Therefore, the method is often called *hill climbing* optimization.

The mathematical background of the method is very simple. If the objective function $f(x)$ is defined and differentiable, its gradient ∇f always indicates the direction of the steepest ascent. Thus, when a guess of x_0 for the local minimum is assumed initially, an improved estimation of the minimum value is achieved if the negative gradient is followed iteratively as

$$x_{n+1} = x_n - \gamma_n \nabla f \quad \text{for} \quad n = 0, 1, 2, \ldots \quad (1)$$

where the parameter γ_n is a value of step size and ∇ is the standard gradient operator. The step size value can change at every iteration. It follows that the sequence

$$f(x_0) \geq f(x_1) \geq f(x_2) \geq \cdots \quad (2)$$

converges and finally the minimum value is obtained. The method is illustrated graphically in Figure 1.

The gradient method is very simple since it requires only a few steps to find a minimum (or minimum) value. However, the major drawback is that the method can get trapped in local extremes and finding global values depends on the starting point. Also, often large jumps are observed when the gradient descent is performed. Smaller step sizes can reduce this effect.

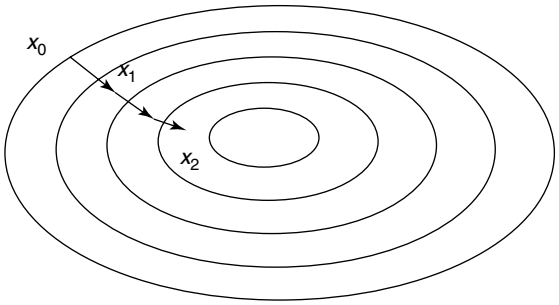


Figure 1. Graphical illustration of gradient descent optimization method.

2.2 Newton's method

Newton's optimization method originates from the iteration method for the approximate determination of zeros of a differentiable function. It is well known that the Taylor expansion of $f(x)$ given by

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2 + \cdots \tag{3}$$

leads to its extreme when $f''(x)$ is positive and Δx solves the following equation:

$$f'(x) + f''(x)\Delta x = 0 \tag{4}$$

Thus, for the multidimensional case and the initial guess x_0 for the minimum value, the optimization iterative scheme can be obtained as

$$x_{n+1} = x_n - \gamma_n [Hf(x_n)]^{-1} \nabla f(x_n) \quad \text{for} \quad n = 0, 1, 2, \ldots \tag{5}$$

where $Hf(x_n)$ is the Hessian matrix of the function $f(x_n)$ representing all its second-order partial derivatives. The convergence of Newton's method to the optimal value is much faster than for gradient descent. However, it is important to note that for multidimensional problems finding the inverse of the Hessian matrix is numerically expensive and can lead to instabilities.

2.3 Simplex method

The simplex algorithm is one of the most widely used linear optimization techniques. The algorithm is used to solve the linear programming problem, which aims to maximize a linear, real-valued objective function give in the form

$$f(x_1, x_2, \ldots, x_n) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + a_0 \tag{6}$$

where $\boldsymbol{x}$ is the vector of variables and $\boldsymbol{a}$ is the vector of coefficients. The maximization is subjected to the constraint represented by the inequality

$$\boldsymbol{Ax} \leq \boldsymbol{b} \tag{7}$$

where $\boldsymbol{A}$ and $\boldsymbol{b}$ are the matrix and vector of coefficients, respectively.

The optimization algorithm used in the above problem utilizes the geometrical concept of a simplex. In geometry, a simplex is a higher-dimensional generalization of a triangle. More precisely, an n-simplex is a geometric figure called a *polytope* built from a set of $n + 1$ affinely independent points, i.e., a 1-simplex is a line segment, a 2-simplex is a triangle and a 3-simplex is a tetrahedron (shown in Figure 2) in Euclidean space. The simplex algorithm defines the objective function given by equation (6) on a polytope. The aim is then to find a geometrical point where the objective function has a maximum value. The process of searching through the polytope vertices leads to the maximum value. The simplex algorithm requires that the linear programming problem is converted into the so-called augmented form by replacing the inequalities with equalities in the constraints.

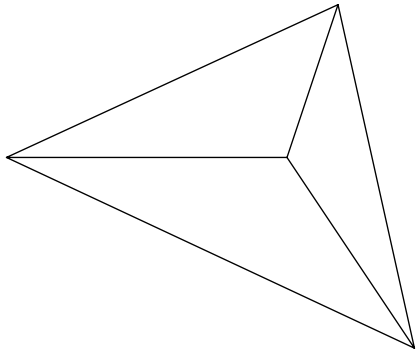


Figure 2. Example of simplex—tetrahedron.

3 SIMULATED ANNEALING

3.1 Basic theory

The SA algorithm is arguably the first and simplest of the heuristics inspired by physical and biological mechanisms and so it is described here first. It is based on an analogy with the physical process of annealing. Normal polycrystalline solids have a microscopic domain structure that is quite disordered; under equilibrium conditions, the solid is only at a *local* minimum of energy. However, if the solid is heated to a high temperature and then cooled slowly, it will arrive at a highly ordered state—the *global* minimum of energy. Consider a random search method for optimization—this attempts moves in random directions in the search space—accepting a trial move only if the objective function decreases. One can think of the method as moving downward gradually on the surface specified by the objective. Because the method accepts all downward moves, it will converge to a local minimum if the starting point for the algorithm is within the basin of attraction of that minimum. To escape from local minima, one can modify the algorithm by adding a certain amount of *thermal motion*, which allows the objective to sometimes increase. This is the basis of SA; the allowed upward movements are large in the early stages of the algorithm when local minima are to be avoided, corresponding to a high-temperature regime. In the later stages, the thermal motion is small to allow convergence, corresponding to low temperature. To decide whether an upward movement is allowed, Boltzman statistics are used. A probability

$$p = \exp\left(\frac{-\Delta E}{T}\right) \tag{8}$$

is computed, and if less than a uniform random deviate generated in parallel on the interval [0,1], the move is accepted. Here, ΔE is the change in the objective function (positive), and T is an external control parameter, the *temperature*, which decreases as the algorithm proceeds. Downward movements are always accepted. The origins of the method are in the Metropolis algorithm [4], although the serious development of the method is associated with Kirkpatrick *et al.* [5].

3.2 A case study: sensor placement

The algorithm is very simply described. It is also very straightforward to implement in computer code. The use of the algorithm is illustrated here via a case study—the optimal placement of sensors on a structure for damage localization. The structure considered here was a finite element (FE) model of a cantilever plate; the FE package used here was LUSAS. The plate was simulated to be of steel and having dimensions of 300 mm $\times$ 200 mm $\times$ 2.5 mm. The fixed edge of the plate was taken on one of the 200-mm edges. The FE mesh used 20×20 elements to produce results with appropriate accuracy. A coarser 4×4 cell mesh was used to place the sensors. Each node of the regular coarse mesh was considered as a candidate sensor location, except for those nodes on the fixed edge. This gave 20 candidate sensor locations indexed between 1 and 20 as shown in any of the figures from this section. Each of the cells of the 4×4 coarse mesh contained a 5×5 array of FE elements; damage was simulated in each of the coarse cells by "deleting" the 9 FE elements in the center of the cell. "Deletion", in this case, corresponded to setting the Young's modulus of the element in question to a very small value. The data generated from the FE model was the mode-shape data for the lowest natural frequencies for each possible damage condition of the plate, of which there were 16. The mode-shape data was then twice differentiated in both the x and y directions to obtain mode-shape curvature data (as described in **Modal–Vibration-based Damage Identification**). A centered-difference formula was used in each direction and the magnitude of the resulting vector was used as a scalar curvature measure. Only the curvature values on the coarse mesh (i.e., corresponding to sensor locations) were saved for damage identification purposes. The curvature data were used as training data to establish a neural network, which could locate the damage to within a cell of the coarse mesh. (Neural networks are described in some detail in **Artificial Neural Networks**.) To obtain sufficient data for the neural networks, multiple copies of each damage *pattern* (curvature vector) were made and each was corrupted with a low rms Gaussian noise vector to simulate the acquisition of real data with measurement noise. More details of this case study can be found in [6].

The neural network used here as a damage locator was a standard multilayer perceptron (MLP) trained using back propagation learning (*see* **Artificial Neural Networks**). In brief, the network was used as a learning machine that could be trained to indicate the position of damage within the structure when presented with a vector of modal curvature values. The number of input nodes to the network was fixed equal to the number of sensors chosen for the optimization and there were 16 outputs—each corresponding to a given damage location. The neural network training used a scheme—the 1 of M scheme—which basically generated the posterior probabilities of the 16 possible damage locations when presented with the curvatures. When a curvature pattern is presented, one chooses the highest output as the indicator of the damage location. The optimization problem illustrated here is simply to choose the best n sensor positions to provide data for the neural network diagnostic. The success of the neural network is judged by considering a set of curvature vectors corresponding to the different damage locations and simply counting the number of correct and incorrect diagnoses. These numbers are then converted into a single probability of misclassification. Optimization over the set of sensor distributions corresponds to minimizing this probability of error.

The implementation of the algorithm SA used here follows that in [3]. The objective function is the probability of misclassification described above. The initial temperature is chosen of the same order of magnitude as the probability of error for a randomly chosen sensor distribution. Each sensor distribution is encoded as a 20-bit binary vector where a 1 corresponds to the presence of a sensor and a 0 corresponds to the absence of a sensor. The bits of the distribution vector correspond to the position indices for the sensors indicated in the figures of this section. SA moves are made by operating on the bit vectors described previously. If a 10-sensor distribution is desired, an initial distribution is chosen randomly with 10 bits set to 1. Trial moves involve changing the position of a 1 randomly. Moves are accepted if they lead to a decrease of the objective function or—in the case of increase of the objective—the probability of acceptance leads to acceptance as described above. Fifty trial moves are taken at each temperature unless ten moves are accepted, and in each case, the temperature is decreased by a multiplicative factor of 0.9. Fifty temperature steps are taken.

The first run of the algorithm considered here was directed to find a 10-sensor distribution. The algorithm converged after 1350 iterations to a solution with a probability of error of 0.0014. In fact, during the run, the algorithm found three solutions with this error and they are given in Table 1.

The second of these distributions is depicted in Figure 3.

For the second illustration, the best three-sensor distribution was sought. In this case, it was actually feasible to find the optimum by exhaustive search as there were only 1140 possible candidate distributions.

Table 1. Best 10-sensor distributions from simulated annealing

10100001001101111001
10000100101101101011
11100000001111101010

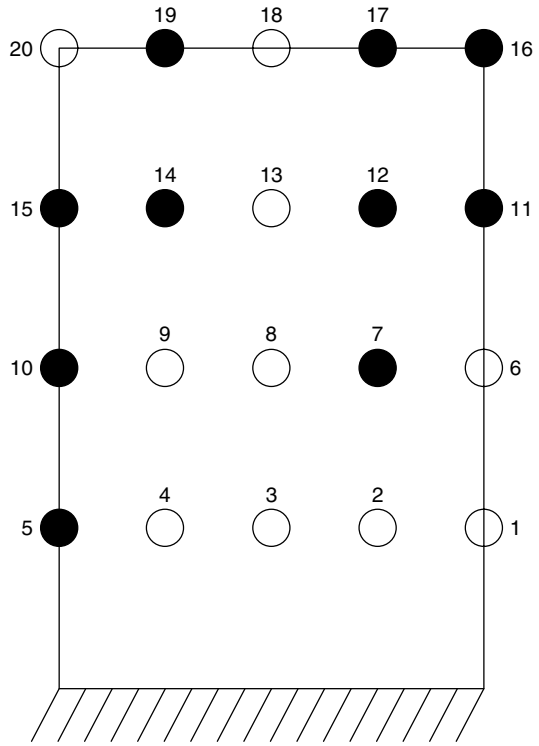


Figure 3. Best 10-sensor pattern for curvature diagnostic from simulated annealing.

Figure 4 shows the evolution of the objective function as the SA algorithm proceeded.

It can be seen that the algorithm found the minimum after approximately 700 iterations (although a fluctuation caused a temporary rise afterward). This is a considerable saving on exhaustive search. Figure 5(a) shows the sensor distribution found by the algorithm. Figure 5(b) shows the optimum distribution found by exhaustive search. It is found that the two distributions are mirror images of each other.

One issue that arises here, as it will with any optimization procedure that uses a neural network to determine the objective function, is that the objective is a random variable conditioned on the initial weights of the network. This means that the fitness of a given sensor distribution will actually be characterized by a distribution function. Now, in the case of sensor networks with a small number of sensors, one might expect that the distribution functions for near-optimal distributions will not overlap. This means that the probability of finding the true optimum is increased. In the three-sensor example shown here, the best SA distribution and that from exhaustive search actually have different objective function values so that it is the mirror image in the exhaustive search that actually corresponds to the SA optimum. It may be desirable to take into account the random nature of the fitness values. One way of doing this would be to use a different (randomly chosen) initial condition for the network every time a given sensor pattern occurs in the process. This was not done here.

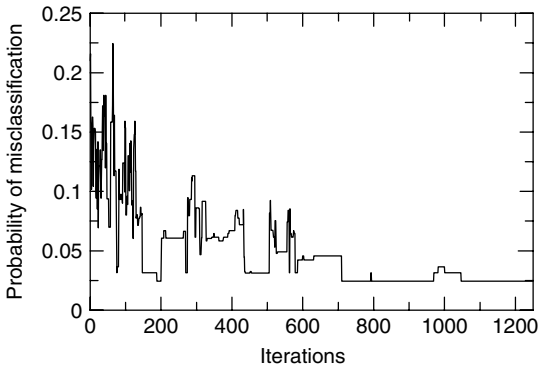


Figure 4. Progress of the SA algorithm toward the optimum three-sensor distribution.

The SA algorithm was initially formulated with combinatorial optimization in mind and that is how it has been used here. There are variants of the algorithm that can deal with real-valued optimization variables, and one example is discussed in [7]. A possible weakness of the algorithm is that it operates on a single point at each iteration and is thus more susceptible to getting trapped in a local minimum. Descriptions of population-based variants

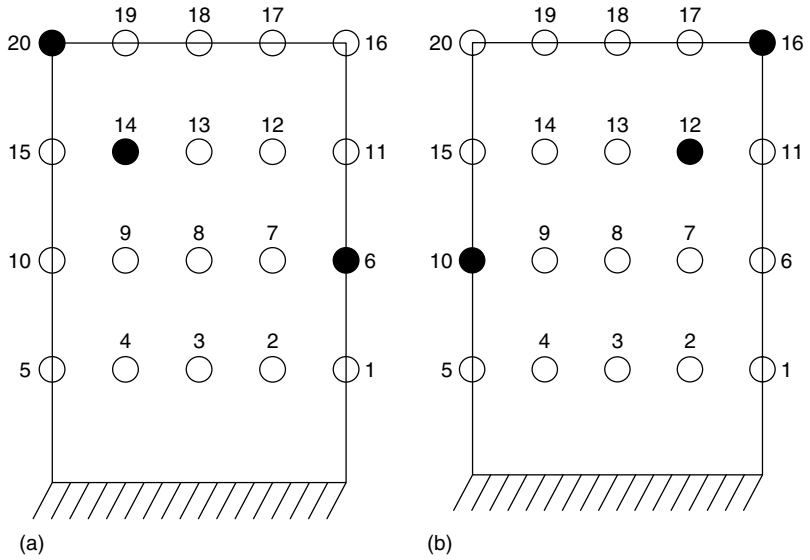


Figure 5. Best three-sensor patterns for curvature diagnostic: (a) SA and (b) exhaustive search.

of the algorithm that avoid this problem to an extent can be found in [8, 9].

4 GENETIC ALGORITHMS

4.1 Basic theory

GAs are a population-based algorithm motivated by biology. They were first introduced in their simplest form by Holland [10]; however, they have been the subject of considerable research since, and now form only one example of a broad class of *evolutionary algorithms* [11]. For the sake of brevity, only a short discussion of the simple GA is given here; for more detail, the reader is referred to the standard introduction to the subject [12].

GAs are optimization algorithms that evolve solutions in a manner analogous to the Darwinian principle of natural selection. They differ from the classical gradient-based optimization techniques in that they work on encoded forms of the possible solutions like the SA example of Section 3. Each possible solution i.e., each set of possible parameters in solution space is encoded as an *individual*. The most usual form for this individual is a binary string e.g., 00011010110. The first hurdle in setting up a problem for solution by GA methods is working out how best to encode the possible solutions as individuals. In the case of the sensor placement problem described earlier, a natural coding for the problem is provided by a bit-vector where the occurrences of a 1 signal the existence of a sensor, i.e., the individual 00100010000000001000 represents a solution in which transducers are placed at positions 3, 7, and 17.

Having decided on a representation, the next step is to generate, at random, an initial population of possible solutions. The number of individuals in a population depends on several factors, including the size of each individual, which itself depends on the dimension of the solution space. For example, if the problem of interest requires the optimization of three control parameters, the bit-vector will be in three segments; if each parameter is required to an accuracy of 1 part in 1000, 10-bits per parameter will be needed and the overall individual will need 30 bits. Larger initial populations will be required if the individuals are larger to have an appropriate initial level of diversity in the population.

Having generated a population of random individuals, it is necessary to decide which of them are fittest in the sense of producing the best solutions to the problem. This is the *selection* process. To do this, a fitness function is required, which operates on the encoded individuals and returns a single number that provides a measure of the suitability of the solution. In the SA example of Section 3, the objective function was the probability of misclassification; this was an object that required minimization; to get a fitness function, such an objective should be inverted to give a maximization problem. The fitter individuals are used for mating to create the next generation of individuals, which hopefully provides a better solution to the problem. Individuals are selected for mating based on their fitnesses. One way of implementing this is so-called roulette-wheel selection; the probability of a particular individual being chosen is equal to its fitness divided by the sum of the fitnesses of all the individuals in the population.

Once sufficient individuals have been selected for mating, they are paired up at random and their genes combined to produce two new individuals. The most common method of combination used is called *crossover*. Here, a position along the individuals is chosen at random and the substrings from each individual after the chosen point are switched. This is illustrated in Figure 6(a). This mechanism is termed *one-point crossover*. In a *two-point crossover*, a second position is chosen and the individual substrings switched again.

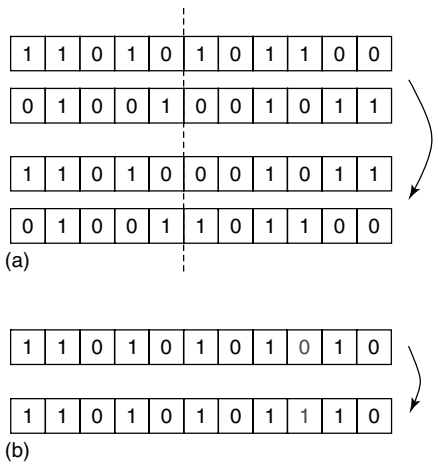


Figure 6. The basic genetic operations: (a) crossover and (b) mutation.

If an individual in a particular generation is extremely fit, i.e., is very close to the required solution, it is almost certain to be selected several times for mating. Each of these matings, however, involves combining the individual with a less fit individual, so the maximum fitness of the population may be lower in the next generation. To avoid this, a number of the most-fit individuals can be carried through unchanged to the next generation. These very fit individuals are called the *elite*.

To prevent a population from stagnating, it can be useful to introduce perturbations into the population. New entirely random individuals may be added at each generation. Such individuals are referred to as *new blood*. Also, by analogy with the biological process of the same name, individuals may be *mutated* by randomly switching one of their binary digits with a small probability. This latter process is one of the more important ones for the simple GA and is illustrated in Figure 6(b).

With genetic methods, it is not always possible to say what the fitness of a perfect individual will be. Thus the iterative process is usually continued until the population is dominated by a few relatively fit individuals. One or more of these individuals will generally be acceptable as solutions.

4.2 A case study: impact location

As in the case of SA, the simple GA is illustrated on a sensor placement problem. In this case, the best set of sensors is chosen to train an impact-locating neural network. The details of this case study can be found in [13].

The structure under examination consisted of a rectangular 530 mm × 300 mm composite plate and four aluminum channels. The top flanges of the channels were attached to the plate by a line of rivets, and the bottom flanges were fixed rigidly with screws to a pneumatic measuring table. This box structure was intended to simulate the skin panel of an aircraft. The composite plate was instrumented with 17 piezoceramics (*PZT Sonox P5*, 15 mm × 15 mm) fixed on the lower surface of the plate, the impacts being performed on the upper surface. The piezoceramics were used as strain sensors. Figure 7 shows the total distribution of sensors used.

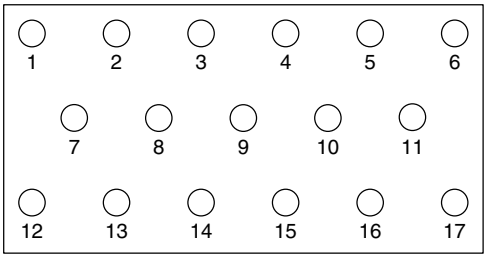


Figure 7. Schematic for sensor positions on a composite plate.

The impacts were applied using a *PCB* instrumented hammer. The levels of force applied were kept below 0.1 N in order not to damage the plate. Two sets of measurements were made. The first comprised 80 impacts at random positions on the plate and was intended for use as network training data. The second set of measurements contained 95 impacts placed on a regular grid and these were divided between the network validation and test sets: 48 in the former, and 47 in the latter.

The strain data were recorded using a DIFA SCADAS II 24 channel measuring system running the LMS 3.4.04 data-acquisition software. For each impact, 8192 samples were recorded at a frequency of 25 kHz.

To design a neural network that could predict the impact location, it was important to extract appropriate features that could be used as information-rich network inputs. A number of different candidates have been considered in the past [13]: (i) time after impact of maximum response, (ii) magnitude of maximum response, (iii) peak-to-trough range of the response, and (iv) real and imaginary parts of the response *spectrum*, integrated over frequency. The features were investigated alone and in combination, and it was found that the best results were obtained using features (i) and (ii) in tandem. This meant that each sensor contributed two features and that, if all sensors were used, the dimension of the pattern vectors for training was 34. This was the maximum, as the object of the exercise was to establish optimal subsets.

As for the SA case study, the neural network paradigm used for this study was the standard MLP trained with backpropagation. A principled approach to training demands the availability of three data sets. The first—the *training* set—is used to determine

the network weights. The second—the *validation* set—is used to investigate the optimum structure for the network and the final *testing* set is used to assess the effectiveness of the optimized network. The number of neurons in the input and output layers of the network was fixed by the number of measurement features and diagnostic outputs, respectively. In this case, the network needed up to 34 inputs; for impact location, the network was required to signal the location of damage and two outputs were required, namely, the x and y coordinates of the impact site. In principle, only one hidden layer is needed to approximate any function; therefore the strategy adopted here assumed a single hidden layer and optimized the performance by using all 34 inputs and varying the number of hidden neurons between 1 and 50. The network was also optimized over initial conditions and regularization was used (*see* **Artificial Neural Networks**). The MLP used hyperbolic tangent activation functions and a bias neuron was connected to all neurons. In all cases, the number of data presentations during training was equal to 100 000.

For the location problem using all sensor data, the optimization procedure produced a minimum error over the validation set when there were eight neurons in the hidden layer. When the corresponding network was evaluated on the testing set, the mean (modulus) of the x error was 23.1 mm and the mean y error was 25.7. This gave an area corresponding to 1.5% of the plate area. (The area error was used as the statistic for the evaluation of the networks, and the best value on the validation set was 1.1%.)

Despite the fact that the description above gave the simple GA as applied to a binary bit-vector representation for the data, it was shown in [6] that the binary representation is suboptimal for the sensor optimization problem. As a result, a modified GA is used, where the gene is a vector of integers, each specifying the position of a sensor; i.e., the gene (2,14,17) represents a three-sensor distribution, with sensors at locations 2, 14, and 17 on the candidate mesh. The operations of reproduction, crossover, and mutation for such a GA are straightforward modifications of those for a binary GA.

The initial population for the GA was generated randomly as standard. The individuals—in this case, sensor distributions—were propagated according to their fitness. For a given sensor distribution, a neural network was trained using the designated features and the area error on the validation set was obtained; this was inverted to give the fitness. Distributions with coincident sensors were penalized. The parameters used for the GA runs were as follows: population size of 50, number of generations 100, probability of crossover 0.8, and probability of mutation 0.1. A single member elite was used and two new blood were added at each iteration.

For the first example, the analysis was restricted to three-sensor distributions. The reason is that the number of candidate three-sensor distributions is 684, and it is feasible to compare the results with those from exhaustive search. Also three sensors are the minimum number that would be needed to locate the impact event from time-of-flight data using triangulation.

Recall that the error of the neural network depends on the starting conditions. To investigate this, the exhaustive search was carried out six times. In each case, different random numbers were used for initializing the networks and also for calculating the various probabilities required by the GA. The best distribution from each run is given in Table 2 together with the associated percentage error area.

The best three-sensor distribution is shown in Figure 8. It agrees strongly with engineering intuition: the sensors are placed in such a way as to effectively triangulate over a large area of the plate. Also, the degradation in performance over the full 17-sensor distribution is hardly significant: the area error is 2.0% instead of 1.1%.

When the GA was applied for the six network start-up conditions, it found the optimal solution in all cases. In most of the cases, it found them faster than the exhaustive search. Figure 9 shows the evolution of the fitness functions over the six runs. The solid lines are the maximum fitnesses of each generation and the dotted lines are the average fitnesses.

Table 2. Best three-sensor distributions from six exhaustive searches

Search number	Distribution	Area
1	3 : 12 : 17	2.16
2	8 : 11 : 12	2.15
3	1 : 3 : 17	2.15
4	3 : 12 : 14	2.19
5	3 : 7 : 11	2.13
6	3 : 10 : 12	1.99

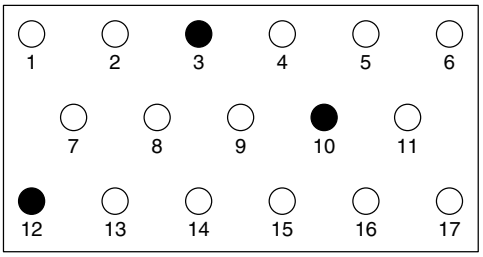


Figure 8. Optimal sensor distribution for impact location from exhaustive search.

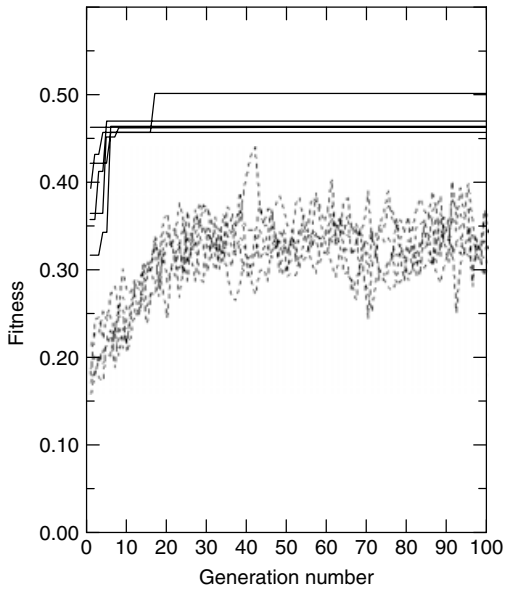


Figure 9. The evolution of the fitness for the three-sensor GA optimizations.

5 DIFFERENTIAL EVOLUTION

5.1 Basic theory

The simple GA described in Section 4.1 used a binary form to encode the optimization parameters. Although this is quite general, it was almost immediately observed that a better encoding of the parameters for a particular problem was actually in terms of integers. This observation leads to the idea of evolutionary algorithms in general [11], where the objective is simply to generate the best approach to solving a given engineering problem. In the drive to find the best algorithm, it is accepted that one might move a distance away from the biological motivations of the algorithms. In the case of an optimization, which requires the search for a number of real parameters, a real-coded algorithm is clearly desirable. Although there are instances of real-coded GAs, the subject of this section is slightly different. The differential evolution (DE) algorithm was developed by Storn and Price [14] and has proved to be an extremely effective heuristic for optimization.

As in the case of the GA, the DE algorithm works with a population of candidate solutions or individuals. Figure 10 shows diagrammatically the procedure for evolving between subsequent populations. The process is repeated for each vector within the current population being the *target vector*. Each of these vectors has an associated cost value (or fitness value in maximization problems) obtained from the cost (or fitness) function. This target vector is pitted against a *trial vector* in a selection process with the vector with the lowest cost (or highest fitness) advancing to the next generation. The process for constructing the trial vector involves mutation and crossover, processes known from the GA to give fast, robust algorithms when used in conjunction. Mutation maintains diversity in the population while crossover builds new parameter combinations from the existing vector parameters.

The mutation procedure used here employs vector differentials. Two vectors (individuals) ($\mathbf{A}$ and $\mathbf{B}$) are randomly chosen from the current population to form a vector differential $\mathbf{A} - \mathbf{B}$. The *mutated vector* is then obtained by adding this differential, multiplied by a scaling factor, F, to a further randomly chosen vector $\mathbf{C}$ to give the overall expression for the mutated vector as $\mathbf{C} + F \cdot (\mathbf{A} - \mathbf{B})$. The scaling factor, F, will have an optimal value for most functions between 0.4 and 1.0.

The trial vector is the child of two vectors, the target vector and the mutated vector, and is obtained via the crossover process. In this work, the process of uniform crossover was used. Uniform crossover decides which of the two parent vectors contributes to each chromosome of the trial vector by a series of $D - 1$ binomial experiments. Each experiment, whose outcome is either success or failure, is mediated by a crossover constant, C, where $0 \leq C \leq 1$. If the random number is greater than C, the trial vector

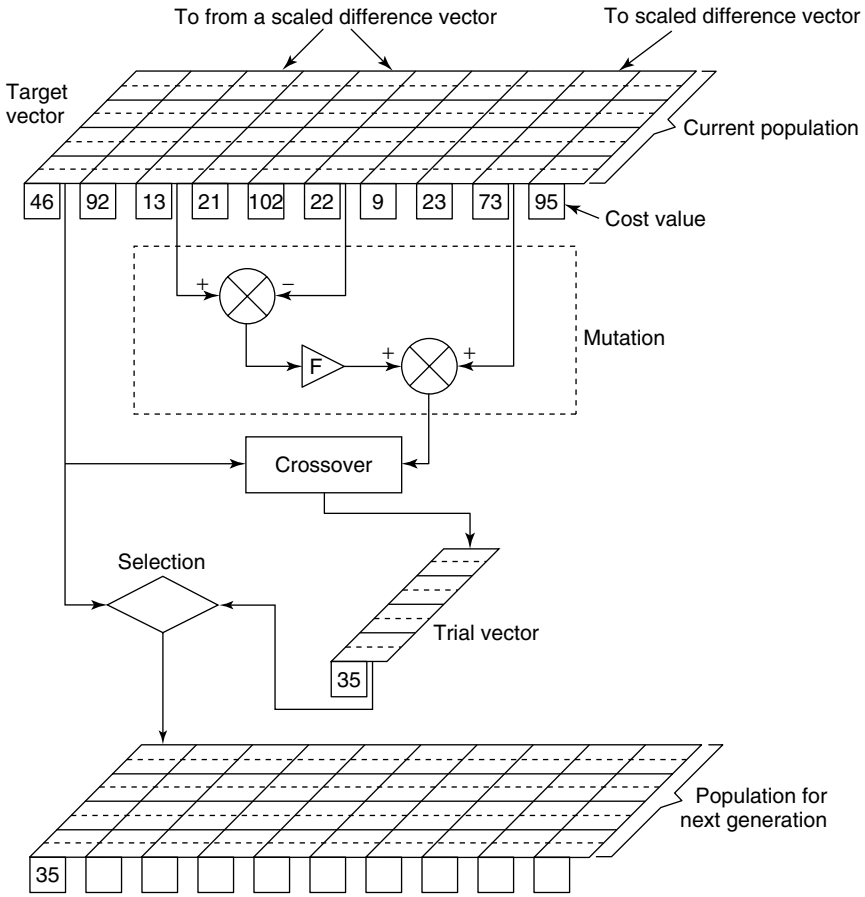


Figure 10. Schematic for differential evolution.

gets its parameter from the target vector; otherwise the parameter comes from the mutated vector.

This process of evolving through generations is repeated until the population becomes swamped by only a few low-cost (or high-fitness) solutions, any of which would be suitable. Now that the general procedure for the DE algorithm has been explained, a case study can be considered.

5.2 A case study: sensor placement for Lamb wave inspection

The case study presented here is based on the idea of inspecting plates using high-frequency Lamb waves. The idea is that waves will be launched from an emitter or actuator and recorded by a distant receiver or sensor. Once the wave profiles have been established for the normal condition of the plate, one can look for evidence of damage in the form of scattered and reflected waves in the wave profiles. As before, the optimization problem is the problem of placing emitters and receivers to best detect damage.

The method used to encode the problem variables as individuals in this case is remarkably straightforward; each chromosome of the gene is simply an x or y coordinate of one of the sensors. Figure 11 shows the situation for one possible arrangement of two emitters and two receivers on a $2\,\mathrm{m} \times 1\,\mathrm{m}$ plate. An eight-chromosome gene uniquely defines the sensor arrangement.

Once a procedure for the encoding of sensor locations has been devised, the next issue that arises

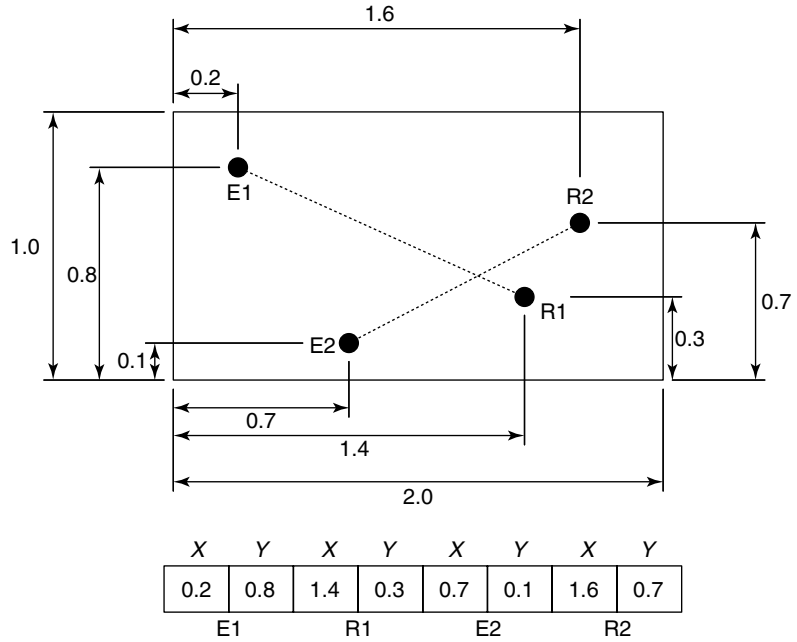


Figure 11. DE encoding for Lamb wave sensor placement problem.

concerns the objective function that is going to be optimized using the DE. For the problem at hand, the maximizing of damage detection coverage on a structure, the approach that will be taken is to try to minimize a cost function based upon angles between emitter, receiver, and damage locations. The reason for this is that the emitter considered here generates a directed wave and the receiver works on the same principles. This means that the greater the angle that a potential damage location makes with the emitter/receiver axis, the less likely that this sensor placement would be to detect said damage. However, if more than one pair of sensors is being used the situation should improve, with the best pair of sensors being used for each potential damage location. This may be easier explained by referring to Figure 12, which shows the situation for only one defect location and one possible placement of two pairs of emitters and receivers on a plate structure. The usefulness of the E1/R1 combination for the detection of the shown defect is dependent upon the greater of the two angles, R1E1D and E1R1D, with the ideal situation being that the defect is located on the axis between E1 and R1. In this case, there exists a second pair of sensors that may be more suitable for detecting the defect and

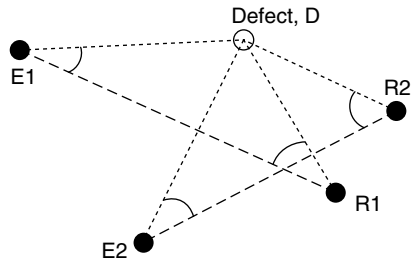


Figure 12. Schematic for Lamb wave inspection showing scattering of waves from a defect.

so the best pair of sensors is selected in each case. The process needs to be repeated over a mesh of all possible damage locations for this sensor placement to arrive at an overall cost value for the individual representing the emitter and receiver positions.

The cost function for the two pairs situation may be obtained as the maximum scattering angle between a sensor pair and a defect, with the maximum taken over all possible defect positions and over the two sensor pairs. The extension to different numbers of sensor pairs is clearly straightforward.

The optimization was carried out for a square plate with numbers of sensor pairs between one and

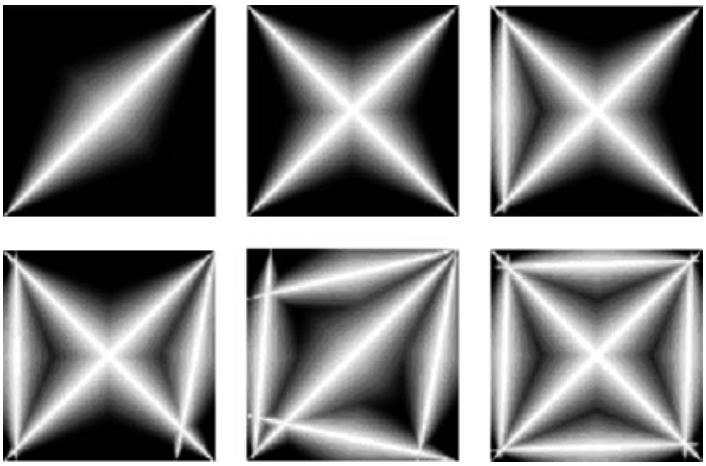


Figure 13. Optimized plate coverage for Lamb wave sensing.

six. Figure 13 shows the solutions obtained from the DE algorithm for the square plate with directional emitters and sensors; the grey-scale plots use white to indicate areas where damage is more likely to be detected and the darker areas indicate areas where detection may be a problem. As expected, in all cases, the algorithm has placed the sensors at the outer edge of the plate. The emitters and receivers associated with one another are obvious from the white paths. The range for the scattering angle in the plots in Figure 13 are between 0° and 25°. Any location shown in black indicates an angle of 25° or more. For the square plate, the cases for between one and four pairs of sensors show little surprise with sensors being located at opposite corners of the plate to begin with and subsequent pairs trying to remove "black spots". It is only the five-pair case that shows some departure from this process, only to return to the high level of symmetry in the six-pair case.

6 MORE METHODS INSPIRED BY BIOLOGY

6.1 Ant colony metaphors

Algorithms based on *ant colony metaphors* are a comparatively recent addition to the group of heuristic optimization algorithms. These are based on the cooperative interaction of simple computational agents termed *ants*. The basic forms of the algorithms—ant-density, ant-quantity, and ant-cycle—were introduced in [15]. Of these basic forms, the ant-cycle algorithm proved to be the most effective, and this was renamed *ant system* and discussed in more detail in [16]. Originally, the algorithms were applied to the travelling salesman problem (TSP), but were then shown to apply to other hard combinatorial optimization problems like the quadratic assignment problem and job-shop scheduling problem. Modifications soon appeared, which could deal with continuous design spaces [17] and constrained problems [18]. The ant algorithms have proved to be a useful addition to the armory of methods, combining aspects of greedy search with population-based cooperative search.

Real ants are well known to be capable of finding the shortest path to a food source from their nest without using visual clues [19–21]. This is done by exploiting *pheromone* information. Ants deposit pheromone as they move and follow pheromone trails deposited by previous ants. If a trail has higher pheromone information, an ant will follow it in preference to other trails with less pheromone. In other words, an ant will follow a trail with higher pheromone, with higher probability.

Consider Figure 14. Initially the ants leave the nest and have no preference as to which path they will follow and they will choose a path with equal probability. Assuming that all ants travel with equal speed, the ants taking the lower path will reach the food before those taking the upper one. When

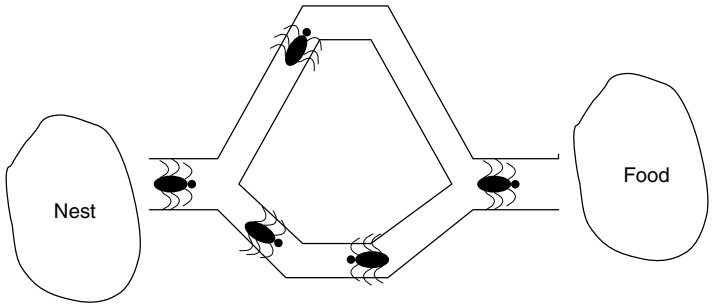


Figure 14. Schematic for ant foraging activity.

returning to the nest, they will choose the lower path with higher probability because the upper path ants have not arrived yet to lay trail. This reinforces the pheromone path on the lower trail. When the lower path ants arrive back at the nest, new ants will initially see twice as much trail on the lower path and choose it with higher probability. It is not difficult to see that a positive feedback mechanism emerges, which reinforces the desirability of the lower trail. Eventually, all ants will follow the shorter path.

This simple example illustrates the idea behind ant algorithms, which is that simple agents can communicate using distributed memory about the problem—pheromone—to cooperate and solve an optimization problem.

Because the aim here and elsewhere in the engineering literature is to solve an engineering problem and not to model real ant colonies, various simplifications are assumed. Artificial ants differ from real ants in the following ways:

1. They are completely blind.
2. They have some memory.
3. They live in a discrete-time environment.

To explain the original ant system algorithm, it is convenient to do so in the context of the TSP, which provided the original motivation for the algorithms. The basic TSP is formulated as follows: Given N cities distributed randomly within the plane, find the shortest tour that visits each city once only and returns to the original city. The distance $d(i, j)$ between cities i and j with coordinates (x_i, y_i) and (x_j, y_j) is computed using the standard Euclidean norm. One can think of the problem in terms of a graph, where each pair of cities is potentially joined by an edge (i, j). The ants move around in this graph laying a pheromone trail with intensity $\tau(i, j)$ for each edge (i, j) they traverse. This trail can be updated locally as the ants move or can be updated globally at the end of an iteration when all ants have completed a tour. The global updating rule, which forms the basis of the *ant system* algorithm, was shown to be superior to local rules [13]. The ants are forced to complete a tour by maintaining a *tabu list* for each ant; this simply contains a list of the cities that the ant has already visited, and the ant is forbidden to revisit any city in the tabu list.

Given a state of the system at time t in a specific iteration, each ant k is at one of the cities. The probability that an ant will make the journey from city i to city j is assumed to be

$$P_k(i, j) = \frac{[\tau(i, j)]^\alpha [\eta(i, j)]^\beta}{\sum_{j \in J_k(t)} [\tau(i, j)]^\alpha [\eta(i, j)]^\beta} \qquad (9)$$

where $J_k(t)$ are the allowed cities for ant k at time t i.e., the complement of the tabu list. If j is not in $J_k(t)$, the transition is forbidden by setting $P_k(i, j) = 0$. In equation (3), $\tau(i, j)$ is the trail intensity associated with the edge (i, j). The *visibility* $\eta(i, j)$ is the inverse of the distance between cities i and j. This is included to allow a degree of greediness in the algorithm; the ants are more likely to move to a nearer city at a given step. This is not guaranteed to lead to a tour with overall minimum length as at step $N - 1$; the ant may be very far from its starting city. The parameters α and β control the relative importance of trail and visibility. Note that the inclusion of visibility means that the ants are not

totally blind in this variant of the algorithm. After N steps when the ants have completed a tour, the pheromone levels for each edge are updated using the rule

$$\tau(i, j) \longrightarrow \rho\tau(i, j) + \sum_{k=1}^{m} \tau_k(i, j) \qquad (10)$$

where

$$\tau_k(i, j) = \frac{Q}{L_k} \qquad (11)$$

if edge (i, j) is in the tour of the kth ant and $\tau_k(i, j) = 0$ otherwise. Q is a user-defined constant and L_k is the total tour length for ant k in that iteration. The total number of ants is m. The parameter ρ is also user defined. A number between 0 and 1, this represents the trail persistence between iterations; $1 - \rho$ is a measure of trail *evaporation*.

This is how the ant system algorithm is used to solve the TSP. To illustrate an optimization problem here, the algorithm was applied to a known TSP problem—the Oliver30 problem. This was chosen as it was the problem used in the original paper on the ant system algorithm [11]. Overall, 25 runs of the ant system algorithm were carried out. The agreement between the runs was impressive. By 500 iterations, 16 of the runs had arrived at the best previously known solution, and this had a tour length of 423.74 (Figure 15). The remaining nine runs had arrived at a local minimum with tour length 423.91. After 1000 iterations, 22 of the runs had arrived at the best solution, with 3 runs remaining at the local minimum. After 1500 runs, only two of the local minima persisted and after 2000 iterations all 25 runs had arrived at the "optimum". The fastest run hit the best solution at 27 iterations and the slowest run took 1981 iterations. The best tour obtained using a greedy nearest neighbor heuristic had a length of 473.33, substantially worse than that from the ant algorithm.

Having established that the ant algorithm works well on a difficult problem, it is straightforward to apply to problems of engineering interest. Elsewhere in this encyclopedia (*see* **Sensor Placement Optimization**), an application to the sensor placement described above in Section 3 is given. The algorithm is shown to outperform a GA.

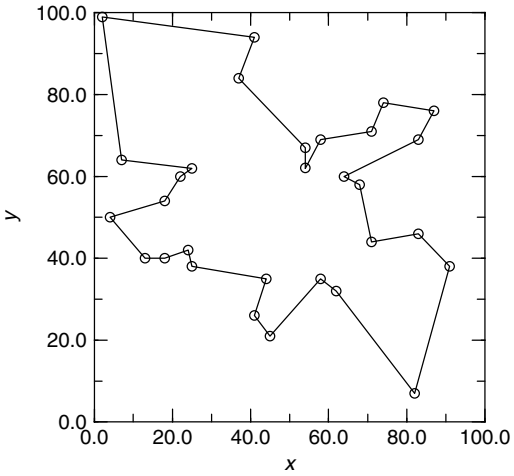


Figure 15. Ant algorithm solution to Oliver30 TSP problem.

6.2 Immune system metaphors

Algorithms based on immune system metaphors are another comparatively recent addition to this group of algorithms. Unlike GAs or SA, the immune system has inspired a diverse range of algorithms designed to address a number of problems including (but not restricted to) pattern recognition, anomaly detection, data analysis, scheduling, machine learning, and optimization. Analogies with the biological immune system have enabled the development of the field of *immune engineering* [22], which encompasses a number of algorithms motivated by more or less detailed aspects of the actual immune system that can be used for the solution of real engineering problems. The field has developed to the point where the first pedagogical treatment has appeared [23]. In the field of SHM, the main applications of immune engineering have been in anomaly or novelty detection, examples of which can be found in [24–26]. The object of this section is to consider an optimization algorithm motivated by the immune system—the clonal selection algorithm (CSA).

The biological immune system is very complex and it would be inappropriate to discuss it in detail here; instead, a brief discussion will explain the motivation of the algorithm, which is the subject of this article. Much of this section is a précis of [22] and the interested reader can refer to the original reference for far more detail.

The human body has two interrelated components for its protection from foreign material. The first of these systems is the *innate immune system*. The body is born with the basic ability to recognize certain invaders and to destroy them. However, this part of the immune system is not relevant for motivating the optimization algorithm required here. The relevant mechanism is the *adaptive immune system*, which is able to optimize its response to microbial agents by modifying certain cells to maximize their ability to recognize chemicals on the surface of the invading cells (called *antigens* here). The adaptive immune system allows the body to learn to recognize any given antigen and mobilize the body's defenses even if it has never seen it before.

The adaptive immune system mainly acts through agents (cells) called *lymphocytes*. These are a heterogeneous set of cells, their numbers being of the order of 10^{12}. The complexity of the immune system thus rivals that of the brain, which typically contains 10^{10} neurons. The lymphocytes can be divided into *B cells* and *T cells*.

The B cells, when activated, differentiate into plasma cells, which can secrete *antibodies*. Each B cell can secrete only one type of antibody. The antibodies are proteins that bind to the antigens and signal other cells to kill, or otherwise deactivate, the invading body. T cells are also vital to the biological immune system; however, one can argue that they play a secondary, regulatory role in motivating the engineering algorithm needed for optimization problems.

The antibodies are composed of *constant regions* and *variable regions*. The constant regions are responsible for a number of functions that are necessary for the effective functioning of the immune system; however, they are again secondary in motivating the optimization algorithm. The variable regions are those that are subject to frequent mutation and can therefore adapt to maximize their ability to bind to the antigen. The actual computational agents or individuals of the optimization routine really represent only the variable regions of the antibodies.

The principle that motivates the optimization routine is the *clonal selection principle*. When exposed to antigens, the B cells respond by producing antibodies. The antigen stimulates the B cells to divide and proliferate into the antibody secreting plasma cells. Only the B cells that recognize the antigen proliferate by cloning. High levels of mutation—*hypermutation*—allow the system to adapt by maximizing the affinity of the antibodies for the antigen. The initial population of B cells has sufficient diversity to effectively adapt to the attack of any antigen. The final result of the adaption is a population of cells adapted to the removal of the antigen; this population endures in the body in the form of *memory cells*, which can then form the defense against a secondary attack involving the same antigen. Another important ingredient of the process, which will not be pursued here, is *self–non-self discrimination*, which ensures that the antibodies do not evolve into forms that may attack the host cells. This is important in motivating algorithms for novelty detection [24–26].

The engineering version of the adaptive immune system is basically constructed by analogy, with the objective of generating the population of memory cells that are optimized for a given antigen. The idea is to begin with an initial population of individuals and then iterate by a process of selection and mutation in much the same way as standard evolutionary algorithms. In fact, the algorithm here—the *CSA*—can broadly be regarded as a type of GA without crossover. The individuals in the algorithm will be identified as the antibodies, essentially ignoring the distinctions between the real entities, the B cells, plasma cells, and biological antibodies. The quantity that is optimized in the real immune system is the affinity between the antibodies and the antigens; the affinity in the CSA is therefore defined as the value of the objective function, which must be maximized. As in the binary GA, the individuals of the population here are binary strings that encode the parameters to be optimized. The steps in the CSA can be summarized as follows (see also Figure 16):

1. Define the problem—objective function and encoding.
2. Generate random initial population of antibodies—$\mathbf{Ab_1}$. The population number of individuals is set at P.
3. Calculate the affinities. This is the standard process of decoding the individuals and substituting into the objective function.
4. Choose the N best individuals—based on their affinities.
5. Cloning by taking C copies of each of the best individuals; these clones form the set $\mathbf{Ab_2}$. The

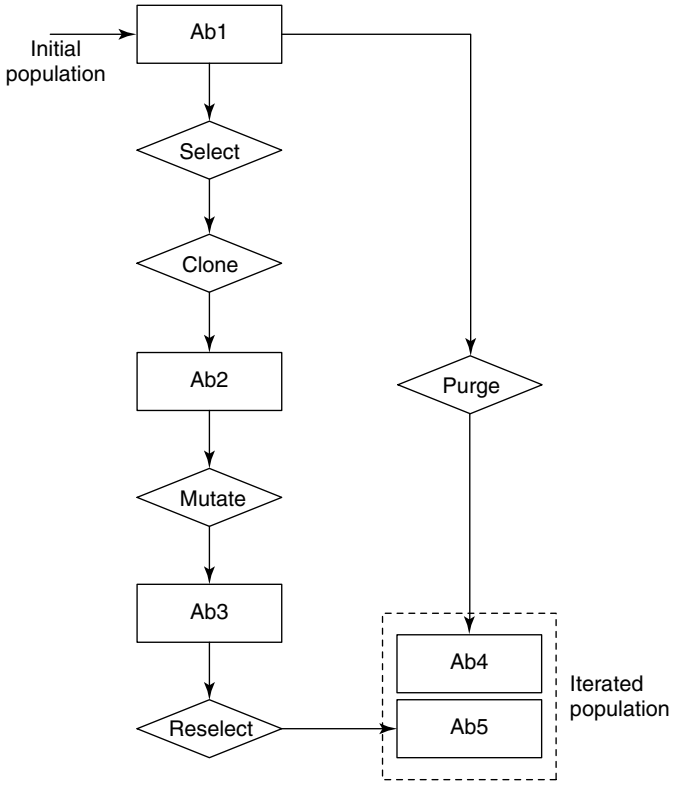


Figure 16. Schematic for the CSA algorithm.

number of clones of an individual can also be proportional to its affinity.

6. The clones are now subjected to mutation—applied in the same way as the binary GA by flipping bits of the binary string that represents the individual. A high mutation probability is used. This forms the set $\mathbf{Ab_3}$.
7. Remove a percentage of the individuals with lowest affinity from $\mathbf{Ab_1}$—this forms the set $\mathbf{Ab_4}$. This step represents the death of the least-stimulated cells, which is a feature of the biological immune system. Alternatively, one can replace the least-stimulated cells with random individuals to maintain diversity.
8. Reselect a proportion of the cells with highest affinity from $\mathbf{Ab_3}$ to form the set $\mathbf{Ab_5}$ (these are the memory cells), which is added to $\mathbf{Ab_4}$ to form the new population.
9. The process is iterated until some performance criterion is satisfied.

The various parameters are balanced to maintain a constant population size, which gets successively better (has higher average affinity) as the algorithm progresses.

An illustration of the use of the CSA algorithm on the sensor placement problem considered in Section 3 of this article can be found in [27].

7 A FINAL CASE STUDY—DAMAGE DETECTION

In this section, the previously discussed SA algorithm and the GA are applied in two cases: first when data obtained from an FE beam simulation with multiple cracks are analyzed; second, when experimental data measured from three cantilever steel beams with two cracks are examined to estimate depth and positions of their cracks. The results obtained with these optimization techniques are discussed to highlight the relative merits and drawbacks.

7.1 Optimization approach

The basis for the damage detection method is to establish a model of the system and then to compute the parameters of the model that gives the best (in some sense) correspondence with experimental data. As usual, one first formulates an objective function, $F(\cdot)$, which gives a scalar measure of the difference between measured and calculated dynamic characteristics of the structure under consideration, $\mathbf{D}^{(\mathrm{m})}$ and $\mathbf{D}^{(\mathrm{c})}$, respectively. An estimate $\mathbf{R}^{(\mathrm{s})}$ of the state of damage is the value $\mathbf{R}$ for which the function $F(\cdot)$ is a minimum, where

$$F(\mathbf{R}) = \|\mathbf{D}^{(\mathrm{m})} - \mathbf{D}^{(\mathrm{c})}(\mathbf{R})\| \tag{12}$$

$\mathbf{R}^{(\mathrm{s})}$ can be estimated using various optimization procedures, but one should be aware that the estimate of the state of damage given by the numerical procedure may be quite different from the correct value. This is usually the case in which the objective function has many local minima and each one is related to an incorrect state of damage. In general, one would hope that just the global minimum provides an accurate estimate of the actual state of damage of the structure.

As an example of damage identification as an optimization problem, Shen and Taylor [28] utilized a min–max optimization to assess the damage in a simply supported beam; they used a cost function with both the natural frequencies and the mode shapes of the structure,

$$\left(\sum_{i=1}^{n_f} \left(\left(f_i^{(\mathrm{c})}(r_0, a) - f_i^{(\mathrm{m})} \right)^2 + \sum_{j=1}^{n_\mathrm{p}} \left(\phi_{ij}^{(\mathrm{c})}(r_0, a) - \phi_{ij}^{(\mathrm{m})} \right)^2 \right) \right)_{\min} \tag{13}$$

where n_f is the number of considered modes, r_0 gives the position of the crack, a its depth, n_p the number of measurement points, and ϕ_{ij} is the ith mode shape at the point j; the superscript (c) refers to the calculated and the superscript (m) to the measured dynamic characteristics of the structure. According to this method, and assuming a known location r_0, the authors obtained a good estimate for the crack depth, while they found very poor results trying to evaluate both the location and the size of the crack, probably due to the presence of local minima.

For maximum generality, it was decided here to use an FE model of the beam in which each element is potentially cracked [29, 30]; in this way the number of cracks in the structure is not limited. As a consequence, the stiffness matrix of the structure can be written as follows:

$$\mathbf{K} = \mathbf{K}(\mathbf{R}) \tag{14}$$

where $\mathbf{R}$ is a vector representing the state of damage,

$$\mathbf{R} = [r_1 r_2 r_3 \ldots r_n]^{\mathrm{T}} \tag{15}$$

n indicates the number of elements into which the structure is subdivided, and $r_i = a_i / h$ with a_i being the depth of the crack in the ith element and h the height of the beam.

In general, the problem of local minima can be avoided to an extent if more knowledge of the structure physical properties is introduced into the objective function and some boundary conditions are included to limit the state-of-damage domain [28]. To compare the performance of GA with SA the following formulation, derived from [29], is used:

$$F(\mathbf{R}) = \frac{1}{\alpha g_1(\mathbf{R}) + v(\mathbf{R})} \tag{16}$$

Table 3. Simulated damage scenarios

Case	First cracked element			r_i			Second cracked element			r_i		
	True	GA	SA	True	GA	SA	True	GA	SA	True	GA	SA
1	5	5	5	0.20	0.197	0.197						
2	3	3	3	0.20	0.197	0.197						
3	1	1	1	0.15	0.150	0.151	4	4	4	0.10	0.102	0.090

where

$$g_1(\mathbf{R}) = \sum_{i=1}^{N} \left(1 - \frac{f_i^{(\mathrm{m})}/f_i^{*(\mathrm{m})}}{f_i^{(\mathrm{c})}(\mathbf{R})/f_i^{*(\mathrm{c})}} \right) \qquad (17)$$

with N being the number of natural frequencies.

$$v(\mathbf{R}) = \sum_{i=1}^{n} r_i \qquad (18)$$

denotes the global damage of the structure as the sum of damage affecting n elements making up the mathematical model.

7.2 Numerical results

The procedure for damage assessment described here was applied to identify the damage in a beam, cracked in one or two locations (Table 3), with the following properties: elastic modulus $E = 2.06 \cdot 10^{11}\,\mathrm{N\,m^{-2}}$, mass density $\rho = 7850\,\mathrm{kg\,m^{-3}}$, length $L = 0.7\,\mathrm{m}$, and cross-sectional area of $0.020 \times 0.020\,\mathrm{m^2}$. The beam was discretized in 10 elements. Both GAs and SA were applied to maximize the objective function, in which α was set equal to 80 000; the relative results are also shown in Table 3 and the evolution of the objective function during the optimization has been plotted in Figure 17, where it is evident that with 10 000 evaluations of the cost function, it is possible to reach the global maximum using both SA and the GA. Furthermore, it may be stated that it is necessary to run the GA several times (five times here) to properly account for the random initializations of the algorithms. This consideration can be important in assessing damage in more complex structures, where a greater number of objective function evaluations is potentially needed to reach an accurate estimate of the state of damage.

7.3 Experimental results

To validate the damage assessment procedure, data obtained by performing tests on three cantilever steel beams have been analyzed. These beams were tested both in the undamaged and in the cracked states, so it was possible to update their FE models in which accelerometer mass, stiffness of the clamping device, and elastic modulus of the beam were assumed as variables. On completion of the updating task, values

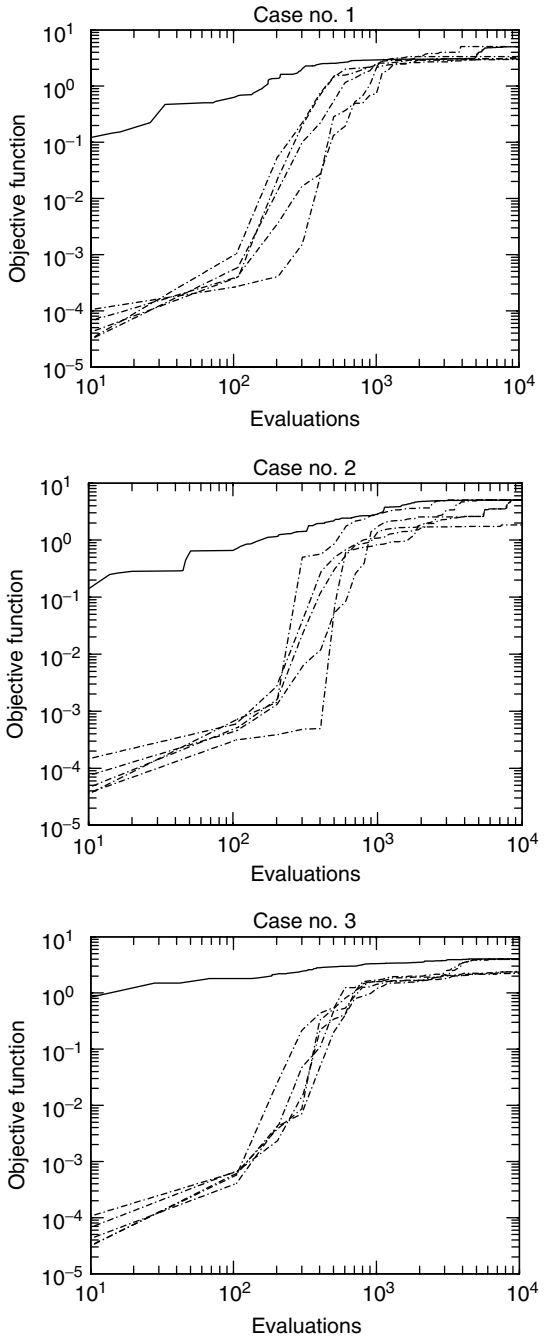


Figure 17. GA fitness evolution (dashed) and SA fitness evolution (solid).

Table 4. Experimental cantilever beams

Case	First cracked element			r_i			Second cracked element			r_i		
	True	GA	SA	True	GA	SA	True	GA	SA	True	GA	SA
C1	4	4	4	0.20	0.207	0.213	8	8	8	0.20	0.207	0.206
C2	4	3	3	0.20	0.127	0.174	8	8	8	0.30	0.367	0.349
C3	4	4	4	0.30	0.267	0.267	8	8	8	0.20	0.258	0.255

for $f_i^{(c)}$ were computed and introduced into the objective function formulation. The results obtained by running both SA and genetic algorithms are shown in Table 4. This table highlights that both the procedures identify the correct state of damage except for beam C2 (where the estimate for the location of one crack is wrong) and give sufficiently accurate results.

8 CONCLUSIONS

Only the briefest of conclusions is warranted here. It is clear that SHM inspires a number of important and complex optimization problems. It is fortunate that the mathematical discipline of optimization provides a battery of powerful techniques that can address these problems. Amongst the most powerful, particularly for combinatorial optimization problems, are the heuristics which have been inspired by physical and biological metaphors and several of these techniques have been illustrated in this article.

REFERENCES

[1] Luenberger DG. *Linear and Nonlinear Programming*. Springer, 2003.

[2] Fletcher R. *Practical Methods of Optimisation*. Wiley-Interscience, 2000.

[3] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical Recipes – The Art of Scientific Computing*. Cambridge University Press, 1986.

[4] Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 1953 **21**:1087–1092.

[5] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimisation by simulated annealing. *Science* 1983 **220**:671–680.

[6] Worden K, Burrows AP. Optimal sensor placement for fault diagnosis. *Engineering Structures* 2001 **23**:885–901.

[7] Corana A, Marchesi M, Martini C, Ridella S. Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Transactions on Mathematical Software* 1987 **13**: 262–280.

[8] Mahfoud SW, Goldberg DE. Parallel recombinative simulated annealing: a genetic algorithm. *Parallel Computing* 1995 **21**:1–28.

[9] Yip PCP, Pao YH. Combinatorial optimisation with use of guided evolutionary simulated annealing. *IEEE Transactions on Neural Networks* 1995 **6**:290–295.

[10] Holland JH. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence, New Edition*. MIT Press, 1992.

[11] Michalewicz Z. *Genetic Algorithms + Data Structures = Evolution Programs*, *Thrid Edition*. Springer-Verlag, 1996.

[12] Goldberg DE. *Genetic Algorithms in Search, Optimisation, and Machine Learning*. Addison-Wesley Publishing: Reading, MA, 1989.

[13] Worden K, Staszewski WJ. Impact location and quantification on a composite panel using neural networks and a genetic algorithm. *Strain* 2000 **36**:61–70.

[14] Storn R, Price K. Differential evolution – a simple and efficient heuristic for global optimisation over continuous spaces. *Journal of Global Optimization* 1997 **11**:341–359.

[15] Dorigo M, Maniezzo V, Colorni A. *Positive Feedback as a Search Strategy*, Technical Report no. 91-016. Politecnico di Milano, 1991.

[16] Dorigo M, Maniezzo V, Colorni A. The Ant system: optimisation by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics. Part B* 1996 **26**:1–13.

[17] Bilchev G, Parmee IC. The ant colony metaphor for searching continuous design spaces. *Evolutionary Computing, Selected Papers From AISB Workshop*. Sheffield, 1995.

[18] Bilchev G, Parmee IC. Constrained optimisation with an ant colony search model. *Proceedings of ACEDC*. Plymouth, 1996; pp. 145–151.

[19] Goss S, Aron S, Deneuborg JL, Pasteels JM. Self-organised shortcuts in the Argentine ant. *Naturwissenschaften* 1989 **76**:579–581.

[20] Höldobler B, Wilson EO. *The Ants*. Springer-Verlag, 1990.

[21] Beckers R, Deneuborg JL, Goss S. Trails and U-turns in the selection of the shortest path by the ant lasius niger. *Journal of Theoretical Biology* 1992 **159**:397–415.

[22] Nunes de Castro L, Von Zuben FJ. *Artificial Immune Systems: Part I – Basic Theory and Applications*, Technical Report TR-DCA 01/99. Unicamp, 1999.

[23] Nunes de Castro L, Timmis J. *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer, 2002.

[24] Dong Y, Dong E, Jia H, Lv W. A biological immunity-inspired novelty detection algorithm for rotor system monitoring. *Key Engineering Materials* 2005 **294**:71–78.

[25] Dong Y, Sun Z, Jia H. A cosine similarity-based negative selection algorithm for time series novelty detection. *Mechanical Systems and Signal Processing* 2006 **20**:1461–1472.

[26] Surace C, Worden K. A negative selection approach to novelty detection in a changing environment. *Proceedings of 3rd European Workshop on Structural Health Monitoring*. Granada, 2006.

[27] Zhang J, Worden K. Sensor optimisation using an immune system metaphor. *Proceedings of 26th International Modal Analysis Conference (IMAC)*, Orlando, FL, 2008.

[28] Shen M-HH, Taylor JE. An identification problem for vibrating cracked beams. *Journal of Sound and Vibration* 1991 **150**:457–484.

[29] Ruotolo R, Surace C. Damage assessment for a beam with multiple cracks. *Proceedings 21st International Seminar Modal Analysis*. Leuven, 1996; pp. 1005–1016.

[30] Ruotolo R, Surace C. Damage assessment of multiple cracks beams: numerical results and experimental validation. *Journal of Sound and Vibration* 1997 **206**:567–588.

Chapter 39
Uncertainty Analysis

Graeme Manson[1], Keith Worden[1], S. Gareth Pierce[2] and Thierry Denoeux[3]

[1] *Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*
[2] *Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK*
[3] *Université de Technologie de Compiègne, Compiègne, France*

1 INTRODUCTION

Future engineering will rely increasingly on computer simulations and modeling for a wide range of processes in the design-to-decommissioning life cycle—from design and prototyping, through production planning, manufacture, and in-service monitoring and ending with a safe decommission. This "virtual" approach usually offers clear financial and/or environmental advantages. However, a fundamental concern is that the modeling process must be robust to uncertainty: a high-fidelity model can become worthless if there is no understanding of how uncertainty is propagated through it. To elaborate on the concept of uncertainty propagation, consider a complex component incorporating inelastic materials, contacts, friction, etc. The model requires an initial specification of material properties, clearances, and friction coefficients. The response of the model will only be as accurate as these initial values, yet the values may be subject to large uncertainty; consider the difficulty in the mechanical characterization of biomaterials or viscoelastic polymers. In a worst-case analysis, the predictions of the model may prove highly sensitive to variations in the input parameters.

In 1999/2000, members of the Engineering Analysis group, amongst others, at Los Alamos National Laboratories (LANL), carried out arguably the most ambitious calculation in the (short) history of structural dynamics. Using one of the most powerful computers in the world at the time, the platform *Blue Mountain*, the computation essentially attempted

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

to quantify the propagation of uncertainty through a nonlinear finite element (FE) model of a weapon component under blast loading [1]. The calculation made use of 3968 processors from the available 6000 and used concurrently 4000 ABAQUS/Explicit licenses. The analysis took over 72 h and would have required 17.8 years of equivalent single-processor time.

The scale of this calculation is testimony to the importance that uncertainty quantification and propagation (UQP) is inevitably going to assume in the immediate future of structural dynamics and also in the long term. One of the reasons for the LANL interest in UQP is that current test bans in the field of nuclear weaponry forbid the type of experimentation which would usually be used to validate physical and computational models. Another compelling reason for UQP is the fact that the environmental and operational conditions for certain engineering projects are unknown and cannot be estimated with certainty. Alternatively, conditions can be estimated, but cannot be recreated exactly in validation experiments. Consider the design and manufacture of an in-orbit facility. In this case, aspects of the true environment, such as weightlessness and absence of drag, can be reproduced to an extent in terrestrial experiments. However, consider the design of an exploration module intended to operate on the surface of Venus. Such systems are extremely expensive and will often have only one chance to succeed. The performance of the Hubble telescope or the Mars Lander Beagle 2 is an object lesson here.

Another prime motivator for uncertainty analysis is the need for risk assessment in safety-critical systems. In fact, one might argue that the modern origins of the subject are here with projects like the Nuclear Reactor Safety Study [2]. One of the main regulators in the design of structural health monitoring (SHM) systems for say, a civil aircraft, will be the need to eliminate false assertions of damage as a result of environmental and operational variations. Such false positives and the ensuing "no fault found" inspections would substantially increase the cost of ownership of the aircraft. Even more critical is the possibility of a false negative as a result of environmental uncertainty, with potential loss of life being the result.

The purpose of this article is to introduce the issue of uncertainty analysis in general and also its relevance to SHM. In the next section, brief descriptions of the most popular of the many frameworks for uncertainty representation are given. Section 3 discusses the three main uncertainty-related problems of relevance to structural dynamics, namely, quantification, fusion, and propagation. In order to illustrate the ideas of the preceding two sections in a realistic scenario, two case studies conducted on an aerospace structure, namely the wing of a Gnat trainer aircraft, are provided. The first case study, in Section 4, considers the issue of attaining certification for artificial neural network (ANN) damage classifiers through the assessment of the network's robustness to uncertainty. This case study involves the propagation of intervals through the network structure and examines whether the networks, which would be considered as optimal, in the traditional sense, are also most robust to uncertainty. In Section 5, the second case study considers evidence-based classifiers as an alternative to probabilistic classifiers for the problem of damage location. The Dempster–Shafer (DS) theory is employed to construct neural network classifiers with the potential to admit ignorance, rather than misclassify. The section considers issues of propagation and fusion in an evidence-based framework and compares the performance of DS neural networks with their probabilistic counterparts. The article ends with some brief conclusions.

2 THEORIES OF UNCERTAINTY

There are many frameworks available through which uncertainty may be represented and this section examines some of the more common theories, considers their pros and cons, and looks for applications in the literature of structural dynamics. It is not the intention to give an exhaustive review of the literature and, in particular, the large body of work in the high-frequency/Statistical Energy Analysis (SEA) arena is ignored.

2.1 Classical probability theory

The classical probability theory is well known. In fact, if the variations in a parameter are random, there is no better specification of the uncertainty than a probability distribution. However, in practice, this is often not available. Engineering analysis is routinely

based on small samples and one might only be able to estimate the low-order moments of a distribution—mean and variance—with any confidence. Arbitrarily imposing a known distribution shape e.g., Gaussian, on the basis of this information is perilous. In particular, the use of such central statistics may result in a distribution radically different in the tails from the true distribution. In risk analysis, where one is concerned with extreme events, the results of such a strategy will probably be meaningless. Another problem, in general, is that a specification of a problem will necessarily include a region of ignorance, and classical probability theory cannot accommodate this. In particular, a statement of the probability of an event automatically fixes the probability that it will not occur. The evidence for the occurrence of the event is essentially the same as the evidence for nonoccurrence.

2.2 Evidence theory

This can be regarded as an extension of the theory of probability (although this interpretation is contested). The main theory of this sort is DS theory [3, 4], although generalizations exist [5]. Essentially, the single probability is replaced here with two quantities. The *belief, Be*, associated with an event is the sum of the evidence in favor of the event. The *plausibility, Pl*, is the complement of the evidence against the occurrence of the event. By the fact that these quantities can be different, DS theory accommodates the idea of ignorance. DS theory thus replaces the single probability with an interval [*Pl, Be*], and these quantities are sometimes called *lower and upper probabilities. Be* can be regarded as the best-case estimate of the probability of an event, corresponding to the case when all the missing evidence (ignorance) turns out to be in support of the event. Applications in engineering of DS theory are rather few and far between, [6] cites 12 application references in total, but interestingly exclude [7, 8] and those that specifically consider problems of SHM.

2.3 Possibility theory

Similar to DS theory, the possibility theory makes use of two complementary uncertainty measures—possibility and necessity [9]. Essentially, some proposition e is mapped into the interval [0,1], which may be divided into the three intervals: necessity $N(e)$, necessity of the contrary proposition $N(e')$, and ignorance $\theta(e)$. Possibility of the proposition is given by $P(e) = N(e) + \theta(e) = 1 - N(e')$. Applications of the possibility theory are even rarer than applications of DS, [6] cites only three. In some ways, the possibility theory can be interpreted in terms of fuzzy sets [10].

2.4 Fuzzy logic

This is one of the elder statesmen of contenders with probability theory [11]. This extends the classical probability theory by relaxing one of the fundamental set-theoretic properties on which it is based. In the classical set theory, an element x is either a member of a set A or a member of its complement A^c. In fuzzy set theory, x may be associated (with given weight) with a number of different sets. Fuzzy logic encodes uncertainty by associating linguistic descriptors with a variable x like *large* or *small*. x may be a member of both sets, large and small, but it is associated to each by a membership function that mediates the likelihood of its membership. There are analogs of all the basic mathematical operations for such fuzzy variables (often based on interval arithmetic) and it is possible to construct fuzzy analogs of many *crisp* theories. The most intensive use of fuzzy logic in engineering is associated with control, although the interesting studies [12–14] construct a fuzzy version of modal analysis. The same authors have also formulated a fuzzy version of FE analysis [15].

2.5 Interval methods

Rather simply stated, the interval method replaces *crisp* numbers and variables with intervals $[\underline{x}, \overline{x}]$ [16]. Uncertainty in a parameter is encoded in the statement that it lies somewhere between two given bounds $\underline{x}$ and $\overline{x}$ with certainty. Note that this information could be incorporated in a probabilistic analysis by giving a distribution for the parameter on the interval, e.g., a uniform distribution. However, interval analysis makes no use of such additional

assumptions. There are again analogs of all the expected arithmetical operations, which thus allow the propagation of interval quantities through various types of models. There are numerous uses of interval arithmetic documented in the literature. In terms of structural dynamics, the previously cited study on fuzzy modal analysis, [13], made extensive use of interval arithmetic. In [17], there is an independent attempt to formulate a fuzzy FE analysis in terms of interval arithmetic. The paper [18] is of interest in that it considers a problem of damage detection. The problem with interval arithmetic is that it is conservative in nature and that the bounds on the calculations expand considerably in practical computations. An attempt to improve on this behavior is under development in the form of *affine arithmetic* [19]. An application to the structural eigenvalue problem can be found in [20].

2.6 Convex models

In a sense, the convex models can be regarded as a generalization of the interval concept of uncertainty (although they are more than this). A given parameter p is associated with a convex set, which may be said to contain the parameter. For example, an *ellipsoidal-bound* convex model takes the form

$$\Im(\alpha, p) = \{p : (p - \overline{p})^T W (p - \overline{p}) < \alpha\} \quad (1)$$

where W specifies the axes of the ellipsoid containing the data and $\overline{p}$ is the mean of the data set. This is a convex set. This approach to uncertainty was pioneered by Ben-Haim and has been applied by him and coworkers in numerous contexts [21]. It is possible to prove numerous theorems such as: if the input to a linear system is a convex model, then so is

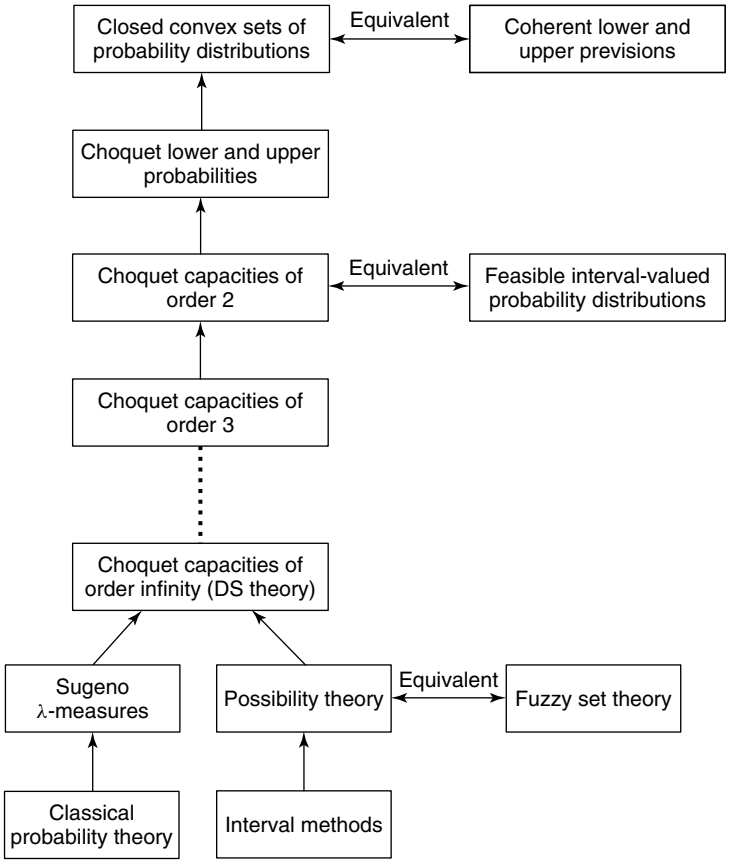


Figure 1. Klir and Smith's hierarchy of theories of uncertainty.

the output. This property does not hold for a nonlinear system, but by relaxing the constraints one obtains an *information-gap* model and recovers the property of invariance under a nonlinear operator [22]. An interesting formulation of an information-gap model on a problem of structural dynamics can be found in [23].

2.7 Klir and Smith's hierarchy of uncertainty theories

In a sense, the approaches described above are lower-level approaches to modeling uncertainty. In the survey [24], Klir and Smith identify a hierarchy of theories of uncertainty as shown in Figure 1.

As one moves up the diagram, one encounters more general representations of uncertainty. In practice, none of the theories above DS theory have so far been applied to engineering problems; however, they have potential for the future. All the theories above the classical probability theory are based on a more general form of measure—*capacity* [25]. The classical probability theory has, for disjoint events A and B, that

$$p(A \cup B) = p(A) + p(B) \tag{2}$$

i.e., the measure p is *additive*. Choquet's capacities g allow

$$g(A \cup B) >< g(A) + g(B) \tag{3}$$

i.e., they account for the possibility of mutual cooperation or inhibition between events. As discussed above, this freedom has so far been neglected in terms of engineering applications.

3 QUANTIFICATION, FUSION, AND PROPAGATION OF UNCERTAINTY

In terms of uncertainty, structural dynamics has arguably three main concerns, which are discussed briefly in this section.

3.1 Quantification

Given a parameter a, associate with this a meaningful uncertainty measure with the framework of a given uncertainty model. For example, assign a probability in the classical probability theory or a basic probability assignment (BPA) in DS theory. The process of quantification may involve several steps, each refining the measure. For example, in [26], the following process is described within the context of the Bayesian probability theory. A prior distribution is defined, for example, for a given model parameter. If it exists, expert opinion data are used to update/refine the priors to obtain a posterior distribution for the parameter. This posterior then becomes the prior for stage 2 whereby the results of computer simulation are used to update. Finally, the posterior from stage 2 becomes the prior for stage 3 where experimental data are used to update. The multistage process may also be thought of as a precursor to fusion.

3.2 Fusion

Given a parameter a, which has a set of uncertainty measures $m(a)_i$ associated with different uncertainty models U_i, how does one refine each measure in the light of the others? For example, if one is given a classical probability 0.75 for an event, and lower and upper probabilities [0.6, 1.0] from DS theory, how can one refine each measure given the other? Associated with this problem is that of *normalization* i.e., which uncertainty measure in the convex model theory corresponds to a given probability or BPA? Or is it even possible to make a quantitative comparison? This is largely an open problem, although some preliminary results are available; [27] sketches a relationship between fuzzy set theory and the classical probability theory.

3.3 Propagation

If a parameter a, with a given uncertainty measure $m(a)$, forms the input to a given physical process, which uncertainty measure should be associated with the output, or a given aspect of the output? As an example, consider the discussion in [28] a control problem. The aim is to establish, given a system with uncertain parameters, the probability that the derived

controller will be stable. An associated problem here is the question of *sensitivity*, i.e., which parameters, with their associated uncertainties, are the main cause of variation in the model/system response.

4 CASE STUDY 1: ASSESSMENT OF NEURAL NETWORK ROBUSTNESS FOR DAMAGE CLASSIFICATION

The first case study considers the application of ANN classifiers to the interpretation of data for the purposes of damage location. ANN classifiers have found diverse applications in many fields including aircraft wing damage detection [29, 30]. Conventional network training can be viewed within a framework of the global optimization (minimization) of the error function between the network output prediction and the true target data. There exist a number of strategies to locate this minimum, the most common being variants of gradient descent such as scaled conjugate gradients [31, 32]. These techniques make use of the local gradient information of the error surface to ascertain the optimum search direction but are susceptible to the danger of locking a solution into a local minimum rather than locating the true global minimum. A number of strategies have been devised to counter this problem [31–33], and new techniques of searching the error function space are currently active areas of research, for example, using genetic algorithms (GAs) [32].

An additional complication to network performance lies in the capability of a network to generalize its classification performance to previously unseen data. There exists the widely recognized problem of overtraining that can occur such that a network starts to learn the noise present in data rather than the underlying data structure. The use of cross-validation and early stopping [33] using an independent validation data set are often used as termination criteria for training to help avoid this problem.

The problems of network overtraining and lack of generalization are central to understanding the inertia to the practical application of ANNs, especially to safety-critical applications. If, for example, one envisages an ANN classifier being used to assess the condition of a major structural airframe component, it is imperative that the performance of the classifier to the most diverse range of inputs is well understood. Poor classifier performance could result in catastrophic failure and possible loss of life. Although a range of techniques have been developed for output confidence interval predications [32, 34–37], they all adopt a probabilistic standpoint and therefore suffer from the common drawback that since the probability distributions are usually estimated from the low-order moments of the data (typically mean and standard deviation), there is often no representation associated with the extremes of the distributions. Unfortunately, it is often the extreme events of the data that are likely to be associated with the unpredictable failure events of greatest interest.

This case study comprises a novel nonprobabilistic approach applied to predicting extreme network outputs in the presence of uncertainty in the input data. This technique is based on the theory of convex models and information-gap uncertainty as pioneered by Ben-Haim [38–41]. Interval-based [42] techniques are extended to investigate the response of a simple multilayer perceptron (MLP) network used for a classification problem to locate damage sites on the wing of a Gnat trainer aircraft. A comparison with conventional network training based on cross-validation is presented. It is shown that the use of interval-based network propagation allows a new criterion for network selection to be established. This technique allows the identification of an ANN classifier, which is intrinsically more robust to noise on the input data than network solutions obtained by conventional maximum likelihood training. Furthermore, by virtue of the conservative nature of interval sets, the reliance on probabilistic-based estimates of confidence bounds on network predictions is obviated. The interval-based worst-case error predictions represent an inclusive bounded solution set given a specified degree of input noise to the classifier.

4.1 Experimental data acquisition and signal processing

The work concerns an SHM strategy to the problem of damage location on an aircraft (Gnat) wing located at DSTL Farnborough, as shown in Figure 2. The wing was instrumented with an array of 12 accelerometers to measure the response to forced

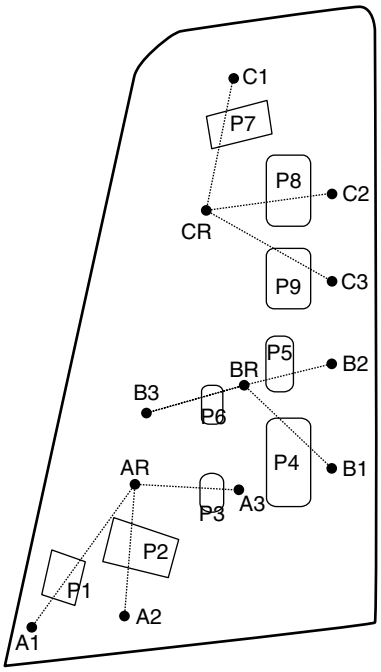


Figure 2. Detail of Gnat wing, showing position of sensors and removable panels.

vibration induced from a shaker mounted to the underside of the wing. The wing had a series of nine removable panels (P1–P9), which could be removed and replaced to provide a reproducible and reversible representation of changing conditions on the wing structure. In this fashion, damage represented by a local change in stiffness properties could be simulated.

Data were collected from all 12 accelerometers for a variety of undamaged (normal condition with all panels in place) and simulated damage data. Note from Figure 2 that the accelerometers were positioned in three distinct groupings (A, B, C) across the plate. Rather than record the individual acceleration responses, the experiment was configured to record the ratio of measured accelerations between transducer pairs AR/A1, AR/A2, AR/A3, etc. in such a way that the transmissibilities between transducer pairs formed the base measurement [29, 30]. In this fashion, there were a total of nine measurement variables recorded. Figure 3 illustrates a typical raw transmissibility measurement with 1024 spectral lines recorded with 1 Hz width in the frequency range 1024–2048 Hz.

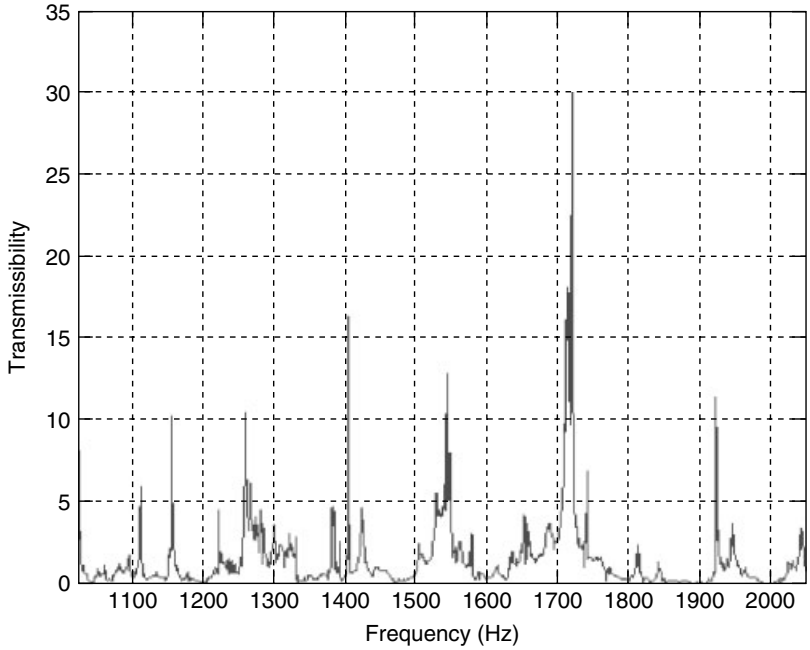


Figure 3. Example of a transmissibility plot spectrum.

By systematic removal of panels P1–P9, the effect of simulated damage on the spectral response of the transmissibility functions could be observed. For each panel removed, 100 individual measurements were recorded for each of the nine separate transmissibilities (labeled T1 through T9) corresponding to removal of panels P1–P9. Two complete runs through the set were completed making a total of 18 runs each with 9 transmissibilities each with 100 measurements; giving a total of 16 200 results. Additionally, 7 normal condition (undamaged) cases were recorded for each of the 9 transmissibilities, again taking 100 individual measurements for each, thus giving a total of 6300 results representing normal condition.

The data were inspected manually to select features (spectral line windows) from the T spectra (transmissibility spectra) that corresponded to particular damage events. This problem was simplified by using a nonlocal argument with respect to the transducer groups [29, 30]. For removal of panels P1–P3, only the T spectra from T1, T2, T3 (corresponding to the accelerometers A1–A3 and AR) were considered relevant, and the spectra from transducer groups B and C were ignored. Similarly, for the removal of panels P4–P6, only measurements from the B set of transducers were considered and likewise only measurements from the C set of transducers for the removal of P7–P9.

From a total of 77 individual features [29, 30], a single feature was selected (by inspection) to correspond to the removal of a single panel. In this fashion, the feature set was reduced to nine individual features (F1–F9). For each individual feature, the data comprised 700 normal condition measurements, and 1800 test measurements. An outlier (novelty) analysis was then performed [43] by computing the Mahalanobis squared distance of the data points with respect to the normal condition data.

$$\Delta^2 = (x - \mu)^T \mathbf{\Sigma} \Sigma^{-1} (x - \mu) \tag{4}$$

where Δ is the Mahalanobis distance, x is the data, μ is the mean of the normal condition data, and $\mathbf{\Sigma}$ is the covariance matrix of the data.

By performing the novelty analysis for each of the nine features, a data matrix of size [1800 by 9] was obtained. This was divided into three equal parts to form separate *training, validation*, and *test* data sets for subsequent network evaluation. Finally, the data sets were logarithmically compressed and normalized to -1 to $+1$ before presentation to the network.

4.2 Network topology and training

The MLP network implementation and training was undertaken in MATLAB™ using the NETLAB toolbox developed by Nabney [32]. The data were presented to a series of MLP networks with different numbers of hidden nodes. Each network had nine input nodes corresponding to the features F1–F9, and nine output nodes corresponding to the classes P1–P9 (Figure 4). The input values were x_i, $i = 1, \ldots, d$. The outputs from the second layer were given by

$$a_k^{(2)} = \sum_{j=1}^{M} w_{kj}^{(2)} \tanh \left[\sum_{i=1}^{d} w_{ji}^{(1)} x_i + b_j^{(1)} \right] + b_k^{(2)} \quad \text{for } j = 1, \ldots, M \quad \text{and} \quad k = 1, \ldots, C \tag{5}$$

where w is the weight matrix, b is the bias matrix, M is number of input nodes, and C is number of output nodes.

The network output was given by transformation of the second layer activations by the output activation function. Since there were a series of C independent output classes, it was appropriate to utilize the *softmax* function [32]:

$$y_k = \frac{\exp[a_k^{(2)}]}{\sum_{k'} \exp[a_{k'}^{(2)}]} \tag{6}$$

The choice of the softmax activation function ensured that the outputs always summed to unity, and thus could be directly interpreted as class conditional probability values.

The number of hidden nodes in the second layer was varied between 1 and 15 hidden units. Each individual network structure was trained with 100 independent training sessions starting at differently randomly chosen points on the error surface so that a total of 1500 independent networks were evaluated. Up to 1000 iterations of a scaled conjugate gradient optimization were implemented within a maximum likelihood training framework using a small hyperparameter $\alpha = 0.001$ to control weight decay.

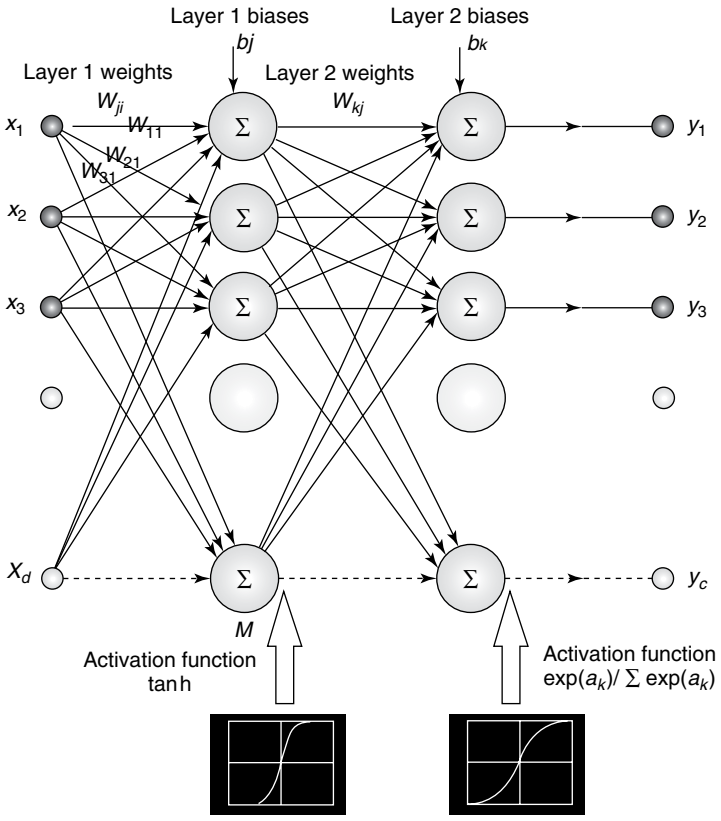


Figure 4. Multilayer perceptron network structure.

4.3 Conventionally trained network results

Traditional network selection was implemented by dividing the data into three equal portions designated the training, validation, and test data sets. The training data were used to train the networks; the best network was then selected from the 1500 possibilities by finding the best classification rate on the validation data set. Finally, the performance of this selected best network was assessed using the test data set. It was found that a network with eight or nine hidden nodes produced an excellent classification performance for a relatively compact network structure.

However, considering that an eight-hidden-node network structure has 161 independent weight and bias components and the total number of training examples was 594 (66 examples of each class), it is very likely that the network structure with eight hidden nodes was likely to possess poor generalization performance. It was for this reason that the maximum network size considered in the present analysis was four hidden nodes, which have 22 independent weight and bias components. This choice of network provides reasonable classification performance on the validation data, whilst maintaining a relatively simple network structure. Using a cross-validation [33] approach to network selection, the best-performing network (with four hidden nodes) from the validation set was selected. This network gave a classification rate of 92.9% when applied to the test data set.

4.4 Interval-based classification networks and network robustness

Having established the network performance to *crisp* (i.e., single-valued) input data, the next step was

to devise a method to investigate the sensitivity of the classification performance of the network to fluctuations in its input data. Quantification of this performance would allow an estimate of the network robustness to be evaluated. Perhaps the most obvious way of performing this analysis would be to use a Monte Carlo approach, randomizing the input data (within certain predefined bounds) and monitoring the associated changes in the output classification performance. This technique has a significant drawback, especially when applied to nonlinear MLP networks, in that it is impossible to be sure of mapping all possible combinations of variation in input space to output space unless an unfeasibly large number of sample points are used. Since interest lies in understanding the worst possible performance of the classifier in the presence of input data uncertainty, a Monte Carlo approach would therefore not provide certainty of behavior under all possible input data conditions. Similarly, the other techniques discussed in the introduction have a similar flaw in that they generate a probabilistic view of the likelihood of the classifier performance with respect to input data fluctuation. It was to circumvent this problem that the input data set was redefined as a series of interval number inputs. Interval numbers [42] occupy a bounded range of the number line, and can be defined as an ordered pair of real numbers $[a, b]$ with $a < b$ such that

$$[a, b] = \{x | a \leq x \leq b\} \tag{7}$$

Interval numbers have specific rules for the standard arithmetic operations of addition, subtraction, multiplication, and division [40].

The MATLAB™ compatible toolbox INTLAB [44] was used to implement the interval calculations required to calculate forward propagation through the MLP networks. This toolbox incorporated a rigorous approach to rounding, which was critical when using finite precision calculations in order to preserve the true conservative interval bounds. Formally, with each network, we associated an input set $I(\beta)$ composed of all possible inputs to the network, the size of the uncertainty was governed by the β parameter. Having defined the input set $I(\beta)$, the output response set $R(\beta)$ of all network outputs was computed. It was then possible to quantify the network reliability in terms of how large a β parameter could be tolerated before a point in the failure set (defined by choosing an appropriate threshold for a performance governing parameter, for example, the worst-case error) was just reached; at this point, β attained a critical value β_{CR}. A large value of β_{CR} was desirable as the network would then be more robust [41].

Each input value x_i of the test data set was intervalized by a parameter β such that

$$[x_{ia}, x_{ib}] = [(x_i - \beta), (x_i + \beta)] \tag{8}$$

Propagation of interval sets through a crisp-valued ANN weight matrix gives rise to interval number outputs. For a regression problem, this is manifest as a set of upper and lower bounds around the true output prediction. However, for a classification problem, it is necessary to introduce some new definitions for the concepts of the confusion matrix and classification rate to be valid.

4.4.1 Defining interval classification rates

For crisp-valued outputs, the designation for a correct classification is if the class with the highest class conditional probability output (the winning class) belongs to the correct target class. If this is not the case, then a misclassification is assigned. The overall classification rate is just the percentage of classifications to total number of data presentations. For an intervalized network output with sufficiently small interval bounds, it would be expected that the classification rate would be the same as in the crisp data situation. The class with the highest bound output remains the winning class (correct or incorrect), and its lower bound is greater than any of the other classes upper bounds. However, as the interval size increases, a point is reached where one (or more) of the losing class upper bounds becomes equal to or greater than the lower bound of the winning class. (The lower bound of the winning class is defined as the *threshold value* for interval-based classification.) At this critical point, it becomes impossible to distinguish between the two (or more) classes, and either (or any) of the classes could be the winner. The best-case classification rate is defined as the percentage of correct classifications, regardless of whether these classifications were unambiguous (only one class identified) or ambiguous (more than one class identified). The worst-case classification rate is

defined as the percentage of unambiguous, correct classifications. In keeping with the framework of Ben-Haim [39, 41], opportunity is defined as the best classification rate minus the crisp classification rate. It is clear that, as the interval size increases, the worst-case classification rate decreases, whilst the best-case classification rate and the opportunity increase.

4.4.2 Interval propagation through conventionally trained network

The network with four hidden nodes, which was previously chosen via the cross-validation approach, was then subjected to the propagation of intervalized data in order to investigate its robustness to data uncertainty. The behavior of this network is shown in Figure 5.

This is a typical information-gap style plot [40, 41], where the opportunity is a measure of how much performance headroom is available to the system if we are prepared to tolerate the presence of uncertainty in the input data. However, for safety-critical systems, we are more likely to be interested in the worst-case classification rate. From such a figure, it is possible to decide on a minimum acceptable classification rate (for example 80%) and then infer the corresponding interval size, which in turn relates to the spread in the input data. For the case of Figure 5, the 80% minimum worst-case classification rate corresponds to an uncertainty in the input data of $\beta_{CR} = 0.01$. Since the interval output set is conservative, we can then guarantee that, if the input set remains within the bound specified by β_{CR}, the output classification rate cannot fall below the minimum specified rate of 80%. In this fashion, the interval forward propagation routine can be used to assess the robustness of an individual network to uncertainty in the input data [45].

4.5 Interval-based network selection

Having established the basis for the definition of classification rates, worst-case error, opportunity, and best-case error applied to interval outputs from an individual network [28], the technique was extended to investigate propagation through multiple networks to appraise the capability of using interval results as a basis for network selection. The same networks trained using the conventional maximum likelihood approach detailed above were investigated. For each particular value of the number of hidden nodes (1–15), the interval forward propagation was evaluated for all 100 individual networks. The range of interval outputs for the worst-case error, opportunity,

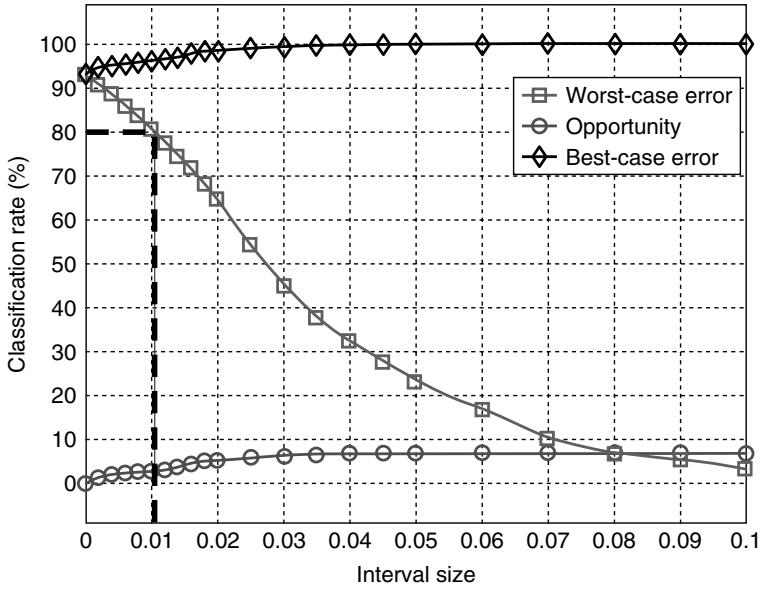


Figure 5. Classification rate as a function of interval size for four hidden node network chosen using cross-validation.

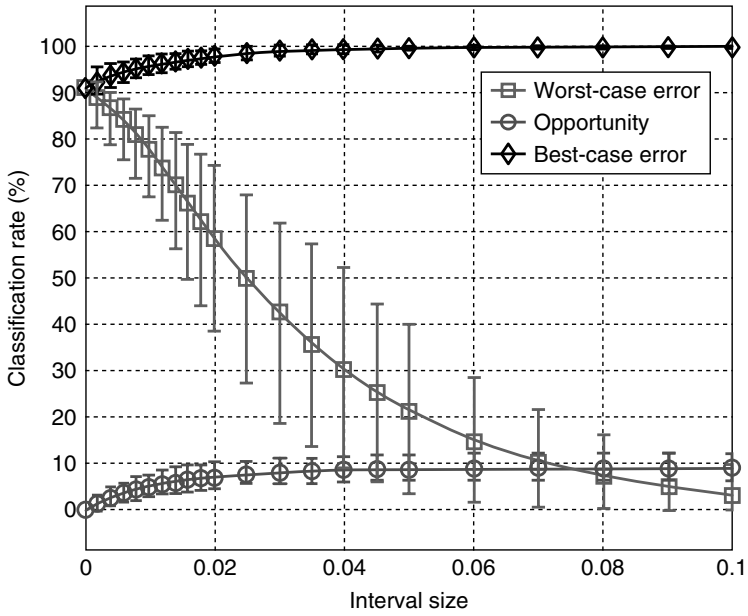


Figure 6. Interval output classification rates as function of interval size for all 100 of the 4 hidden node networks.

and best-case error were then calculated. The results for all 100 of the 4 hidden node networks are shown in Figure 6, where the markers indicate the mean values and the error bars show the range (i.e., the maximum and minimum values) of the three functions.

The most interesting feature of Figure 6 is the large spread in the values of the worst-case error function. For example, at an interval size of 0.02, the worst-case error falls between classification rates of 38.7 and 74.6%. The best-performing network using conventional cross-validation on crisp data, whilst not the worst-performing network, was certainly not the best. The best-performing network in terms of highest worst classification rate varied, depending upon the interval size but a few networks performed consistently well over the entire interval range.

In general, to select the best network in an interval tolerant sense, it is required to set a sensible upper limit on the input uncertainty (the interval size. The best network would then be the one giving the highest worst-case error at this uncertainty level. It is imperative to then check to ensure that this network also provides good performance at lower interval values. In this fashion, when considering interval input data, it is possible to determine a more appropriate choice of network selection criteria than is available using the conventional maximum likelihood training applied to crisp input data.

The merit of the interval-based forward propagation technique lies in two distinct areas. Firstly, it allows the definition of a critical size of uncertainty in the input data set that guarantees that the output classification performance does not fall below a predefined level. This is useful to apply to a single network structure (which could be obtained from any general training technique) to evaluate the robustness of that particular network to uncertainty or noise in the input data. Secondly, and more importantly, an extension of the first technique can be used as an alternative criterion for network selection. It was demonstrated that the worst-case interval output classification rates from conventional MLP networks can be used to find network structures with significantly improved classification performances over their conventional crisp-output counterparts in the presence of input uncertainty or noise. In general, it seems that prior knowledge of the maximum size of the input uncertainty is required to select the optimum network structure.

5 CASE STUDY 2: EVIDENCE-BASED DAMAGE CLASSIFICATION FOR AN AIRCRAFT STRUCTURE

The recent past has seen considerable use of machine learning techniques for SHM. The basic idea of the approach is to use data measured from undamaged and damaged structures in order to train a learning machine to assign a condition label to previously unseen data. The simplest problem of SHM is arguably *damage detection*. This is most easily carried out in the machine-learning context by using a novelty detector [46]. Novelty detection involves the construction of a model of the normal condition of a system or structure, which can then be used in a hypothesis test on unseen data to establish whether the new data corresponds to normal condition or not. The advantage of the novelty detection approach is that it can be carried out using unsupervised learning, i.e., with only samples of undamaged data. If a more detailed diagnosis of a system is required, e.g., if it is necessary to specify the type or location of damage in a structure, this can still be done using machine-learning methods. For higher levels of diagnostics, algorithms based on classification or regression are applicable; however, these must be applied in a supervised learning context and examples of data from both the undamaged and damaged conditions can be used [47].

The most popular classifiers for damage location and quantification so far have been those based on MLP neural networks [31] (although there is growing popularity for classifiers based on the concepts of statistical learning theory—like support vector machines [47]). Training of MLP networks as classifiers is usually accomplished by using the 1 *of M* strategy [31], which implicitly assumes a Bayesian probabilistic basis for the classification. Whilst probability theory is only one (but arguably the most important) of a group of theories that can quantify and propagate uncertainty, other theories of uncertainty, perhaps with the exception of fuzzy set theory, have been largely unexplored in the context of damage identification. The object of this case study is to design a classifier for damage location based on DS theory of evidence [3, 4]. The reason for exploring the possibilities of DS theory is that it *extends* probability in a number of ways, which are potentially exploitable in an SHM context. The current case study is looking only to demonstrate that the method is competitive with the probability-based MLP approach on an experimental case study of an aircraft wing. The DS classifier here is also implemented using a neural network structure [48].

5.1 Dempster–Shafer reasoning

DS theory is a means of decision fusion, which is formulated in terms of probability-like measures but extends probability theory in a number of important respects. The basic idea of *belief* was introduced by Dempster in [3] and extended in Shafer's treatise [4].

The basic model is formulated in similar terms to probability. In the place of the sample space is the *frame of discernment* Θ, which is the set spanning the possible events for observation A_i, where $i = 1, \ldots, N$. On the basis of sensor evidence, each event or union of events is assigned a degree of probability mass or *basic belief assignment* (*BBA*) m such that,

$$0 \leq m(A_i) \leq 1 \quad \forall A_i \subseteq \Theta \tag{9}$$

$$m(\phi) = 0 \tag{10}$$

$$\sum_{A_i \subseteq \Theta} m(A_i) = 1 \tag{11}$$

where ϕ is the empty set. (Note that normalization, as in equation (11), is not always required in belief function theory.) The difference between this *evidential* theory and probability theory is that the total probability mass need not be exhausted in the assignments to individual events. There is allowed to be a degree of *uncertainty* or *ignorance*. This is sometimes denoted by a probability mass assignment to the *whole* frame of discernment $m(\Theta)$ or $m(A_1 \cup \ldots A_N)$.

The *belief* in an event B is denoted by $Be(B)$ and is defined by

$$Be(B) = \sum_{A_i \subseteq B} m(A_i) \tag{12}$$

and this is the total probability, which is committed to the support of the proposition that B has occurred.

The *doubt* in the proposition B is denoted by $Dou(B)$ and is defined by

$$Dou(B) = Be(\sim B) \tag{13}$$

i.e., the doubt is the total support for the negation of the proposition B (negation is denoted by a tilde).

One of the fundamental differences between the DS and probability theory is that the belief and doubt *do not necessarily sum to unity* i.e., it is not certain that $B \cup \sim B$ is true. This is illustrated diagrammatically in Figure 7.

The uncertainty Un in the proposition B is that portion of the probability mass which does not support B *or* its negation. If further evidence were provided, some of the uncertainty could move in support of B but the mass assigned to the doubt *cannot* move. This means that the possible belief in B is bounded above by the quantity $Be(B) + Un(B) = 1 - Dou(B)$ and this quantity is termed the *plausibility* of B and is denoted by $Pl(B)$. The plausibility can also be defined by

$$Pl(B) = \sum_{A_i \cap B \neq \phi} m(A_i) \tag{14}$$

A concrete example will be useful at this point.

Example *Consider a composite structure which may have sustained damage at one of two internal sites A and B which are indistinguishable. It is known that the only possible damage mechanism at site A is delamination (denoted D), but site B may fail by delamination or fiber fracture (denoted F) and the relative probabilities of the damage mechanisms are unknown. It is further known that failure at A is twice as likely as failure at B. What can one say about the likely damage type if a fault is found?*

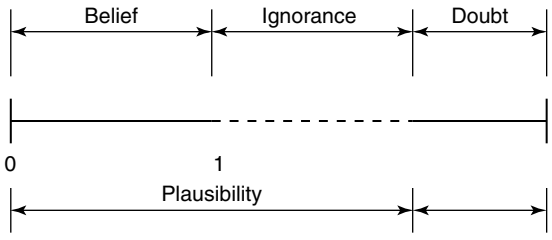


Figure 7. The Dempster–Shafer uncertainty interval.

First of all, if damage occurs at A, it is certainly by delamination and this forces the mass assignment,

$$m(D) = \frac{2}{3} \tag{15}$$

the remaining mass cannot be assigned with certainty, so it is assigned to the frame of discernment,

$$m(\Theta) = m(D \cup F) = \frac{1}{3} \tag{16}$$

The belief in the delamination is simply $Be(D) = 2/3$ *as this is the only basic mass assignment to B. There is no such assignment to F so the belief* $Be(F) = 0$. *The plausibility in D is given by*

$$Pl(D) = m(D) + m(D \cup F) = 1 \tag{17}$$

and the plausibility of F is similarly calculated as 1/3. *The uncertainty interval for D is* [2/3,1] *and that for F is* [0,1/3].

Note that it is not possible to use probability theory here directly as the relative probabilities at site B are not known. It is possible to construct bounds on the probabilities though. Suppose delamination were impossible at site B, then the overall probability of delamination would be 2/3 *and this would be a lower bound. If delamination were certain at B, the overall probability would be* 1. *Note that these quantities are the belief and plausibility respectively. For this reason, the belief and plausibility are sometimes termed the lower and upper probabilities.*

The interpretation of some common instances of the uncertainty interval for a proposition B is as follows:

[0,0] B is impossible.
[1,1] B is certain.
[0.75,0.75] There is no uncertainty. B has a true probability of 0.75.
[0,1] There is total ignorance regarding B.
[0.25,1] B is plausible, there is no support for $\sim B$.
[0,0.75] $\sim B$ is plausible, there is no support for B.
[0.25,0.75] Both B and $\sim B$ are plausible.

All this suffices to establish terminology, to explain how to compute belief functions, and how to interpret the results. It does not provide a means of data

fusion—that requires the use of *Dempster's combination rule* [49].

Suppose that one has two sensors 1 and 2. BPAs are possible on the basis of either sensor alone, denoted m_1 and m_2. Belief functions Be_1 and Be_2 can be computed. Dempster's rule allows the calculation of an overall probability assignment m_+ and a corresponding overall belief function Be_+, where this *direct sum*,

$$Be_+ = Be_1 \oplus Be_2 \tag{18}$$

is induced by m_+. Suppose that sensor 1 makes assignments $m_1(A_i)$ to the proposition A_i (which can, and usually does, include the frame of discernment), and sensor 2 makes assignments $m_2(B_j)$, then Dempster's rule makes assignments as follows.

Consider a matrix with row entries labeled by i and column entries j, then the (i, j)th element of the matrix is an assignment of probability mass $m_1(A_i) \times m_2(B_j)$ to the proposition $A_i \cup B_j$.

This is best understood by an example:

Example *Suppose a classifier is required which can assign a damage type to data from a composite structure. The possible damage types are delamination D, fiber fracture F, matrix cracking M, or fiber pullout P. Two different classifiers are trained, A and B, which produce different probability mass assignments. Classifier A makes the assignments,*

$$m_A(D) = 0.25, \quad m_A(F \cup M) = 0.5, \quad m_A(\Theta) = 0.25 \tag{19}$$

and classifier B returns,

$$m_B(D \cup M) = 0.3, \quad m_B(D \cup F) = 0.4, \quad m_B(\Theta) = 0.3 \tag{20}$$

Dempsters rule induces a mass assignment matrix,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$
$m_B(D \cup F)$	$0.5 \times 0.3 = 0.15$	$0.5 \times 0.4 = 0.2$	$0.5 \times 0.3 = 0.15$
$m_B(\Theta)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$

to the propositions,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	D	M	$D \cup M$
$m_B(D \cup F)$	D	F	$D \cup F$
$m_B(\Theta)$	D	$F \cup M$	Θ

and the direct sum m_+ assigns support to the propositions D, M, F, $D \cup M$, $D \cup F$, $F \cup M$, and $\Theta = D \cup M \cup F \cup P$.

The belief in delamination is $Be_+(D) = 0.075 + 0.15 + 0.075 = 0.3$; similarly, $Be_+(M) = 0.1$, $Be_+(F) = 0.2$, and $Be_+(P) = 0$. The plausibility of delamination D is given by

$$Pl_+(D) = m_+(D) + m_+(D \cup M) + m_+(D \cup F) = 0.3 + 0.075 + 0.15 = 0.525 \tag{21}$$

similarly,

$$Pl_+(M) = 0.275, \; Pl_+(F) = 0.45 \; \text{and} \; Pl_+(P) = 0.075 \tag{22}$$

In summary, the uncertainty intervals for the fused belief function are

D	[0.3,0.525]
M	[0.1,0.275]
F	[0.2,0.45]
P	[0,0.075]

The most plausible diagnosis is clearly delamination.

In mathematical terms, Dempster's combination rule is expressed as

$$m_+(C) = \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j) \tag{23}$$

and

$$Be_+(C) = \sum_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j) \qquad (24)$$

Unfortunately, things are not quite as straightforward as this. Problems arise in using Dempster's rule if the intersection between supported propositions A_i and B_j is empty. In this circumstance, a nonzero mass assignment is made to the empty set ϕ and this contradicts the basic definition of the mass assignment, which demands that $m_+(\phi) = 0$. In order to preserve this rule, Dempster's rule *must* assign zero mass to nonoverlapping propositions. However, if this is the case, the probability mass is lost and the total mass assignment for m_+ is less than unity, contradicting another rule for probability numbers. A valid mass assignment is obtained by *rescaling* m_+ to take account of the lost mass. If the mass lost on nonoverlapping propositions totals k, the remaining mass assignments should be rescaled by a factor $K = 1/(1-k)$. The combination rule (23) is modified to

$$m_+(C) = K \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j) \qquad (25)$$

Example *Consider the last example. Suppose the assignments made by sensor A were as before, but those of sensor B were now,*

$$m_B(D \cup M) = 0.3, \quad m_B(P \cup F) = 0.4,$$
$$m_B(\Theta) = 0.3 \qquad (26)$$

Dempster's rule gives the same assignments,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$
$m_B(P \cup F)$	$0.5 \times 0.3 = 0.15$	$0.5 \times 0.4 = 0.2$	$0.5 \times 0.3 = 0.15$
$m_B(\Theta)$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.4 = 0.1$	$0.25 \times 0.3 = 0.075$

but this time to the propositions,

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	D	M	$D \cup M$
$m_B(P \cup F)$	ϕ	F	$P \cup F$
$m_B(\Theta)$	D	$F \cup M$	Θ

and a total mass of 0.15 *is lost on the empty set. This means that the assignments should be rescaled by a factor* $K = 1/0.85 = 1.1765$ *(to four decimal places). The mass matrix becomes,*

	$m_A(D)$	$m_A(F \cup M)$	$m_A(\Theta)$
$m_B(D \cup M)$	0.0882	0.1176	0.0882
$m_B(P \cup F)$	0.1765	0.2353	0.1764
$m_B(\Theta)$	0.0882	0.1176	0.0882

and the calculation for the belief functions and uncertainty intervals proceeds exactly as before.

The differences between the DS approach and the probabilistic are now manifest. First of all, probabilistic—or rather Bayesian—approaches are unable to accommodate ignorance. All probability must be assigned to the set of propositions under consideration. Secondly, the Bayesian approach is unable to meaningfully assign probabilities to the union of propositions. If the uncertainty for all propositions is zero and the mass assigned to unions is zero, DS is reduced to Bayesian probability reasoning.

There are other frameworks that seek to extend Bayesian methods in a similar manner to DS such as the generalized evidence processing (GEP) approach of [5] and those proposed in [50, 51].

5.2 The DS neural network

The object of this section is to briefly describe the neural network implementation of the DS-based

classifier. More detail can be found in the original reference [46].

The basic idea is to assign one of M classes $C_1, \ldots, C_M$ (these form the frame of discernment) to a feature vector $\underline{x}$ on the basis of a set of N training examples $\underline{x}_1, \ldots, \underline{x}_N$. Suppose the vector $\underline{x}$ is close to a training example $\underline{x}_i$ with respect to an appropriate distance measure d ($d_i = \|\underline{x} - \underline{x}_i\|$). It is then appropriate that the class of the vector $\underline{x}_i$ influences ones beliefs about the class of $\underline{x}$. One has evidence about the class of $\underline{x}_i$. The approach to the classification taken in [48] is to allocate belief to the event C_q (the class of $\underline{x}_i$, according to the distance d_i).

$$m^i(C_q) = \alpha\phi_q(d_i) \tag{27}$$

where $0 < \alpha < 1$ is a constant and ϕ_q is an appropriate monotonically decreasing function. Each training vector close to $\underline{x}$ contributes some degree of belief. For each training vector, a degree of belief is also assigned to the whole frame of discernment Θ as follows:

$$m^i(\Theta) = 1 - \alpha\phi_q(d_i) \tag{28}$$

The function ϕ used here is the basic Gaussian,

$$\phi_q(d_i) = \exp(-\gamma_q(d_i)^2) \tag{29}$$

where γ_q is a positive constant associated with class q. To simplify matters, one confines the construction of the belief assignment for the vector $\underline{x}$ to a sum of the beliefs induced by its nearest neighbors. The sum is computed using Dempster's combination rule as described in the previous subsection. Actually, a further simplification is made to speed up the processing. Rather than summing over the nearest neighbors from the whole training set in order to assign the belief, one sums over a set of *prototypes* constructed from the training set by a clustering algorithm. Each prototype $\underline{p}_i$ is assigned a degree of membership to the class q denoted by u_q^i with the constraint $\sum_{q=1}^{M} u_q^i = 1$. These are used to compute the belief in the class q for $\underline{x}$ given the distances d_i from the prototypes

Although it is a gross simplification, the algorithm can be summed up as follows:

1. Construct the prototypes $\underline{p}_i$ from the training data using a clustering algorithm.
2. Given a vector $\underline{x}$, compute the distances from the vector to the prototypes. Using the parameters d_i and u_q^i, assign a degree of belief for each class q.
3. Use Dempster's combination rule to compute the total belief in each class from all the contributing prototypes.

The algorithm extends the probabilistic classifier by also making an assignment to the frame of discernment and this quantifies the degree of uncertainty of the classification. Reference [48] explains how the algorithm can be implemented in terms of a four-layer neural network. Unlike an MLP, the network is not a simple feed-forward structure.

In order to assign a class to the vector $\underline{x}$, one selects that with the largest overall belief assignment induced by the training data.

5.3 Damage location example

The same problem that was investigated in the first case study, namely damage location on a Gnat aircraft wing (Figure 2), is revisited in order to compare the evidence-based approach to damage classification with the probabilistic approach. The only significant difference between the data acquisition and signal processing methodology outlined in the previous case study and that used in the current study related to the number of features used. In the previous study, the candidate features were reduced to form a feature set comprising of nine features. In the current study, the candidate features were reduced to form a feature set of four individual features. The best four features were selected from the initial set using a GA to optimize the classification error of an MLP classifier [52]. The reduction to four features was made to ensure that the MLP network used for comparison with the DS network later was unlikely to suffer from overtraining.

By performing the novelty analysis for each of the four features, a data matrix of size $[1800 \times 4]$ was obtained. As before, this was divided into three equal parts to form separate *training, validation*, and *test* data sets for subsequent network evaluation. Finally, the data sets were logarithmically compressed and normalized to -1 to $+1$ before presentation

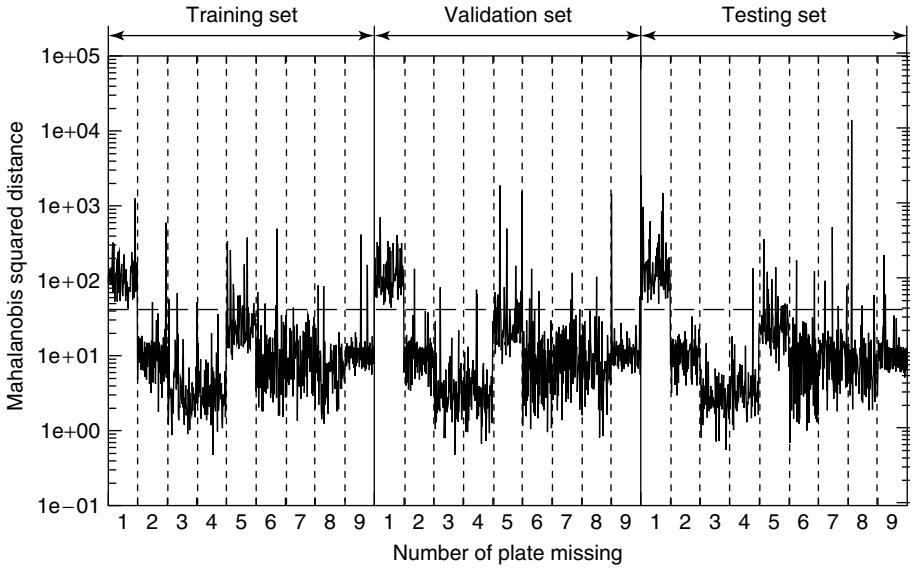


Figure 8. Outlier statistic for all damage states for the novelty detector trained to recognize panel 1 removal.

to the network. Figure 8 shows the results of the outlier analysis for a feature that was clearly able to recognize the removal of inspection panel 1.

The plot in Figure 8 shows the discordancy (novelty index) values returned by the novelty detector over the whole set of damage states. The horizontal dashed lines in the figures are the thresholds for 99% confidence in identifying an outlier; they are calculated according to the Monte Carlo scheme described in [53]. The novelty detector substantially fires only for the removal of panel for which it has been trained. The outputs from the four novelty detectors are then used to form the data with which the damage location classifiers are trained.

5.4 DS network results

The final stage of the analysis was to produce a classifier based on the DS neural network algorithm, which could serve as a damage location system. As with a standard MLP network, the specification of the DS network structure requires hyperparameters; in this case, the number of prototypes (analogous to the number of hidden units in the first layer of the network) and the starting values of the weights before training. These were computed by a cross-validation procedure as for the MLP [54]. Many neural networks were trained with the same training data but with differing numbers of prototypes and initial weights. Up to 30 prototypes were considered, and in each case 10 randomly chosen initial conditions were used. The best network was selected by observing which of them produced the minimum misclassification error on the validation set. The final judgment of the network capability was made by using the independent testing set.

The results for the presentation to the classifier are summarized in the confusion matrix given in Table 1. The best DS network used 29 prototypes. The probability of correct classification was 89.7%. There were four events associated with the frame of discernment, corresponding to a probability mass of 0.007. This means that, allowing for the fact that the network indicates when it has insufficient evidence to make a classification, the classification error is 9.6%. The main source of confusion is in locating damage to the two smallest panels, 3 and 6, which was anticipated.

In order to make a comparison with the standard approach, the algorithm chosen was a standard MLP neural network. The neural network was presented with the same four novelty indices at the input layer and required to predict the damage class at the output layer. The procedure for training the neural

Table 1. Confusion matrix for the best DS network using four log features–testing set

Prediction	1	2	3	4	5	6	7	8	9	Θ
True class 1	54	5	5	0	0	0	2	0	0	0
True class 2	0	63	0	0	2	0	0	0	0	1
True class 3	6	1	56	2	0	0	0	0	0	1
True class 4	5	0	1	55	0	3	0	2	0	0
True class 5	0	0	0	0	65	0	0	1	0	0
True class 6	2	2	2	4	0	54	1	0	0	1
True class 7	0	1	1	0	0	0	61	2	1	0
True class 8	0	0	1	0	1	0	0	62	1	1
True class 9	0	0	0	0	0	0	0	3	63	0

network again followed the guidelines in [54]. The training set was used to establish weights, whilst the network structure and training time etc. were optimized using the validation set. The testing set was then presented to this optimized network to arrive at a final classification error. For the network structure, the input layer necessarily had four neurons, one for each novelty index, and the output layer had nine nodes, one for each class. The training phase used the same 1 of M strategy as in the previous case study.

The best neural network had 19 hidden units and resulted in a testing classification error of 0.108 i.e., 89.2% of the patterns were classified correctly. This means that the misclassification probability is 10.8%. The confusion matrix is given in Table 2. Again, the main errors were associated with the two small panels P3 and P6.

The main conclusion here is that the DS approach to classification implemented as a neural network gives comparable results to the standard analysis using an MLP neural network. The data used are from a full-scale experimental test on an aircraft wing and is therefore a stringent test of algorithms from the point of view of SHM. In fact, the DS network shows a slight improvement, giving a classification probability of 89.7% compared to 89.2% for the MLP. One possible advantage of the DS approach is the fact that it can assign patterns to the frame of discernment and thus indicate to the analyst that there is insufficient evidence to make a classification. Another advantage of using the DS approach might be to provide methodological diversity in order to use data fusion. In this case, the errors made by the DS network and the MLP network are slightly different and therefore there is some advantage to fusion—this is considered in [55].

6 CONCLUSIONS

The purpose of this article was to introduce the many frameworks for the representation of uncertainty and the three uncertainty-related issues of interest in the field of structural dynamics. In order to put these ideas into context, and to showcase the merits of nonprobabilistic treatments of uncertainty, within the area of SHM, two SHM-related case studies were discussed.

Table 2. Confusion matrix for best MLP neural network using four log features–testing set

Prediction	1	2	3	4	5	6	7	8	9
True class 1	61	1	0	0	0	0	1	0	0
True class 2	0	63	0	0	3	0	0	0	0
True class 3	1	0	48	8	0	5	7	2	0
True class 4	0	1	3	56	0	2	0	4	0
True class 5	0	0	0	0	66	0	0	0	0
True class 6	4	1	0	9	0	52	0	0	0
True class 7	1	0	0	0	0	0	59	5	1
True class 8	1	0	0	0	1	0	1	59	4
True class 9	0	0	0	0	0	0	0	5	61

The first case study illustrated how, by taking an alternative approach to the convention technique of neural network selection via cross-validation, the robustness of a network to data uncertainty could be drastically improved. This finding should have implications related to the issue of certification of neural network classifiers and regressors for the purpose of monitoring safety-critical structures. The second case study illustrated that the use of an evidence-based classifier, with its ability to admit ignorance, could reduce the likelihood of misclassifications within an SHM system. This will clearly have both economic and safety-related benefits.

REFERENCES

[1] Hemez FM. *Uncertainty, Validation of Computer Models and the Myth of Numerical Predictability*, LA-UR-01-2492. Los Alamos National Laboratory Report, 2001.

[2] Rasmussen NC, *et al. Reactor Safety Study: An Assessment Accident Risks in U.S. Commercial Nuclear Power Plants*. US Nuclear Regulatory Commission, NUREG-75/014, WASH-1400, 1975.

[3] Dempster AP. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics* 1967 **38**:325–339.

[4] Shafer G. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[5] Thomopoulos SCA. Theories in distributed decision fusion: comparison and generalisation. *Sensor Fusion III: 3D Perception and Recognition. Proceedings of SPIE*.1383 1990.

[6] Oberkampf WL, Helton JC, Sentz K. Mathematical representation of uncertainty. *AIAA Non-Deterministic Approaches Forum*, AIAA 2001-1645. Seattle, WA, 2001.

[7] Osegueda RA, Lopez H, Pereyra L, Ferregut CM. Localisation of damage using fusion of modal strain energy differences. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 695–701.

[8] Pereyra L, Osegueda RA, Carrasco C, Ferregut CM. Detection of damage in a stiffened plate from fusion of modal strain energy differences. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 1556–1562.

[9] Dubois D, Prade H. *Possibility Theory: An Approach to Computerised Processing of Uncertainty*. Plenum Press: New York, 1986.

[10] Klir GJ. On fuzzy-set interpretation of possibility theory. *Fuzzy Sets and Systems* 1999 **108**:263–273.

[11] Zadeh LA. Fuzzy sets. *Information and Control* 1965 **8**:338–353.

[12] Lallemand B, Plessis G, Tison T, Level P. Modal behaviour of structures defined by imprecise geometric parameters. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 1422–1428.

[13] Plessis G, Lallemand B, Tison T, Level P. A fuzzy method for the modal identification of uncertain experimental data. *Proceedings of 18th International Modal Analysis Conference*, IMAC XVIII. San Antonio, TX, 2000; pp. 1831–1837.

[14] Plessis G, Lallemand B, Tison T, Level P. Fuzzy modal parameters. *Journal of Sound and Vibration* 2000 **233**:797–812.

[15] Lallemand B, Cherki A, Tison T, Level P. Fuzzy modal finite element analysis of structures with imprecise material properties. *Journal of Sound and Vibration* 1999 **220**:353–364.

[16] Moore RE. *Methods and Applications of Interval Analysis*. SIAM: Philadelphia, PA, 1979.

[17] Muhanna RL, Mullen RL. Formulation of fuzzy finite element methods for mechanics problems. *Computer-Aided Civil and Infrastructure Engineering* 1999 **14**:107–117.

[18] Worden K, Osegueda R, Ferregut C, Nazarian S, George DL, George MJ, Kreinovich V, Kosheleva O, Cabrera S. Interval methods in non-destructive testing of material structures. *Reliable Computing* 2001 **7**:341–352.

[19] Comba JLD, Stolfi J. Affine arithmetic and its applications to computer graphics. *Proceedings of SIBGRAPH '93*, Pernambuco, 1993; pp. 9–18.

[20] Manson G. Sharper eigenproblem estimates for uncertain multi degree of freedom systems. *Proceedings of 21st International Modal Analysis Conference*. Orlando, FL, 2003.

[21] Ben-Haim Y, Elishakoff I. *Convex Models of Uncertainty in Applied Mechanics*. Elsevier Science, 1990.

[22] Ben-Haim Y. Set-models of information-gap uncertainty: axioms and an inference scheme. *Journal of the Franklyn Institute* 1999 **336**:1093–1117.

[23] Hemez FM, Ben-Haim Y, Cogan S. *Information-Gap Robustness for the Test-Analysis Correlation of*

a Non-Linear Transient Simulator. LA-UR-02-3538. Los Alamos National Laboratory Report, 2002.

[24] Klir GJ, Smith RM. On measuring uncertainty and uncertainty-based information. *Annals of Mathematics and Artificial Intelligence* 2001 **32**(1–4):5–33.

[25] Choquet G. Theory of capacities. *Annales de L'Institut Fourier* 1953-54 **5**:132–295.

[26] Reese CS, Wilson AG, Hamada MS, Martz HF, Ryan KJ. *Integrated Analysis of Computer and Physical Experiments*. LA-UR-00-2915. Los Alamos National Laboratory Report, 2000.

[27] Bement TR, Booker JM, Sellers KF, Singpurwalla ND. *Membership Functions and Probability Measures of Fuzzy Sets*. LA-UR-00-3660. Los Alamos National Laboratory Report, 2000.

[28] Bergman LA. Uncertainty modelling in dynamical systems: a perspective. In *Structural Dynamics @ 2000: Current Status and Future Directions*, Ewins DJ and Inman DJ (eds). Research Studies Press: Baldock, 2001, pp. 9–16.

[29] Manson G, Worden K, Allman DJ. Experimental validation of a structural health monitoring methodology. Part II. Novelty detection on a Gnat aircraft. *Journal of Sound and Vibration* 2003 **259**(2):345–363.

[30] Manson G, Worden K, Allman DJ. Experimental validation of a structural health monitoring methodology. Part III. Damage location on an aircraft wing. *Journal of Sound and Vibration* 2003 **259**(2):365–385.

[31] Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[32] Nabney IT. *Netlab-Algorithms for Pattern Recognition*. Springer-Verlag, 2002.

[33] Haykin S. *Neural Networks, a Comprehensive Foundation*, *Second Edition*, Prentice Hall, 1999.

[34] MacKay DJC. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[35] MacKay DJC. The evidence framework applied to classification networks. *Neural Computation* 1992 **4**(5):720–736.

[36] Papadopoulos G, Edwards PJ. Confidence estimation methods for neural networks: a practical comparison. *IEEE Transactions on Neural Networks* 2001 **12**(6):1278–1287.

[37] Lowe D, Zapart C. Point-wise confidence interval estimation by neural networks: a comparative study based on automotive engine calibration. *Neural Computing and Applications* 1999 **8**:77–85.

[38] Ben-Haim Y, Elishakoff I. *Convex Models of Uncertainty in Applied Mechanics*. Elsevier, 1990.

[39] Ben-Haim Y. *Robust Reliability in the Mechanical Sciences*. Springer-Verlag, 1996.

[40] Hemez FM, Ben-Haim Y, Cogan S. Information gap robustness for the test-analysis correlation of a nonlinear transient simulation. *Proceedings of the 9th AIAA/ISSMO Symposium on Multi-disciplinary Analysis and Optimisation*. Atlanta, GA, 2002.

[41] Ben-Haim Y. *Information-Gap Decision Theory*. Academic Press, 2001.

[42] Moore RM. *Interval Analysis*. Prentice Hall, 1966.

[43] Worden K, Tomlinson GR. *Nonlinearity in Structural Dynamics*. Institute of Physics Publishing, 2001.

[44] Rump SM. INTLAB – INTerval LABoratory. In *Developments in Reliable Computing*, Csendes T (ed). Kluwer Academic Publishers, 1999, pp. 77–104.

[45] Pierce SG, Worden K, Manson G. Information-Gap analysis of a neural network damage locator, *IMechE meeting on Pattern Recognition-Detection, Classification and Monitoring*. University of Liverpool: Liverpool, 2004.

[46] Hayton P, Utete S, King D, King S, Anuzis P, Tarassenko L. Static and dynamic novelty detection methods for jet engine health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 2007 **365**:493–514.

[47] Worden K, Manson G. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 2007 **365**:515–537.

[48] Denoeux T. A neural network classifier based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* 2000 **30**:131–150.

[49] Shafer G, Logan R. Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence* 1987 **33**:271–298.

[50] Barnett JA. Computational methods for a mathematical theory of evidence. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, BC, 1981; pp. 868–875.

[51] Yen J. A reasoning model based on an extended Dempster-Shafer theory. *Proceedings of the 5th*

AAAI-86 National Conference on Artificial Intelligence, Philadelphia, 1986; pp. 125–131.

[52] Worden K, Manson G, Hilson G, Pierce SG. Genetic optimisation of a neural damage locator. *Journal of Sound and Vibration* 2008 **309**(3–5):529–544.

[53] Worden K, Manson G, Fieller NR. Damage detection using outlier analysis. *Journal of Sound and Vibration* 2000 **229**:647–667.

[54] Tarassenko L. *A Guide to Neural Computing Applications*. Arnold, 1998.

[55] Worden K, Manson G, Denoeux T. *An Evidence-Based Approach to Damage Localisation on an Aircraft Structure*. Accepted for publication in Mechanical Systems and Signal Processing 2008.

FURTHER READING

Goldberg DE. *Genetic Algorithms in Search, Optimisation and Machine Learning*. Addison-Wesley, 1989.

Chapter 40

Piezoceramic Materials—Phenomena and Modeling

Jayabal Kaliappan and Srinivasan M. Sivakumar

Department of Applied Mechanics, Indian Institute of Technology, Chennai, India

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

Transducers such as piezoceramics play the core role in structural health monitoring (SHM) applications where a sensing cum actuating element that can be embedded to the structure could help. Piezoceramics convert mechanical energy into electrical energy, called the *direct effect*, and electrical to mechanical energy, called the *converse effect* [1]. Their superiority in being tunable to very high rates of loading and being shapeable to any arbitrary shape dominates the transducer market today, which is billing over $10 billion worldwide. The awareness in the requirement of monitoring a structure for its health for both safety and longevity assessment makes these a much sought after material today.

This article deals with piezoceramics focusing on the behavior characteristics, the phenomena realized by these characteristics, the various available forms of the material, and modeling such phenomena. The effects of different loading conditions and varieties of models to describe the reversible and irreversible effects are also presented. Important parameters pertaining to piezoceramics that play a role in design and how they are computed for various shapes of piezoceramics for various applications are explained.

Finally, the article closes with throwing some light on a few critical issues that mar piezoceramics and the research and development efforts that are going on in addressing these issues.

2 BEHAVIOR CHARACTERISTICS—PHENOMENA

In piezoceramics, the piezoelectric property is realized because of the crystal structure the material exhibits. In general, the ABX_3–perovskite structure is observed, where A and B represents cations alkaline earth (Ca, Ba, Sr, etc.) and transition metals (Fe, Ti, Ni, etc.), respectively, and X is oxide halide ion: for example, $BaTiO_3$–Ti^{4+} at the center, Ba^{2+} at the corners and O^{2-} at the center of faces (Figure 1).

The centrosymmetric structure with the transition metal ion present at the center balances the polarization effects that may occur otherwise. For example, at higher temperatures, it exhibits, for example, a cubic structure that is of a high symmetry (centrosymmetry) and exhibits no polarization since the positive metal ion (charge center) coincides with the center of the structure. Such a zero polarization phase is called a *paraelectric phase*. In general, these materials exhibit paraelectric state above a certain temperature called the *Curie temperature*, T_C. Below Curie temperature, the cubic structure becomes unstable giving way to another crystal structure (for e.g., tetragonal structure) of lower symmetry, which is more stable at such temperatures. These structures are typically noncentrosymmetric. The more stable state (maximum entropic state) is achieved in these structures by slight movement away from the center of positive and negative charges leading to remnant polarization. For example, in the barium titanate crystals, at low temperatures, a tetragonal structure becomes more stable and at its natural state, the O^{2-} ions assume a slightly downward position while the Ti^{4+} assumes an upward position. The separation that occurs in the two charge centers (for example, O^{2-} ion charge center and Ti^{4+}, Ba^{2+} ions charge center are separated in the tetragonal state as shown in Figure 1) leads to a spontaneous polarization in the crystal and, therefore, the unit cell. This also induces an associated spontaneous strain. An extensive review of the underlying materials science of piezoceramic materials is available in [2, 3].

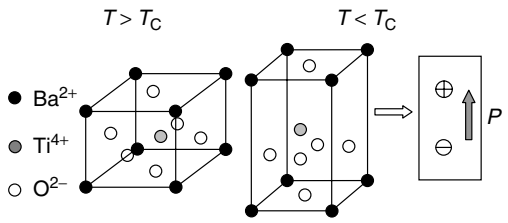


Figure 1. Perovskite structure unit cell—paraelectric and ferroelectric phases.

In the example shown in Figure 1, the lattice length of each side of the cubic structure above the Curie temperature, T_C, is given by a_o, while the lattice lengths of the tetragonal structure that is stable below T_C, are given by a and c. Assuming that the total strain is in the small strain range, the strain that occurs due to spontaneous transformation from the cubic to the tetragonal stage can be written as spontaneous strain $= (c - a_o)/a_o$.

Such a spontaneous strain is realizable only if all the unit cells are aligned along a single direction. In a polycrystal material (described in a later part in this article), the cumulative sum of the spontaneous strains achievable by means of poling (the process of poling is explained in a later section) is less than the spontaneous strain. Such a strain in the polycrystal after poling is called the *remnant strain*. The most popular commercially available piezoelectric ceramic is the lead Zirconate Titanate (PZT) i.e., $Pb(Zr,Ti)O_3$ in which the composition with Zr and Ti components vary in percentage [2]. Depending on the Zr/Ti ratio, the crystal structure of the ceramic, the Curie temperature, and the piezoelectric coupling coefficient, i.e., maximum strain to electrical voltage ratio, vary.

Apart from the PZT ceramics, there are other compositions that show piezoelectric effect and are used in specific applications or for understanding the characteristics of the piezo/ferroelectric phenomena. A lanthanum doping is found to improve the piezo coupling in PZT and such materials are referred to as *PLZT* (lead lanthanum zirconate titanate). One of the main problems with these materials is the lead content, which may prevent its application in biocompatibility applications. The stability of the composition in terms of lead erosion in these materials is important in such applications. One such composition is $BaTiO_3$ (barium titanate).

3 PHENOMENA REALIZED BY THE ABOVE CHARACTERISTICS OF THE FERROELECTRIC MATERIAL

The above description of characteristics provides an understanding of the various ways in which the material could be stimulated for a particular response. In the following, specific effects that come into play at various stimuli conditions and states of material are described. The significant ones are the electrostriction, direct and converse piezoelectric effect, domain switching, etc. While the electrostriction effect is prominent in the paraelectric phase, the other effects such as direct and converse piezoelectricity and domain switching take place in the ferroelectric phase.

3.1 Direct piezoelectric effect

One of the important effects that has significant role in the applications is the direct piezoelectric effect. The first demonstration of the direct piezoelectric effect was done by the brothers Pierre Curie and Jacques Curie. They demonstrated the effect using crystals of tourmaline, quartz, and Rochelle salt. Later, artificial piezoelectric materials were developed to improve the piezoelectric constants many folds and they are called *ferroelectric materials*.

Piezoelectric effect occurs primarily in the ferroelectric phase in which the remnant polarization is present in the lattice structure as shown in Figure 1. The tensile or compressive stress brings about a shift in the charge center (either in terms of further shift for the tensile stress case or closer shift for the compressive stress case). This shift, in turn, alters the polarization in the material (Figure 2). Macroscopically, a charge accumulation occurs at the surface of the material as in the dielectric material, resulting in an electric field. The phenomenon is called the *direct piezoelectric effect*. When the surfaces of positive and negative charges are connected to a potentiometer, a transient potential change is observed. This change per unit separation of the surfaces is a measure of the average electric field produced due to the applied stress. The higher the stress, the higher is the potential change observed. Upon removal of the stress, the charge centers are brought back to their original equilibrium positions. Thus, this effect is reversible. However, in a later section, we have dealt with a critical situation of high stress during which this reversibility could be lost resulting in domain switching. This effect is best understood as a linear and reversible response, i.e.,

$$P = h\,\varepsilon \quad \text{(or)} \quad P = d\,\sigma \tag{1}$$

where h and d are piezoelectric constants. Invoking the relation between stress, σ, and strain, ε, a relationship between the constants h and d can be obtained. This direct piezoelectric effect is used in force/pressure sensor applications owing to noticeable and measurable change in polarization.

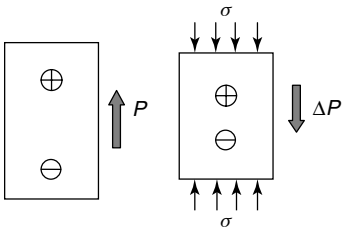


Figure 2. Direct piezoelectric effect.

3.2 Converse piezoelectric effect

As noted earlier, the converse of the above effect is also observed in these materials. When an electric field is applied on a ferroelectric material in its poled ferroelectric phase, a strain is produced because of increased or decreased charge separation that occurs owing to the field (Figure 3). This effect is often referred to as *converse piezoelectric effect*. Depending on the direction of remnant polarization in the ferroelectric phase due to poling, the electric field applied may produce positive or negative strain. It should be noted here that this is a reversible effect meaning that once the field is removed, the strain produced goes to zero. There is, however, a

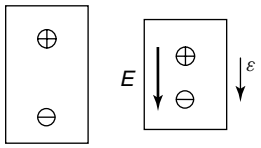


Figure 3. Converse piezoelectric effect.

critical value of strain or stress beyond which the reversibility may be lost. Such a situation is discussed separately as domain switching phenomenon in a forthcoming section. This reversible effect, for a first approximation, can be mathematically expressed as

$$\varepsilon = dE \tag{2}$$

where d is the piezoelectric constant. The linear relationship between the stain, ε, and electric field, E, is found sufficient in most of the situations to describe the converse piezoelectric effect since the strains involved are in the small strain regime. For a detailed treatment of classical linear piezoelectricity, the reader is referred to [4–7].

This effect occurs in the ferroelectric phase, i.e., when there is a spontaneous polarization in the unit cell (due to an existing separation in the charge centers). This charge separation is important in this effect since a nearly linear and reversible effect is realized. Depending on the direction of the forcing electric field, the strain is either positive or negative. However, this effect is quite different from another effect commonly known as the *electrostrictive effect* encountered in these materials in their paraelectric phase. In the electrostrictive effect, which is described in the next section, strains are always positive, since there is no directional effect found in that phase unlike the directional effect produced by the spontaneous polarization in the ferroelectric phase. Converse effect is used in many actuator applications.

3.3 Electrostriction

Electrostriction is noticed prominently in the paraelectric phase where charge centers coincide even without external field. The deformation induced is due to the separation of the positive and negative charge centers from each other that happens owing to an external electric field under equilibrium conditions. It is interesting to note that the reverse phenomenon does not occur here, i.e., upon application of a deformation, no electric displacement is produced.

When the electric field, E, is applied in one direction, the positive and the negative charge centers separate and this separation causes strain in the material to maintain equilibrium. When the electric field is removed, the configuration reverts back to the original one where the charge centers coincide. Thus, this effect is reversible even for fairly large electric fields. Since the reversal of electric field also produces charge separation from zero separation, the resulting configuration due to the applied electric field is always of one sign as shown in Figure 4. Thus, deformation is the same in both cases for a given magnitude of electric field, E. This effect can be expressed by

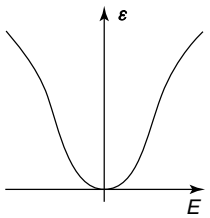


Figure 4. Electrostrictive effect.

$$\varepsilon = qE^2 \tag{3}$$

where q is the *electrostrictive constant.* This can be directly derived from the Coulomb's law of electrostatics.

Though the electrostrictive effect is observed in most of the materials, they are too weak for technical applications. In some applications, electrostrictive transducers are preferred over piezoelectric transducers owing to their minimal hysteresis and aging effects. However, their temperature-dependent and nonlinear saturation effects are to be taken into account while designing control systems.

3.4 Domain switching

The phenomenon of domain switching described here plays a major role in the nonlinear and hysteretic behavior of the material. When the temperature falls below the Curie temperature, the material transforms from cubic crystallographic structure to another structure such as tetragonal, orthorhombic, and rhombohedral, depending on the temperature and the composition of the material. While the cubic state is stable above the Curie temperature, the other less-symmetric state is stable below the Curie temperature. During this transition, the material transforms from a paraelectric phase to a ferroelectric phase as described in the introductory section. In general,

a single cubic unit cell can transform into any of multiple configurations of the other less-symmetric state that is energetically motivated [8]. One such example involving cubic to tetragonal transformation is discussed below. In the absence of any forcing field, electrical and stress, during the transition, every one of the configurations of the less-symmetric state formed is a preferred configuration and, therefore, this leads to the formation of equal proportions of each of these configurations leading to a net zero polarization although each of the configurations or the variants exhibit a spontaneous polarization. Such a state of the material is usually referred to as a *thermally depoled ferroelectric state*. The group of unit cells with the same polarization direction is referred to as a *domain*.

We use the transition from a cubic to a tetragonal state as an example to explain this. When the cubic structure transforms to tetragonal structure, it can assume any of the six orientations as shown in Figure 5. Since there is no preference to any one of the orientations under the purely thermal cooling, equal proportions of the six orientations are expected. Generally, as noted earlier, groups of unit cells of the same polarization (domains) form separated by domain walls. Domain walls separating two orientations of opposite polarizations and domain walls separating domains with polarization 90° apart are shown in Figure 6. Application of a forcing field large enough will induce a domain to switch to another. This can also be thought of as movement of the domain wall in the crystal. Such a phenomenon is called *domain switching*. This domain switching brings about permanent irreversible change in the net polarization and net strain of a group of domains

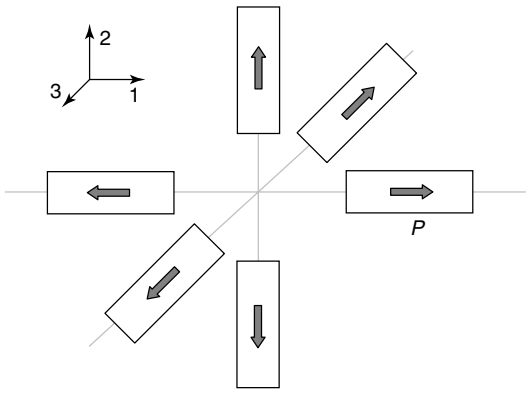


Figure 5. Domain switching for a tetragonal lattice.

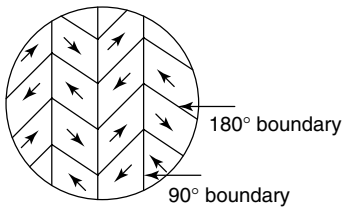


Figure 6. Domains and their boundaries.

in question. During this process of switching, some amount of energy is dissipated in the movement of the domain walls and related mechanisms.

When the domain switching is induced by the electric field, a domain may switch to any of the other five domains such that the polarization direction of the switched domain is the most favorable to the external electric field. Hence, an electric field can induce either a 90° switching or a 180° switching in a unit cell depending upon the direction of the electric field and the orientation of polarization in the unit cell. The electric field at which domain switching is initiated in a ferroelectric material from a poled state is referred to as *coercive electric field* and it is a material-dependent property. However, the electric field producing zero net polarization that is easily obtainable without ambiguity from the experimental plots is often referred to as the *coercive electric field* in some of the literature [9]. The other way to induce domain switching in a ferroelectric material is the application of stress. The stress sufficient enough will force the domains to reorient in such a way that their polarization direction is normal to the direction of stress. Hence, unlike the electric field, stress can produce only 90° domain switching and for any given domain all the four 90° switching are equally probable on application of stress. The compressive stress at which domain switching starts occurring in the ferroelectric material is called the *coercive stress*. This coercive stress is an important property of the material that should be paid enough attention to while designing the transducers.

The crystallographic orientation of tetragonal unit cell remains the same and only the polarization direction inside the unit cell reverses during 180° switching. Hence, 180° switching is associated with only polarization change and the strain due to switching is zero. However, during 90° switching, since both the crystallographic orientation and the polarization direction of the unit cell change, 90°

switching causes both strain and polarization change [10]. Domain switching is primarily responsible for nonlinear behavior of ferroelectric material.

4 BEHAVIOR UNDER DIFFERENT LOADING CONDITIONS

Response of ferroelectric materials under different loading conditions assumes significance from a design point of view. When the material is in a completely depoled state, it can be considered to be macroscopically, though not microscopically, isotropic. As discussed before, an electric field can pole the material in a specific direction from a depoled state. The response of ferroelectric materials under cyclic electric fields with superimposed constant stresses are reported in the literature [10–12] and one such behavior in terms of electric displacement versus electric field is shown in Figure 7. For small electric fields, the ferroelectric material behaves linearly and as the electric field reaches the coercive electric field, nonlinearity is introduced through microscopic domain switching. Once the domain switching saturates, the linear relationship is retained again where the increment in polarization is very small with respect to electric field. During unloading, the ferroelectric material reaches the remnant polarization at zero electric field. In this state, the ferroelectric material can be treated to be a transversely isotropic material since the plane normal to the remnant polarization direction of the ferroelectric material is the plane of symmetry or the transverse plane. Reversing the electric field

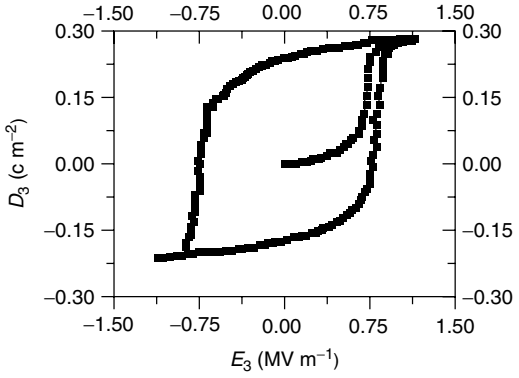


Figure 7. Response regimes and the related switched domains in hysteresis response [11].

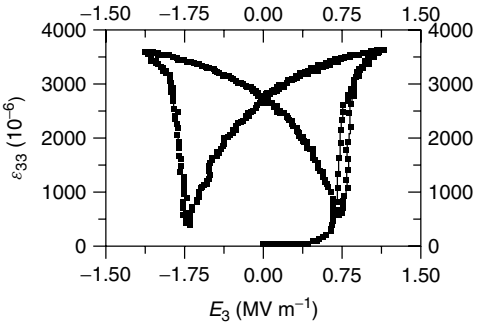


Figure 8. Response regimes and the related switched domains in the butterfly curve [11].

causes the domains to switch in the opposite direction causing a macroscopic polarization change. Remnant polarization obtained during reversal of negative electric field equals the magnitude of the positive remnant polarization but with opposite sense. Hence, the piezoelectric constants determined at the positive and negative remnant polarization states will have the same magnitude but opposite sense.

During cyclic electric field, the dependence of the normal strain on the applied field is shown in Figure 8 [11]. The virgin state of the material is taken as the reference state and the subsequent change in dimensions are measured with respect to that reference state. For small increments in the field, the strain is negligible. As the electric field crosses the coercive electric field, appreciable strain is observed. The increment in the normal strain is due to two reasons: irreversible reorientation of unit cells in the direction of the electric field and the reversible piezoelectric effect of the switched domains. After the switching reaches the saturation state, only reversible strain is present in the material and this is realized during unloading. The strain present in the material after unloading represents the remnant strain and the ferroelectric material behaves linearly in this state for small magnitudes of electric field. This property has widely been exploited for transducer applications.

5 EFFECT OF GRAIN SIZE, TEMPERATURE, AND FREQUENCY

With the advancement that has taken place in the manufacturing process, ferroelectrics possessing

a variety of grain sizes can be developed. It is observed that the ferroelectric property depends also on the grain size. The domains become smaller with decrease in grain size and the width of the domains is roughly proportional to the square root of the grain size for a specific range (between 1 and 10 μm). As the grain size decreases, the multiple domain state may become a single domain state. Finer grain ferroelectrics produce decreased hysteresis and the maximum strain value decreases monotonically with reduction in grain size for the same electric field [13, 14]. This may be due to the pinning effect by the grain boundaries that prevents the easy movement of domain walls. It has been observed in pure $BaTiO_3$ at room temperature that when the particle size decreases below a critical value, about 0.2 μm, c/a ratio of the tetragonal structure falls drastically and the tetragonality is lost at about 0.12 μm [15]. With reduction in grain size, the electrostrictive strain also decreases in relaxor ferroelectrics (discussed later in this section) for the same applied electric field [16]. Hence, the grain size is an important determinant that would influence the electrical and mechanical properties of the ferroelectric and efforts are made to control it depending upon the requirement. Particle size, chemical composition, and heating process are primarily responsible for controlling grain size and the optimum conditions are to be experimentally determined.

As mentioned before, Curie temperature of ferroelectrics vary depending upon its composition. Below Curie temperature, with the ferroelectric phase, the material has appreciable hysteresis and as it moves closer to the paraelectric phase, the hysteresis decreases. On reaching Curie temperature, the hysteresis completely disappears as shown in Figure 9. Hence, when the ferroelectric phase is used in the transducer design for a specific application, considerable care is to be taken to ensure that the operating temperature does not reach the Curie temperature. Otherwise, the ferroelectric will lose its piezoelectric behavior resulting in the failure of the device. Increase in the rate of application of electric field results in a higher coercive electric field. With the loading rate, the hysteresis loop between the polarization and the electric field widens and the butterfly loop between the strain and electric field gets broader with a decreased strain [17]. The reason for the ferroelectrics experiencing lesser strain under higher loading rate than the quasi-static loading is due to the less availability of time for the domains to switch completely in the former case. Even with the mechanical stress, the response is rate dependent, i.e., faster the loading rate bigger is the hysteresis in the stress–strain loop [17].

For most compositions of piezoceramics, the dielectric properties do not vary greatly with frequency, but a special type of ferroelectrics called the *relaxor ferroelectrics* has the tendency to vary the permittivity as a function of temperature and frequency [18]. The composition of a widely used relaxor ferroelectric is $Pb(Mg_{1/3}Nb_{2/3})O_3$ (PMN). Relaxor ferroelectrics have high permittivity and are hence primarily used in compact chip capacitors. Their high electrostriction coefficient with less hysteretic dissipation renders them more attractive to actuator applications.

6 MODELS FOR FERROELECTRIC MATERIALS

The response of ferroelectric materials under different loading conditions should be accurately predictable to incorporate them in the measurement type devices and precision instruments. Among the many models proposed for describing the behavior of the material, linear models suffice for most of the applications. For low-to-moderate input fields, the dielectric and strain relations with the electric field and mechanical stress can be considered to be practically linear. The

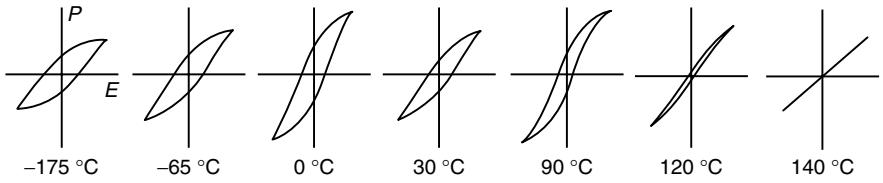


Figure 9. Variation of the size of the hysteresis curve with temperature [Courtesy: DoITPoMS, University of Cambridge].

linear model representing the piezoelectric behavior is represented by

$$P_k = d_{kij}\sigma_{ij} + \kappa_{km}E_m$$
$$\varepsilon_{ij} = S_{ijkl}\sigma_{kl} + d_{mij}E_m \tag{4}$$

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \hline P_1 \\ P_2 \\ P_3 \end{bmatrix} = \left[\begin{array}{cccccc|ccc} S_{11} & S_{12} & S_{13} & 0 & 0 & 0 & 0 & 0 & d_{31} \\ S_{12} & S_{11} & S_{13} & 0 & 0 & 0 & 0 & 0 & d_{31} \\ S_{13} & S_{13} & S_{33} & 0 & 0 & 0 & 0 & 0 & d_{33} \\ 0 & 0 & 0 & S_{44} & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & 0 & S_{44} & 0 & d_{15} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{66} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & d_{15} & 0 & \kappa_{11} & 0 & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 & 0 & \kappa_{11} & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 & 0 & 0 & \kappa_{33} \end{array}\right] \times \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \\ \hline E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{5}$$

where $\boldsymbol{S}$ is the elastic compliance tensor, $\boldsymbol{d}$ the piezoelectric coupling tensor, and $\boldsymbol{\kappa}$ the dielectric susceptibility tensor. As can be seen in equation (4), the number of constants involved in the material properties is high and they can be considerably reduced by invoking elastic and electric symmetries. For the lowest macroscopic symmetry, i.e., a thermally or electrically depoled state, there are 21 elastic constants, 18 piezoelectric constants, and 6 dielectric constants considering a tetragonal crystal structure. Since the ferroelectric materials are primarily used in their piezoelectric region, i.e., after poling is done in a specific direction, they can be considered to be transversely isotropic, which is a higher symmetry state macroscopically. Hence, the total number of constants required in the piezoelectric linear analysis can be reduced to just 10 consisting of 5 elastic constants, 3 piezoelectric constants, and 2 dielectric constants. Thus, the linear relationship for piezoelectric ceramics described in equation (4) can be represented in a compact form as given in equation (5).

As discussed earlier, the ferroelectric material behaves nonlinearly under higher electric field and stress. Owing to complex geometry requirements in the design of ferroelectric devices and operating conditions, the material may be exposed to electric field and stress, beyond their critical values, affecting the performance of the devices. In addition, owing to lack of fracture resistance, the ferroelectric ceramic experiences cracks at electrode interfaces, within the ceramic layers, and near the electrode tips. In linear analysis, the effect of switching on the fracture toughness of the material is ignored [19]. Hence, the electromechanical fields that are determined using linear analysis are considerably different from those of nonlinear analysis since the latter includes the switching process happening near the crack tip [20–23]. Hence, the nonlinear models are necessary to capture the ferroelectric behavior more realistically in order to design the ferroelectric devices more accurately.

To capture the nonlinearity, many models have been proposed and reported in literature [9, 24, 25]. They can broadly be classified as phenomenological models and micromechanical models. Phenomenological models are generally derived within the framework of irreversible thermodynamics. The state of the material is defined by internal variables at any given time and the evolution of these internal variables is determined by the kinetic equations [26–30]. Remnant strain and remnant polarization are the generally used internal variables in this approach and they define the irreversible state of the material. Some of the phenomenological models use plasticity approach in which the electric yield surface and stress yield surface are developed and the material's macroscopic reversible behavior occurs within these surfaces. When domain switching takes place in the material, the switching surface displays an irreversible change due to variation in the remnant strain and remnant polarization. During switching, the switching surface expands or moves depending upon the hardening rule that determines the evolution of the new surfaces. The phenomenological models are computationally effective and the implementation of these into the finite element formulation is simple. These finite element formulations are then used to predict the performance of ferroelectric devices and to analyze the fracture process in the material. The flip side of the phenomenological models is that they need experiment-dependent constants as

they do not involve microscopic mechanism of the material.

The other classification of models, micromechanical models, is based on the internal microstructure and the microscopic switching mechanisms [10, 31–35]. In these models, each grain with a single domain or multiple domains is considered for modeling and is allowed to switch when the ferroelectric grain meets some switching criterion. The criterion may be based on work done, total potential, Gibb's energy, and internal energy density [10, 32, 36–39]. The driving force for each increment of electric field and stress is calculated and checked for the switching criterion and on satisfying the criterion, the existing domains are converted to another set of domains that are favorable to the external forcing fields. The macroscopic material behavior is obtained by microscopic averaging of the individual grains or domains. Though it is computationally costly preventing it from being as popular as phenomenological models, it has some advantages over the latter models. Since the material constants are related to the microstructure of the ferroelectric materials, the microscopic models are independent of the experiments. The material response under loading conditions that are difficult to conduct in the lab can be predicted through these models. A combination of phenomenological and micromechanical models may deliver the advantages of both, for instance, the saturation strain and saturation polarization of the ferroelectric materials under specific loading conditions are calculated using micromechanical models and are given as input to the phenomenological models to gain the computational edge.

However, in general, as discussed earlier, the piezoceramics are employed within their linear region only. When unconstrained, the piezoceramic produces strain in all the directions under the influence of electric field, and hence, the material experiences no stress; however, as the deformation of the material is constrained, it exhibits enormous force on the constraining element, which is used for actuation purposes. By rearranging the linear relationship discussed in the preceding section, the stresses developed in the piezoceramic can be determined as follows:

$$\sigma_{ij} = S^{-1}_{ijkl}\varepsilon_{kl} - \boldsymbol{e}_{ijk}E_k \tag{6}$$

Table 1. Material constants of a few ferroelectric materials widely studied in literature [43]

	PZT-4	PZT-5H	PZT-6B	$BaTiO_3$
$\boldsymbol{Y}_{11}$	139	126	168	150
$\boldsymbol{Y}_{12}$	77.8	79.1	84.7	65.3
$\boldsymbol{Y}_{13}$	74.0	83.9	84.2	66.2
$\boldsymbol{Y}_{33}$	115	117	163	146
$\boldsymbol{Y}_{44}$	25.6	23.0	35.5	43.9
$\boldsymbol{Y}_{66}$	30.6	23.5	41.7	42.4
$\boldsymbol{e}_{15}$	12.7	17	4.6	11.4
$\boldsymbol{e}_{31}$	−5.2	−6.5	−0.9	−4.3
$\boldsymbol{e}_{33}$	15.1	23.3	7.1	17.5
$\boldsymbol{\kappa}_{11}$	6.46	15.05	3.60	9.87
$\boldsymbol{\kappa}_{33}$	5.62	13.02	3.42	11.16
ρ	7500	7500	7550	5700

Y, elastic modulus (GPa); e, piezoelectric coefficient ($\mathrm{C\,m^{-2}}$); κ, dielectric permittivity ($\mathrm{nF\,m^{-1}}$); ρ, density ($\mathrm{kg\,m^{-3}}$).

where, $\boldsymbol{e} = \mathrm{S}^{-1}d^T$, the modified piezoelectric coupling tensor. From equation (6), the stress developed in the piezoceramic can be determined when a specific quantity of strain is restrained. Also, it is apparent from equation (6) that the stress in the piezoceramic becomes zero when the deformation is allowed to take place freely. Some typical properties of various parameters involved in the constitutive models are given Table 1.

7 IMPORTANT DESIGN PARAMETERS IN PIEZOCERAMICS

Many parameters defining the piezoelectric properties relevant to practical applications play an important role in the piezoelectric devices. A few of those figures of merit are discussed here [40]. For the given applied electric field, E, the value of strain, ε, produced is determined by *piezoelectric strain constant*, d, which is an important property for actuator applications.

$$\varepsilon = dE \tag{7}$$

The electric field developed in the material due to external stress, σ, is represented by *piezoelectric voltage constant*, g, which is used for sensor

applications.

$$E = g\sigma \tag{8}$$

The relationship between the piezoelectric strain and piezoelectric voltage constants can be obtained as $g = d/\kappa$, where κ is called *dielectric permittivity*.

Another important figure of merit is the *electromechanical coupling factor*, f, which is defined as the ratio of the stored mechanical energy in the piezoelectric material to the given input electrical energy. It can also be defined as the ratio between the stored electrical energy and the given mechanical energy input. For mathematical ease, the electromechanical coupling factor is taken as

$$f^2 = \text{stored mechanical energy/input electrical energy} \tag{9}$$

$$\text{Mechanical energy stored} = \frac{1}{2}\varepsilon\sigma = \frac{1}{2S}\varepsilon^2 = \frac{1}{2S}(dE)^2 \tag{10}$$

$$\text{Input electrical energy} = \frac{1}{2}\kappa E^2 \tag{11}$$

$$f^2 = \frac{\frac{1}{2S}(dE)^2}{\frac{1}{2}\kappa E^2} = \frac{d^2}{\kappa S} \tag{12}$$

Electromechanical coupling factor indicates only the energy stored in the material and not the useful work done by it. Hence, another figure of merit is required to denote the amount of actual work done for the given input energy and that is referred to as the *energy transmission coefficient*, λ.

$$\lambda = \text{output mechanical energy/input electrical energy (or)} = \text{output electrical energy/input mechanical energy} \tag{13}$$

The relationship between the electromechanical coupling factor (f) and the energy transmission coefficient (λ_{max}) can be obtained as

$$\lambda_{max} = \left[\frac{1}{f} - \sqrt{\frac{1}{f^2} - 1}\right]^2 \tag{14}$$

Mechanical quality factor, Q, is an important parameter to evaluate the resonant strain and resonance spectrum of the given ceramic. The off-resonance strain can also be determined using the mechanical quality factor at the resonance frequency and the inverse of mechanical quality factor defines the *mechanical loss*. The mechanical quality factor is defined with the resonance frequency, ω_0, as

$$Q = \frac{\omega_0}{2\Delta\omega} \tag{15}$$

When the wave energy is transferred between two mediums, a part of it is transmitted and the rest is reflected back. The ratio of reflected energy to transmitted energy depends upon an important parameter called *acoustic impedance*, Z, and the transmission is complete if the acoustic impedance of the materials is matching.

$$Z = \sqrt{\rho M} \tag{16}$$

where ρ is the density and M is the elastic stiffness of the material.

A simple application of piezoelectric ceramic is the gas igniter. When force is applied on piezoelectric material, it produces very high voltage that is sufficient to ignite the gas. Given the resonance time period of the piezoelectric material considered for the igniter, its required length can be determined by a simple calculation as follows. The resonance frequency, ω_r, and the length, L, are related by $\omega_r = 1/L$, where ω_r is expressed in kilohertz and L in meter. For example, if the resonance time period, t_r, is 20 μs, then

$$\omega_r = \frac{1}{t_r} = \frac{1}{20\,\mu s} = \frac{1}{20 \times 10^{-6}\,s} = 50\,\text{kHz} \tag{17}$$

$$L = \frac{1}{\omega_r} = \frac{1}{50} = 0.02\,\text{m} = 20\,\text{mm} \tag{18}$$

When the electric field is applied, the piezoelectric ceramic experiences a strain, and on alternating the

field, the material vibrates. As the frequency of the driving electric field matches the resonance frequency of the device, the magnitude of the strain of the ceramic becomes very large and this phenomenon is known as *piezoelectric resonance*. This phenomenon is exploited in vibrators. Piezoelectric resonance is understood due to the accumulation of the input energy in the material. The internal energy, U, of the piezoelectric vibrator is the addition of mechanical and electrical energies.

$$\begin{aligned} U_{\text{Total}} &= U_{\text{M}} + U_{\text{E}} \\ &= \frac{1}{2}\varepsilon\sigma + \frac{1}{2}PE \end{aligned} \tag{19}$$

Using the linear relationship of ε and P on E and σ as discussed before,

$$U_{\text{Total}} = \frac{1}{2}[S\sigma^2 + dE\sigma] + \frac{1}{2}[d\sigma E + \kappa E^2] \tag{20}$$

where, pure mechanical energy, U_{M}, and pure electrical energy, U_{E}, are represented by

$$U_{\text{M}} = \frac{1}{2}S\sigma^2 \quad \text{and} \quad U_{\text{E}} = \frac{1}{2}\kappa E^2 \tag{21}$$

and the energy transduced, U_{trans}, from mechanical to electrical or vice versa is represented by

$$U_{\text{trans}} = \frac{1}{2}dE\sigma \tag{22}$$

The electromechanical coupling factor for piezoelectric vibrator is defined as

$$f = \frac{U_{\text{trans}}}{\sqrt{U_{\text{M}}U_{\text{E}}}} \tag{23}$$

For example, the electromechanical coupling factor with certain elastic boundary conditions for the resonator shape shown in Figure 10 can be derived as follows. The displacement of the ferroelectric material along the axis 1 is used for actuation as the electric field is applied along the axis 3. The ceramic exhibits strain in all the directions, but the stress will be developed only along the axis 1, since the actuation is done in that direction. Hence, the electromechanical coupling factor for this specific actuation or vibration mode can be obtained as

$$f_{31} = \frac{\frac{1}{2}d_{31}E_3\sigma_1}{\sqrt{\frac{1}{2}S_{11}\sigma_1^2 \times \frac{1}{2}\kappa_{33}E_3^2}} = \frac{d_{31}}{\sqrt{S_{11}\kappa_{33}}} \tag{24}$$

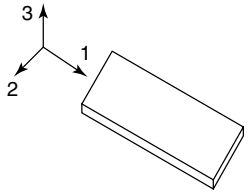


Figure 10. Piezoelectric resonator shape.

The electromechanical coupling factor for various resonator shapes under different elastic boundary conditions are listed in [40]. In piezoelectric vibrators, the size and shape of the device are also significant in addition to the piezoelectric material and the vibration mode at which the resonance is produced. In general, the piezoelectric ceramic with higher mechanical quality factor is desirable for vibrators and hence, the harder piezoelectric materials.

Current trends in the actuator applications are in exploring and using the electrostrictive effect of high permittivity materials since they have specific advantages over piezoelectric actuators. As discussed earlier, the strain produced by the electrostrictive effect is always positive irrespective of the direction of the electric field applied. The electrostrictive actuators, operating above Curie temperature, do not contain domains and hence, return to their original position after the removal of the electric field and do not suffer from aging problems. Hence, the materials showing strong electrostrictive effects are gaining more interest despite the fact that their electrostrictive effect is not as strong as the piezoelectric effect.

8 APPLICATIONS

The piezoelectric behavior of ferroelectric ceramics, i.e., generation of high voltage and detection of acoustic and ultrasonic energy, are widely exploited in developing electromechanical actuators and sensors [8, 40, 41]. A piezoelectric solid-state ceramic actuator converts the electrical energy directly into a linear motion. A piezoelectric ceramic sensor converts the mechanical energy directly into an electrical charge. Large force actuation, microsecond

range response, and insensitivity to magnetic field are some of the major advantages of piezoceramic transducers. High-resolution ferroelectric ceramic actuators can provide incremental motion ranging from a few microns to subnanometers with almost negligible backlash. Micro- and nanopositioning are the key factors in semiconductor industry and in medical diagnostics. Another important characteristic of PZT actuators is that they exhibit consistent performance even after a large number of cycles. Ferroelectric ceramics are used from simple gas igniters to complex scanning microscopy and precision machining. The ultimate usage of piezoceramics lies in *SHM,* which is a nondestructive evaluation technique. The recent developments in the piezo devices and piezo composites cause a remarkable improvement in SHM systems [42].

The main challenge in incorporating the piezoelectric or electrostrictive materials in the actuator devices is the small displacements produced by them on application of low or medium electric fields. For example, 1 kV applied across a poled ferroelectric strip of 1 mm thickness with a piezoelectric coupling coefficient of 0.5×10^{-9} C/N will produce only 0.5-μm displacement. Hence, with these small displacements, the ferroelectric actuator devices should be able to produce the required actuation strain. One way of achieving the more demanding strain is through stacking of ferroelectric materials by connecting the alternate layers without allowing a short circuit between the adjacent layers as shown in Figure 11. The main advantage of this multilayer arrangement, which is widely used in actuator devices, is the generation of high displacement with low driving voltage and high electromechanical coupling.

Except for some applications, even this summed up displacement is not sufficient to produce the required strain. Also, the stack actuation results in larger size of the device, which is undesirable in some applications like MEMS (micro-electromechanical systems) where the size of the device counts. To overcome the drawback of the low actuation strain in multilayer actuators, the piezoelectric plates are bonded with the elastic plates to produce large strains utilizing the bending displacement. In such a bimorph, the ferroelectric beams may be poled in the same direction or in the opposite direction and the placement of electrodes in between the beams and the measurement of voltage through them depends on the way the ferroelectric beams are poled. Bimorph actuators produce more strain than the stack actuators, but with low response time and generative forces. A compromise between stack actuation and bimorph is obtained through the composite actuator structure called *Moonie*, which develops the strain that is one magnitude higher than that of stack actuators and larger generative forces in less response time when compared to bimorphs.

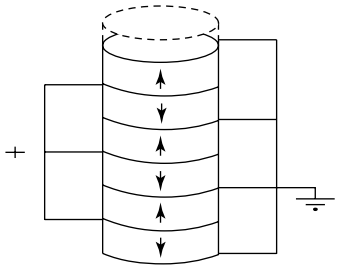


Figure 11. Stack actuation.

When the ferroelectric ceramics are used as sensors, they are generally brought under two major classes as passive and active sensing/sampling. In passive sampling, the ferroelectric sensors detect the perturbations due to ambient conditions that are not introduced artificially. These measurements may be for a single value that occurs as maximum during the operation like the maximum strain or a continuously varying data like acceleration and ambient frequency response of dynamic systems. Active-sensing systems require the externally supplied energy that is generated specifically for sensing purpose.

An important field effectively using ferroelectric ceramics both as sensors and actuators is *nondestructive evaluation techniques*. Owing to the impurities present in the material and the manufacturing process, flaws are likely to occur in the materials and it is mandatory to check for the magnitude of defects before the material is put in use. Also, when the material or structure is in service, it is essential to check for any possible degradation of its performance to avoid drastic failure of the system. The manufacturing process of the material can be optimized and the quality of service of the structure can be ensured through nondestructive techniques. Nondestructive techniques involving ferroelectric materials make use of different concepts like ultrasonic, eddy current, and radiography to examine the state of the structure. The piezoceramic transducers may be

pasted on the surface of the material or embedded within the material through complex manufacturing process. When the embedded piezoceramic is driven by an electric pulse, it transmits ultrasonic waves in a wide range of frequency. A conventional scanning probe receives the signal from the structure to check for flaws developed owing to environmental and operational reasons. This way, not only the condition of the materials under probe but also the piezoceramic actuators can be checked. In case of radiography, a harmonic excitation is given to the piezoceramic and the thermal field output is visualized through an infrared camera. In eddy-current method, the actuator, driven by a harmonic electric source, produces electromagnetic field, which is detectable by external sensors.

SHM is the extension of nondestructive evaluation techniques and is the process of implementing active damage-detection strategy for aerospace, civil, and mechanical engineering infrastructures. SHM involves the measurements of inputs to and the response of the structure using piezoelectric transducers in order to predict the onset of damage and deterioration in the structural condition. This provides a path for the transition from a schedule-based maintenance practice to condition-based maintenance.

9 CONCLUSIONS

The wide range of applications of piezoelectric ceramics in sensing and actuation devices demands more usage of them, especially the *PZT*. The main challenge in employing piezoceramics for actuation over other smart materials like shape memory alloys lies in the quantity of strain they produce under external loads. A simple way to overcome this concern is through stack actuation, but at the cost of the size of the device. Also, there is an increasing demand for the restricted use of piezoceramics containing lead owing to the growing awareness on health and environmental problems associated with lead content in materials. One possible solution that addresses both problems lies with the utilization of the nonlinear range that enhances greater potential for their use in applications that demand higher performance. To deal with the nonlinearity in the design of the devices in the applications, models that are capable of predicting the piezoelectric ceramics more precisely and accurately in their nonlinear region, under variety of loading conditions, becomes necessary. This also helps in overcoming the hesitation in designer's approach to dealing with the nonlinear behavior, providing a new dimension to the usage of piezoceramic materials.

RELATED ARTICLES

Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators

Piezoelectricity Principles and Materials

Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors

REFERENCES

[1] Cady WG. *Piezoelectricity*. Dover: New York, 1964.

[2] Jaffe B, Cook Jr WR, Jaffe H. *Piezoelectric Ceramics*, Academic Press: New York, 1971.

[3] Lines ME, Glass AM. *Principles and Applications of Ferroelectrics and Related Materials*. Oxford University Press: Oxford, 1977.

[4] Maugin GA. *Continuum Mechanics of Electromagnetic Solids*. Amsterdam: North-Holland, 1988.

[5] Parton VZ, Kudryavtsev BA. *Electromagnetoelasticity. Piezoelectrics and Electrically Conductive Solids*. Gordon & Breach: New York, 1988.

[6] Eringen AC, Maugin GA. *Electrodynamics of Continua I. Foundations and Solid Media*. Springer: New York, 1989.

[7] Ikeda T. *Fundamentals of Piezoelectricity*. Oxford University Press: Oxford, 1990.

[8] Moulson AJ, Herbert JM. *Electroceramics: Materials, Properties, Applications*. Chapman & Hall: New York, 1990.

[9] Kamlah M. Ferroelectric and ferroelastic piezoceramics—modeling of electromechanical hysteresis phenomena. *Continuum Mechanics and Thermodynamics* 2001 **13**:219–268.

[10] Hwang SC, Lynch CS, McMeeking RM. Ferroelectric/Ferroelastic interactions and a polarization switching model. *Acta Metallurgica et Materialia* 1995 **43**:2073–2084.

[11] Fang D, Changqing Li. Nonlinear electric-mechanical behavior of a soft PZT-51 ferroelectric ceramic. *Journal of Materials Science* 1999 **34**:4001–4010.

[12] Lynch CS. The effect of uniaxial stress on the electro-mechanical response of 8/65/35 PLZT. *Acta Materialia* 1996 **44**:4137–4148.

[13] Yamaji A, Enomoto Y, Kinoshita K, Tanaka T. *Proceedings of the First Meeting on Ferroelectric Materials and Their Applications*. Kyoto, 1997; pp. 269.

[14] Uchino K, Takasu T. Evaluation method of piezoelectric ceramics from a viewpoint of grain size. *Inspec* 1986 **10**:29.

[15] Uchino K, Sadanaga E, Hirose T. Dependence of the crystal structure on particle size in barium titanate. *Journal of the American Ceramic Society* 1989 **72**:1555–1558.

[16] Shrout TR, Kumar U, Megherhi M, Yang N, Jang SJ. Grain size dependence of dielectric and electrostriction of $Pb(Mg_{1/3}Nb_{2/3})O_3$ based ceramics. *Ferroelectrics* 1987 **76**:479–487.

[17] Zhou D, Kamlah M, Munz D. Rate dependence of soft PZT ceramics under electric field loading. *Proceedings of SPIE* 2001 **4333**:64–70.

[18] Smolenskij GA, Isupov VA, Agranovskaya AI, Popov SN. Composition dependence of the phase transition temperature in $Ba_xSr_{1-x}Nb_2O_6$. *Soviet Physics—Solid State* 1961 **2**:2584–2588.

[19] Zhang TY, Zhao M, Tong P. Fracture of piezoelectric ceramics. *Advances in Applied Mechanics* 2002 **38**:147–289.

[20] Zhu T, Yang W. Toughness variation of ferroelectrics by polarization switch under non-uniform electric field. *Acta Materialia* 1997 **45**:4659–4702.

[21] Yang W, Zhu T. Switch toughening of ferroelectrics subjected to electric fields. *Journal of the Mechanics and Physics of Solids* 1998 **46**:291–311.

[22] Chen W, Lynch CS. Finite element analysis of cracks in ferroelectric ceramic materials. *Engineering Fracture Mechanics* 1999 **64**:539–562.

[23] Kreher W. Influence of domain switching zones on the fracture toughness of ferroelectrics. *Journal of the Mechanics and Physics of Solids* 2002 **50**:1029–1050.

[24] Landis CM. Non-linear constitutive modeling of ferroelectrics. *Current Opinion in Solid State and Materials Science* 2004 **8**:59–69.

[25] Huber JE. Micromechanical modeling of ferroelectrics. *Current Opinion in Solid State and Materials Science* 2005 **9**:100–106.

[26] Cocks ACF, McMeeking RM. A phenomenological constitutive law for the behavior of ferroelectric ceramics. *Ferroelectrics* 1999 **228**:219–228.

[27] Kamlah M, Jiang Q. A constitutive model for ferroelectric PZT ceramics under uniaxial loading. *Smart Materials and Structures* 1999 **8**:441–459.

[28] Landis CM, McMeeking RM. A phenomenological constitutive law for ferroelastic switching and a resulting asymptotic crack tip solution. *Journal of Intelligent Material Systems and Structures* 2000 **10**:155–163.

[29] McMeeking RM, Landis CM. A phenomenological multiaxial constitutive law for switching in polycrystalline ferroelectric ceramics. *International Journal of Engineering Science* 2002 **40**:1553–1577.

[30] Shieh J, Huber JE, Fleck NA. An evaluation of switching criteria for ferroelectrics under stress and electric field. *Acta Materialia* 2003 **51**:6123–6137.

[31] Michelitsch T, Kreher WS. A simple model for the nonlinear material behavior of ferroelectrics. *Acta Materialia* 1998 **46**:5085–5094.

[32] Chen W, Lynch CS. A micro-electro-mechanical model for polarization switching of ferroelectric materials. *Acta Materialia* 1998 **46**:5303–5311.

[33] Hwang SC, Huber JE, McMeeking RM, Fleck NA. The simulation of switching in polycrystalline ferroelectric ceramics. *Journal of Applied Physics* 1998 **84**:1530–1540.

[34] Lu W, Fang DN, Li CQ, Hwang KC. Nonlinear electric-mechanical behavior and micromechanics modeling of ferroelectric domain evolution. *Acta Materialia* 1999 **47**:2913–2926.

[35] Huber JE, Fleck NA, Landis CM, McMeeking RM. A constitutive model for ferroelectric polycrystals. *Journal of the Mechanics and Physics of Solids* 1999 **47**:1663–1697.

[36] Sun CT, Jiang LZ. Domain switching induced stress at the tip of a crack in piezoceramics. *Proceedings of the 4th ESSM 2nd MIMR*. Harrogate, 1998; pp. 715–722.

[37] Hwang SC, McMeeking RM. A finite element model of ferroelectric/ferroelastic polycrystals. *Proceedings of SPIE* 2000 **3992**:404–417.

[38] Sun CT, Achuthan A. Domain switching criteria for piezoelectric materials. *Proceedings of SPIE* 2001 **4333**:240–249.

[39] Shaikh MG, Phanish S, Sivakumar SM. Domain switching criteria for ferroelectrics. *Computational Material Science* 2006 **37**:178–186.

[40] Uchino K. *Ferroelectric Devices*. Marcel Dekker: New York, 2000.

[41] Leo DJ. *Smart Material Systems*. John Wiley & Sons: New Jersey, 2007.

[42] Los Alamos National Laboratory report. *A Review of Structural Health Monitoring Literature: 1996–2001*. LA-13976-MS, 2003.

[43] Jaffe H, Berlincourt DA, Piezoelectric transducer materials. *Proceedings of IEEE* 1965 **53**:1372–1386.

Chapter 41

Constitutive Modeling of Magnetostrictive Materials

Srinivasan Gopalakrishnan

Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

1 INTRODUCTION

Some magnetic materials show elongation and contraction in the magnetization direction owing to an induced magnetic field. This is called *magnetostriction*, which is due to the switching of a large amount of magnetic domains caused by spontaneous magnetization below the Curie point of temperature. Thus, magnetostrictive materials have the ability to convert magnetic energy into mechanical energy and vice versa. This coupling between magnetic and mechanical energies represents the transudation capability and this allows a magnetostrictive material to be used in both actuation and sensing devices. Owing to magnetostriction and its inverse effect (also called *Villery effect*) [1], magnetostrictive materials can be used both as an actuator and as a sensor. The theoretical and experimental study of magnetostrictive materials has been the focus of considerable research for many years. However, only with the recent development of giant magnetostrictive materials (e.g., Terfenol-D), it is now possible to produce sufficiently large strains and forces to facilitate the use of these materials in actuators and sensors. This has led to the application of magnetostrictive materials to devices such as micropositioners, vibration controller, sonar projectors and insulators, etc. Magnetostrictive material has found its way in many structural applications such as vibration control, noise control, and structural health monitoring.

The constitutive relationship of magnetostrictive materials consists of a sensing and an actuation equation. In the sensing equation, magnetic flux density is a function of applied magnetic field and stress, whereas, in actuation equation, strain is a function of applied magnetic field and stress. Both sensing and actuation equations are coupled through applied magnetic field and mechanical stress level. The compliance matrix in the actuation law depends on the magnetic field, while the permeability matrix in the sensing law depends on the measured stress.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Hence, unlike the piezoceramic material, both the sensing and actuation law cannot be uncoupled. For evaluating all the material constants in the constitutive law, both sensing and actuation laws have to be solved simultaneously. For each magnetic field, there are different stress–strain curves and for each stress level, there are different magnetostriction–magnetic field curves, which are highly nonlinear. There are very few works reported in the literature toward this direction. Small sets of data provided by the manufacturer [2] are highly inadequate for modeling and analysis of structures with built-in magnetostrictive sensors. Hence, there is need to numerically characterize the constitutive model so that the material data can be generated for all magnetic fields and stress levels. This is the main focus of this article.

Analysis of smart structures with magnetostrictive material is generally performed using uncoupled models. Uncoupled models are based on the assumption that the magnetic field within the magnetostrictive material is proportional to the electric coil current times the number of coil turns per unit length (Ghosh and Gopalakrishnan, 2002). Owing to this assumption, the actuation and the sensing equations get uncoupled. For the actuator, the strain due to magnetic field (which is proportional to coil current) is incorporated as the equivalent nodal load in the finite element model for calculating the block force. Thus, with this procedure, analysis can be carried out without taking the smart degrees of freedom in the finite element model. Similarly, for sensor, where generally coil current is assumed as zero, the magnetic flux density is proportional to mechanical stress, which can be calculated from the finite element results through postprocessing. This assumption on magnetic field leads to the violation of flux line continuity, which is one of the four Maxwell's equations in electromagnetism. On the other hand, in the coupled model, it is considered that magnetic flux density and/or strain of the material are functions of stress and magnetic field, without any additional assumption on magnetic field, like the uncoupled model. Benbouzid *et al.* modeled the static [3] and dynamic [4] behavior of the nonlinear magnetoelastic medium for magnetostatic case using finite element method. Magnetomechanical coupling was incorporated considering both permeability and elastic modulus as functions of stress and magnetic field. However, all these works do not provide a convenient way for analysis of magnetostrictive smart structure considering coupled magnetomechanical features. This article deals with the numerical characterization of the constitutive relationship considering the coupled features of magnetostrictive materials, which can be directly used in any mathematical/numerical formulation such as finite element method for structures with magnetostrictive material considering both magnetic and mechanical degrees as unknown degrees of freedom. In addition, it is shown that the magnetic field is not proportional to applied coil current (which is the assumption of the uncoupled model) and depends on the mechanical stress on the magnetostrictive material. This study here also demonstrates how the coupled model preserves the flux line continuity, which is one of the drawbacks of the uncoupled model.

The constitutive relations of magnetostrictive materials are essentially nonlinear [2]. The prediction of the behavior of magnetostrictive material, in general, is extremely complicated owing to its hysteretic nonlinear character. In structural application, owing to these nonlinear material properties, modeling of the system becomes nonlinear, for which exact nonlinear constitutive relationships are essential. Toupin [5] and Maugin [6] had done extensive work related to electrostrictive and piezoelectric phenomena, which have similarities in form with the magnetostriction phenomena. Earlier study to model uncoupled nonlinear actuation of magnetostrictive material was done by Krishnamurty *et al.* [7] by considering a fourth-order polynomial of magnetic field for each stress level. In this approach, for the stress level for which the curve is not available, the coefficients of the curve have to be interpolated from the coefficients of the nearest upper and lower stress level curves.

The numerical characterization of the nonlinear-coupled constitutive model can be performed by expanding the magnetic flux density and strain variations as higher order polynomials, the constants of which can be computed from the data supplied by the manufacturer [2] for different magnetic fields and stress levels through an iterative procedure. Alternatively, the numerical characterization of the constitutive law can be performed using a back propagation (BP) artificial neural network (ANN) and this uses the data supplied by the manufacturer. This procedure completely avoids the nonlinear iteration, and is readily amenable for implementation alongside any finite element code. Here, one three-layer ANN is

trained using the data supplied by the manufacturer, to get this nonlinear mapping directly. ANN is a universal approximator, which can give a nonlinear parameterized mapping from a given input data to an output data. In this study, ANN is used to get the direct mapping for the constitutive relationship of magnetostrictive materials, where inputs in the network are magnetic field and applied stress level and outputs in the network are the strain and the magnetic flux density. Hence, nonlinearity in elastic modulus and permeability is replaced by this trained network. Both these approaches are demonstrated in this article.

The article is organized as follows. In the next section, the complete architecture of the ANN developed for this work is explained, which includes the process of training and validation of network. This is followed by a detailed section on the numerical characterization of linear and nonlinear coupled and uncoupled constitutive models. Finally, the developed models are validated with each other and also with available results.

2 ARTIFICIAL NEURAL NETWORK (ANN)

ANNs can provide nonlinear parameterized mapping between a set of inputs and a set of outputs with unknown function relationship. A three-layer network (Figure 1) with the sigmoid activation functions can approximate any smooth mapping. A typical supervised feedforward multilayer neural network is called as a *back propagation* neural network. The structure of a BP neural network shown in Figure 1 mainly includes an input layer for receiving the input data; some hidden layer for processing data; and an output layer to indicate the identified results. In this study, the task of identifying nonlinear magnetostriction through ANN is performed by training the neural network using the known samples.

2.1 Training of the network

The training of a BP neural network is a two-step procedure [8]. In the first step, the network propagates input through each layer until an output is generated. The error between the output and the target output is then computed. In the second step, the calculated error is transmitted backward from the output layer and the weights are adjusted to minimize the error. The training process is terminated when the error is sufficiently small for all training samples. In practical applications of the BP algorithm, learning is the result from many presentations of these training examples to the multilayer perceptron. One complete presentation of the entire training set during the learning process is called an *epoch*. The learning process is maintained on an epoch-by-epoch basis until the synaptic weights and bias levels of the network stabilize and the averaged squared error over the entire training set converges to some minimum value. For a given

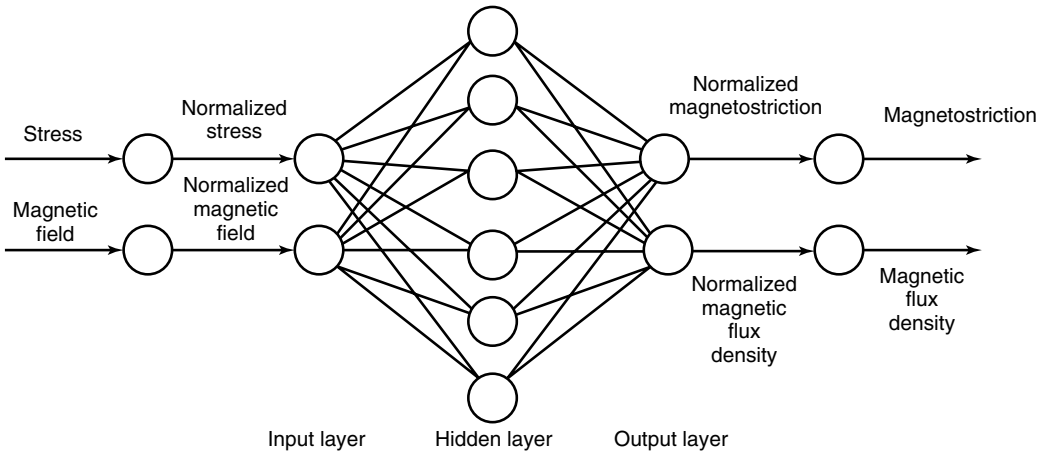


Figure 1. Artificial neural network architecture.

training set, BP learning can be done in a sequential or batch mode.

2.2 Validation of trained network

To validate the trained network, the data set is separated into two parts, one for training and the other for testing the network performance. The network is trained using training sample and the trained network is validated with the test sample. A network is said to generalize well when the input–output mapping computed by the network is corrected with the test data that was never used in creating or training the network. Although the network performs useful interpolation, because of multilayered perceptrons with continuous activation functions, it leads to output functions that are also continuous.

3 ANHYSTERETIC-COUPLED CONSTITUTIVE MODEL

Here, experimental data is taken from Etrema manual [2] for Terfenol-D, a giant magnetostrictive material to verify the proposed model. Experimental data of the magnetostriction versus magnetic field for different stress level given in the manual is reproduced in Figure 2 and the stress versus strain curves for different magnetic field level are reproduced in Figure 3.

The application of magnetic field causes strain in the magnetostrictive material (Terfenol-D) and hence the stress, which changes magnetization of the material. As described by Butler [2], Moffett *et al.* [9], and Hall and Flatau [10], the three-dimensional coupled constitutive relationship between magnetic and mechanical quantities for magnetostrictive material is given by

$$\{\varepsilon\} = [S^{(H)}]\{\sigma\} + [d]^T\{H\} \tag{1}$$

$$\{B\} = [\mu^{\sigma}]\{H\} + [d]\{\sigma\} \tag{2}$$

where $\{\varepsilon\}$ and $\{\sigma\}$ are strain and stress, respectively. $[S^{(H)}]$ represents elastic compliance measured at constant $\{H\}$ and $[\mu^{(\sigma)}]$ represents the permeability measured at constant stress $\{\sigma\}$. Here, $[d]$ is the magnetomechanical coupling coefficient, which provides a measure of the coupling between the mechanical strain and magnetic field. In general, $[S]$, $[d]$, and $[\mu]$ are nonlinear as they depend upon $\{\sigma\}$ and $\{H\}$. Equation (1) is often referred to as the *direct effect* and equation (2) is known as the *converse effect*. These equations are traditionally used for actuation and sensing purpose, respectively. It should be noted that the elastic constants used correspond to

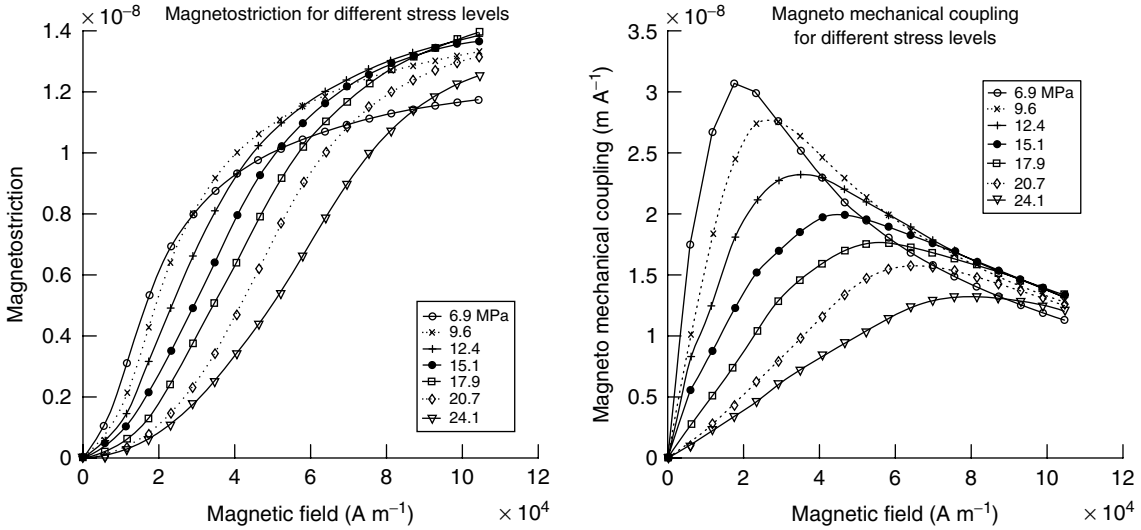


Figure 2. Magnetostriction and magnetomechanical coupling as a function of magnetic field data plotted from Etrema data.

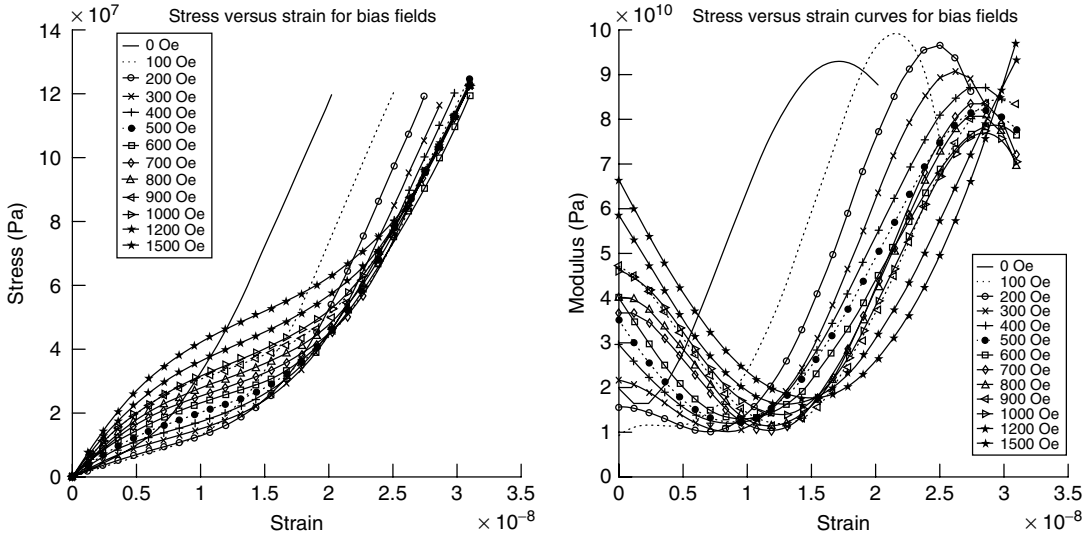


Figure 3. Stress–strain relationship for different magnetic field data plotted from Etrema data.

the fixed magnetic field values and the permeability corresponds to the fixed stress values.

3.1 Coupled constitutive model

Analysis of smart structures using magnetostrictive materials as either sensors or actuators has traditionally been performed using uncoupled models. Uncoupled models make the assumption that the magnetic field within the magnetostrictive material is constant and proportional to the electric coil current times the number of coil turns per unit length [11]. Hence, actuation and sensing problems are solved by two uncoupled equations, which are given by the last part of equations (1) and (2), respectively. This makes the analysis relatively simple; however, this method has its limitations. It is quite well known that $[S]$, $[d]$, and $[\mu]$ all depend on the stress level and magnetic field. In the presence of mechanical loads, the stress changes and so does the magnetic field. Estimating the constitutive properties using the uncoupled model in such cases gives inaccurate predictions. Hence, the constitutive model should be represented by a pair of coupled equations given by equations (1) and (2) to predict the mechanical and magnetic response. It is therefore necessary to simultaneously solve for both the magnetic response and the mechanical response, regardless of whether the magnetostrictive material is being used as a sensor or an actuator. Owing to in-built nonlinearity, the uncoupled model may not be capable of handling certain applications such as (i) modeling passive damping circuits in vibration control and (ii) development of self-sensing actuators in structural health monitoring. In these applications, the coupled equations require to be solved simultaneously. The simultaneous solution of coupled equations is a necessity for general-purpose analysis of adaptive structures built with magnetostrictive materials.

In general, the errors that result from using uncoupled models, as opposed to coupled ones, are problem dependent. There are some cases where very large differences exist in situations where an uncoupled model is used over a coupled model [11]. In this work, the coupled case is analyzed with both linear and nonlinear models. In the linear-coupled model, magnetomechanical coefficient, elasticity matrix, and permeability matrix are assumed as constants. In the nonlinear-coupled model, mechanical and magnetic nonlinearity are decoupled in their respective domains. The nonlinear stress–strain relationship is generally represented by the modulus of elasticity and the nonlinear magnetic flux–magnetic field relationship is represented by the permeability of the material. Magnetomechanical coupling coefficient is assumed as constant in this case.

3.1.1 Linear model

From equation (1) and equation (2), the 3-D constitutive model for the magnetostrictive material can be written as

$$\{\sigma\} = [Q]\{\varepsilon\} - [e]^T\{H\} \tag{3}$$

$$\{B\} = [e]\{\varepsilon\} + [\mu^\varepsilon]\{H\} \tag{4}$$

where $[Q]$ is elasticity matrix, which is the inverse of compliance matrix $[S]$, and $[\mu^\varepsilon]$ is the permeability at constant strain. $[\mu^\varepsilon]$ and $[e]$ are related to $[Q]$ through

$$[e] = [d][Q] \tag{5}$$

$$[\mu^\varepsilon] = [\mu^\sigma] - [d][Q][d]^T \tag{6}$$

For ordinary magnetic materials, where magnetostrictive coupling coefficients are zero, $[\mu^\varepsilon] = [\mu^\sigma]$, the permeability.

Consider a magnetostrictive rod element of length L, area A, with Young's modulus Q. If a tensile force F is applied, the rod develops a strain ε, and hence stress σ. Total strain energy in the rod is

$$V_s = \frac{1}{2}\int \varepsilon\sigma \,\mathrm{d}v = \frac{1}{2}\int \varepsilon\{Q\varepsilon = e\mathbf{H}\}\,\mathrm{d}v$$
$$= \frac{1}{2}(ALQ\varepsilon^2 = ALe\varepsilon\mathbf{H}) \tag{7}$$

Magnetic potential energy in the magnetostrictive rod is

$$V_M = \frac{1}{2}\int B\mathbf{H}\,\mathrm{d}v = \frac{1}{2}\int \{e\varepsilon + \mu\mathbf{H}\}\mathbf{H}\,\mathrm{d}v$$
$$= \frac{1}{2}AL\mathbf{H}e\varepsilon + \frac{1}{2}AL\mu^\varepsilon\mathbf{H}^2 \tag{8}$$

The magnetic and mechanical external work done for N number of coil turns with coil current I is

$$W_m = IN\mu^\sigma\mathbf{H}A, \tag{9}$$

$$W_e = F\varepsilon L \tag{10}$$

The total potential energy of the system is $T_p = -(V_e - W_e) + (V_m - W_m)$

$$T_p = -\frac{1}{2}ALQ\varepsilon^2 + \frac{1}{2}ALe\varepsilon\mathbf{H} + \frac{1}{2}AL\mathbf{H}e\varepsilon$$
$$+ \frac{1}{2}AL\mathbf{H}\mu^\varepsilon\mathbf{H}^2 - IN\mu^\sigma\mathbf{H}A + F\varepsilon L \tag{11}$$

Using Hamilton's principle, two equations in terms of $\mathbf{H}$ and ε can be written as

$$-ALQ\varepsilon + ALe\mathbf{H} + FL = 0 \tag{12}$$

$$ALe\varepsilon + ALe\mathbf{H} + FL = 0 \tag{13}$$

Dividing both equations by AL, the equations will become

$$-Q\varepsilon + e\mathbf{H} = -\frac{F}{A} \tag{14}$$

$$e\varepsilon + \mathbf{H}\mu^\varepsilon = -\frac{F}{A} \tag{15}$$

As the right-hand side of equation (15) is not a function of ε and the left-hand side is magnetic flux density (equation 4), the magnetic flux density in this model is not a function of ε. Hence, it is preserving the flux line continuity. Eliminating $\mathbf{H}$ from equation (14) and substituting this in equation (15), stress–strain relationship of the magnetostrictive material can be written as follows:

$$\mathbf{H} = (Q\varepsilon - F/A)/e \tag{16}$$

$$\varepsilon = \frac{IN\mu^\sigma Ae + L\mu^\varepsilon F}{ALe^2 + AL\mu^\varepsilon Q} = \frac{IN\mu^\sigma eA + F\mu^\varepsilon L}{AL\mu^\sigma Q} \tag{17}$$

From equation (17), the total strain for applied coil current I and tensile stress F/A can be written as

$$\varepsilon = \lambda + \varepsilon\sigma \tag{18}$$

where λ is the strain due to coil current, which is called the *magnetostriction* and ε_σ is the strain due to tensile stress (elastic strain).

$$\lambda = \frac{IN\mu^\sigma Ae}{AL\mu^\sigma A} = \frac{INd}{L} \tag{19}$$

$$\varepsilon_\sigma = \frac{L\mu^\varepsilon F}{AL\mu^\sigma Q} = \frac{F}{AQ^*} \tag{20}$$

Let Q^* be the modified elastic modulus, and substituting the value of e and μ^ε from equation (5)

and equation (6), Q^* will be

$$Q* = \frac{Q\mu^\sigma}{\mu^\varepsilon} = \frac{Q\mu^\sigma}{\mu^\sigma - d^2 Q} = Q + \frac{e^2}{\mu^\varepsilon} \quad (21)$$

If the value of μ^σ is much greater than d^2Q, μ^ε can be assumed as equal to μ^σ and $Q*$ can be assumed as equal to Q. If the value of μ^σ is much greater than d^2Q, the total strain of the rod is the same as for the uncoupled model. The first term in the above expression is the strain due to magnetic field, and the second term is the strain due to the applied mechanical loading. However, for Terfenol-D [2], the value of d^2Q is comparable with μ^σ. Substituting the value of strain from equation (17) in equation (16), the value of magnetic field is

$$\mathbf{H} = \frac{F}{Ae}\left(1 - \frac{\mu^\varepsilon}{\mu^\sigma}\right) + \frac{IN}{L} \quad (22)$$

Note that although the magnetostriction value (INd/L) in equation (19) is the same for coupled and uncoupled cases, the value of magnetic field is different. Let r be the ratio of two permeabilities or two elastic modulii. From equation (21), r can be written as

$$r = \frac{\mu^\sigma}{\mu^\varepsilon} = \frac{Q^*}{Q} \quad (23)$$

If the value of r is 1, the results of coupled analysis are similar with uncoupled analysis. In Figure 4, the value of r is shown in contour plot for different values of constant strain permeability and modulus of elasticity for coupling coefficient of $15 \times 10^{-9}\,\mathrm{m\,A^{-1}}$. In Figure 4(a), the value of r is shown for different values of permeability and elastic modulus. In Figure 4(b), the value of r is shown for different values of permeability and modified elasticity. In these plots, it is clear that for a particular value of elasticity, if the value of permeability increases, the value of r will decrease. However, for a particular value of permeability, if the value of elastic modulus increases, the value of r will increase. In Figure 5, the value of r is shown in contour plot for different value of permeabilities and coupling coefficient, considering the modulus of elasticity as 15 GPa. In Figure 5(a), value of r is given for different values of constant strain permeability and coupling coefficient. In Figure 5(b), the value of r is shown for different values of constant stress permeability and coupling coefficients. In these plots, it is clear that for a particular value of permeability, if the value

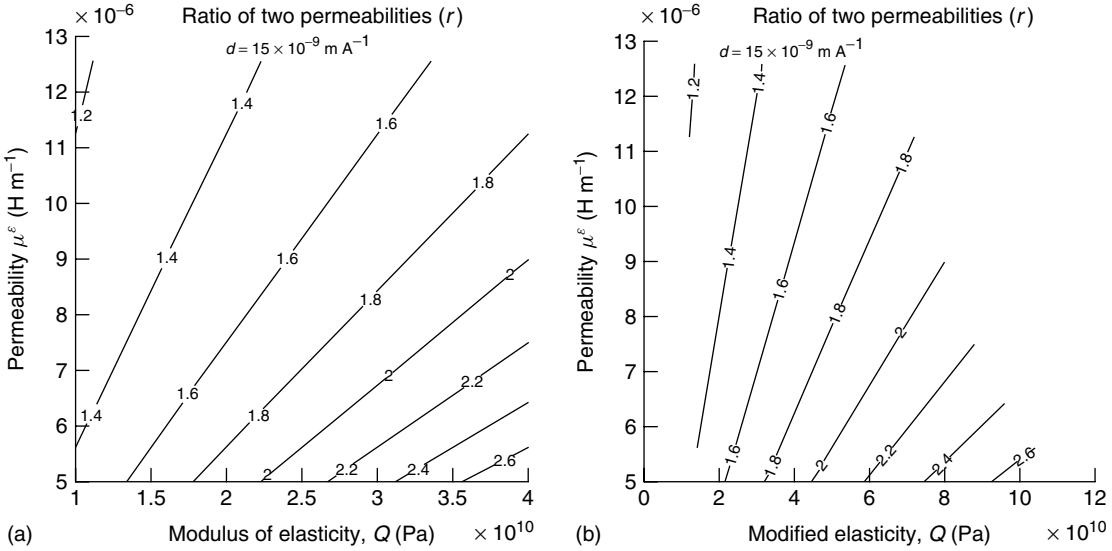


Figure 4. Ratio of two permeabilities (r): permeability versus modulus of elasticity (a) and permeability versus modified elasticity (b), for $d = 15 \times 10^{-9}\,\mathrm{m\,A^{-1}}$.

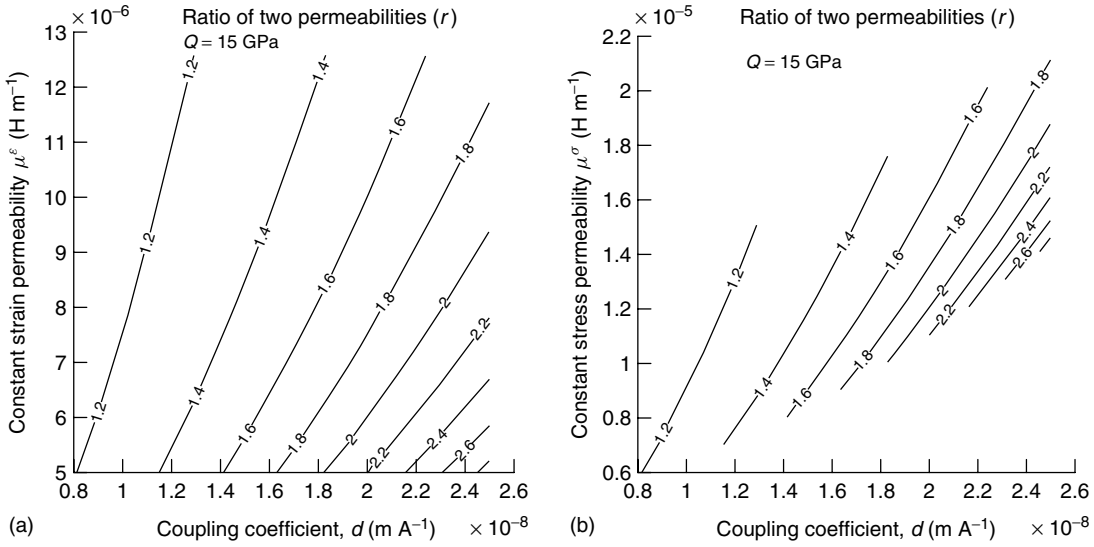


Figure 5. Ratio of two permeabilities (r): permeability versus coupling coefficients (a) and constant stress versus coupling coefficients (b), for $Q = 15$ Gpa.

of coupling coefficient increases the value of r will increase. However, for a particular value of coupling coefficient, if the value of permeability increases the value of r will decrease.

In Figure 6, the value of r is shown in contour plot for different value of elastic modulus and coupling coefficient considering constant strain permeability as $7 \times 10^{-6}\,\mathrm{H\,m^{-1}}$. In Figure 6(a), the value of r is

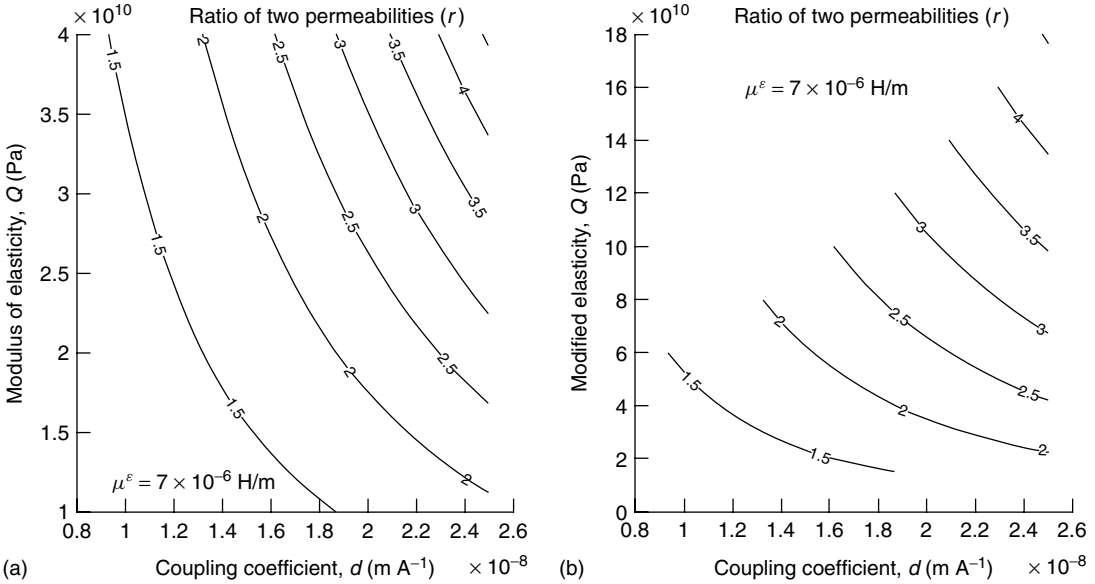


Figure 6. Ratio of two permeabilities (r): modulus Q versus coupling coefficient (a) and modulus Q^* versus coupling coefficient (b), for $\mu = 7 \times 10^{-6}\,\mathrm{H\,m^{-1}}$.

given for different values of modulii of elasticity and coupling coefficients. In Figure 6(b), the value of r is shown for different values of modulii of elasticity and coupling coefficients. In these plots, it is clear that for a particular value of elastic modulus, if the value of coupling coefficient increases, the value of r will increase. Similarly, for a particular value of coupling coefficient, if the value of elastic modulus increases, the value of r will increase.

From the experimental data given in Etrema manual [2], the best value of Q, μ^{σ}, and d is calculated by minimizing the difference between the experimental data and the data according to equation (17), by least square approach. In the first set of experimental data, only magnetostriction values were reported, which is given in equation (19). The value of coupling coefficient is calculated by minimizing the total square error λ_{error} as given by

$$\lambda_{\text{error}} = \sum \left(\lambda_{\text{exp}} - \lambda\right)^2 \tag{24}$$

Similarly, in the second set of experimental data, the strain due to compressive stress (ε_{σ}) was reported. The expression for the value of elastic strain ε_{σ} is given in equation (20). In this expression, the value of Q^* is calculated by minimizing the total square error $\varepsilon_{\sigma}^{\text{error}}$:

$$\varepsilon_{\sigma}^{\text{error}} = \sum \left(\varepsilon_{\sigma}^{\text{exp}} - \varepsilon_{\sigma}\right)^2 \tag{25}$$

From equation (24), using the first set of experimental data (plotted in Figure 2), the value of d was calculated as 14.8×10^{-9} (m A^{-1}). From equation (25), using the second set of experimental data (plotted in Figure 3), the value of Q^* *is* 33.4 GPa. Assuming that the constant strain permeability (μ^{ε}) of the material is $7 \times 10^{-6}\,\text{H m}^{-1}$, the value of r is 1.6, constant stress permeability (μ^{σ}) is $11.2 \times 10^{-6}\,\text{H m}^{-1}$, and Q is 20.8 GPa. From these studies, it is clear that for a giant magnetostrictive material, like Terfenol-D, the coupled analysis gives better result than the uncoupled analysis. However, for a magnetostrictive material with low coupling coefficient, the uncoupled analysis gives a similar result as the coupled analysis.

The coupled-linear model cannot model the high nonlinearity of magnetostriction λ, which is required for the design of actuators. Even considering nonlinear magnetic (magnetic field–magnetic flux) and mechanical (stress–strain) relationships with linear-coupling coefficient, nonlinear relationships of magnetostriction cannot be modeled, as it is a function of coil current, coil turns per unit length of actuator, and magnetomechanical coefficient (equation (19)). In the next section, we introduce a nonlinear model with a constant coupling coefficient, which can model the nonlinear constitutive model exactly for constant magnetic coupling.

3.1.2 Nonlinear-coupled model

The model developed in this section is based on a coupled magnetomechanical formulation, which allows accurate prediction of both the mechanical and magnetic responses of a magnetostrictive device with nonlinear magnetic and mechanical properties. Nonlinearity in this model is introduced using two nonlinear curves, one for the stress–strain relationship and the second for magnetic field–magnetic flux relation, which enables to decouple the nonlinearity in the mechanical and magnetic domains. Magnetomechanical coefficient is considered as a real parameter scalar value. Two-way coupled magnetomechanical theory is used to model magnetostrictive material. The formulation starts with the constitutive relations.

In the earlier linear-coupled model, stress (σ) and magnetic flux density (**B**) were expressed as a function of the components of strain (ε) and magnetic field (**H**) as per equation (3) and equation (4), respectively. A main drawback with such an approach is that the nonlinearity between magnetic domains (μ) and mechanical domains (Q) is not uncoupled. Hence, it is difficult to model nonlinearity in earlier representations. To address these issues, a different approach is used in which equation (3) and equation (4) are rearranged in terms of the mechanical strain (ε) and the magnetic flux density (**B**). In doing so, the mechanical nonlinearity is limited to the stress–strain relationship and the magnetic nonlinearity is limited to magnetic field and magnetic flux relationship. One-dimensional nonlinear modeling is again studied using one-dimensional experimental data from Etrema manual. The constitutive equation can now be rewritten in terms of magnetic flux density (**B**) and strain (ε), as

$$\sigma = E\varepsilon - f^T\mathbf{B} \tag{26}$$

$$H = -f\varepsilon + g\mathbf{B} \tag{27}$$

where

$$g = (\mu^{\varepsilon})^{-1}$$
$$f = gdQ = e/\mu^{\varepsilon}$$
$$E = Q + Qdf = Q^{*} \tag{28}$$

Like the linear case, considering a magnetostrictive rod element of length L, area A, applied tensile force F, strain ε, stress σ, and elastic modulus E, the total strain energy in the rod is

$$V_e = \frac{1}{2}AL\varepsilon\sigma = \frac{1}{2}\varepsilon(E\varepsilon - f\mathbf{B})$$
$$= \frac{1}{2}ALE\varepsilon^2 - \frac{1}{2}AL\varepsilon f\mathbf{B} \tag{29}$$

Magnetic potential energy in the magnetostrictive rod is

$$V_m = \frac{1}{2}AL\mathbf{B}\mathbf{H} = \frac{1}{2}AL(-F\varepsilon + b\mathbf{B})\mathbf{H}$$
$$= -\frac{1}{2}AL\mathbf{B}f\varepsilon + \frac{1}{2}ALg\mathbf{B}^2 \tag{30}$$

Magnetic external work done for N coil turns with coil current I is

$$W_m = IN\mathbf{B}A \tag{31}$$

Mechanical external work done is

$$W_E = F\varepsilon L \tag{32}$$

Total potential energy of the system is equal to $T_p = -V_e - V_m + W_m + W_e$, which in expanded form becomes

$$T_p = -\frac{1}{2}ALE\varepsilon^2 + ALf\mathbf{B} - \frac{1}{2}ALg\mathbf{B}^2$$
$$+ IN\mathbf{B}A + F\varepsilon L \tag{33}$$

Using Hamilton's principle, we get two equations for $\mathbf{B}$ and ε as

$$-ALE\varepsilon + ALf\mathbf{B} + FL = 0 \tag{34}$$

$$ALE\varepsilon f + ALg\mathbf{B} + INA = 0 \tag{35}$$

Dividing by volume AL, equations (34) and (35) become

$$E\varepsilon - f\mathbf{B} = \frac{F}{A} \tag{36}$$

$$-f\varepsilon + g\mathbf{B} = \frac{F}{A} \tag{37}$$

Eliminating $\mathbf{B}$ from equation (36) and substituting this in equation (37), the stress–strain relationship for the magnetostrictive material can be obtained as

$$\mathbf{B} = \frac{E\varepsilon - F/A}{f} \tag{38}$$

$$\varepsilon = \frac{F/A + INf\,(gL)}{E - f^2/g} \tag{39}$$

Assuming E^* as the magnetically free elastic modulus

$$E^* = E - f^2/g - Q \tag{40}$$

Total strain for applied coil current I and tensile force F is given as

$$\varepsilon = \frac{INf}{(gLE^*)} + \frac{F}{AE^*} \tag{41}$$

Here $E = Q^*$ is the elastic modulus for a magnetically stiffened rod and $Q = E^*$ is for magnetically flexible rod. Magnetically stiffened means that the magnetic flux $\mathbf{B} = \mathbf{0}$ inside the rod as the rod is wound by short-circuited coils. Magnetically flexible means the rod is free from any coil. E–Q relation can be obtained form equation (28). To model the one-dimensional nonlinear magnetostrictive stress–strain and magnetic field–magnetic flux relationships, equations (26) and (27) can be written as

$$E(\varepsilon) - f\mathbf{B} = \sigma \tag{42}$$

$$-f\varepsilon + g(\mathbf{B}) = \frac{IN}{L} \tag{43}$$

where f is the real parameter of scalar value and $\varepsilon - E\ (\varepsilon)$, $\mathbf{B} - g(\mathbf{B})$ are two real parameter nonlinear curves. The basic advantage of this model is that only two nonlinear curves are required for representing nonlinearity reported in different stress levels. As

opposed to this approach, in straightforward polynomial representation of magnetostriction [7], one requires single nonlinear curve for every stress level. To get the coefficients of two nonlinear curves and the value of real parameter f, experimental data from Etrema manual [2] is used. From strain, applied coil current, and stress level available in the manual, these coefficients are evaluated. Considering modulus elasticity as 30 GPa, and f as $75.3 \times 106\,\mathrm{m\,A^{-1}}$ as an initial guess, the values of magnetic flux density $\mathbf{B}$ are calculated from equation (42). Similarly, from equation (43), values of $g(\mathbf{B})$ are evaluated. From these values of $\mathbf{B}$ and $g(\mathbf{B})$, the curve of $\mathbf{B} - g(\mathbf{B})$ is computed.

This curve is used to get the mechanical relationship. Here, the value of $\mathbf{B}$ is computed from the $\mathbf{B} - g(\mathbf{B})$ relationship. From this value of $\mathbf{B}$, using equation (42), value of $E(\varepsilon)$ is calculated. From the $E(\varepsilon)$ and ε values, the mechanical nonlinear curve of $\varepsilon - E(\varepsilon)$ relationship is computed. In summary, first the magnetic nonlinear curve is evaluated from the mechanical nonlinear curve and the mechanical nonlinear curve is evaluated from the magnetic nonlinear curve with the help of experimental data given in the Etrema manual [2]. This iteration continues till both the mechanical curve and magnetic curve converge. Thus, with the help of experimental data given in the Etrema manual [2] and equations (42) and (43), nonlinear mechanical and magnetic relationships are evaluated. Initial values of modulus of elasticity and f are computed on a trial and error basis.

For sensor device, where coil current is assumed as zero, strain and the value of magnetic flux due to the application of stress are given by

$$\mathbf{B} = \frac{E(\varepsilon) - \sigma}{f} \tag{44}$$

$$\varepsilon = \frac{g(\mathbf{B})}{f} \tag{45}$$

The magnetic field is approximated by a sixth-order polynomial of magnetic flux density and the modulus is approximated by a sixth-order polynomial of the strain, which is given by

$$g(\mathbf{B}) = c_5 \times \mathbf{B}^5 + \cdots + c_1 \times \mathbf{B} + c_0 \tag{46}$$

$$\mathbf{B} = d_5 \times g(\mathbf{B})^5 + \cdots + d_1 \times g(\mathbf{B}) + d_0 \tag{47}$$

$$E(\varepsilon) = a_6 \times \varepsilon^6 + \cdots + a_1 \times \varepsilon + a_0 \tag{48}$$

$$\varepsilon = b_6 \times E(\varepsilon)^6 + \cdots + b_1 \times E(\varepsilon) + b_0 \tag{49}$$

These curves are shown in Figure 7. Coefficients of these polynomial curves are given in Table 1, where

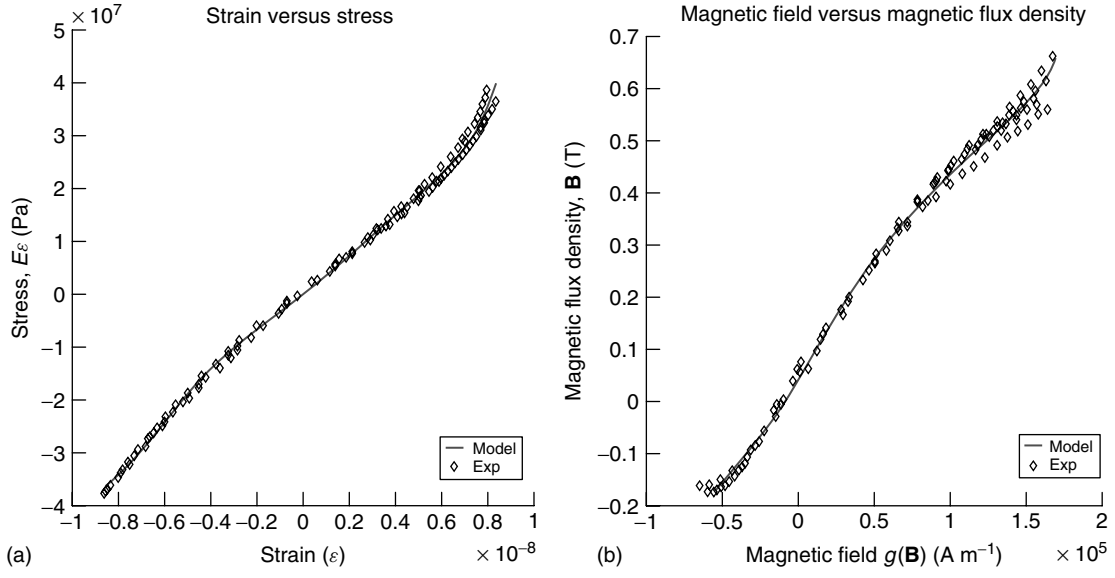


Figure 7. Nonlinear stress–strain curves (a) and magnetic flux and magnetic field curves (b).

Table 1. Coefficients for sixth-order polynomial

	c	*d*	*a*	*b*
6	0	0	4.5419e+28	1.5853e−50
5	−2.1687e+06	−1.9526e−27	7.6602e+25	−6.1288e−44
4	1.5211e+06	1.1589e−21	−4.2662e+22	−1.0355e−34
3	3.5828e+05	−2.0047e−16	−6.6788e+19	−4.0508e−27
2	−2.1062e+05	2.0096e−12	2.3911e+16	1.0806e−27
1	2.2754e+05	4.7789e−06	1.2539e+13	2.7977e−11
0	−8.8129e+03	4.4239e−02	3.3893e+10	−1.9704e−05

the unit of **B** is T, $g(\mathbf{B})$ is $\mathrm{A\,m^{-1}}$, and $E(\varepsilon)$ is Pa. The value of magnetomechanical coupling parameter (f) is $75.3 \times 106\,\mathrm{m\,A^{-1}}$, which is the reciprocal of $13.3 \times 10^{-9}\,\mathrm{A\,m^{-1}}$.

On the basis of the two curves given by equations (46) and (48) and parameter (f), strain and magnetostriction versus applied magnetic field for different stress levels are plotted in Figure 8. The experimental data of strain–magnetic field relationships for different stress levels almost match with this model. Similarly, strain-compressive force and elastic modulus for different magnetic field levels are plotted in Figure 9. Elastic modulus initially decreases and then increases for each magnetic field level, which is also reported in Butler [2]. The magnetic flux, strain, stress, and coil current are computed as follows. As two nonlinear curves are related in these relationships, the calculation of magnetic flux and strain from stress and coil current is an iterative procedure. Initially, the value of magnetic flux **B** is assumed as a certain value. From curve $\mathbf{B} - g(\mathbf{B})$, the value of $g(\mathbf{B})$ is evaluated. From equation (37), the value of strain is evaluated considering the magnetic field as coil current times coil turns per unit length of the actuator. Using this strain, from the $\varepsilon - E(\varepsilon)$ curve the value of $E(\varepsilon)$ can be found. From equation (36), the value of **B** can be determined. If this **B** value is not the same as assumed, this iteration will be continued until the value converges.

3.1.3 ANN model

To avoid the iterative procedure mentioned for the nonlinear model, one three-layer ANN is developed,

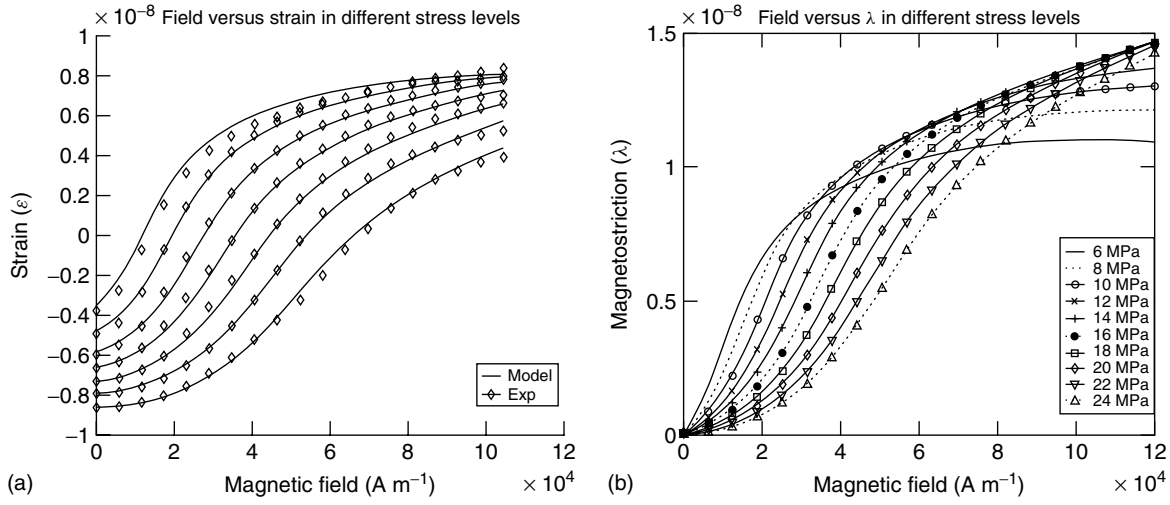


Figure 8. Nonlinear strain–magnetic field curves (a) and magnetostriction–magnetic field curves (b).

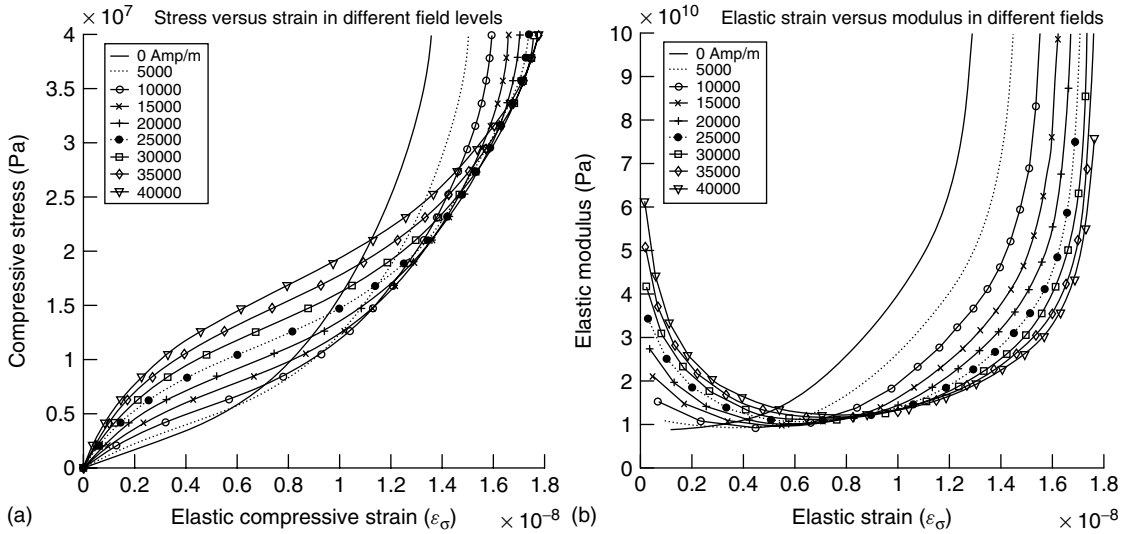


Figure 9. Nonlinear stress–strain curves (a) and elastic modulus–elastic strain curves (b).

which gives direct nonlinear mapping from magnetic field and stress to magnetic flux density and strain. Standard logistic function $y = 1/(1 + e^{-1.7159v})$ is used in hidden layer as activation function with linear output layer. Input (stress and magnetic field) and output (strain and magnetic flux density) data is normalized for a better performance of network.

$$\sigma_n = \frac{(\sigma - \sigma_{\text{mean}})}{(\max|\sigma| - \sigma_{\text{mean}})} \tag{50}$$

$$\mathbf{H}_n = \frac{(\mathbf{H} - \mathbf{H}_{\text{mean}})}{(\max|\mathbf{H}| - \mathbf{H}_{\text{mean}})} \tag{51}$$

$$\varepsilon_n = \frac{(\varepsilon - \varepsilon_{\text{mean}})}{(\max|\varepsilon| - \varepsilon_{\text{mean}})} \tag{52}$$

$$\mathbf{B}_n = \frac{(\mathbf{B} - \mathbf{B}_{\text{mean}})}{(\max|\mathbf{B}| - \mathbf{B}_{\text{mean}})} \tag{53}$$

The normalized stress, σ_n, is calculated using equation (50). The value of σ_{mean} and $(\max|\sigma| - \sigma_{\text{mean}})$ are -1.57966×10^7 and 0.830345×10^8 Pa, respectively. Similarly, $\mathbf{H}^n$, the normalized magnetic field, is calculated from equation (51). The values of $\mathbf{H}_{\text{mean}}$ and $(\max|\mathrm{H}| - \mathbf{H}_{\text{mean}})$ in equation (51) are both 750 Oe. Normalized strain ε_n is the output of network, from which strain is calculated using equation (52). The value of $\varepsilon_{\text{mean}}$ and $(\max|\varepsilon| - \varepsilon_{\text{mean}})$ are 0.203245×10^{-3} and 0.106504×10^{-2}, respectively. In equation (53), the value of $\mathbf{B}_{\text{mean}}$ is 0.385126 T and the value of $(\max|B| - \mathbf{B}_{\text{mean}})$ is 0.319818 and 0.499656 T, respectively. To train this network, some training and validation samples are generated through the iterative process stated earlier. Weight and bias parameter of the trained network are given in Tables 2 and 3. Different validation studies are also carried out.

Table 2. Connection between input layer and hidden layer

Input layer Neurons	Hidden Node 1	Hidden Node 2	Hidden Node 3	Hidden Node 4
$\mathbf{H}_n$	4.6187	−2.1241	−0.15066	0.38726
σ_n	1.2866	−1.3195	−0.43318	0.031765
Input bias	2.2483	−0.25721	−0.61579	0.088719

Table 3. Connection between hidden layer and output layer

Output layer neurons	ε_n	$\mathbf{B}_n$
Hidden node 1	0.63447	0.67409
Hidden node 2	−0.53313	−0.23172
Hidden node 3	−0.69546	−0.071654
Hidden node 4	0.48235	2.1900
Output bias node	−0.29930	−1.5467

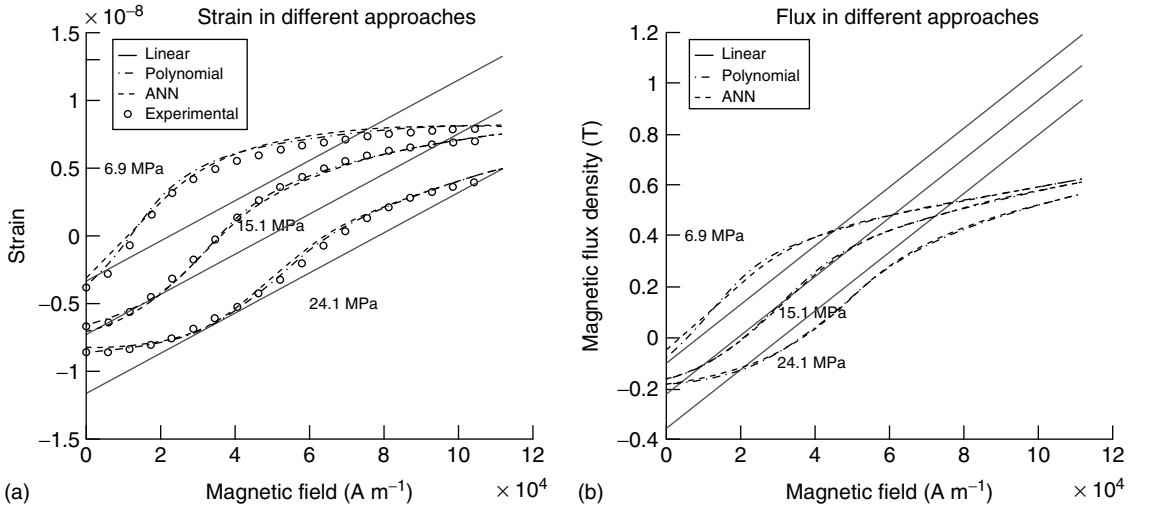


Figure 10. Comparison of different models: strain–magnetic field curves (a) and flux density–magnetic field curves (b).

3.2 Comparison between different coupled models

A comparative study of different models is done using a magnetostrictive rod with varying magnetic field and stress level. Three different stress levels (6.9, 15.1, and 24.1 MPa) are used to compute the total strain and magnetic flux density in the rod for varying magnetic field levels and shown in Figure 10. In Figure 10(a), total strain is shown according to linear, polynomial, ANN, and experimental approaches. Both polynomial and ANN approaches show close results with the experimental data throughout the magnetic field range. As in the case of the linear model, the results do not match with the experimental data throughout the magnetic field range. However, this model can be used in low magnetic field level for medium stress level, and in medium field level for high stress level. For a low stress level, the linear model can be used on an average sense. In the absence of experimental data (Etrema manual) of magnetic flux density, only computational results are shown in Figure 10(b). Magnetic flux density is shown according to linear, polynomial, and ANN approaches. Similar to the strain results, the results of the ANN model and polynomial model are in excellent agreement. However, the results of the linear model do not match throughout the magnetic field range. In the linear model, for medium stress level in low magnetic field level, magnetic flux density matches with the nonlinear model. For high and low stress level, the linear model can be used in the average sense.

4 SUMMARY

This article is mainly intended for anhysteretic linear and nonlinear, coupled constitutive relationships of magnetostrictive material. The coupled model is studied without assuming any direct relationship of magnetic field unlike the uncoupled model. In the linear-coupled model, the elastic modulus, permeability, and magnetoelastic constant are considered as constants. However, this model cannot predict the highly nonlinear properties of magnetostrictive material. In the nonlinear-coupled model, nonlinearity is decoupled in the magnetic and the mechanical domain using two nonlinear curves for stress–strain and magnetic flux density and magnetic field intensity. In this model, the computation of magnetostriction requires the value of magnetic flux density, which is obtained through an iterative process for nonlinearity of curves. To avoid this iterative computation, one three-layer ANN is developed, which will give nonlinear mapping from stress level and magnetic field to strain and magnetic flux density. Finally, a comparative study of linear, polynomial, and ANN

approaches is done, which shows that the linear-coupled model can predict the constitutive relationships in an averaged sense only. The nonlinear models are shown to predict experimental results exactly throughout the magnetic field range.

ACKNOWLEDGMENTS

The author wishes to thank his graduate student Dr Debiprasad Ghosh for performing the numerical simulation addressed in this article.

REFERENCES

[1] Villery E. Change of magnetization by tension and by electric current. *Annals of Physical Chemistry* 1865 **126**:87–122.

[2] Butler JL. *Application Manual for the Design of TERFENOL-D Magnetostrictive Transducers*, Technical Report TS 2003. Edge Technologies: Ames, IA, 1988.

[3] Benbouzid MEH, Reyne G, Meunier G. Nonlinear finite element modeling of giant magnetostriction. *IEEE Transactions on Magnetics* 1993 **29**: 2467–2469.

[4] Benbouzid MEH, Kvarnsjo L, Engdahl G. Dynamic modeling of giant magnetostriction in TERFENOL-D rods by the finite element method. *IEEE Transactions on Magnetics* 1995 **31**:1821–1823.

[5] Toupin RA. The elastic dielectric. *Journal of Rational Mechanics and Analysis* 1956 **5**:849–915.

[6] Maugin GA. *Nonlinear Electromechanical Effects and Applications*. World Scientific, 1985.

[7] Krishnamurty AV, Anjanappa M, Wang Z, Chen X. Sensing of delaminations in composite laminates using embedded magnetostrictive particle layers. *Journal of Intelligent Material Systems and Structures* 1999 **10**:825–835.

[8] Rumelhart D, Hinton G, Williams R. Learning representations by back propagation error. *Nature* 1986 **323**:533–536.

[9] Moffett MB, Clark AE, Wun-Fogle M, Linberg J, Teter JP, McLaughlin EA. Characterization of TERFENOL-D for magnetostrictive transducers. *Journal of the Acoustical Society of America* 1989 **86**: 1448–1455.

[10] Hall D, Flatau A. One-dimensional analytical constant parameter linear electromagnetic magneto-mechanical models of a cylindrical magnetostrictive TERFENOL-D transducer. *Proceedings of the ICIM94: 2nd International Conference on Intelligent Materials*. Williamsburg, VA, 1994; pp. 605–616.

[11] Ghosh DP, Gopalakrishnan S. Structural health monitoring in a composite beam using magnetostrictive material through a new FE formulation. In *Proceedings of SPIE on Smart Materials, Structures, and Systems*, Mohan S, Dattaguru B, Gopalakrishnan S (eds). SPIE: Bellingham, WA, 2002; Vol. 5062, pp. 704–711.

FURTHER READING

Anjanappa M, Bi J. Magnetostrictive mini actuators for smart structures applications. *Smart Materials and Structures* 1994 **3**:383–390.

Pelinescu I, Balachandran B. Analytical study of active control of wave transmission through cylindrical struts. *Smart Materials and Structures* 2001 **10**:121–136.

Reddy JN, Barbosa JI. On vibration suppression of magnetostrictive beams. *Smart Materials and Structures* 2000 **9**:49–58.

Roy Mahapatra DP, Gopalakrishnan S, Balachandran B. Active feedback control of multiple waves in helicopter gearbox support struts. *Smart Materials and Structures* 2001 **10**:1046–1058.

Saidha E, Naik GN, Gopalakrishnan S. An experimental investigation of a smart laminated composite beam with magnetostrictive patch for health monitoring applications. *Structural Health Monitoring* 2003 **2**:273–292.

Chapter 43

Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators

Srinivasan Gopalakrishnan

Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

1 INTRODUCTION

Finite element (FE) is a powerful numerical tool to model any complex system that is governed by a partial differential equation. Modeling and analysis of structures through FE procedures are now quite well established and matured to such an extent that there are many general-purpose FE codes, commercially available, that satisfy most industrial standards specified for modeling, analysis, and design. With increasing emphasis on incorporating health monitoring concepts at the design stage itself for the present-day systems, it is necessary to build a new or modify the existing FE code to handle this additional complexity. One of the ways or means to incorporate the health monitoring concepts at the design stage itself is by introducing smart material patches anywhere in the structures (if the structure is of laminated composite construction), or select a proper location to surface bond these smart material patches. This article deals with the FE modeling of such structures with built-in smart material patches. However, this article only gives a brief outline of how some FEs are formulated with smart degrees of freedom. Mathematical aspects of FEs, detailed method of formulation of FEs, method of solution of FE equations, etc. can be obtained from many classic text books in [1–3]. Some aspects of modeling of smart sensors and actuators using FEs can also be found in [4].

The main function of the smart material patches is to introduce additional functionalities, namely, the sensing and actuation functions. Sensing function is normally used in the health monitoring application, whereas in applications such as vibration or noise control, actuation mechanism is used. Such functionalities are possible using smart material patches due to the presence of coupling parameter in their constitutive model. Some of the smart materials

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

have two different constitutive laws: one is used for sensing while the other is used for actuation purpose. In piezoceramic material, for example, which is a class of smart material used extensively in structural application, the constitutive law couples the mechanical motion and the electrical field through an electromechanical coupling coefficient. Some of the elementary concepts of behavior of smart materials in relation to structures are given in [5–7]. From FE point of view, this additional complexity arising owing to this coupling results in additional degrees of freedom and hence increased size of the system matrices. However, the general procedure for the formulation of FEs and their synthesis remains the same and these are covered extensively in many classic text books in [1–3].

The general procedure for the formulation of FEs involves the following:

- Develop a suitable constitutive model for the combined material consisting of the host material (say metallic or laminated composite structure) and the smart material patch. The constitutive matrix for metallic isotropic structure is quite well known. A laminated composite normally exhibits anisotropic behavior due to arbitrary orientation of the fiber. The constitutive law for composites can be found in [8, 9]. When the composite constitutive law is coupled with the constitutive law of the smart material (say piezoceramic), the resulting matrix, not only couples the different mechanical motions but also the smart degrees of freedom.
- Formulate the weak form of the governing differential equation. This will involve the energy expressions due to elasticity, inertia, surface and body forces, and external forces. When smart material patches are introduced (say piezoceramic or magnetostrictive material), then the energy due to electrical or magnetic motion should also be included in the weak form of the equations.
- Decide on the number of degrees of freedom and assume a suitable interpolating function for each of the degrees of freedom. Note that if the structure has smart material patches, then the smart degree of freedom (say electrical degree of freedom for piezoceramic material) also needs to be introduced and interpolated over an element using a suitable function, as is done for the other degrees of freedom.
- These interpolating functions are then substituted for the different field variable in the weak form of the equation and the resulting equation is minimized to obtain the discritized equation of motion, which involves computing the stiffness, the mass, and the damping matrices. The discritized governing equations are solved using many standard methods, the details of which can be obtained from many standard FE text books.

This article is organized as follows. The next section explains briefly the constitutive law for two different classes of smart materials, namely, the piezoceramic material and magnetostrictive materials. Although the detailed treatment of their constitutive models are dealt in detail in earlier articles (*see* **Piezoceramic Materials—Phenomena and Modeling**; **Constitutive Modeling of Magnetostrictive Materials**), here, only those details that are required for FE formulation are given. This section is followed with a detailed derivation of constitutive law for a laminated composite with embedded piezoelectric material patch. This method then extends to piezo fiber composites (PFCs). Next, the detailed 1-D and 2-D FE formulation with embedded smart senor/actuator patches are outlined, which is followed by some numerical examples.

2 CONSTITUTIVE MODELS FOR SMART SENSORS AND ACTUATORS

2.1 Piezoelectric sensors/actuators

Piezoelectric or magnetostrictive materials have two constitutive laws, one of which is used for sensing and the other for actuation purposes. For 2-D state of stress, the constitutive model for the piezoelectric material is of the form

$$\{\sigma\}_{3\times1} = [C]^{(E)}_{3\times3}\{\varepsilon\}_{3\times1} - [e]_{3\times2}\{E\}_{2\times1} \quad (1)$$

$$\{D\}_{2\times1} = [e]^{\mathrm{T}}_{2\times3} + [\mu]^{(\sigma)}_{2\times2}\{E\}_{2\times1} \quad (2)$$

The first of this constitutive law is called *the actuation law*, while the second is called *the sensing law*. Here, $\{\sigma\}^{\mathrm{T}} = \{\,\sigma_{xx} \quad \sigma_{yy} \quad \tau_{xy}\,\}$ is the stress vector; $\{\varepsilon\}^{\mathrm{T}} = \{\,\varepsilon_{xx} \quad \varepsilon_{yy} \quad \gamma_{xy}\,\}$ is the strain vector;

$[e]$ is the matrix of piezoelectric coefficients of size 3×2, which has a unit of N/V-mm; and $\{E\}^{\mathrm{T}} = \{E_x \quad E_y\} = \{V_x/t \quad V_y/t\}$ is the applied field in two coordinate directions, where V_x and V_y are the applied voltages in the two coordinate directions, and t is the thickness parameter. It has a unit of V/mm. $[\mu]$ is the permittivity matrix of size 2×2, measured at constant stress and has a unit of N/V/V and $\{D\}^{\mathrm{T}} = \{D_x \quad D_y\}$ is the vector of electric displacement in two coordinate directions. This has a unit of N/V-mm. $[C]$ is the mechanical constitutive matrix measured at constant electric field. Equation (1) can also be written in the form

$$\{\varepsilon\} = [S]\{\sigma\} + [d]\{E\} \tag{3}$$

In the above expression, $[S]$ is the compliance matrix, which is the inverse of the mechanical material matrix $[C]$ and $[d] = [C]^{-1}[e]$ is the electromechanical coupling matrix, where the elements of this matrix have a unit mm/V and the elements of this matrix are direction dependent. In most analyses, it is assumed that the mechanical properties will change very little with the change in the electric field and, as a result, the actuation law (equation 1) can be assumed to behave linearly with the electric field, while the sensing law (equation 2) can be assumed to behave linearly with the stress. This assumption will considerably simplify the process of analyses.

The first part of equation (1) represents the stresses developed due to mechanical load, while the second part of the same equation gives the stresses due to voltage input. From equations (1) and (2), it is clear that the structure is stressed due to the application of electric field even in the absence of mechanical load. Alternatively, when the mechanical structure is loaded, it generates an electric field. In other words, the above constitutive law demonstrates the electromechanical coupling, which is exploited for a variety of structural applications, such as vibration control, noise control, shape control, or structural health monitoring. Piezoceramic materials are a class of piezoelectric materials, which are available in different forms, namely, ceramic, polymer, or crystal forms. The most commonly used piezoceramic material is the PZT (lead–zirconate–titanate) material, which is extensively used as bulk actuator material as they have high electromechanical coupling factor. Owing to the low electromechanical coupling factor, piezo polymers polyvinylidene fluoride (PVDF) are extensively used only as sensor material. With the advent of smart composite structures, a new brand of material called *piezofiber composite* is found to be a very effective actuator material for use in vibration/noise control applications. The form of the constitutive model of this material is very similar to that of the piezoelectric material and since this is always a part of the composite structure, its constitutive model is established in this context at a later part of this article.

2.2 Magnetostrictive sensors/actuators

Some magnetic materials (magnetostrictive) show elongation and contraction in the magnetization direction due to an induced magnetic field. This is called *the magnetostriction*, which is due to the switching of a large amount of magnetic domains caused by spontaneous magnetization, below the Curie point of temperature. Thus, magnetostrictive materials have the ability to convert magnetic energy into mechanical energy and vice versa. This coupling between magnetic and mechanical energies represents what is called *the transudation capability*, which allows a magnetostrictive material to be used in both actuation and sensing applications. The most extensively used magnetostrictive material for structural applications is the Terfenol-D. The constitutive laws (both actuation and sensing) for magnetostrictive materials such as Terfenol-D are much more complex than the piezoelectric materials. They are highly nonlinear and have the similar form as that of the piezoelectric material, which is given by

$$\{\varepsilon\} = [S]^{(H)}\{\sigma\} + [d]^{\mathrm{T}}\{H\} \tag{4}$$

$$\{B\} = [d]\{\sigma\} + [\mu]^{\{(\sigma\}}\{H\} \tag{5}$$

Here, $[S]$ is the compliance matrix measured at constant magnetic field H; $[d]$ is the magnetomechanical coupling matrix, the elements of which has units of meter per ampere; and $\{B\}$ is the vector of magnetic flux density in the two coordinate directions. It has a unit called *tesla*, which is equal to weber per cubic meter. $\{H\}$ is the magnetic field intensity vector in the two coordinate directions and it has a unit called *oersted,* which is equal to *Ampere per meter.* It is related to the ac current $(I(t))$ through the relation $H = nI$, where n is the number of turns in the actuator. $[\mu]$ is the matrix of magnetic permeability

measured at constant stress and it has a unit of weber per ampere-meter. As in the case of piezoelectric material, the first equation (equation 4) is the actuation constitutive law, while the second equation (equation 5) is the sensing law. The stress–strain relations are different for different magnetic field intensities. The strain is linear to stress only for small magnetic field intensities. For higher magnetic field intensities, both sensing and actuator equations require to be simultaneously solved to arrive at the correct stress–strain relation. This is because change in the magnetic field changes the stress, which changes the magnetic permeability. Hence, characterization of material properties of Terfenol-D is more difficult compared to the piezoelectric material. These aspects were clearly highlighted in the previous article on constitutive model for magnetostrictive material. The reader is also encouraged to refer [8, 9] for the constitutive model of the above material.

In this article, we assume only the linear behavior of these materials and proceed with FE modeling of these smart sensors and actuators based on this assumption. Here, the FE modeling of some of the 1-D and 2-D structures with both piezo and magnetostrictive material patches embedded in a composite material is given.

3 CONSTITUTIVE MODEL FOR A COMPOSITE MATERIAL WITH EMBEDDED PIEZOELECTRIC SENSORS/ACTUATORS

Fundamental to any FE modeling is to first establish the constitutive model and this also holds true for a 3-D laminate with embedded piezoelectric sensors/actuators. Here, we can take the same approach as undertaken to establish the constitutive model for conventional composite structure. That is, we first establish the constitutive model at the lamina level in the fiber coordinate system, which is then transformed to the global coordinate system. These relations are then synthesized for all the laminae to establish the constitutive model of the laminate. However, additional matrices arise in this case due to the presence of electromechanical coupling. Consider a lamina with the piezoelectric layer as shown in Figure 1. The constitutive model in directions 1, 2, and 3 for such a lamina is given by equations (1) and (2), respectively. In matrix form, it is given by

$$\begin{Bmatrix} \{\sigma\} \\ \{D\} \end{Bmatrix} = \begin{bmatrix} [C] & -[e] \\ [e]^{\mathrm{T}} & [\mu] \end{bmatrix} \begin{Bmatrix} \{\varepsilon\} \\ \{E\} \end{Bmatrix} \tag{6}$$

$$\text{or } \{\overline{\sigma}\} = [\overline{C}]\{\overline{\varepsilon}\} \tag{7}$$

Expanding the above relation, we get

$$\begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{12} \\ D_1 \\ D_2 \\ D_3 \end{Bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 & 0 & 0 & -e_{31} \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 & 0 & 0 & -e_{32} \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 & 0 & 0 & -e_{33} \\ 0 & 0 & 0 & C_{44} & 0 & 0 & 0 & -e_{24} & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 & -e_{15} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e_{15} & 0 & \mu_{11} & 0 & 0 \\ 0 & 0 & 0 & e_{24} & 0 & 0 & 0 & \mu_{22} & 0 \\ e_{31} & e_{32} & e_{33} & 0 & 0 & 0 & 0 & 0 & \mu_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ E_1 \\ E_2 \\ E_3 \end{Bmatrix} \tag{8}$$

Here, $E_i = -\nabla\Phi$, where Φ is the electric potential vector. The above constitutive model is then transformed to the global $x-y-z$ coordinate system using

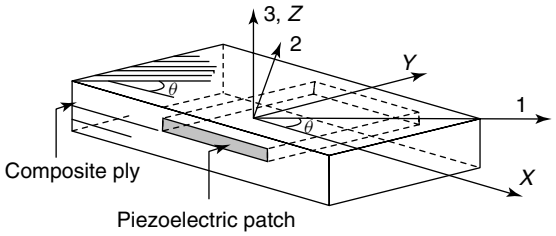


Figure 1. Local and global coordinate system for a lamina with embedded piezoelectric patch.

the transformation matrix, which is given by

$$[T] = \begin{bmatrix} [T_{11}] & [0] \\ [0] & [T_{22}] \end{bmatrix} \tag{9}$$

where

$$[T_{11}] = \begin{bmatrix} C^2 & S^2 & 0 & 0 & 0 & -2CS \\ S^2 & C^2 & 0 & 0 & 0 & 2CS \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & C & S & 0 \\ 0 & 0 & 0 & S & C & 0 \\ CS & -CS & 0 & 0 & 0 & C^2 - S^2 \end{bmatrix} \tag{10}$$

$$[T_{22}] = \begin{bmatrix} C^2 & S^2 & 0 \\ S^2 & C^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

$$C = \cos(\theta),\ S = \sin(\theta) \tag{12}$$

Here θ is the fiber orientation of the lamina. The constitutive model in the global $x-y-z$ direction is then given by

$$\begin{aligned} \{\sigma\} &= [T]^{\mathrm{T}} \begin{bmatrix} [C] & -[e] \\ [e]^{\mathrm{T}} & [\mu] \end{bmatrix} [T]\{\varepsilon\} \\ &= \begin{bmatrix} [\overline{\overline{C}}] & -[\overline{\overline{e}}] \\ [\overline{\overline{e}}] & [\mu] \end{bmatrix} \{\varepsilon\} \end{aligned} \tag{13}$$

In expanded form, the above equation becomes

The elements of $[\overline{\overline{C}}]$ and $[\overline{\overline{e}}]$ are given by

$$\begin{aligned}
\overline{\overline{C}}_{11} &= 4C_{66}C^2S^2 + C^2(C_{11}C^2 + C_{12}S^2) \\
&\quad + S^2(C_{12}C^2 + C_{22}S^2) \\
\overline{\overline{C}}_{12} &= -4C_{66}C^2S^2 + S^2(C_{11}C^2 + C_{12}S^2) \\
&\quad + C^2(C_{12}C^2 + C_{22}S^2), \\
\overline{\overline{C}}_{13} &= C_{13}C^2 - C_{23}S^2 \\
\overline{\overline{C}}_{16} &= -2C_{66}CS(C^2 - S^2) \\
&\quad + CS(C_{11}C^2 + C_{12}S^2) - CS(C_{12}C^2 + C_{22}S^2), \\
\overline{\overline{C}}_{21} &= \overline{\overline{C}}_{12} \\
\overline{\overline{C}}_{22} &= 4C_{66}C^2S^2 + S^2(C_{11}S^2 + C_{12}C^2) \\
&\quad + C^2(C_{12}S^2 + C_{22}C^2), \\
\overline{\overline{C}}_{23} &= C_{23}C^2 + C_{13}S^2 \\
\overline{\overline{C}}_{26} &= 2C_{66}CS(C^2 - S^2) \\
&\quad + CS(C_{11}S^2 + C_{12}C^2) - CS(C_{12}S^2 + C_{22}C^2), \\
\overline{\overline{C}}_{31} &= \overline{\overline{C}}_{13},\ \overline{\overline{C}}_{32} = \overline{\overline{C}}_{23} \\
\overline{\overline{C}}_{33} &= C_{33}, \quad \overline{\overline{C}}_{36} = CS(C_{13} - C_{23}), \\
\overline{\overline{C}}_{44} &= C_{44}C^2 + C_{55}S^2, \\
\overline{\overline{C}}_{45} &= CS(C_{55} - C_{44}), \\
\overline{\overline{C}}_{54} &= \overline{\overline{C}}_{45}, \quad \overline{\overline{C}}_{55} = C_{44}S^2 + C_{55}C^2,
\end{aligned}$$

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{zx} \\ \sigma_{xy} \\ D_x \\ D_y \\ D_z \end{Bmatrix} = \begin{bmatrix} \overline{\overline{C}}_{11} & \overline{\overline{C}}_{12} & \overline{\overline{C}}_{13} & 0 & 0 & 0 & 0 & 0 & -\overline{\overline{e}}_{31} \\ \overline{\overline{C}}_{12} & \overline{\overline{C}}_{22} & \overline{\overline{C}}_{23} & 0 & 0 & 0 & 0 & 0 & -\overline{\overline{e}}_{32} \\ \overline{\overline{C}}_{13} & \overline{\overline{C}}_{23} & \overline{\overline{C}}_{33} & 0 & 0 & 0 & 0 & 0 & -\overline{\overline{e}}_{33} \\ 0 & 0 & 0 & \overline{\overline{C}}_{44} & 0 & 0 & 0 & -\overline{\overline{e}}_{24} & 0 \\ 0 & 0 & 0 & 0 & \overline{\overline{C}}_{55} & 0 & -\overline{\overline{e}}_{15} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \overline{\overline{C}}_{66} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \overline{\overline{e}}_{15} & 0 & \overline{\overline{\mu}}_{11} & 0 & 0 \\ 0 & 0 & 0 & \overline{\overline{e}}_{24} & 0 & 0 & 0 & \overline{\overline{\mu}}_{22} & 0 \\ \overline{\overline{e}}_{31} & \overline{\overline{e}}_{32} & \overline{\overline{e}}_{33} & 0 & 0 & 0 & 0 & 0 & \overline{\overline{\mu}}_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ 2\varepsilon_{yz} \\ 2\varepsilon_{zx} \\ 2\varepsilon_{xy} \\ E_x \\ E_y \\ E_z \end{Bmatrix} \tag{14}$$

$$
\begin{aligned}
&\overline{\overline{C}}_{61} = \overline{\overline{C}}_{16}, \overline{\overline{C}}_{62} = \overline{\overline{C}}_{26}, \quad \overline{\overline{C}}_{63} = \overline{\overline{C}}_{36} \\
&\overline{\overline{C}}_{66} = C_{66}(C^2 - S^2)^2 + C^2S^2(C_{11} - C_{12}) \\
&\qquad - C^2S^2(C_{12} - C_{22}) \\
&\overline{\overline{e}}_{31} = (e_{31}C^2 + e_{32}S^2), \\
&\overline{\overline{e}}_{32} = (e_{31}S^2 + e_{31}S^2), \\
&\overline{\overline{e}}_{33} = e_{33}, \overline{\overline{e}}_{14} = (e_{15}C^2S + e_{24}CS^2), \\
&\overline{\overline{e}}_{24} = (e_{24}C^3 + e_{15}S^3), \quad \overline{\overline{e}}_{15} = (e_{15}C^3 - e_{24}S^3), \\
&\overline{\overline{e}}_{25} = (e_{24}C^2S - e_{15}CS^2), \\
&\overline{\overline{e}}_{36} = CS(e_{31} - e_{32}), \quad \overline{\overline{\mu}}_{11} = \mu_{11}C^4 + \mu_{22}S^4, \\
&\overline{\overline{\mu}}_{12} = C^2S^2(\mu_{11} + \mu_{22}), \\
&\overline{\overline{\mu}}_{22} = \mu_2C^4 + \mu_{11}S^4, \quad \overline{\overline{\mu}}_{33} = \mu_{33}
\end{aligned} \tag{15}
$$

For 2-D analysis, we normally employ either plane stress or plane strain assumptions. For plane stress assumption in the $x-z$ plane, we substitute $\sigma_{yy} = \sigma_{xy} = \sigma_{yz} = D_x = D_y = 0$ in equation (14). Simplifying this, we can write the constitutive model for a 2-D piezoelectric composite as

$$
\begin{Bmatrix} \{\sigma\} \\ D_z \end{Bmatrix} = \begin{bmatrix} [\hat{C}] & -[\hat{e}] \\ [\hat{e}]^{\mathrm{T}} & \hat{\mu} \end{bmatrix} \begin{Bmatrix} \{\varepsilon\} \\ E_z \end{Bmatrix} = \begin{Bmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \sigma_{xz} \\ D_z \end{Bmatrix}
$$
$$
= \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} & 0 & -\hat{e}_{31} \\ \hat{C}_{13} & \hat{C}_{33} & 0 & -\hat{e}_{32} \\ 0 & 0 & \hat{C}_{55} & 0 \\ \hat{e}_{31} & \hat{e}_{32} & 0 & \hat{\mu}_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{zz} \\ 2\varepsilon_{xz} \\ E_z \end{Bmatrix} \tag{16}
$$

where,

$$
\begin{aligned}
\hat{C}_{11} &= \overline{\overline{C}}_{11} + \frac{1}{\Delta}\Big[\overline{\overline{C}}_{12}\Big(\overline{\overline{C}}_{26}\overline{\overline{C}}_{16} - \overline{\overline{C}}_{66}\overline{\overline{C}}_{12}\Big) \\
&\quad + \overline{\overline{C}}_{16}\Big(\overline{\overline{C}}_{26}\overline{\overline{C}}_{16} - \overline{\overline{C}}_{22}\overline{\overline{C}}_{16}\Big)\Big] \\
\hat{C}_{13} &= \overline{\overline{C}}_{13} + \frac{1}{\Delta}\Big[\overline{\overline{C}}_{12}\Big(\overline{\overline{C}}_{26}\overline{\overline{C}}_{36} - \overline{\overline{C}}_{66}\overline{\overline{C}}_{23}\Big) \\
&\quad + \overline{\overline{C}}_{16}\Big(\overline{\overline{C}}_{26}\overline{\overline{C}}_{23} - \overline{\overline{C}}_{22}\overline{\overline{C}}_{36})\Big)\Big] \\
\hat{C}_{33} &= \overline{\overline{C}}_{33} + \frac{1}{\Delta}\Big[\overline{\overline{C}}_{23}\Big(\overline{\overline{C}}_{26}\overline{\overline{C}}_{36} - \overline{\overline{C}}_{66}\overline{\overline{C}}_{23}\Big) \\
&\quad + \overline{\overline{C}}_{36}\Big(\overline{\overline{C}}_{26}\overline{\overline{C}}_{23} - \overline{\overline{C}}_{22}\overline{\overline{C}}_{36}\Big)\Big] \\
\hat{C}_{55} &= \overline{\overline{C}}_{55} - \frac{\overline{\overline{C}}_{45}^2}{\overline{\overline{C}}_{44}} \\
\hat{e}_{31} &= \overline{\overline{e}}_{31} + \frac{1}{\Delta}\Big[\overline{\overline{e}}_{32}\Big(\overline{\overline{C}}_{12}\overline{\overline{C}}_{66} - \overline{\overline{C}}_{16}\overline{\overline{C}}_{26}\Big) \\
&\quad + \overline{\overline{e}}_{36}\Big(\overline{\overline{C}}_{16}\overline{\overline{C}}_{22} - \overline{\overline{C}}_{12}\overline{\overline{C}}_{26}\Big)\Big] \\
\hat{e}_{32} &= \overline{\overline{e}}_{32} + \frac{1}{\Delta}\Big[\overline{\overline{e}}_{32}\Big(\overline{\overline{C}}_{23}\overline{\overline{C}}_{66} - \overline{\overline{C}}_{26}\overline{\overline{C}}_{36}\Big) \\
&\quad + \overline{\overline{e}}_{36}\Big(\overline{\overline{C}}_{36}\overline{\overline{C}}_{22} - \overline{\overline{C}}_{23}\overline{\overline{C}}_{26}\Big)\Big] \\
\hat{\mu}_{33} &= \overline{\overline{\mu}}_{33} + \frac{1}{\Delta}\Big[\overline{\overline{e}}_{36}\Big(\overline{\overline{C}}_{22}\overline{\overline{C}}_{69} + \overline{\overline{C}}_{66}\overline{\overline{C}}_{29}\Big) \\
&\quad + \overline{\overline{e}}_{32}\Big(\overline{\overline{C}}_{29}\overline{\overline{C}}_{66} - \overline{\overline{C}}_{26}\overline{\overline{C}}_{69}\Big)\Big]
\end{aligned} \tag{17}
$$

In the next subsection, we derive the constitutive of the PFCs, wherein the piezo fibers together with the interdigital electrodes (IDEs) are embedded in the composite structures to perform a variety of applications.

3.1 Constitutive model for piezofiber composite (PFC) structures

PFC sensors and actuators are a class of multifunctional composites, wherein the piezo fibers, the associated IDEs are embedded in the composites at the manufacturing stage as shown in Figure 2. In this section, we outline the derivation of the constitutive model of PFC sensors and actuators.

The constitutive model can be derived by considering a rectangular representative volume element (RVE) shown in Figure 2. The rectangular cross section of the fibers can provide maximum volume fraction of ceramic, which is preferable for actuation. The configuration can be obtained using fibers that have been tape-cast and diced, extruded, or cast into a mould. Figure 3 also shows an actuator element (say qth) with its host composite structure and having local coordinate system (X_a^q, Y_a^q, Z_a^q). The RVE of the two-phase ceramic–matrix composite system is described by one quadrant axisymmetric model about the x_3 axis. Here, h is the total depth of a single PFC layer, p is the uniform spacing of the IDEs spanning along x_1, and b is the width of each electrode. The constitutive relations for orthotropic active ceramic bulk form [10] can be represented as

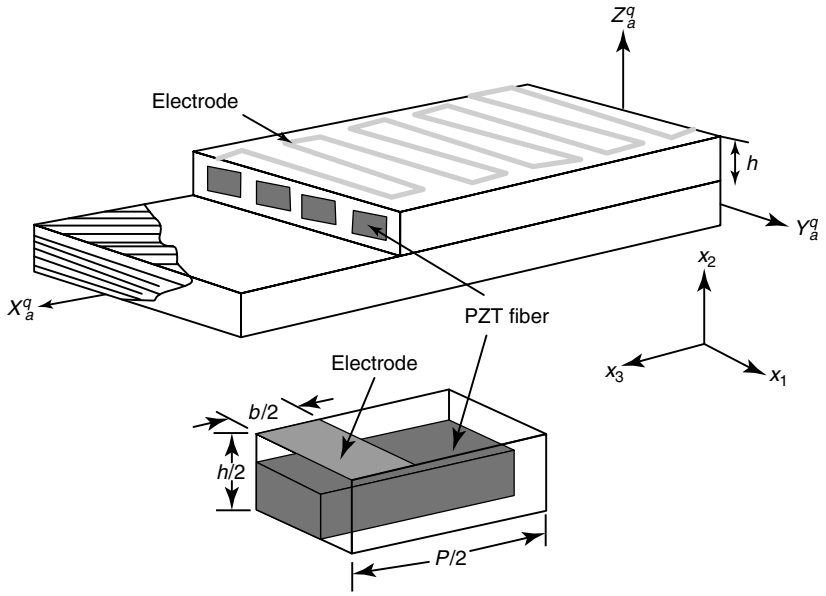


Figure 2. Configuration of piezoelectric fiber composite (PFC) for composite beam actuation.

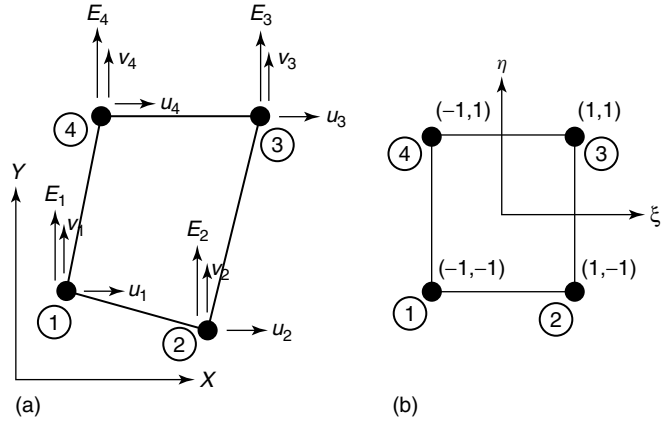


Figure 3. Element degrees of freedom and the isoparametric coordinate system.

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \tau_{xz} \\ D_z \end{Bmatrix} = \begin{bmatrix} C^E{}_{11} & C^E{}_{12} & C^E{}_{13} & -e_{31} \\ C^E{}_{12} & C^E{}_{22} & C^E{}_{23} & -e_{32} \\ C^E{}_{13} & C^E{}_{23} & C^E{}_{33} & -e_{33} \\ e_{31} & e_{32} & e_{33} & \mu^s{}_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{zz} \\ \gamma_{xz} \\ E_z \end{Bmatrix} \tag{18}$$

This relation is of a very similar form as that of PZT sensor/actuator. This relation can be directly used in equation (41) by substituting equation (18) for matrices $[\hat{C}]$ and $[e]$. For 1-D analysis, this requires reduction into single equivalent constitutive law by considering the volume fraction of the piezo fiber (PZT) to the total volume of the laminate. For pure piezoceramic material, $C_{11}^E = C_{22}^E$, $C_{23}^E = C_{13}^E$, $e_{32} = e_{31}$. For matrix phase, all e_{ij} are zero, and their mechanical and dielectric properties are represented without superscripts. Assuming negligible distortion of the equipotential lines and electric fields beneath the electrodes, and imposing proper field continuity between the ceramic and matrix phases, the effective unidirectional constitutive law for a PFC

beam structure can be expressed as

$$\sigma_{zz} = C^{eff}{}_{33}\varepsilon_{zz} - e^{eff}{}_{33}E_z \tag{19}$$

where,

$$C_{33}{}^{eff} = \left(\overline{\overline{C}}_{33}V_1{}^p + C_{22}V_1{}^m\right) - \frac{V_1{}^pV_1{}^m\left(C_{12} - \overline{\overline{C}}_{13}\right)^2}{C_{22}V_1{}^p + \overline{\overline{C}}_{11}V_1{}^m} \tag{20}$$

$$e^{eff}{}_{33} = \overline{\overline{e}}_{33}V_1{}^p + \frac{\overline{\overline{e}}_{31}V_1{}^pV_1{}^m\left(C_{12} - \overline{\overline{C}}_{13}\right)}{C_{22}V_1{}^p + \overline{\overline{C}}_{11}V_1{}^m} \tag{21}$$

$$\overline{\overline{C}}_{11} = \left(\overline{C}_{33}V_2{}^p + C_{11}V_2{}^m\right) - \frac{V_2{}^pV_2{}^m\left(C_{12} - \overline{C}_{23}\right)^2}{C_{11}V_2{}^p + \overline{C}_{22}V_2{}^m} \tag{22}$$

$$\overline{\overline{e}}_{31} = \overline{e}_{31}V_2{}^p + \frac{\overline{e}_{32}V_2{}^pV_2{}^m\left(C_{12} - \overline{C}_{12}\right)}{C_{11}V_2{}^p + \overline{C}_{22}V_2{}^m} \tag{23}$$

$$\overline{\overline{e}}_{33} = \overline{e}_{33}V_2{}^p + \frac{\overline{e}_{32}V_2{}^pV_2{}^m\left(C_{12} - \overline{C}_{23}\right)}{C_{11}V_2{}^p + \overline{C}_{22}V_2{}^m} \tag{24}$$

$$\overline{C}_{jk} = C_{jk}{}^E + \frac{V_3{}^me_{3j}e_{3k}}{V_3{}^m\mu_{33} + V_3{}^m\mu^s{}_{33}} \tag{25}$$

$$\overline{e}_{3j} = \frac{\mu_{33}e_{3j}}{V_3{}^p\mu_{33}V_3{}^m\mu^s{}_{33}} \tag{26}$$

Here, v_i^p and v_i^m, for $i = 1, 2$ represents the length fraction of ceramic and matrix phase along direction i, respectively, and

$$V_3{}^p = \frac{\dfrac{p}{h}}{\dfrac{p}{h} + \left(1 - V_2{}^p\right)} \quad b \ll p \tag{27}$$

represents the volume fraction of ceramic phase in RVE. The details of the above derivation can be found in [11]. Similar models for uniform packing circular fiber can be found in [12]. Essentially, these models provide dominant electromechanical coupling in direction 3, which can be aligned along the local host beam axis during bonding or embedding. This is unlikely in uniformly electroded PZT plate structure.

4 FINITE ELEMENT MODELING

In this section, FE modeling of laminated composite structures with embedded/surface-bonded smart sensors/actuators is given. First, a 2-D FE model for a composite beam with embedded piezoelectric sensor/actuator is given. Since the FE formulation of PFCs is very similar to the piezoelectric laminate case, we will not repeat the formulation here. This is later followed by the modeling of composites with embedded magnetostrictive sensors/actuators. Some numerical examples are also given to demonstrate the use of the formulated element.

4.1 Four-noded isoparametric finite element formulation

Isoparametric FE formulation is a class of FE formulation, wherein an arbitrary element shape is mapped onto a square of dimension 2 units. Hence, in such formulation, in addition to mapping the field variables (displacements and electric field), the geometry is also mapped using the same functions.

Let us consider a 2-D composite laminate under plane stress in the $x-z$ plane. Let $u(x, y, t)$ and $w(x, y, t)$ be its displacement components. The strong form of the governing equation essentially gives the governing partial differential equation for the given problem. This form of the governing equation is not readily amenable for numerical solution and, in addition, the exact analytical solution is available only for a few simple problems. On the other hand, the weak form of solution is readily amenable for numerical solution. Both forms can be derived using the Hamilton's principle. First the strong form of the governing differential equation is derived. This requires that the energy associated with the problem be written in terms of displacements. The kinetic, strain, and the electrical energy for the smart laminated having a volume V is given by

$$T = \frac{1}{2}\int_V \rho\{\dot{u}\}^{\mathrm{T}}\{\dot{u}\}\,\mathrm{d}V \tag{28}$$

$$U = \frac{1}{2}\int_V \{\sigma\}^{\mathrm{T}}\{\varepsilon\}\,\mathrm{d}V \tag{29}$$

$$U_e = \frac{1}{2}\int_V E_z\mu_z\,\mathrm{d}V \tag{30}$$

where

$$\{u\}^{\mathrm{T}} = \{u \;\; v\} \tag{31}$$

$$\{\sigma\}^{\mathrm{T}} = \{\sigma_{xx} \;\; \sigma_{zz} \;\; \sigma_{xz}\} \tag{32}$$

Here, ρ is the density of the smart composite. The stresses are related to strain using equation (16), which can be expressed in terms of displacement using strain–displacement relationship. Using these energies in the Hamilton's principle, we get the following strong form of the governing equation and its associated boundary conditions

$$\rho\frac{\partial^2 u}{\partial t^2} - \hat{C}_{11}\frac{\partial^2 u}{\partial x^2} - \hat{C}_{13}\frac{\partial^2 w}{\partial x \partial z} - \hat{C}_{55}\left(\frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 w}{\partial x \partial z}\right) + \hat{e}_{31}\frac{\partial E_z}{\partial x} = F_{sx} + \hat{e}_{31}\frac{\partial E_s}{\partial x}$$

$$\rho\frac{\partial^2 w}{\partial t^2} - \hat{C}_{33}\frac{\partial^2 w}{\partial z^2} - \hat{C}_{13}\frac{\partial^2 u}{\partial x \partial z} - \hat{C}_{55}\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 u}{\partial x \partial z}\right) + \hat{e}_{33}\frac{\partial E_z}{\partial z} = F_{sz} - \hat{e}_{33}\frac{\partial E_s}{\partial z}$$

$$\hat{e}_{31}\frac{\partial u}{\partial x} + \hat{e}_{33}\frac{\partial w}{\partial z} + \overline{\overline{\mu}}_{33}E_z = -D_s - \overline{\overline{\mu}}_{33}E_s \tag{33}$$

where $\{F_s\}^{\mathrm{T}} = \{F_{sx} \;\; F_{sz}\}$ is the surface force vector in the two directions and E_s, D_s are the residual electrical field and the electrical displacement in the smart composite. In the above equation, the first two represent the force equilibrium in the x and z directions, while the third equation represents the equation for electrical field equilibrium. The associated force boundary conditions on the edge parallel to z axis are

$$\hat{C}_{11}\frac{\partial u}{\partial x} + \hat{C}_{13}\frac{\partial w}{\partial z} - \hat{e}_{31}E_z = F_{cx} + \hat{e}_{31}E_s \quad \text{or} \quad u \text{ prescribed} \tag{34}$$

$$\hat{C}_{55}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) = F_{cz} \quad \text{or} \quad w \text{ prescribed} \tag{35}$$

Similarly, the force boundary conditions on the edge parallel to x-axis are

$$\hat{C}_{13}\frac{\partial u}{\partial x} + \hat{C}_{33}\frac{\partial w}{\partial z} - \hat{e}_{33}E_z = F_{cz} + \hat{e}_{33}E_s \quad \text{or} \quad w \text{ prescribed} \tag{36}$$

$$\hat{C}_{55}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right) = F_{cx} \quad \text{or} \quad u \text{ prescribed} \tag{37}$$

where F_{cx} and F_{cz} are the sum of all point loads in x and z directions, respectively. The main goal here is to solve equation (33). These are very difficult to solve exactly. Hence, it is necessary to recast the above equilibrium equation in its weak form and best fit an approximate solution using FE procedure.

The weak form of the governing differential equation is obtained next. This is obtained by performing a variational minimization (Hamilton's principle) of the total energy, which can be written as

$$\delta\left(\frac{1}{2}\int_{t_1}^{t_2}\int_V \{\dot{u}\}^{\mathrm{T}}\rho\{\dot{u}\}\,\mathrm{d}V\,\mathrm{d}t + \frac{1}{2}\int_{t_1}^{t_2}\int_V \{\sigma\}^{\mathrm{T}}\{\varepsilon\}\,\mathrm{d}V\,\mathrm{d}t + \frac{1}{2}\int_{t_1}^{t_2}\int_V E_z D_z\,\mathrm{d}V\,\mathrm{d}t + \int_{t_1}^{t_2}\{u\}^{\mathrm{T}}\{F_c\}\,\mathrm{d}t + \int_{t_1}^{t_2}\int_{S_1}\{u\}^{\mathrm{T}}\{F_s\}\,\mathrm{d}S_1\,\mathrm{d}t + \int_{t_1}^{t_2}\int_{S_2} E_z D_s\,\mathrm{d}S_2\,\mathrm{d}t\right) = 0 \tag{38}$$

where S_1 and S_2 are the surfaces in the structure where the surface forces and residual displacements act. Substituting for stresses and electrical displacements from equation (16), the weak form of the differential equation becomes

$$\int_{t_1}^{t_2}\int_V \{\delta\dot{u}\}^{\mathrm{T}}\rho\{\dot{u}\}\,\mathrm{d}V\,\mathrm{d}t + \int_{t_1}^{t_2}\int_V \{\delta\varepsilon\}^{\mathrm{T}}[\overline{\overline{C}}]\{\varepsilon\}\,\mathrm{d}V\,\mathrm{d}t - \int_{t_1}^{t_2}\int_V \delta E[\overline{\overline{e}}]^{\mathrm{T}}\{\varepsilon\}\,\mathrm{d}V\,\mathrm{d}t + \int_{t_1}^{t_2}\int_V \delta E_z[\overline{\overline{e}}]^{\mathrm{T}}\{\varepsilon\}\,\mathrm{d}V\,\mathrm{d}t + \int_{t_1}^{t_2}\int_V \delta E_z\mu E_z\,\mathrm{d}V\,\mathrm{d}t + \int_{t_1}^{t_2}\{\delta u\}^{\mathrm{T}}\{F_c\}\,\mathrm{d}t$$

$$+\int_{t_1}^{t_2}\int_{S_1}\{\delta u\}^{\mathrm{T}}\{F_s\}\,\mathrm{d}S_1\,\mathrm{d}t$$
$$+\int_{t_1}^{t_2}\int_{S_2}\delta E_z D_s\,\mathrm{d}S_2\,\mathrm{d}t=0 \tag{39}$$

The above equation is the weak form of the governing equation (equation 33) for a composite laminate with piezoelectric smart sensors. This is the starting point for the FE formulation.

Next, we outline the procedure for formulating a four-node isoparametric plane stress smart composite FE. The element configuration is shown in Figure 3. This element will have two mechanical degrees of freedom, namely, the two displacement components $u(x, y, t)$ and $w(x, y, t)$, respectively, and a single electrical degree of freedom $E_z(x, y, t)$ in the z direction. The electric field in this direction will induce stresses in the x direction. Thus, this element will have a total of 12 degrees of freedom. Here, we adopt isoparametric formulation. Since the proposed element is four noded, the bilinear shape functions for the mechanical displacements will be required, which can be written as

$$u(x, y, t)=\sum_{i=1}^{4}N_i(\xi,\eta)u_i(t),$$
$$w(x, y, t)=\sum_{i=1}^{4}N_i(\xi,\eta)w_i(t) \tag{40}$$

where ξ and η are the isoparametric coordinates and $u_i(t)$ and $w_i(t)$ are the nodal mechanical degrees of freedom. The four bilinear shape functions are given by

$$N_1=\frac{1}{4}(1-\xi)(1-\eta),\quad N_2=\frac{1}{4}(1+\xi)(1-\eta),$$
$$N_3=\frac{1}{4}(1+\xi)(1+\eta),\quad N_4=\frac{1}{4}(1-\xi)(1+\eta) \tag{41}$$

Now to choose the interpolating polynomial for the electrical degrees of freedom, we look at the strong form of the governing equation (equation 33). By substituting linear variation for the mechanical degrees of freedom, we find that for consistency of the displacement field, we also require a bilinear variation of the electrical field. Hence, we assume the electric field to vary as

$$E_z(x, y, t)=\sum_{i=1}^{4}N_i(\xi,\eta)E_{zi}(t) \tag{42}$$

where the same shape function given in equation (41) is also used here, and E_{zi} are the nodal electrical degrees of freedom at the four nodes.

In isoparametric formulation, we map the actual geometry of the element to a square of size 2 defined in the generalized coordinate system (ξ, η) through a Jacobian transformation. This requires the variation of the coordinate system in the generalized coordinates in terms of the nodal coordinates of the actual element geometry. Hence, one can use the same displacement shape functions to describe this variation and can be written as

$$x(x, y)=\sum_{i=1}^{4}N_i(\xi,\eta)x_i \tag{43}$$

$$z(x, y)=\sum_{i=1}^{4}N_i(\xi,\eta)z_i \tag{44}$$

The strains are evaluated by using strain–displacement relationship. That is,

$$\begin{Bmatrix}\varepsilon_{xx}\\ \varepsilon_{xx}\\ 2\varepsilon_{xz}\\ E_z\end{Bmatrix}=\begin{bmatrix}\partial/\partial x & 0 & 0\\ 0 & \partial/\partial z & 0\\ \partial/\partial z & \partial/\partial x & 0\\ 0 & 0 & 1\end{bmatrix}\begin{Bmatrix}u\\ w\\ D_z\end{Bmatrix} \tag{45}$$

Using equations (40) and (42) in the above equation enables expressing the strains in terms of nodal displacement vector $\{u\}_e=\{u_1\ \ w_1\ \ u_2\ \ w_2\ \ u_3\ \ w_3\ \ u_4\ \ w_4\}^{\mathrm{T}}$ and electric field vector $\{E_z\}_e=\{E_{z1}\ \ E_{z2}\ \ E_{z3}\ \ E_{z3}\}^{\mathrm{T}}$. That is, strain can be written as

$$\{\varepsilon\}=[B]\{u\}=\begin{bmatrix}[B]_{u(3\times 8)} & 0\\ 0 & [B]_{E(1\times 4)}\end{bmatrix} \tag{46}$$

where $[B]$ matrix is given by

$$[B] = \begin{bmatrix} \frac{\partial N_1}{\partial x} & 0 & \frac{\partial N_2}{\partial x} & 0 & \frac{\partial N_3}{\partial x} & 0 & \frac{\partial N_4}{\partial x} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial N_1}{\partial z} & 0 & \frac{\partial N_2}{\partial z} & 0 & \frac{\partial N_3}{\partial z} & 0 & \frac{\partial N_4}{\partial z} & 0 & 0 & 0 & 0 \\ \frac{\partial N_1}{\partial z} & \frac{\partial N_1}{\partial x} & \frac{\partial N_2}{\partial z} & \frac{\partial N_2}{\partial x} & \frac{\partial N_3}{\partial z} & \frac{\partial N_3}{\partial x} & \frac{\partial N_4}{\partial z} & \frac{\partial N_4}{\partial x} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & N_1 & N_2 & N_3 & N_4 \end{bmatrix} \tag{47}$$

Using equations (46) and (42) in the weak form of the equation (equation 24), and performing variational minimization, we get

$$\begin{aligned} &\{\delta u\}_e^{\mathrm{T}} \left(\int_V [N]^{\mathrm{T}} \rho [N] \, \mathrm{d}V \right) \{\ddot{u}\}_e \\ &\quad + \{\delta u\}_e^{\mathrm{T}} \left(\int_V [B_u]^{\mathrm{T}} [\hat{C}] [B_u] \, \mathrm{d}V \right) \{u\}_e \\ &\quad - \{\delta u\}_e^{\mathrm{T}} \left(\int_V [B_u]^{\mathrm{T}} [\hat{e}] [B_E] \, \mathrm{d}V \right) \{E_z\}_e \\ &\quad - \{\delta E_z\}_e^{\mathrm{T}} \left(\int_V [B_E]^{\mathrm{T}} [\hat{e}]^{\mathrm{T}} [B_u] \, \mathrm{d}V \right) \{u\}_e \\ &\quad - \{\delta E_z\}_e^{\mathrm{T}} \left(\int_V [B_E]^{\mathrm{T}} \hat{\mu}_{33} [B_E] \, \mathrm{d}V \right) \{E_z\}_e \\ &\quad - \{\delta u\}_e^{\mathrm{T}} \{F_c\} - \{\delta u\}_e^{\mathrm{T}} \int_{S_1} [N]^{\mathrm{T}} \{F_s\} \, \mathrm{d}S_1 \\ &\quad - \{\delta E_z\}_e^{\mathrm{T}} \int_{S_2} [B_E]^{\mathrm{T}} D_s \, \mathrm{d}S_2 = 0 \end{aligned} \tag{48}$$

Since $\{\delta u\}_e$ and $\{\delta E_z\}_e$ are arbitrary, the above expression can be written in a concise matrix form as

$$\begin{aligned} &\begin{bmatrix} [M_{uu}] & [0] \\ [0] & [0] \end{bmatrix} \begin{Bmatrix} \{\ddot{u}\}_e \\ \{\ddot{E}_z\}_e \end{Bmatrix} \\ &\quad + \begin{bmatrix} [K_{uu}] & [K_{uE}] \\ [K_{uE}]^{\mathrm{T}} & [K_{EE}] \end{bmatrix} \begin{Bmatrix} \{u\}_e \\ \{E_z\}_e \end{Bmatrix} = \begin{Bmatrix} \{F\}_e \\ \{q\}_e \end{Bmatrix} \end{aligned} \tag{49}$$

The above equation is the elemental equilibrium in the discritized form, where $[M_{uu}]$ is the mass matrix, $[K_{uu}]$ is the stiffness matrix corresponding to mechanical degrees of freedom, $[K_{uE}]$ is the stiffness matrix due to electromechanical coupling, and $[K_{EE}]$ is the stiffness matrix due to electrical degrees of freedom alone. Note that all these matrices require the volume integral to be evaluated. Since the exact integration of these is most difficult to achieve, we resort to numerical integration (see [1–3] for more details). Here, $\{F\}_e$ is the elemental nodal vector and $\{q\}_e$ is the elemental charge vector. These matrices are given by

$$\begin{aligned} [M_{uu}] &= t \int_{-1}^{1} \int_{-1}^{1} [N]^{\mathrm{T}} \rho [N] |J| \, \mathrm{d}\xi \, \mathrm{d}\eta, \\ [K_{uu}] &= t \int_{-1}^{1} \int_{-1}^{1} [B_u]^{\mathrm{T}} [\hat{C}] [B_u] |J| \, \mathrm{d}\xi \, \mathrm{d}\eta, \\ [K_{uE}] &= -t \int_{-1}^{1} \int_{-1}^{1} [B_u]^{\mathrm{T}} [\hat{e}] [B_E] |J| \, \mathrm{d}\xi \, \mathrm{d}\eta, \\ [K_{EE}] &= t \int_{-1}^{1} \int_{-1}^{1} [B_E]^{\mathrm{T}} \hat{\mu}_{33} [B_E] |J| \, \mathrm{d}\xi \, \mathrm{d}\eta \end{aligned} \tag{50}$$

The elemental load and charge vectors are given by

$$\{F\}_e = \{F\}_c + \int_{S_1} [N]^{\mathrm{T}} \{F\}_s \, \mathrm{d}S_1 \tag{51}$$

$$\{q\}_e = -\int_{S_2} [N]^{\mathrm{T}} D_s \, \mathrm{d}S_2 \tag{52}$$

The matrices in equation (49) are then assembled to obtain their global counterparts and solved for obtaining solutions for displacements and electric field. Note that it has a zero diagonal block in the mass matrix. The method of solution for sensing and actuation problem is quite different. For sensing problem, for a given mechanical loading, we need to determine the voltage developed across the smart patch. This is done by first obtaining the mechanical displacement due to the given mechanical load, which is then used to obtain the electric field and hence the voltage developed in the sensor patch. In order to solve this, the global matrix equation can be expanded

and written as

$$[M_{uu}]\{\ddot{u}\} + [K_{uu}]\{u\} + [K_{uE}]\{E_z\} = \{F\}$$
$$[K_{uE}]^{\mathrm{T}}\{u\} + [K_{EE}]\{E_z\} = \{q\} \quad (53)$$

We can write the second part of the above equation as

$$\{E_z\} = [K_{EE}]^{-1}\{q\} - [K_{EE}]^{-1}[K_{uE}]^{\mathrm{T}}\{u\} \quad (54)$$

Using the above equation in the first of equation (53) and simplifying, we get

$$[M_{uu}]\{\ddot{u}\} + [\overline{K}_{uu}] = \{\overline{F}\} \quad (55)$$

where

$$[\overline{K}_{uu}] = [K_{uu}] - [K_{uE}][K_{EE}]^{-1}[K_{uE}]^{\mathrm{T}} \quad (56)$$
$$\{\overline{F}\} = \{F\} - [K_{EE}]^{-1}\{q\} \quad (57)$$

Note that equation (55) is only in terms of mechanical displacements, which is solved. Using this solution, electrical fields are obtained using equation (54), from which voltages can be obtained. This procedure is adopted for sensing problem, which is used in the health monitoring studies. For actuation problem, the voltages and hence the electric field goes as input. That is, the second part of equation (49) is not required. Hence, the equation that requires solution becomes

$$[M_{uu}]\{\ddot{u}\} + [K_{uu}]\{u\} = \{F\} - [K_{uE}]\{E_z\} = \{F^*\} \quad (58)$$

More details on this formulation can be obtained from [13].

4.2 Numerical example

4.2.1 Actuation example

The main objective of this example is to demonstrate the actuation capability of the formulated FE with piezoelectric actuators and validate the formulated element in the process by comparing the results obtained from the standard published results. Here, a piezoelectric bimorph composite beam is considered for analysis. Chen [14] presented the comparative study of the bending of a bimorph beam due to external applied voltage as a part of verifying the accuracy of the piezoelectric FE solution and this bimorph beam configuration was adopted from Hwang [15]. In this study, the same configuration of the bimorph beam is considered. The bimorph beam consists of two identical PVDF beams laminated together with opposite polarities. The schematic diagram of the bimorph beam is as shown in Figure 4. The dimensions of the beam are taken as 100 mm × 5.0 mm × 0.5 mm. The material properties of the PVDF bimorph beam are taken same as that of Chen [14] and are as follows: Young's modulus E_{11} is equal to $0.2e^{10}$ N m^{-2}, shear modulus G_{12} is equal to $0.775e^{10}$ N m^{-2}, Poisson's ratio ν_{12} is equal to 0.29, piezoelectric constant e_{31} is equal to 0.046 C m^{-2}, and piezoelectric constant e_{32} is equal to e_{31}. The theoretical solution for transverse displacement for the above problem is given in [14], which

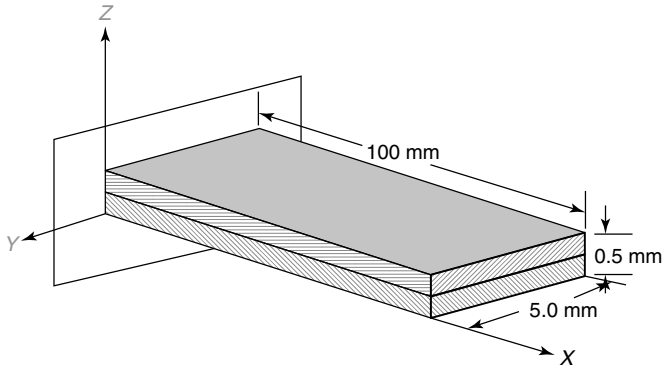


Figure 4. Schematic diagram of the piezoelectric PVDF bimorph cantilever beam.

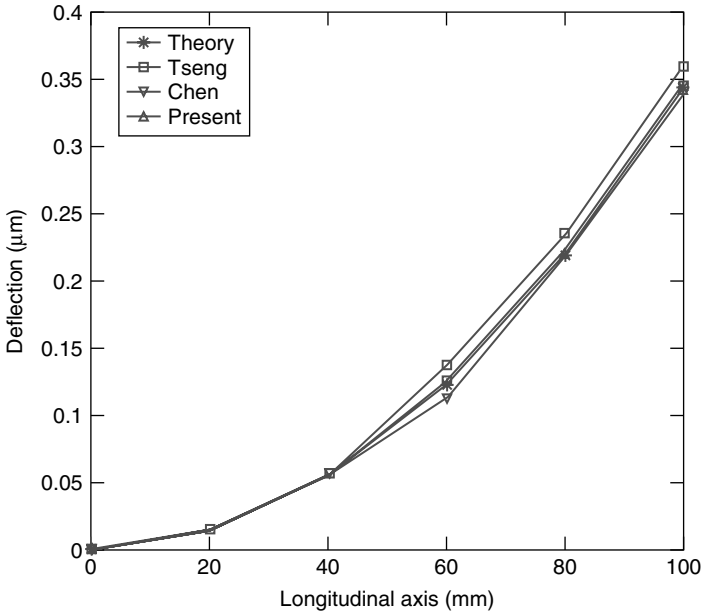


Figure 5. Centerline deflections for a PVDF bimorph beam under a unit voltage.

is given by

$$w(x) = 0.375\frac{e_{31}}{E}V\left(\frac{x}{t}\right)^2 \tag{59}$$

where V is the applied voltage and t is the thickness of the beam. This beam is modeled using 200 formulated elements along the $x-z$ plane. When an external active voltage is applied across the thickness, the induced strain generates control forces that bend the bimorph beam. A unit voltage is applied across the thickness and the deflections at the nodes are computed. The deflection of the beam along the central longitudinal axis obtained from the present formulation is compared with the theoretical value and the works of Chen [14], and Tzou and Tseng [16], respectively. Figure 5 shows the comparison of deflection along the length of the beam for unit voltage applied across the thickness.

Next, the deflection of the beam is calculated for different applied voltage in the range of 0–100 V. The objective here is to actuate the beam to the extent so as to make the tip deflection, due to applied load, negligible. This is shown in Figure 6. From the figure, we can clearly see that with increase in voltage, the tip deflection reduces gradually toward zero value.

5 FINITE ELEMENT MODELING OF MAGNETOSTRICTIVE SENSORS AND ACTUATORS

Analysis of structures with magnetostrictive sensors/actuators is generally performed using uncoupled models. Uncoupled models are based on the assumption that the magnetic field within the magnetostrictive material is proportional to the electric coil current times the number of coil turns per unit length. Owing to this assumption, actuation and sensing equations get uncoupled. For actuator, the strain due to magnetic field (which is proportional to coil current) is incorporated as the equivalent nodal load in the FE model for calculating the block force. Thus, with this procedure, analysis is carried out without taking smart (magnetic) degrees of freedom in the FE model. Similarly for sensor, where generally coil current is assumed zero, the magnetic flux density is proportional to mechanical stress, which can be calculated from the FE results through post-processing. This assumption on magnetic field, leads to the violation of flux line continuity, which is one of the four Maxwell's equations in electromagnetism.

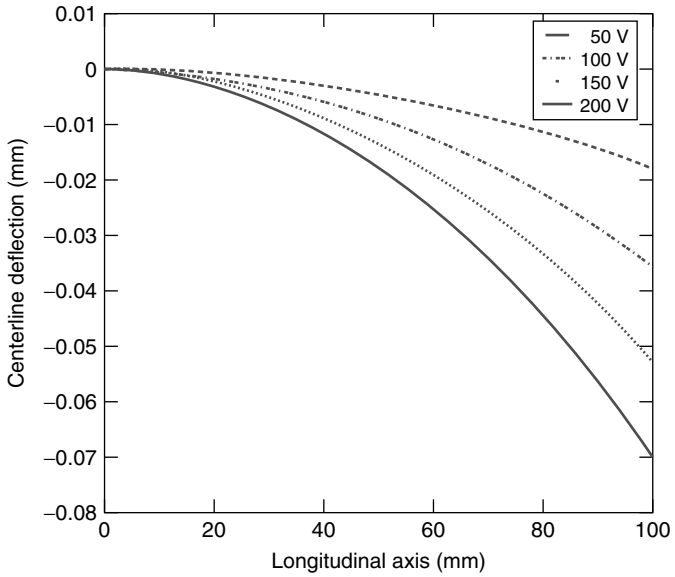


Figure 6. Deflection profile for a PVDF bimorph beam for various voltages.

On the other hand, in coupled model, it is considered that magnetic flux density and/or strain of the material are functions of stress and magnetic field, without any additional assumption on magnetic field, like in uncoupled model. The procedure to obtain the constitutive model for a composite with magnetostrictive sensors/actuators is very similar to what was derived for the piezoelectric sensors/actuators and hence is not repeated here. We directly go to the FE formulation.

The FE formulation of structures with embedded magnetostrictive patches differs depending upon whether the constitutive model is coupled or uncoupled. For the uncoupled model, the magnetic field is assumed proportional to the coil current and, as a result, the sensing and actuation constitutive relations are solved independently. That is, there is no need to have magnetic field as independent degrees of freedom. However, in a coupled model, both the sensing and actuation constitutive laws need to be solved simultaneously, since no explicit variation of the magnetic field with respect to magnetostrictive patch parameters are available. This requires that the magnetic field be taken as independent degrees of freedom. Here, we outline the procedure of modeling magnetostrictive sensors/actuators for both uncoupled and coupled constitutive law.

FE formulation begins by writing the associated energy in terms of nodal degrees of freedom by assuming the displacement and magnetic field variation in three coordinate directions over each element. That is, the displacement field can be written as

$$\{U\} = \{\, u \quad v \quad w \,\} = [N_u]\{U\}_e \tag{60}$$

Here, $u(x, y, z, t)$, $v(x, y, z, t)$, and $w(x, y, z, t)$ are the displacement components in the three coordinate directions, $[N_u]$ is the shape functions associated with mechanical degrees of freedom, and $\{U\}_e$ is the nodal displacement vector. If isoparametric formulation is used, then the conventional isoparametric shape functions in natural coordinate could be adopted. The strains can be expressed in terms of displacement through strain–displacement relationship. That is,

$$\begin{aligned}\{\varepsilon\} &= \{\varepsilon_{xx} \quad \varepsilon_{yy} \quad \varepsilon_{zz} \quad \gamma_{yz} \quad \gamma_{xz} \quad \gamma_{xy}\}^{\mathrm{T}} \\ &= [\overline{B}]\{U\}_e \end{aligned} \tag{61}$$

where $[\overline{B}]$ is the strain–displacement matrix. For coupled analysis, we need to take magnetic field as independent degrees of freedom. In such cases, the magnetic field in the three coordinate directions can

be written as

$$\{H\} = \{H_x \quad H_y \quad H_z\} = [N_H]\{H\}_e \tag{62}$$

where $\{H\}_e$ is the nodal magnetic field vector and $[N_H]$ is the shape function associated with magnetic field degrees of freedom.

The strain energy in a structure with magnetostrictive patches over a volume V is given by

$$V_e = \frac{1}{2}\int_V \{\varepsilon\}^{\mathrm{T}}\{\sigma\}\,\mathrm{d}V \tag{63}$$

Substituting for $\{\sigma\}$ from equation (4), we can write the above equation in terms of strains and magnetic field vector. In this equation, the strains are expressed in terms of displacement using equation (61) and magnetic field in terms of nodal magnetic field vector using equation (62).

The resulting expression for the strain energy becomes

$$V_e = \frac{1}{2}\{U\}_e{}^{\mathrm{T}}[K_{uu}]\{U\}_e - \frac{1}{2}\{U\}_e{}^{\mathrm{T}}[K_{uH}]\{H\}_e \tag{64}$$

where

$$[K_{uu}] = \int_V [\overline{B}]^{\mathrm{T}}[Q][\overline{B}]\,\mathrm{d}V,$$

$$[K_{uH}] = \int_V [\overline{B}]^{\mathrm{T}}[e]^{\mathrm{T}}[N_H]\,\mathrm{d}V \tag{65}$$

where $[K_{uu}]$ is the stiffness matrix associated with mechanical degrees of freedom and $[K_{UH}]$ is the coupling stiffness matrix, which couples the mechanical and magnetic degrees of freedom. Kinetic energy is given by

$$T_e = \frac{1}{2}\int_V \{\dot{U}\}^{\mathrm{T}}\rho\{\dot{U}\}\,\mathrm{d}V \tag{66}$$

Here, $\{\dot{U}\}$ is the velocity vector and ρ is the average density of the host material. Using equation (60) in the above equation, we can write the kinetic energy as

$$T_e = \frac{1}{2}\{\dot{U}\}^{\mathrm{T}}[M_{uu}]\{\dot{U}\} \tag{67}$$

where $[M_{uu}]$ is the mass matrix associated only with mechanical degrees of freedom and is given by

$$[M_{uu}] = \int_V [N_u]^{\mathrm{T}}\rho[N_u]\,\mathrm{d}V \tag{68}$$

The magnetic potential energy for the system can be written as

$$V_m = \frac{1}{2}\int_V [B]^{\mathrm{T}}\{H\}\,\mathrm{d}V \tag{69}$$

Substituting for $[B]$ from equation (4), and $\{H\}$ from equation (62), we can write the magnetic potential energy as

$$V_m = \frac{1}{2}\int_V \{[e]\{\varepsilon\} + [\mu^{\varepsilon}]\{H\}\}^{\mathrm{T}}\{H\}\,\mathrm{d}V \tag{70}$$

Substituting for strains from equation (61), we get

$$V_m = \frac{1}{2}\{U\}_e{}^{\mathrm{T}}[K_{uH}]^{\mathrm{T}}\{H\}_e + \frac{1}{2}\{H\}_e{}^{\mathrm{T}}[K_{HH}]\{H\}_e \tag{71}$$

where,

$$[K_{HH}] = \int_V [N_H]^{\mathrm{T}}[\mu^{\varepsilon}]\left[N_H\right]\,\mathrm{d}V \tag{72}$$

When an applied current I ampere (ac or dc) is fed to the patch having N number of coils, it creates a magnetic field, which, in turn, introduces an external force in the patch. The external work done over a magnetostrictive patch of area A due to this field is given by

$$W_m = IN\int_A [\mu^{\sigma}]\{H\}\,\mathrm{d}A \tag{73}$$

Here, $[\mu^{\sigma}]$ is the permeability matrix measured at constant stress. It is necessary to convert this area integral into a volume integral. If n is the number of coil turns per unit length and $\{l_c\}$ is the direction cosine vector of the coil axis, the above area integral can be converted into volume integral by replacing N by $n\{l_c\}^{\mathrm{T}}$. Substituting for the magnetic field from equation (62), equation (73) becomes

$$W_m = \{F_H\}^{\mathrm{T}}\{H\}_e,$$

$$\{F_H\} = In\{l_c\}^{\mathrm{T}}\int_V [\mu^{\sigma}][N_H]\,\mathrm{d}V \tag{74}$$

The external mechanical work done owing to body force or surface traction vector can be written in the form

$$W_e = \{R\}^{\mathrm{T}}\{U\}_e \tag{75}$$

Using Hamilton's principle, $\delta \int_{t_1}^{t_2} (T_e - V_e + V_m + W_m + W_e)\,\mathrm{d}t = 0$ gives the necessary FE governing equation. This takes a varied form for uncoupled and coupled models, which are given below.

In uncoupled model, the magnetic field is assumed proportional to the coil current and hence a variation with respect to magnetic field is not performed. That is, the magnetic field is normally equal to $H = nI$, where n is normally the coil turns per unit length of the magnetostrictive material patch. With this assumption, the Hamilton's principle gives the equation of motion as

$$[M_{uu}]\{\ddot{U}\}_e + [K_{uu}]\{U\}_e = [K_{uH}]\{H\}_e - \{R\} \quad (76)$$

where $\{\ddot{U}\}_e$ is the elemental acceleration vector and the magnetic field in the above equation is obtained by $\{H\}_e = n_e I$ with n_e being the elemental coil turns per unit length.

In the case of coupled model, one has to also take a variation on the magnetic field as there is no explicit relation of this with respect to any of the parameters. Hence, both mechanical and magnetic degrees of freedom are considered as unknown. The Hamilton's principle gives the following coupled set of equation.

$$\begin{bmatrix} [M_{uu}] & [0] \\ [0] & [0] \end{bmatrix} \begin{Bmatrix} \{\ddot{U}\}_e \\ \{\ddot{H}\}_e \end{Bmatrix} + \begin{bmatrix} [K_{uu}] & -[K_{uH}] \\ [K_{uH}]^{\mathrm{T}} & [K_{HH}] \end{bmatrix} \begin{Bmatrix} \{U\}_e \\ \{H\}_e \end{Bmatrix} = \begin{Bmatrix} -\{R\} \\ \{F_H\} \end{Bmatrix} \quad (77)$$

Note that the stiffness matrix is not symmetric and have a block zero diagonal matrix in the mass matrix as magnetic field does not contribute to the inertia of the composite. For effective solution, the above equation is expanded and the magnetic degrees of freedom are condensed out. The reduced equation of motion can be written as

$$[M_{uu}]\{\ddot{U}\}_e + [K^*{}_{uu}]\{U\}_e = \{R^*\} \quad (78)$$

where,

$$[K^*{}_{uu}] = \left[[K_{uu}] + [K_{uH}][K_{HH}]^{-1}[K_{uH}]^{\mathrm{T}}\right] \quad (79)$$

$$\{R^*\} = [K_{uH}][K_{HH}]^{-1}\{F_H\} - \{R\} \quad (80)$$

Before solution of equation (78), all the matrices are formed for each element and assembled.

After the computation of nodal displacement and velocities, we can compute the sensor open circuit voltage. This is particularly of great interest in structural health monitoring studies. The process of computing this for coupled and uncoupled models is quite different. Using Faraday's law, open circuit voltage V_v in the sensing coil can be calculated from magnetic flux passing through the sensing patch.

In uncoupled model, nodal magnetic field is assumed constant over the element and with zero sensor coil current. To get open circuit voltage, magnetic flux density can be expressed in terms of strain from sensing equation (equation 5), which is given by

$$\{B\} = [d]\{\sigma\} = [d][Q]\{\varepsilon\} = [e]\{\varepsilon\} = [e][\overline{B}]\{U\}_e \quad (81)$$

Now using Faraday's law, open circuit voltage of the sensor having N_S turns and area A can be calculated from the expression

$$V_v = -N_S \int_A \frac{\partial}{\partial t}\{[e]\{\varepsilon\}\}\,\mathrm{d}A \quad (82)$$

The above integral can be converted to volume integral as before by multiplying it with the direction cosine vector, and the open circuit voltage can now be written as

$$V_v = \{F_v\}^{\mathrm{T}}\{U_e\} \quad (83)$$

$$\{F_v\}^{\mathrm{T}} = -n_s\{l_c\}^{\mathrm{T}} \int_V [e][\overline{B}]\,\mathrm{d}V \quad (84)$$

Here, n_s is the coil turns per unit length of the sensor.

In the case of coupled model, the magnetic flux density is computed from nodal magnetic field, which is obtained from FE analysis. Thus, open circuit voltage in the sensor takes a different expression and can be calculated from the expression

$$V_v = -N_s \int_A \frac{\partial}{\partial t}[\mu^\sigma]\{H\}\,\mathrm{d}A \quad (85)$$

This can again be converted into volume integral as before. Substituting for $\{H\}$ from equation (62), in terms of nodal magnetic degrees of freedom, for

which the second part of equation (77) is used and finally after simplification, the open circuit voltage can be written as

$$V_v = \{F_v\}^{\mathrm{T}}\{U_e\} \tag{86}$$

$$\{F_v\}^{\mathrm{T}} = -n_s\{l_c\}^{\mathrm{T}}\left[\int_V [\mu^{\sigma}][N_H]\,\mathrm{d}V\right] \times [K_{HH}]^{-1}[K_{uH}]^{\mathrm{T}} \tag{87}$$

The above formulation is a general 3-D composite structure with embedded/surface-bonded magnetostrictive patches. In the next subsection, we generalize this to a beam structure.

5.1 Modeling of laminated composite beam with embedded Terfenol-D patch

The displacement field for the beam based on first-order shear deformation theory (FSDT) is given by

$$u(x, y, x, t) = u_0(x, t) - z\phi(x, t) \tag{88}$$

$$v(x, y, z, t) = 0 \tag{89}$$

$$w(x, y, x, t) = w(x, t) \tag{90}$$

As before, the magnetic field (H) is only in the axial direction. Next, we need to assume the necessary polynomials for the mid-plane axial displacement u_0, lateral displacement w_0, and the slope ϕ. Since the slopes are independent and not derivable from the lateral displacement, C^0 continuous formulation is sufficient and hence, we can use linear polynomials. All the formulated matrices are numerically integrated. However, this formulation is prone to exhibit what is called the *shear locking* problem. That is, when this element is used for a beam that is very thin, then the shear strains should go to zero. Since this element is based on FSDT, this does not happen, and in the process produces results that are many orders smaller than the actual result. More details on shear locking problems and their alleviation can be obtained in [17–20]. One of the simplest ways to eliminate locking is to reduce the integrated part of the mechanical stiffness matrix contributed by the shear stress. This is undertaken in this formulation.

The approach is very similar to the formulation of the quadrilateral element in the previous section. First, using the assumed polynomial for axial, transverse, and rotation degrees of freedom, the shape functions are constructed. Using these, the strain–displacement matrix is constructed, which relates the strains to the nodal displacements. The stiffness and mass matrix is then evaluated using equations (65) and (68), respectively. Note that while evaluating the stiffness matrix, isoparametric shape function is used and the matrix is numerically integrated. In addition, to alleviate the shear locking problem, the matrix is reduced integrated. We write down only the final form of elemental matrices involved. The mechanical stiffness matrix $[K_{uu}]$ is 6×6 and is given by

$$[K_{uu}] = \begin{bmatrix} \frac{A_{11}}{L} & \frac{A_{15}}{L} & \left(\frac{A_{15}}{2} - \frac{B_{11}}{L}\right) & -\frac{A_{11}}{L} & -\frac{A_{15}}{L} & \left(\frac{A_{15}}{2} + \frac{B_{11}}{L}\right) \\ & \frac{A_{55}}{L} & \left(\frac{A_{55}}{2} - \frac{B_{15}}{L}\right) & -\frac{A_{15}}{L} & -\frac{A_{55}}{L} & \left(\frac{A_{55}}{2} + \frac{B_{15}}{L}\right) \\ & & \left(\frac{A_{55}L}{4} - B_{15}\right) & \left(-\frac{A_{15}}{2} + \frac{B_{11}}{L}\right) & \left(-\frac{A_{55}}{2} + \frac{B_{15}}{L}\right) & \left(\frac{A_{55}L}{4} - \frac{D_{11}}{L}\right) \\ & & & \frac{A_{11}}{L} & \frac{A_{15}}{L} & \left(-\frac{A_{15}}{2} - \frac{B_{11}}{L}\right) \\ & sym & & & \frac{A_{55}}{L} & \left(-\frac{A_{55}}{2} - \frac{B_{15}}{L}\right) \\ & & & & & \left(\frac{A_{55}L}{4} + B_{15} + \frac{D_{11}}{L}\right) \end{bmatrix} \tag{91}$$

where,

$$[A_{ij}, B_{ij}, D_{ij}] = \int_A Q_{ij}[1, z, z^2]\,\mathrm{d}A \qquad (92)$$

The magnetomechanical coupling matrix is given by

$$[K_{uH}]^{\mathrm{T}} = \frac{1}{2}\begin{bmatrix} -e^0{}_{11} & 0 & e^1{}_{11} & e^0{}_{11} & 0 & -e^1{}_{11} \\ -e^0{}_{11} & 0 & e^1{}_{11} & e^0{}_{11} & 0 & -e^1{}_{11} \end{bmatrix} \qquad (93)$$

where,

$$[\, e^0{}_{11} \quad e^1{}_{11}\,] = \int_A e_{11}[\,1 \quad z\,]\,\mathrm{d}A \qquad (94)$$

Matrix $[K_{HH}]$ is given by

$$[K_{HH}] = \frac{L\mu^0}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad (95)$$

$$\mu^0 = \int_A \mu^{\varepsilon}\,\mathrm{d}A \qquad (96)$$

In the case of coupled model, we require force vector $\{F_H\}$, which is given in equation (74). This vector becomes

$$\{F_H\} = \frac{In\mu^0 L}{2}\begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \qquad (97)$$

The mass matrix is given by

$$[M_{uu}] = \begin{bmatrix} 2I_0 & 0 & -2I_1 & I_0 & 0 & -I_1 \\ & 2I_0 & 0 & 0 & I_0 & 0 \\ & & 2I_2 & -I_1 & 0 & I_2 \\ & & & 2I_0 & 0 & -2I_1 \\ & sym & & & 2I_0 & 0 \\ & & & & & 2I_2 \end{bmatrix} \qquad (98)$$

where,

$$[I_0, I_1, I_2] = \int_A \rho[1, z, z^2]\,\mathrm{d}A \qquad (99)$$

5.2 Numerical examples

A composite magnetostrictive bimorph beam shown in Figure 7 is analyzed to verify the effectiveness of the formulated element. A dynamic response analysis is performed on this beam to bring out the effects of coupling in the constitutive model. Note that the coupling aspects were addressed in detail on the article on constitutive model of magnetostrictive materials. In this beam example, length and width of the beam are 500 and 50 mm, respectively. The beam is made of 12 layers with thickness of each layer being 0.15 mm. Surface-mounted magnetostrictive patches at the top and bottom layers are considered as sensor and actuator, respectively. Elastic modulus of composite is assumed as 181 and 10.3 GPa in parallel (E_1) and perpendicular (E_2) direction of fiber. Density (ρ) and shear modulus

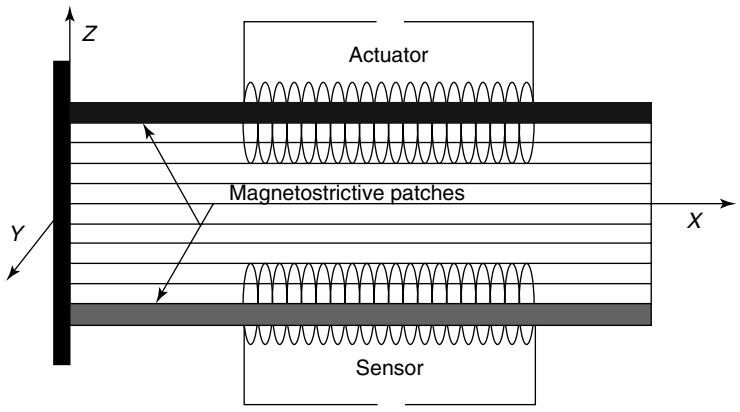


Figure 7. Cantilever magnetostrictive bimorph beam.

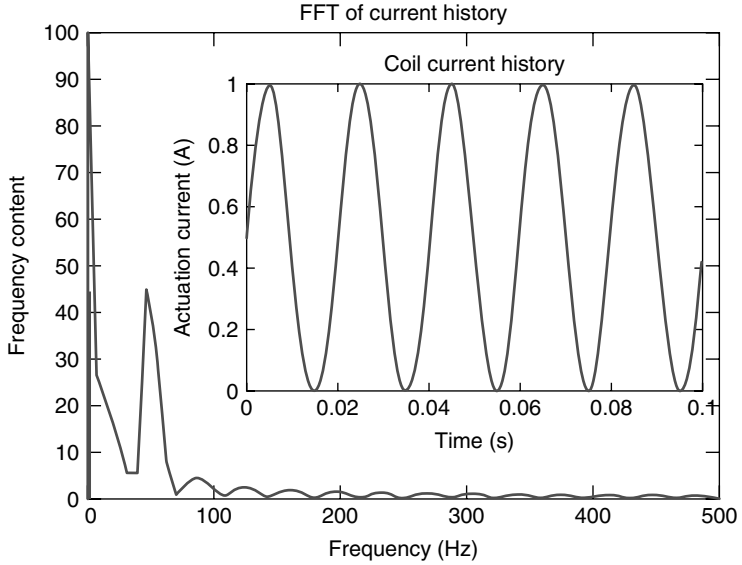


Figure 8. Sinusoidal input force (50 Hz) of in-time and frequency domain.

(G_{12}) of composite are $1.6\,\mathrm{g\,cm^{-3}}$ and 28 GPa, respectively. Elastic modulus (E_m), shear modulus (G_m), and density (ρ_m) of the magnetostrictive material are taken as 30 GPa, 23 GPa, and $9.25\,\mathrm{g\,cm^{-3}}$, respectively. Magnetomechanical coupling coefficient is taken as 15E-09 $\mathrm{m\,A^{-1}}$. Permeability at vacuum or air is assumed to be 400 π $\mathrm{nH\,m^{-1}}$. Constant stress relative permeability for magnetostrictive material is assumed to be equal to 10. The number of coil turns per meter (n) in sensor and actuator is assumed to be 20 000.

To observe the effects of coupled analysis on structure with magnetostrictive material, time-domain analysis is carried out for both low- and high-frequency actuation for the same cantilever composite beam. Here, the structure is actuated through time-domain signal with sinusoidal actuation current I of 0.5-A ac current (I_0) and 0.5-A bias current (I_{dc}) at 50-Hz frequency. The equation of current is of the form

$$I = I_0 \sin(\omega t) + I_{\mathrm{dc}} \tag{100}$$

Figure 8 shows the time and frequency-domain representation of a 50-Hz sinusoidal signal.

Direct transient dynamic analysis is performed for 200 time steps to calculate open circuit voltage of the sensor and beam-tip velocity. The time step considered is equal to 50 ms. Figure 9(a) and (b) shows the tip velocity for unidirectional laminate (0° ply angle) and 90° cross ply laminate, respectively. The figures show the responses for both coupled and uncoupled constitutive models for magnetostrictive material. Following are the observations that can be made from these figures: (i) for 0° laminate, the effect of coupling in the constitutive model is not very significant, while in the 90° laminate, we see significant difference; (ii) coupled models tend to give lower amplitudes; and (iii) the coupled model, in addition to the given lower amplitudes, also shifts the phase. Next, we compute the open circuit voltage developed across the sensor patches and this is shown in Figure 10(a) and (b), respectively. These figures show that the effect of coupling on the voltages is significant for 90° composite. One important aspect to be noticed is that the voltage developed is significant (of the order of millivolts) considering that a small amplitude current was only fed into the actuator.

6 SUMMARY

In this article, FE modeling of composite structures with embedded smart sensors/actuators is presented. Three different smart sensors/actuators, namely, the

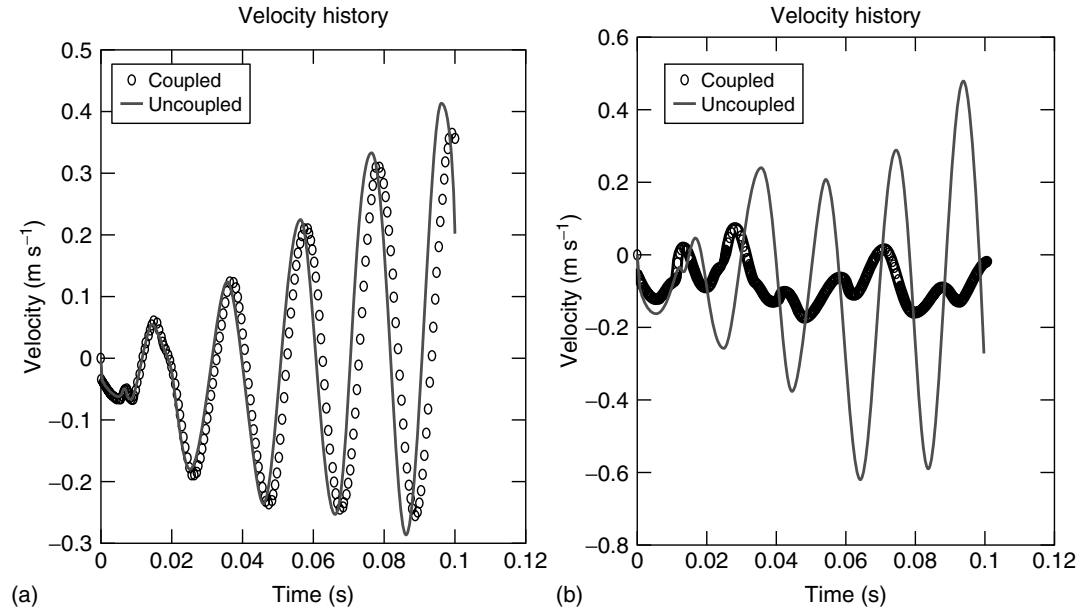


Figure 9. (a) Tip velocity for 0° laminate, (b) tip velocity for 90° laminate.

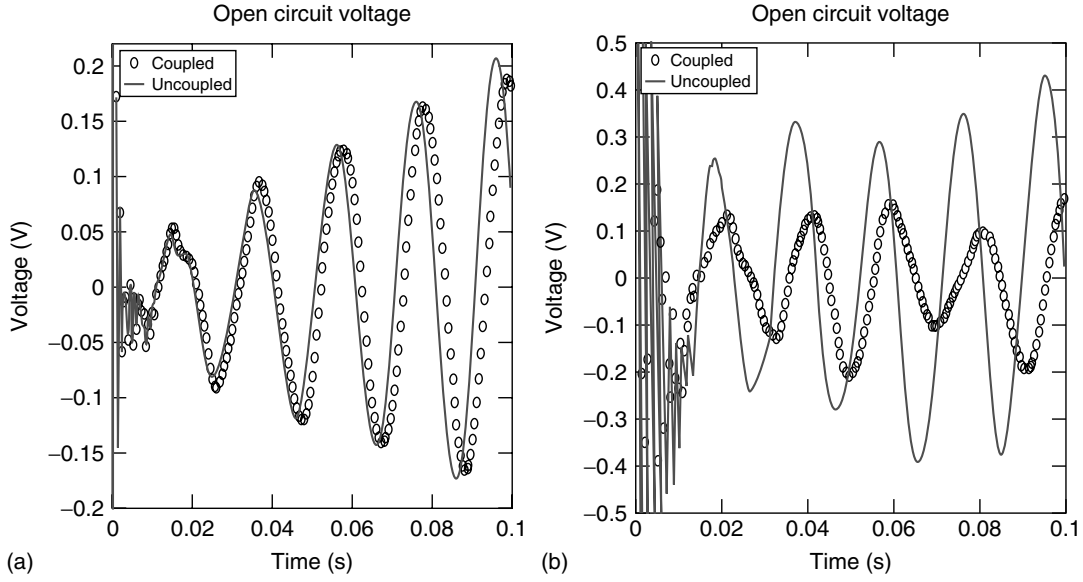


Figure 10. (a) Open circuit voltage for 0° laminate and (b) open circuit voltage for 90° laminate.

PZT, Terfenol-D, and PFCs are considered for FE modeling. The general FE procedure is outlined and specialized to a few specific cases such as the quadrilateral element for the PZT sensor/actuator case and the beam element formulation for the Terfenol-D sensor/actuator. The efficiency of the formulated element is demonstrated using some numerical examples.

REFERENCES

[1] Bathe KJ. *Finite Element Procedures*, *Third Edition*, Englewood Cliffs, NJ: Prentice Hall, 1996.

[2] Cook RD, Malkus RD, Plesha ME. *Concepts and Applications of Finite Element Analysis*. John Wiley & Sons: New York, 1989.

[3] Reddy JN. *Energy Principles and Variational Methods in Applied Mechanics*, *Second Edition*, John Wiley & Sons: New Jersey, 2002.

[4] Varadan VK, Vinoy KJ, Gopalakrishnan S. *Smart Material Systems and MEMS: Design and Development Methodologies*. John Wiley & Sons: UK, 2006.

[5] Srinivasan AV, McFarland DM. *Smart Structures: Analysis and Design*. Cambridge University, USA Press, 2000.

[6] Culshaw B. *Smart Structures and Materials*. Artech House: UK, 1996.

[7] Gandhi MV, Thompson BS. *Smart Materials and Structures*. Kluwer Academic Publishers: Netherlands, 1992.

[8] Jones RM. *Mechanics of Composites Materials*. McGraw Hill: New York, 1975.

[9] Tsai SW. *Introduction to Composite Materials*. Technomic Publishing Company: Connecticut, 1980.

[10] IEEE Standard 176-1978, *IEEE Standard on Piezoelectricity*. The Institute of Electrical and Electronics Engineers, 1978.

[11] Roy Mahapatra DR, *Spectral Element Models for Wave Propagation Analysis, Structural Health Monitoring and Active Control of Waves in Composite Structures*, Ph.D. Thesis, Indian Institute of Science: Bangalore, April 2004.

[12] Bent AA. *Active Fiber Composites for Structural Actuation*, Ph.D. Thesis, Massachusetts Institute of Technology, 1997.

[13] Sastry CVS, Roy Mahapatra D, Gopalakrishnan S, Ramamurthy TS. Distributed sensing of static and dynamic fracture in self-sensing piezoelectric composite: finite element simulations. *International Journal for Intelligent Materials and Systems* 2004 **15**:339–354.

[14] Chen SH, Wang ZD, Liu XH. Active vibration control and suppression for intelligent structures. *Journal of Sound and Vibration* 1997 **200**(2):167–177.

[15] Hwang W-S, Park HC. Finite element modeling of piezoelectric sensors and actuators. *AIAA Journal* 1993 **31**(5):930–937.

[16] Tzou HS, Tseng CI. Distributed piezoelectric sensor/actuator design for dynamic measurement/control of distributed parameter systems: a piezoelectric finite element approach. *Journal of Sound and Vibration* 1990 **138**(1):17–34.

[17] Prathap G, Bhashyam GR. Reduced integration and shear flexible beam element. *International Journal for Numerical Methods in Engineering* 1982 **18**:211–243.

[18] Hughes TGR, Taylor RL, Kanoknukulchal W. A simple and efficient finite element for plate bending. *International Journal for Numerical Methods in Engineering* 1977 **11**:1529–1543.

[19] Prathap G. *The Finite Element in Structural Mechanics*, Kluwer Academic Press: Dordrecht, 1993.

[20] Gopalakrishnan S. Behaviour of isoparametric quadrilateral family of lagrangian fluid finite elements. *International Journal for Numerical Methods in Engineering* 2002 **54**(5):731–761.

Chapter 42
Modeling Aspects in Finite Elements

Srinivasan Gopalakrishnan
Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

1 INTRODUCTION

Modeling is one of the key elements in structural health monitoring (SHM) studies. Proper mathematical models are required to postprocess the measured output to predict the damage location and its extent. Some of the commonly used mathematical models are the finite element methods (FEMs), spectral element methods, and boundary element methods. Some of these techniques are highlighted in articles **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators**; **Modeling for Detection of Degraded Zones in Metallic and Composite Structures**; **Damage Detection Using Piezoceramic and Magnetostrictive Sensors and Actuators**; **Modeling of Lamb Waves in Composite Structures**. Among them, FEM is the most powerful method and has great ability to model complex geometries with relative ease.

FEM is a powerful numerical technique to solve problems governed by partial differential equations over complex domains. It is normally adopted to solve forward problems in structures, that is, for a given a loading (input), one can easily determine the deformations the structures undergo (output). SHM, however, requires estimating the state of the structure from the measured output (deformation, velocities, acceleration, voltages, etc.) for a given predefined input (force) [1]. Hence, SHM falls under the realm of system identification problem. Such problems are also called the *inverse problems*.

Many researchers have used FEM to model flaws. For example, Yang *et al.* [2] used FEM to study the Lamb wave propagation in composite plates. Powar

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

and Ganguli [3] used FEM to model matrix cracks in composites beam and to study the performance of a rotating helicopter blade. FEM was used to obtain crack parameters in [4]. A new method to model cracks under the finite element (FE) framework using the Heaveside function is proposed in [5]. Many researchers have used FEs to model cracks to perform fracture mechanics studies. The related literature is very large and hence is not covered in this article.

Performing inverse problems using finite elements is not straightforward. This requires the determination of frequency response functions (FRFs) directly from the analysis. FEM, being predominantly a time domain method, is a time-consuming process for the estimation of FRF. Frequency domain methods, such as spectral FEMs [6, 7], are ideally suited for solving inverse problems since FRF, also called the *system transfer function* (STF), which directly relates the input and the output, is a direct by-product of such an analysis. Hence, for studying SHM problems through FE, one cannot employ methods requiring the determination of STF.

Alternatively, one can perform an SHM study by treating the problem as a forward problem and study the signature of the output to a set of predefined inputs to characterize the flaw.

Use of modal methods is one such example. There are quite a few works in this area [8–10], wherein the damage was predicted on the basis of the change in the natural frequencies. However, for very small damages, the change in stiffness is so insignificant that the changes in the first few natural frequencies are hardly noticeable. In such a case, one has to find different signatures that are sensitive for small stiffness changes. One such signature is to see the change in the curvature modes [11, 12]. These modes are very difficult to determine from the measured responses. Hence, modal analysis–based methods as such are not recommended for predicting very small damages.

Damage prediction is successful only if one can see a significant change in the output of the structure in a damaged state compared to the output of the structure in a healthy state. To entice a change in the output, especially in structures having very small damage, one has to fine-tune the frequency content of the predefined input signal. This is required to make the wavelengths comparable to the damage sizes. A change in the frequency content of the signal requires a change in the FE mesh. If all these are performed in an ad hoc manner, the process of analysis of a damaged structure takes significant amount of time. This article basically gives some ideas as to how to choose the mesh size for a given flaw size or for a given frequency content of the input signal.

Another factor that makes the interpretation of the signatures obtained from a cracked structure due to a predefined input difficult is the scattering of waves due to boundary reflections.

Although techniques such as time reversal methods [13, 14] are available for identifying the scattered components of wave response due to boundaries, these methods are difficult to apply under the FE computational frame work. Hence, for the damage detection to be successful, it is necessary to differentiate between the reflections arising due to the damage and the reflections from the boundary so that the boundary reflections are filtered out. Knowing the speed of the medium, it is possible to locate the flaw location from the filtered measured responses. This is the most difficult operation to accomplish especially for the short-duration broadband input signal on short finite structures. The problem gets amplified if the medium is dispersive wherein the wave forms of the input signal change its profile as it propagates, making it difficult to identify any reflections. If the medium is nondispersive, reflection also retains the input profile and makes identification of these a lot simpler. A tone burst signal, which is essentially a sine wave modulated over a small time window and travels nondispersively even in a dispersive medium, helps alleviate this problem to a little extent and hence is extensively used in SHM studies. Therefore, the choice of the type of the predefined input signal also plays a significant role in the SHM studies [7].

Failure modes in composites are many compared to metallic structures. The common failure mechanisms in composites are the delamination of the plies, the fiber breakage, and the matrix cracks. In built-up composite structures, debonds are also a common type of failure. Modeling these flaws efficiently is very crucial for their detection using the measured responses. The most common way of modeling flaws, such as cracks in FEM, is to introduce duplicate nodes along the crack front, which have the same nodal coordinates but a different node number. This procedure ensures discontinuity in the medium arising due to the presence of a crack. Such modeling procedures are extensively used in modeling fracture mechanics

problems. Although such models can very accurately model the flaw, the huge model size limits its use in real-time health monitoring. One of the important differences in the response behavior of the cracked structure as compared to the uncracked structures is the presence of a phenomenon of mode conversion [15, 16] present in the former. If such a phenomenon can be built in a simpler model, one can use such models effectively in SHM studies. One such model is the kinematics-based model, wherein the damaged structure is split up into multiple waveguides along the crack front. This process reduces the cracked structure into multiply connected framed structures. By enforcing kinematic relationship among the nodes of the waveguides, one can eliminate the intermediate nodes of the frame structure, thereby obtaining an FE-type model with a built-in damage. References 17 and 18 use such models under a spectral FE environment to study the wave propagation responses in delaminated beams and beams with fiber breakage. This method can be readily extended to model structures with such damage under an FE environment. This is discussed in detail under the modeling section of this article.

This article is organized as follows. First the relationship between the flaw size, frequency content of the signal, and the FE mesh is established. This is followed by a detailed section on different flaw models available under FEM. A number of numerical examples are given to highlight the importance of mesh size, frequency content of the input signal, and the type of forcing function in the SHM studies.

2 RELATIONSHIP BETWEEN MESH SIZE, FLAW SIZE, AND FREQUENCY CONTENT OF THE INPUT SIGNAL

SHM requires a powerful modeling tool to post-process the experimentally measured responses, due to a predefined input (forcing), to predict the location, extent, and severity of the flaws present in the structure. The predefined input is normally triggered through a piezoceramic actuator either statically or dynamically. To rapidly assess the health of the structure, the chosen mathematical model should be of smaller size and, in addition, it should be able to predict the location and extent of the flaws using a smaller number of measured sensor responses. The choice of the method depends upon the structure geometry, size of the flaw, and the frequency content of the input.

FE is a powerful modeling tool to solve structure problems defined by a partial differential equation over any complex domain. Although modeling flaws such as cracks or delamination are straightforward in FEs (which is addressed in the next section), the stress singularity near the crack tip requires a very fine mesh near the crack tip even for a static loading, which makes the model sizes enormously large. If the loading is dynamic, the mesh sizes required are even larger and it depends on the frequency content of the predefined input.

Frequency content of the predefined input is yet another parameter that should be carefully chosen, depending upon the flaw sizes. When the flaw sizes are larger compared to the dimension of the structure, then even static loading is sufficient. However, such flaws will be visible to the naked eye and as such SHM is not needed. However, if the flaws sizes are reasonably small compared to the smallest dimension of the structure, then one requires that the predefined input loading be dynamic in nature and the frequency content be of the order of few hundred hertz. Such low-frequency content problem comes under the category of structural dynamics. On the other hand, if the flaw sizes are very small (not visible to the naked eye), then the frequency content of predefined input signal should be of the order of few hundred kilohertz. These problems come under the category of wave propagation. The main difference between these two is that the latter is a multimodal phenomenon, wherein phase information becomes very important. Cracks in structures act like a boundary and induce a very small impedance mismatch at the crack boundary, which will induce reflections for a high-frequency input signal. These reflected signals can be effectively used to characterize, predict, and locate the cracks in a structure and also its extent. Hence, the choice between the structural dynamics or wave propagation analysis to be adopted depends upon the frequency content of the input signal, which in turn depends on the flaw size. All these have a bearing on the mesh sizes to be chosen for the FE analysis.

Now, two questions need to be answered: what should be the mesh size for a given input loading

and what should be the frequency content of the input signal for a given flaw size? For a given input loading, the mesh sizes should be chosen such that they are comparable to the wavelength. When the frequency content of the signal is high, the wavelengths are very small and hence the mesh sizes have to be small, which in turn increases the problem of sizes enormously. To determine the mesh sizes, first the predefined input signal is transformed into a frequency domain using fast Fourier transform (FFT), and a plot of the amplitude (in the frequency domain) and the frequency gives the frequency content of the signal. Let us denote this by ω (radian per second). If C_0 is the wave speed of the given wave mode, then the wavelength λ is given by

$$\lambda = \frac{2\pi C_0}{\omega} \tag{1}$$

Typically, the mesh sizes should be of the order of $\lambda/8$ [19]. If the mesh sizes are larger than what is given in equation (1), then the mesh boundaries start reflecting the input signal, thereby giving spurious indication of the presence of the crack.

Mesh sizes depend on the speed of the medium in which the signal is propagating. In order to make this statement clear, let us consider two mediums namely aluminum and composite. The wave speed in composite is approximately $3850\,\mathrm{m\,s^{-1}}$, while in aluminum it is $6000\,\mathrm{m\,s^{-1}}$. Let these mediums be subjected to an input pulse having a frequency content of 50 kHz. From equation (1), the wavelength in composites is about 77 mm, while in aluminum it is about 120 mm. Hence, for a given input, composite requires a more dense mesh.

Now, the second question is what should be the frequency content of the predefined input signal for a given flaw size for such a signal to induce an impedance mismatch at the crack boundary and induce a reflection. For this to happen, the wavelengths should be comparable to the flaw sizes. For example, we again consider the same example considered previously, that is, aluminum and composites with damage. The wave speed in composite is approximately $3850\,\mathrm{m\,s^{-1}}$, while in aluminum it is $6000\,\mathrm{m\,s^{-1}}$. Let us assume that both these mediums have a small crack of size 20 mm. Equation (1) can be rewritten as

$$f = \frac{C_0}{a} \tag{2}$$

where f is the frequency in hertz and a is the size of the flaw. From equation (2), substituting the speed of the composite and the aluminum medium, we see that, for inducing a reflection from the damage in composites, the frequency content of the input signal should be 192.5 kHz, while in aluminum medium the frequency content of the predefined input signal should be 300 kHz. In addition, if one needs the signal to travel nondispersively, one can create a tone burst signal created using a sine wave of the above-calculated frequencies. From the above discussion, it is clear that the frequency content of the input pulse and the mesh size are highly dependent upon the medium in which the waves are propagating.

The success of this analysis requires that the speed of the medium be known. The calculation of speed, especially for a dispersive system, is an involved process. The calculation of speed requires a deep understanding of the spectral analysis. References 6 and 7 give a good account of deriving the dispersion relation (speed vs frequency plots) for metallic and composite structures for both 1-D and 2-D waveguides. In most second-order systems, the waves are usually nondispersive and the speeds do not vary with the frequencies. However, for fourth-order systems like beams, the waves are highly dispersive and hence, at different speeds, the speeds will be different. Then the question that arises is, at what frequency should the speed be used in equation (1) to compute the mesh sizes? To answer this question, let us consider the dispersion plot of a composite beam for different ply orientations [6] shown in Figure 1. The plot shows the group speeds for axial and bending waves for different ply orientations of a laminated composite beam. The parameter r is the bending–axial coupling factor [6] for the beam. The speeds are normalized with the axial speed of an isotropic rod ($C_0 = \sqrt{E/\rho}$). The three horizontal lines are the axial wave speeds for three different ply orientations. One can clearly see that the axial speeds are constant for all frequencies and hence they are nondispersive. The figure also shows three nonlinear curves for the bending wave for three different ply orientations. These waves are highly dispersive in nature and hence their speeds are different at different frequencies. If the input signal is broadbanded, then the choice of wave speeds to be used in equation (1) becomes very difficult. In such cases, modulated signals are very useful since their active energies are

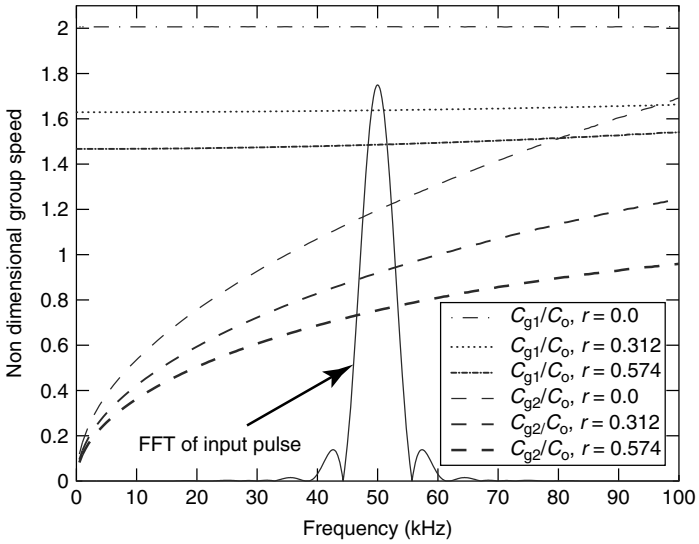


Figure 1. Dispersion plot for an elementary composite beam.

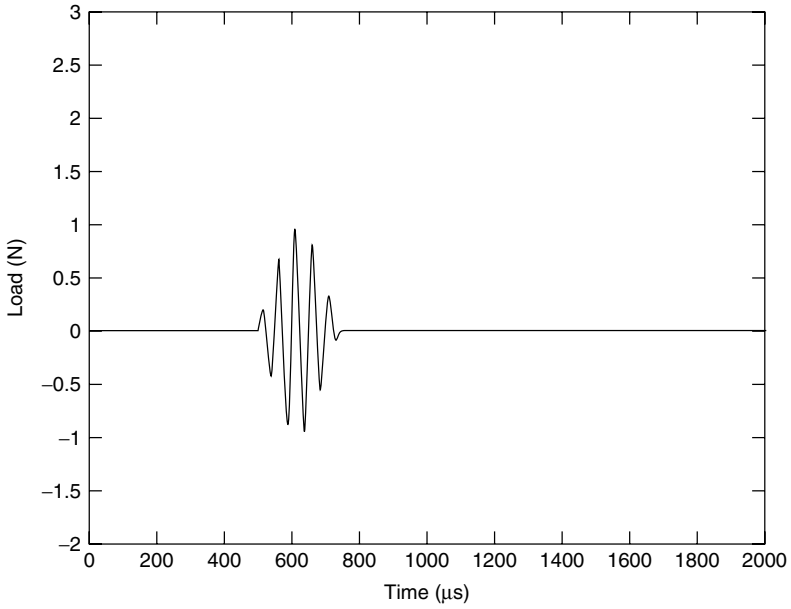


Figure 2. A tone burst signal modulated at 50 kHz frequency.

band limited over a small frequency band. One such modulated signal, also called the *tone burst signal*, is shown in Figure 2. This signal is created using a sine wave of frequency 50 kHz and modulated using a Hanning window. The FFT amplitude of this signal is superimposed in Figure 1 along with the dispersion plots. Through this superimposition, the wave speeds corresponding to 50 kHz can be easily computed for various propagating modes and this value can be used in equation (1) for calculating the wave speeds. In addition, modulated input pulse can force the wave to be nondispersive even in a dispersive medium. It can

be seen from Figure 1 that the modulated frequency of the input signal is 50 kHz. That is, such a pulse excites only those modes that are lying very close to 50 kHz and all other modes do not participate in the response. As a result, the group speeds of the wave are a function of only a small set of frequencies close to the modulated frequency, which forces the response to be nondispersive even in a dispersive medium.

An alternate way of fixing the mesh sizes is by looking at the stiffness of the structure. It is well known that the presence of a flaw reduces the stiffness of the structure. This stiffness reduction depends on the size of the flaw. If the flaw size is small, it causes an insignificant change in the stiffness of the structure and hence an insignificant change in the first few natural frequencies of the structure. However, for large flaw sizes, the stiffness change is significant, and hence the modal frequencies. This is shown in Figure 3 for a laminated composite beam of 20 cm length with two different delamination sizes, where the bending stiffness is plotted as a function of the frequency. For a delamination of 1 cm, one can notice the frequency change only beyond 14 kHz, while for a delamination of 5 cm one can see that the frequency change happens much earlier at 3 kHz. That is, for small flaw sizes, only higher modes get excited and hence, to capture all higher modes accurately, one needs a very fine mesh. Therefore, for such small flaw sizes, modal methods, which are extensively used in SHM studies, are not suitable. One needs wave propagation analysis.

Accuracy of the response to high-frequency input depends on the density of the FE mesh. For a reasonably dense mesh, the wave response predicted may be accurate; however, it may show period error. To reinforce these ideas better, let us consider a simple aluminum rod of 2.0 m length and 0.01 m^2 cross section, with Young's Modulus (E) of 70 GPa and a density (ρ) of 2600 kg m^{-3}. The wave speeds in the aluminum can be calculated from the formula $C_0 = \sqrt{E/\rho} = 5189$ m s^{-1}. This rod is subjected to an input signal, as shown in Figure 4 (inset), which has a frequency content of 46 kHz. From equation (1), the wavelength can be calculated, which is equal to 0.11 m. In order to capture the wave behavior accurately, at least eight elements per wavelength are required, that is, an element length of 0.014 m. Hence, for a length of 2.0 m, at least 145 1-D FEs are required for modeling. Figure 5 shows the axial wave responses at the cantilever tip for different number of elements. We have used 250 1-D rod elements to get a fully converged solution, which shows an initial pulse at around 100 μs and a reflected pulse from the boundary at around 420 μs. Figure 5 also shows (in the inset) the period error when the mesh sizes are inadequate. Inadequate FE mesh for a high-frequency input pulse results in mass or inertial

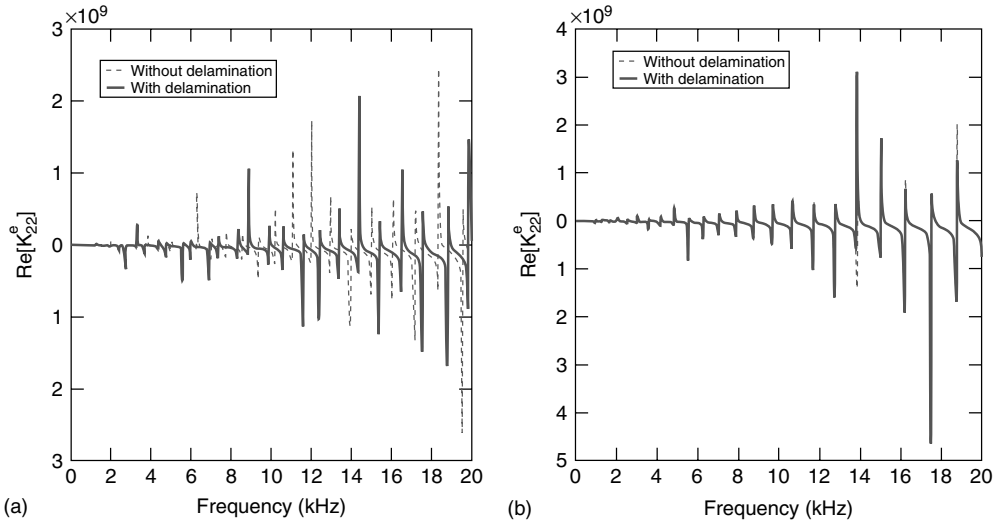


Figure 3. Stiffness changes as a function of frequency for different delamination sizes: (a) 5 cm and (b) 1 cm.

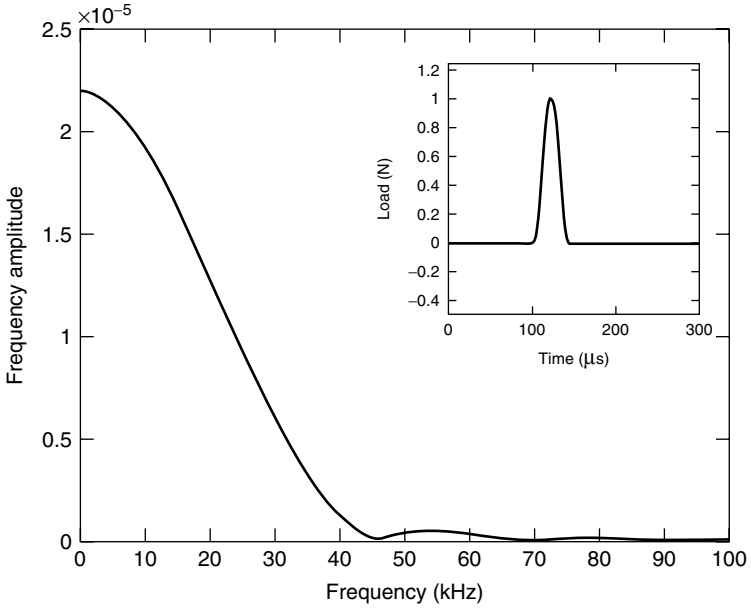


Figure 4. High-frequency input load with its Fourier transform.

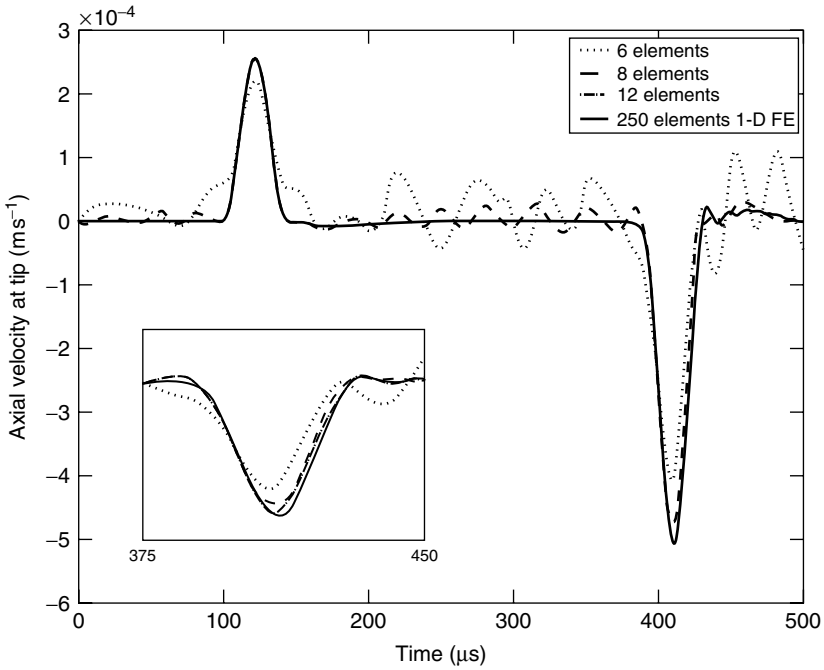


Figure 5. Longitudinal wave responses in a rod for different element discretizations.

distribution not being accurate. As a result, the wave speeds predicted by FE analysis are highly inaccurate, resulting in period error. In addition, if the mesh sizes are much smaller than what is required, then the mesh boundaries act as a fixed boundary and start reflecting responses from these boundaries. These are clearly seen in Figure 5 for very small mesh density. Hence, for very high frequency content input pulse, which is normally the case for most SHM problems, a fine mesh is an absolute necessity.

3 MODELING OF FLAWS IN FINITE ELEMENT METHOD

Materials such as composites have many modes of failures. Among these, delamination and fiber breakages are the important modes of failure. These failure modes correspond to horizontal and vertical cracks in metallic structures. Modeling of these is quite different in some of the methods. This section outlines some of the methods adopted to model the above failure modes in composites and metallic structures under an FE environment. These methods can be classified as follows:

1. stiffness reduction method (SRM),
2. duplicate node method (DNM), and
3. kinematics-based method (KBM).

These methods are explained in detail in the subsequent paragraphs.

1. Stiffness reduction method

It is quite well known that the presence of flaws causes the reduction of stiffness in the structure. A simple way of modeling flaws is to incorporate the stiffness loss in the region of the flaw by modifying the material properties P (where P can signify Young's modulus, shear modulus, density, etc.) to αP where, $\alpha < 1$. This concept is demonstrated in Figure 6.

Figure 6(a) shows the actual model of a laminated beam with a through-width delamination and Figure 6(b) shows the equivalent stiffness-reduced model. This procedure could be adopted to model any number of flaws in the structure. However, such a modeling, although good in estimation of the remaining life of the structure, is not suitable for determining the extent of damage, which is paramount in SHM studies. In [20], such a model was used to perform SHM studies on large civil structures.

2. Duplicate node method

A better way of modeling flaws in FE is to completely model the entire crack front. This can be performed in the following way. Modeling can be done using either the 1-D beam element or the 2-D plane stress/strain FEs using the concept of duplicate nodes. In the case of beams, the modeling of a flawlike delamination is done by keeping the two nodes in the same place, one for the elements above the flaw and the other for those below it. This is shown in Figure 7(a). Here, elements 1 and 2 are at the left part of the flaw,

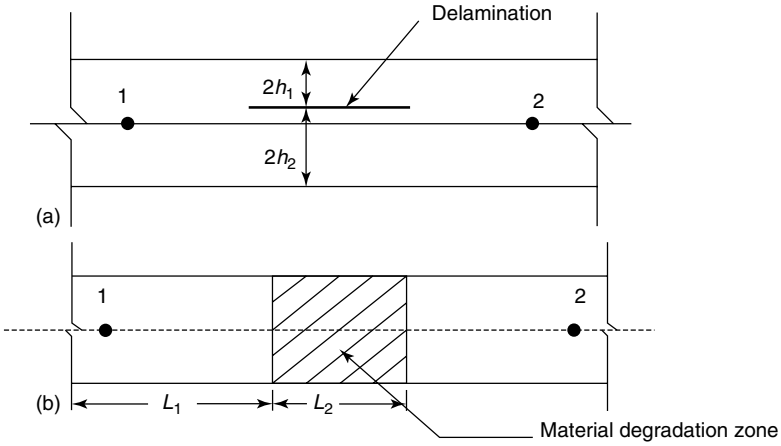


Figure 6. Modeling of flaws using the stiffness reduction method: (a) actual model and (b) stiffness-reduced model.

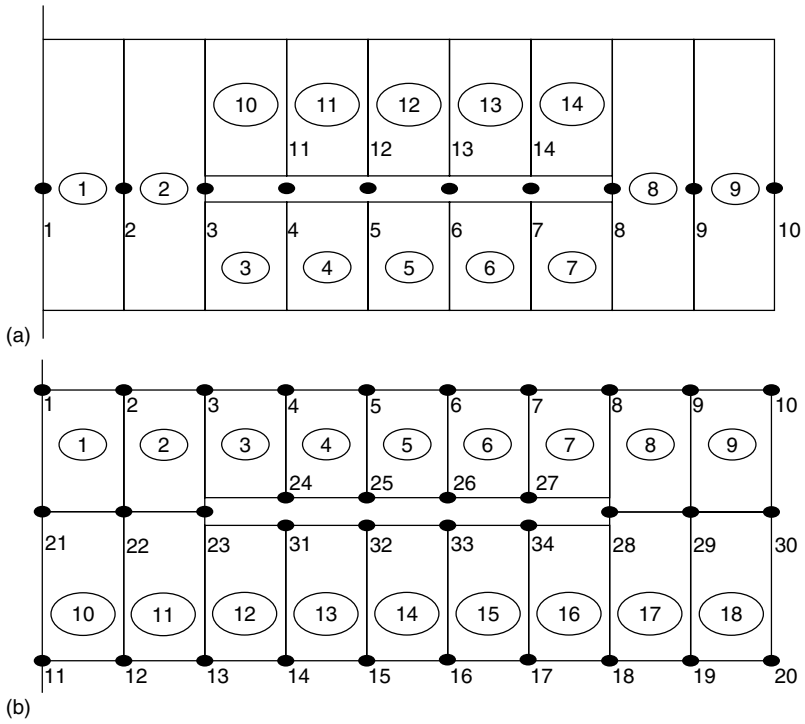


Figure 7. Modeling of flaws using the duplicate node method: (a) using beam elements and (b) using plane stress/strain elements.

while elements 8 and 9 are to the right part of the flaw. All other elements are in the flawed zone, either above or below the flaw. The zone below the flaw is modeled with elements 3–7 and the one above by elements 10–14. The flaw is modeled through proper nodal connectivity of these elements. That is, for the healthy zone, element 1 is connected with nodes 1 and 2 and element 9 is connected with nodes 9 and 10. All these nodes are in the midplane of the corresponding beam elements. For elements below the flaw, say, element 3 is connected with nodes 3 and 4, where nodes 3 and 4 are not in the midplane of this element. This may create high bending–stretching coupling, which is the objective of the modeling process to induce mode conversion. Similarly, for elements above the delamination, say, element 10 is connected through nodes 3 and 11, where nodes 3 and 11 are not in the midplane of this element. As mentioned earlier, nodes 4 and 11 are in the same place and are not connected with any direct element. These are termed as *duplicate nodes*. Similarly, nodes 5 and 12, nodes 6 and 13, and nodes 7 and 14 are duplicate nodes. In the flaw zone, the lower part of the elements connects with nodes 4, 5, 6, and 7 and the upper part of the elements connects nodes 11, 12, 13, and 14. If these nodes are merged with their corresponding duplicate nodes, the beam will be healthy. In addition, with these duplicate nodes, different kinds of contact or gap elements can be used in FE analysis to prevent interpenetration of the crack front.

A similar procedure could be adopted in modeling a flaw using 2-D plane stress/strain elements, which are shown in Figure 7(b) using duplicate nodes. The figure is self-explanatory. The extending of this procedure to model fiber breakages and multiple delaminations in composites is quite similar and straightforward.

3. Kinematics-based method

This is quite a powerful method for modeling delaminations, fiber breakage, and multiple delaminations.

This approach was used under the spectral FE environment for modeling and detection of different types of damages in 1-D waveguides (see [17, 18, 21]). In this article, we extend this method to work with FEM. Its use in SHM studies till now has been explored only to model delaminations and fiber breakages in 1-D laminated composite structures. This method of modeling is limited to through-width straight-line cracks.

The main idea behind this modeling approach is to cut the beam structure into multiple elements (domains) along the crack front. The stiffness and mass matrices for each of the subdomains is generated. The intermediate nodes away from the crack front arising due to this splitting are then connected to the nodes along the crack front through rigid links in order create a coupling between the axial and transverse displacements. The procedure for modeling delaminations and fiber breakage are quite different and hence described separately.

1. Modeling of single and multiple delamination

Here, let us consider a composite beam with a delamination, the dimensions of which are shown in Figure 8(a). This delaminated beam is split into base laminates and sublaminates, as shown in Figure 8(b). The equivalent beam model with eight nodes makes up this delaminated beam. Each subdomain (element) is indicated by a number within a circle. Let each node support three degrees of freedom, namely the axial deformation $\boldsymbol{u}$, the transverse deformation $\boldsymbol{w}$, and the slope φ. These are represented by a vector $\{\mathbf{u}\}_i$, where $\boldsymbol{i}$ represents the domain (element) of interest. Note that the through-width delamination lies between nodes 4 and 7. It is assumed in this modeling that there is no contact action between

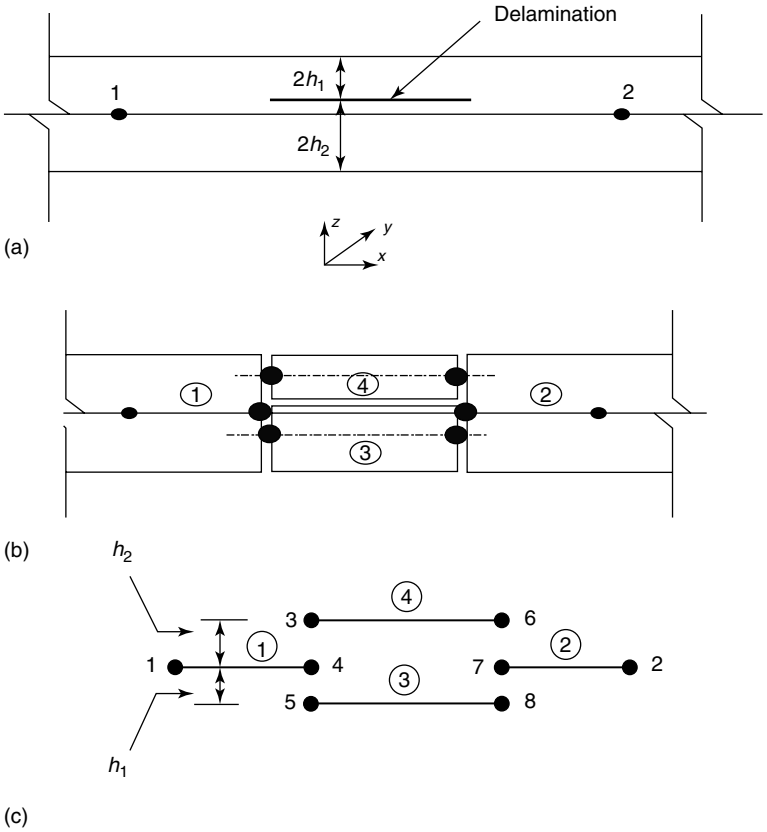


Figure 8. (a) Original delaminated beam, (b) splitting of the damaged beam into base and sub laminates, and (c) equivalent beam models.

the sublaminates at the plane of delaminations and the cross sections are perfectly straight at the interfaces. The connections between nodes 3-4, 4-5, 6-7, and 7-8 are made with the help of the rigid links to simulate the bending–axial coupling. This model does not take care of crack tip stress singularity as it is not of importance in the context of damage detection. A similar approach can be adopted when the laminated composite structures has multiple delaminations. Some examples on the use of this model in the context of SHM are given in the next section.

2. Modeling of fiber breakage

Here, we have adopted the same procedure that was used for modeling delamination. Let us consider a beam with a fiber breakage represented in Figure 9(a). The split-up model is shown in Figure 9(b) and the

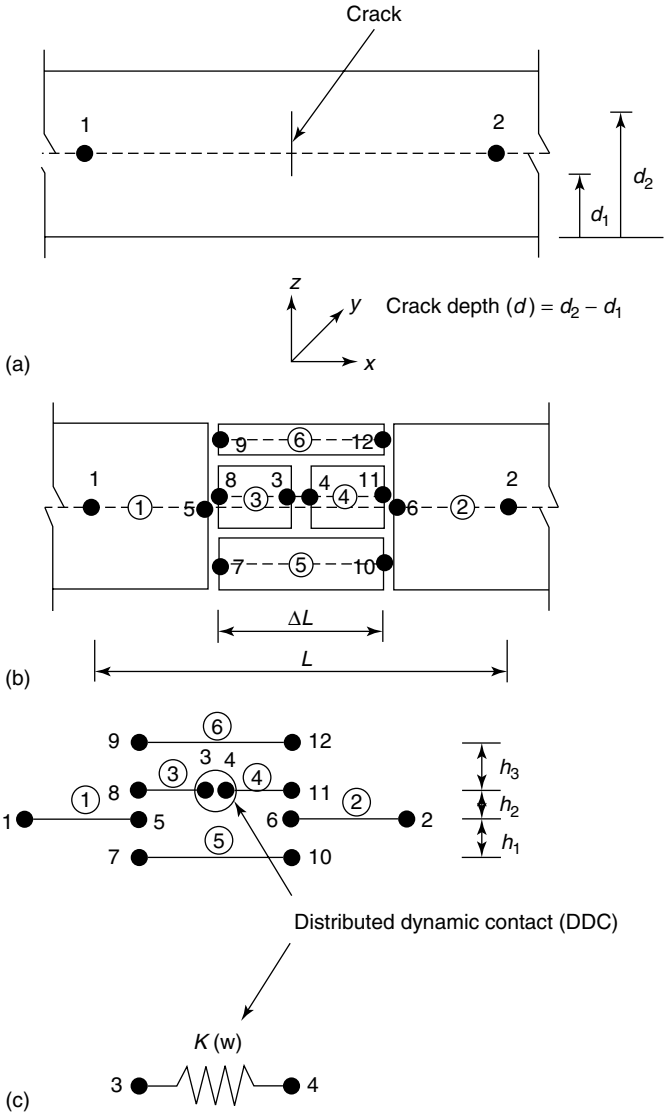


Figure 9. (a) Original beam with fiber breakage, (b) splitting of the damaged beam into base and sub laminates, and (c) equivalent beam models.

beam model representation is shown in Figure 9(c). In this model, we assume that there exists a distributed dynamic contact at crack surfaces and the crack surface remains perfectly straight. The nodes along the left and the right side of the crack are connected by rigid links to simulate the bending–axial coupling behavior. That is, nodes 9-8, 8-5, 5-7, 10-6, 6-11, and 11-122 are connected by a rigid link. Unlike the previous case of delamination, there is a hanging interface between nodes 3 and 4, which are connected by a nonlinear spring to simulate the distributed dynamic contact. The spring constant had to be chosen in such a manner that this simplified beam model simulates the actual waveguide behavior at high frequencies. As before, the effects of crack tip stress singularity is ignored in this model. Some examples of this model usage are given in the next section.

4 NUMERICAL EXAMPLES

A few numerical examples are given in this section, wherein some of the modeling methods described in the previous section are used in the context of SHM. The main objective of any SHM analysis is to determine the presence of damages and their extent and severity from the measured responses. For the above analysis, the SRM method may not be useful. If one is required to predict the location and also the extent of damage, then the method based on duplicate nodes is very useful. In addition, if the flaw is 2-D in nature, the KBM method is not useful. There are many situations wherein the designer wants to estimate the overall life of the structure due to the presence of a number of damages, in which case, the SRM method is quite sufficient for modeling. Hence, the choice of the model entirely depends upon the level of sophistication required for the analysis. The next few paragraphs describe numerical examples related to static, free vibration, and wave propagation in the context of SHM using the models described above.

4.1 Static and free vibration analysis of a cantilever beam using DNM

Here, two different studies are performed. In the first case, the results from two different DNM models (1-D beam and 2-D plane stress models) are compared to see the effectiveness of each of these models for SHM studies. In the next case, the results from the DNM- and KBM-based models are compared.

Here, a unidirectional 12-layer laminated composite beam of total depth 1.8 mm and length 500 mm is considered for the study. A through-width delamination is symmetrically placed with respect to both top and bottom and the sides. The lamina is assumed to be orthotropic with the Young's Modulus in the two perpendicular directions being E_1 equal to 181 GPa, E_2 equal to 10.3 GPa, and the shear modulus G_{12} equal to 28 GPa. The FE model details are as shown in Figure 7 and Table 1 gives the percentage increase in the tip deflection for a tip unit load. From the results, we find that the agreement between the two DNM models is closer for large delaminations.

Next, a free vibration study is performed to assess the performance of the two different DNM models. These are plotted in Figure 10. Both the models (beam and 2-D) show more or less similar results. When the delamination is 10% of the length of the cantilever, the 15th natural frequency reduces 20%, corresponding to the corresponding frequency

Table 1. Comparison of static results for two different DNM models

Delamination (%)	Beam DNM model (% increase in tip deflection)	Plane stress DNM model (% increase in tip deflection)
10	0.39	0.08
20	0.91	0.61
30	2.35	2.03
40	5.15	4.80
50	9.77	9.37
60	16.63	16.24
70	26.13	25.75
80	38.90	38.32

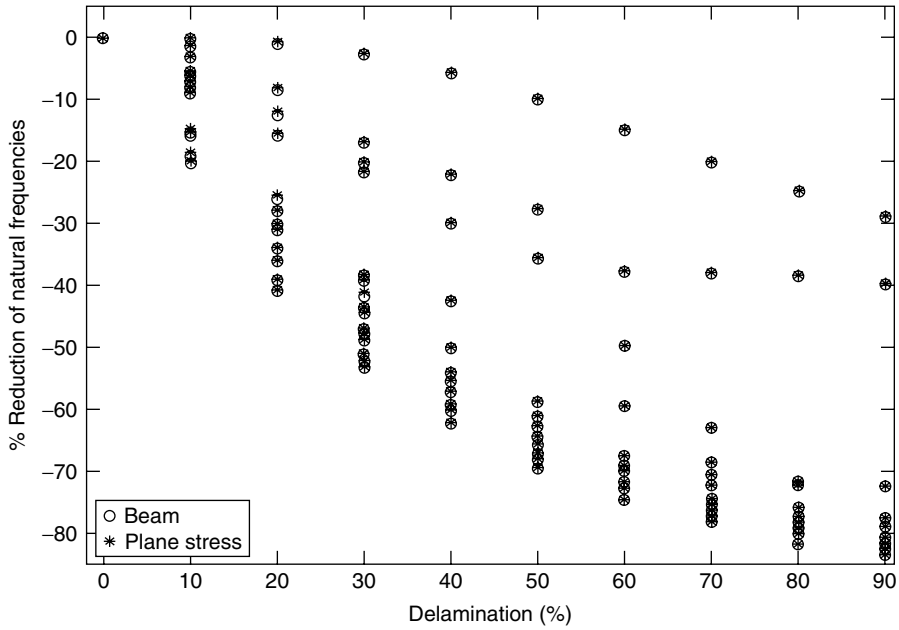


Figure 10. Comparison of free vibration results for two different DNM models.

of the healthy beam. Similarly, for 90% delamination, the first natural frequency reduces 29%, the second natural frequency reduces 40%, and the third natural frequency reduces 72%. From these studies, it is quite clear that beam DNM models can give comparable results with the plane stress DNM models. This has huge implications in SHM studies as beam models can give substantially smaller problem sizes, which is one of the key requirements for SHM simulations.

In the next exercise, the static analysis is performed using the DNM (plane stress model) and the KBM model on a metallic cantilever beam of length 635 mm, width 25.4 mm, depth 25.4 mm, the Young's modulus of the beam assumed to be 70 GPa, and shear modulus 27 GPa. The results are generated for a given crack length of $L/10$ located at a different location along and across the beam. The DNM model has 100 elements with due care taken to have a finer mesh near the crack tip. The KBM model has only four elements, as shown in Figure 8(c). The beam is subjected to a load of 5000 N. The results of these two models are tabulated in Table 2. From Table 2, it is clear that the four-element KBM model is able to capture the damaged behavior of the cantilever metallic beam quite accurately. This again results in substantial reduction in problem sizes. However, most SHM problems require high-frequency content loads for damage detection, which needs wave propagation analysis. This is addressed in the next subsection.

4.2 Response analysis of a delaminated cantilever composite beam

The aim of this numerical example is to bring out the differences in the responses predicted by the KBM model as opposed to the DNM model (2-D plane stress model). Here, we consider a delaminated AS/3505-6 graphite epoxy composite cantilever beam of length 800 mm, with a midplane delamination of size 50 mm, introduced at 400 mm from the root. The beam is 16 mm thick and 10 mm wide. The beam is subjected to a transverse tone burst signal of 20 kHz frequency content. The flexural group speed in composites is around 3850 m s^{-1}. As per equation (1), the wavelengths at 20 kHz frequency is around 192.5 mm, which is more than the size of the damage of 50 mm. Hence, in order to induce an impedance mismatch at the crack tip, the frequency content of the pulse needs to be increased. Alternatively, the mesh size can be computed by assuming the wavelength as 50 mm, which is the size of the

Table 2. Comparison of static results for DNM and KBM models for different crack locations

Location of the crack of size $L/10$ in the cantilever beam	Tip deflection based on plane stress DNM model (mm)	Tip deflection based on four-element KBM (mm)	% difference
Crack located symmetrically lengthwise and depthwise	1.78×10^{-4}	1.804×10^{-4}	1.37
Crack located symmetrically lengthwise and 17.78 mm from the bottom	1.778×10^{-4}	1.90×10^{-4}	5.92
Crack located symmetrically depth wise and 228.6 mm from the fixed end	1.786×10^{-4}	1.77×10^{-4}	1.02
Crack located symmetrically depth wise and 342.9 mm from the fixed end	1.773×10^{-4}	1.84×10^{-4}	3.78

damage. For accurate modeling, we require eight elements per wavelength, which requires the element size to be approximately equal to 6.25 mm. In order to model 800 mm, one requires a total of 128 beam elements to model this problem. The KBM model is simulated using 250 elements, while the DNM model contains 2560 plane stress triangular elements. A small portion of the 2-D DNM model mesh near the crack tip is shown in Figure 11(a). The KBM model near the crack tip is simulated using rigid connections, as explained in the previous section. Figure 11(b) shows the transverse velocities predicted by these two models. One can clearly see two reflections: the first from the damage and the second from the fixed boundary. It can be clearly seen that the KBM model predicts the reflections from the crack tip as accurately as the DNM model at a fraction of the computational cost.

In this last example, all the three models, namely the SRM, DNM, and KBM, are compared on a composite cantilever beam that was considered in the last example, and subjected to a transverse broadband impact load shown in Figure 4. Two different damage types, namely the delamination and fiber breakage, both of which are of 20 mm size, are considered on this beam and are symmetrically placed as shown in Figure 12. The beam with 20-mm delamination is modeled by the DNM model, while the beam with 20-mm fiber breakage is modeled using the KBM model. The results from these are compared with the 1-D beam FE model with stiffness reduced by 50% in the small region close to the damage. The main objective of this example is to not only to compare the responses predicted by these models but to also look at the possibility of substituting the detailed KBM model with the SRM model for SHM analysis.

The damage is modeled as a horizontal crack (delamination) using the DNM model, where a total of 3500 2-D triangular elements are used. The vertical crack or the fiber breakage is modeled using the KBM method, while the SRM model is obtained from the healthy beam 2-D model by reducing the stiffness (composite modulus $\overline{Q}_{11}$) by around 50%. This input has a frequency content of 44 kHz. For this force, with a wave speed of 3850 m s^{-1}, it requires nearly 320 beam elements. However, this simulation is performed using 500 elements for the SRM and KBM models, respectively. The hanging interface in case of the KBM model of the spring is simulated with a spring constant of 1×10^{-3} times the $\overline{Q}_{11}$. The different beam configurations with the associated damages are shown in Figure 12 and the transverse response histories are shown in Figure 13.

From Figure 13, we can clearly see that each of these models pick up the reflection from the damage at around 0.5 ms and the responses show very little change between different damage types of the same size and location. Hence, it can be concluded that the SRM model, which is the easiest to model, can indeed be used to get the response estimates on a damages structure. However, if one needs to further predict the location and extend of damage, more local models such as the DNM method may be required.

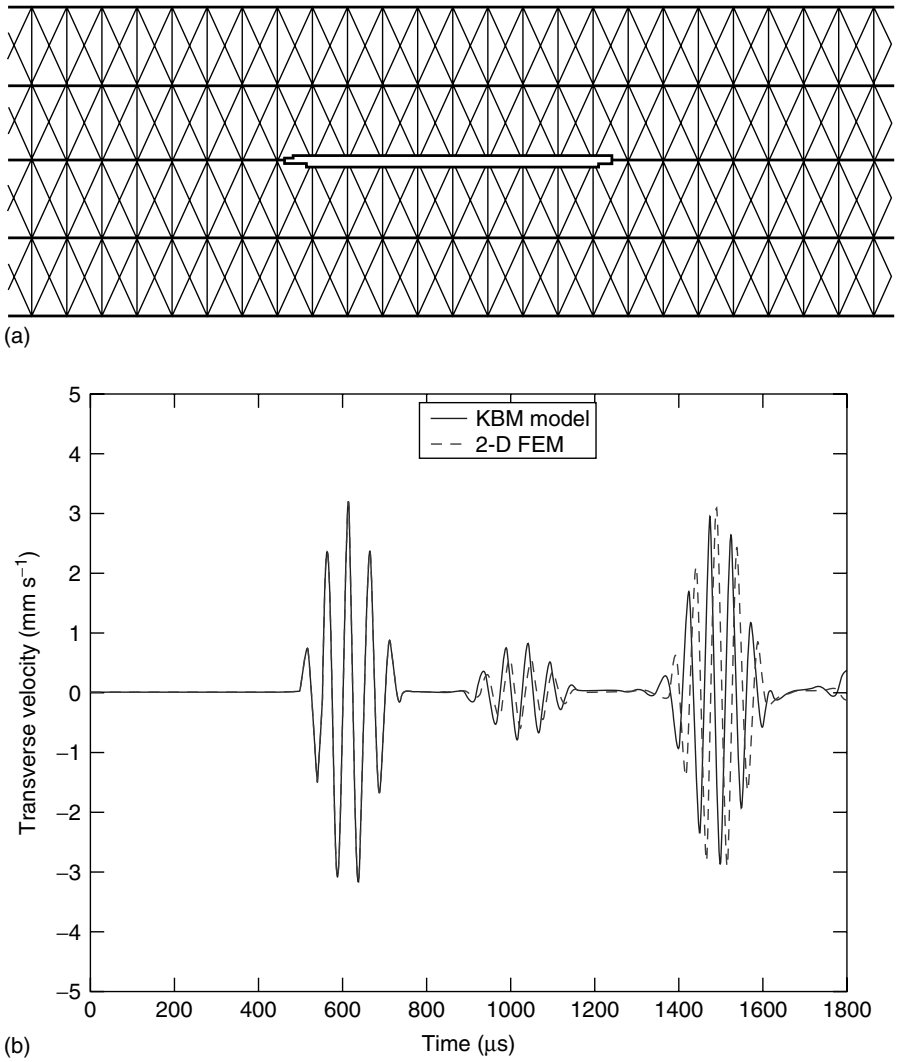


Figure 11. Comparison of wave propagation responses for DNM and KBM models for a delaminated composite beam. (a) DNM model near the crack tip and (b) transverse response history at the tip.

5 MODELING SUGGESTIONS FOR 2-D AND 3-D STRUCTURES

All the examples shown so far dealt with 1-D waveguides. This is because the modeling concepts outlined are better understood with 1-D waveguides. However, most practical structures are either two or three dimensional in nature. Here we outline some suggestions for modeling flaws in general and in 2-D and 3-D structures in particular:

1. The procedure for deciding mesh sizes and the frequency content of the input signal for 2-D and 3-D structures is the same as that outlined in Section 2 of this article. However, the mesh size obtained from equation (1) should be applied for all the dimensions of the structure.
2. For modeling 1-D, straight-line cracks for SHM studies, the KBM method is quite sufficient for damage detection purposes. However, a more detailed DNM model will be required if one needs to perform life estimation studies.

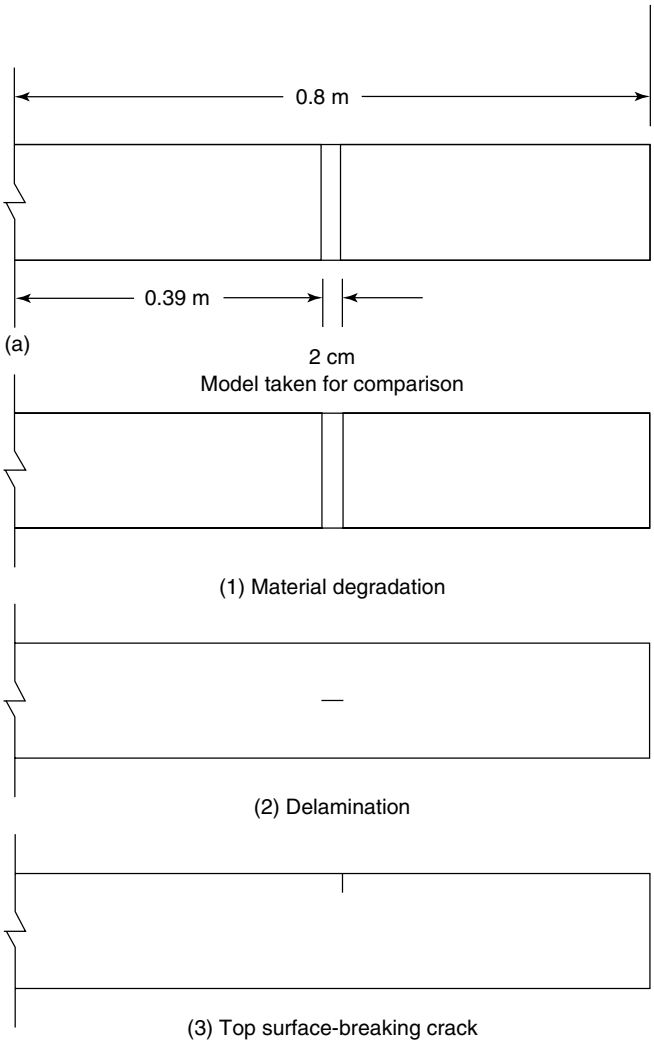


Figure 12. (a) Basic model showing the damage: (1) SRM model with 50% stiffness reduction in the cracked region, (2) delamination modeled using DNM, and (3) vertical surface breaking crack modeled using KBM with a spring constant of 1×10^{-3} of $\overline{Q}_{11}$.

3. Deriving FEs based on KBM is not straightforward and hence this method is seldom used in 2-D or 3-D structures with flaws. However, for certain crack orientations, some crack functions are derived, which capture the mode coupling aspects in 2-D structures. These functions can be readily incorporated into the FE formulation to obtain a simplified damage model. This is outlined in [22].
4. If the frequency content of the signal is large, that is, if the mesh sizes are very small, then we seldom use a graded mesh as used in static analysis of a cracked structure. We normally choose very small but equal mesh sizes as dictated by equation (1). This was used in some of the numerical simulations in the last section (see Figure 11a).
5. All the examples demonstrated the use of various modeling methods for either perfectly horizontal or perfectly vertical cracks, which is normally the case in composites. However, in metals, the commonly occurring cracks are inclined in

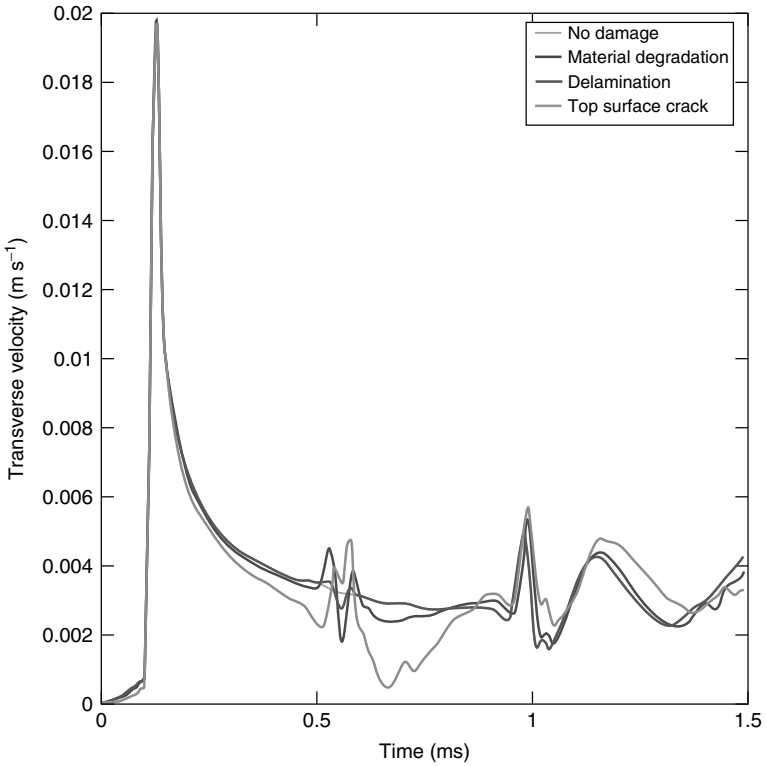


Figure 13. Comparison of wave propagation responses for different models.

nature. For such structures with inclined cracks, the DNM method is normally used, although the SRM method may also be employed. However, for matrix cracking type in damages, where the determination of the location is not that important, the SRM method is ideally suited.

6 MODELING PITFALLS IN FEM FOR SHM AND THEIR REMEDIES

One of the fundamental objectives of the SHM is to rapidly obtain the state of the structure. It was mentioned earlier that the FE model sizes depend on the size of the flaw. For very small flaw sizes, the frequency content of the signal should be large, which leads to enormously large problem sizes, which takes many hours to obtain the solution. This aspect defeats the very purpose of SHM. This is one of the bottlenecks in SHM modeling.

Unlike in a beam, wherein most damages are through-width, single-dimension cracks, in most 2-D and 3-D structures, the damages have an area. For modeling these type of flaws, the DNM method is ideal, although the mesh sizes will be very large. One way to reduce the size of the model is to adopt reduced-order FE models. Reduced-order FE models have only a few nodes retained, which correspond to either sensor locations or the locations where forces are applied, while all other nodes are condensed out. For static analysis, simple static condensation as outlined in [23] is sufficient. For dynamic loading, many reduced-order models are available, which are outlined in [24]. The procedure to model and its application to SHM studies are reported in [25].

Spectral FEM [6] can model very large structures with very limited model sizes irrespective of the frequency content of the input signal. However, modeling arbitrary geometries in spectral FEM is practically impossible. The solution to such problems is to marry these two methods. For example, in the

region very close to the crack, FE modeling can be used and, in the region far away from the crack, spectral FEM [6] could be used. This is because spectral FEM can model long and straight edge surfaces with only one element. Such an approach was adopted in [26] to model cracks in isotropic beams. Alternatively, a new FE formulation based on "partition of unity" [27] can be used to model the zone near the crack and, far away, one can use spectral FEM. Such a method will be very useful for SHM analysis of large-scale structures.

A complete SHM analysis requires modeling of flaws, their detection from the measured responses, and their location, extent, and severity. Today, many general-purpose modeling tools such as NASTRAN, ANSYS, ABAQUES, and ALGOR are available for modeling cracks using any of the above methods described in this article and also for assessing its severity. Some of these software can also perform optimization studies to decide on the number of sensors required and their locations. However, most of these tools cannot perform SHM-centric analysis as they do not have the other components of SHM, namely the damage-detection algorithms and the signal-processing algorithms. Since SHM concepts are built in the design of today's civil, aerospace, mechanical, and ship structures, the need of the hour is to develop SHM-centric FE software that integrates many different modules and new modeling concepts. Such an effort will go a long way toward taking SHM development to a higher level.

7 CONCLUSIONS

This article presents some of the modeling concepts for analysis of SHM-related problems. The work presented here can serve as a useful guideline for those involved with SHM modeling. It clearly brings out the relationship that exists between the predefined input, its frequency content, the wavelength, and mesh size. The article describes various modeling methods available under the FE environment and clearly brings out the situation wherein such models can be used. A number of numerical examples are provided to elucidate the concepts brought out in this article.

The aim of this article is to help the analyst in obtaining fast and accurate damage estimates, location, extent, and severity. These guidelines are not complete by themselves. There are many other rules that govern the accuracy of FE solutions under dynamic loading, and these are not covered here. For example, in situations where the gradients of the field variables such as strains are steep, as in the case of structures with cracks, one needs to have finer mesh discretization near the crack tip. Although one can use a graded mesh for solving such problems under static loading, for wave propagation problems, it is preferred to have a very small size uniform mesh so that response distortions do not occur at the mesh boundaries. For structural dynamics and wave propagation problems involving defects, the choice of mass matrix also plays a very important part in obtaining accurate damage estimates. If the frequency content of input loading is high, as in the case of wave propagation problems, the use of lumped mass matrix as opposed to consistent mass matrix is recommended. This would not only reduce the matrix storage requirement but would also accelerate the convergence of the solution. Modal methods are seldom used to obtain time histories for problems involving high-frequency content input. This is because, for such problems, one has to obtain main eigen frequencies for response estimation, which will be highly time consuming. Instead, we use time integration methods. In addition to the above, other FE rules for node numbering, element numbering, treatment of boundaries, choice of solvers, etc. should be rigorously followed in modeling.

REFERENCES

[1] Doebling SW, Farrar CR, Prime MB, Shevitz DW. A review of damage identification methods that examine changes in dynamic properties. *Shock and Vibration Digest* 1998 **30**(2):91–105.

[2] Yang CH, Ye L, Su Z, Bannister M. Some aspects of numerical simulation for Lamb wave propagation in composite laminates. *Composite Structures* 2006 **75**(1-4):267–275.

[3] Powar PM, Ganguli R. On the effect of matrix cracks in composite helicopter rotor blade. *Composite Science and Technology* 2005 **65**(3-4):581–594.

[4] Dado MHF, Shpli OA. Crack parameter estimation in structures using finite element modeling. *International Journal for Solids and Structures* 2003 **40**(20):5389–5408.

[5] Ghoshal A, Kim HS, Kim J, Choi SB, Prosser WH, Tai H. Modeling delamination in composite structures by incorporating the Fermi–Dirac distribution function and hybrid damage indicators. *Finite Elements in Analysis and Design* 2006 **42**(8-9): 715–725.

[6] Gopalakrishnana S, Roy Mahapatra D, Chakraborty A. *Spectral Finite Element methods*. Springer-Verlag: London, 2008.

[7] Doyle JF. *Wave Propagation in Structures*. Springer-Verlag: New York, 1997.

[8] Abe M, Fujino Y, Yanagihara M, Sato M. Monitoring of a Hakucho Suspension Bridge by ambient vibration measurement, *Proceedings of SPIE—The International Society for Optical Engineering*, 2000 **3995**:237–244.

[9] Adams RD, Cawley P, Pye CJ, Stone BJ. A vibration technique for non-destructively assessing the integrity of structures. *Journal of Mechanical Engineering Science* 1978 **20**:93–100.

[10] Peter Carden E, Fanning P. Vibration based condition monitoring: a review. *Structural Health Monitoring* 2004 **3**:355–377.

[11] Pandey AK, Biswas M, Samman MM. Damage detection from changes in curvature mode shapes. *Journal of Sound and Vibration* 1991 **145**(2): 321–332.

[12] Cornwell P, Doebling SW, Farrar CR. Application of the strain energy damage detection method to plate-like structures. *Journal of Sound and Vibration* 1999 **224**(2):359–374.

[13] Dominguez N, Gibiat V, Esquerre Y. Time domain topological gradient and time reversal analogy: an inverse method for ultrasonic target detection. *Wave Motion* 2005 **42**:31–52.

[14] Park HW, Sohn H, Law KH, Farrar CF. Time reversal active sensing for health monitoring of a composite plate. *Journal of Sound and Vibration* 2007 **302**:50–66.

[15] Chakraborty A, Gopalakrishnan S. A spectrally formulated finite element for wave propagation analysis in layered composite media. *International Journal for Solids and Structures* 2004 **41**(18):5155–5183.

[16] Sharma V, Ruzzene M, Hanagud S. Damage index estimation in beams and plates-using laser vibrometry. *AIAA Journal* 2006 **44**:919–923.

[17] Nag A, Roy Mahapatra D, Gopalakrishnan S, Sankar TS. A spectral finite element with embedded delamination for modeling of wave scattering in composite beams. *Composite Science and Technology* 2003 **63**:2187–2200.

[18] Sreekanth Kumar D, Roy Mahapatra D, Gopalakrishnan S. A spectral finite element for wave propagation and structural diagnostic analysis in a composite beam with transverse cracks. *Finite Elements in Analysis and Design* 2004 **40**:1729–1751.

[19] Kuhlemeyer RL, Lysmer J. Finite element accuracy for wave propagation problems. *ASCE Journal of Soil Mechanics* 1973 **99**:421–427.

[20] Hu XF, Shenton HW. Dead load based damage identification method for long-term structural health monitoring. *Journal of Intelligent Material Systems and Structures* 2007 **18**(9):923–938.

[21] Roy Mahapatra D, Gopalakrishnan S. Spectral finite element analysis of coupled wave propagation in composite beams with multiple delaminations and strip inclusions. *International Journal for Solids and Structures* 2004 **41**:1173–1208.

[22] Chakraborty A, Gopalakrishnan S. A spectral finite element model for wave propagation analysis in laminated composite plate. *ASME Journal of Vibration and Acoustics* 2006 **128**(4):477–488.

[23] Cook RD, Malkus RD, Plesha ME. *Concepts and Applications of Finite Element Analysis*. John Wiley & Sons: New York, 1989.

[24] Sastry CVS, Roy Mahapatra D, Gopalakrishnan S, Ramamurthy TS. An iterative system equivalent reduction process for extraction of high frequency response from reduced order finite element model. *Computer Methods in Applied Mechanics and Engineering* 2003 **192**(15):1821–1840.

[25] Priyank Gupta, Gridhara G, Gopalakrishnan S. Damage detection based on damage force indicator, using reduced order FE models. *International Journal for Computational and Engineering Mechanics* 2008 **9**(3):154–170.

[26] Gopalakrishnan S, Doyle JF. Spectral super-elements for wave propagation in structures with local non-uniformities. *Computer Methods in Applied Mechanics and Engineering* 1995 **121**:77–90.

[27] Hu N, Fukunaga H, Kameyama M, Roy Mahapatra D, Gopalakrishnan S. Analysis of wave propagation in beams with transverse and lateral cracks using a weakly formulated spectral method. *ASME Journal of Applied Mechanics* 2007 **74**(1): 119–127.

Chapter 46
Fatigue Life Assessment of Structures

Thomas Bruder

Fraunhofer Institute for Structural Durability and System Reliability (LBF), Darmstadt, Germany

1 INTRODUCTION

Fatigue refers to the effect due to which a component fails under repeated or cyclic loading, with the maximum loads not reaching static strength criteria. Even today, components and systems fail because of fatigue cracks. Failures often happen due to misinterpretation of service loading/usage, the component-related material strength, or environmental conditions (e.g., corrosion).

Evaluating the fatigue life of structures experimentally typically takes a lot of time. Sometimes, components exist only once and fatigue testing prior to service is impossible, e.g., for structures of big power plants. Numerical fatigue analyses, load and structural health monitoring, as well as non-destructive testing techniques help to prevent failure of components and systems due to fatigue.

This article focuses on engineering approaches to fatigue life assessment based on finite element (FE) results. Anyhow, by applying these basic methods, many applications of interest in the context of structural health monitoring (SHM) can be covered, such as

- detecting fatigue critical areas, which may be the subject of monitoring;
- assessing fatigue impact of the loading applied; and
- deriving a relationship between fatigue impact at a location, which is difficult to access and locations where stress or strain data can be acquired.

A general introduction into fatigue and fatigue damage models is given in **Damage Evolution Phenomena and Models**, which outlines "Fatigue Damage Models" in Section 3 and "Load History Effects" in Section 4. For more detailed information, see Refs. 1–6. Corrosion, creep, buckling, as well as global yielding of the component's net section is not covered in this article. **Static Damage Phenomena and Models** describes static damage phenomena.

Fatigue damage is especially sensitive to actual usage and is difficult to analyze without detailed knowledge of the service loading experienced by the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

component of interest. For a typical notched component manufactured from medium-strength steel, an increase in the loading amplitude by 15% leads to a reduction in fatigue life by 50%. Since loading plays an important role in components' fatigue life assessment, this article describes how to deal with loading without loosing the fatigue-relevant content. In this context, fatigue-relevant loads are determined by the number, amplitude (size), and mean value (position) of load cycles.

In general, for deriving service loading, extensive full system tests or simulations should be avoided owing to time and cost constraints. A common method to reduce efforts is to perform measurements at various specific conditions/events only. Typically, the time histories are transformed to more abstract histograms, which allow efficient data handling as well as durability-focused visualization. The final testing target is then derived by mixing various load scenarios and extrapolating to longer usage. Such a target statistically represents the usage of the component during a certain lifetime. Nondamage relevant parts of the load-time histories are removed and the remaining signal can be used for numerical analyses or be replayed on a test rig system.

Since fatigue is a local phenomenon, finite element or boundary element methods (FEMs/BEMs) are powerful tools that are used to derive stresses and strains at fatigue-relevant locations.

Generally, various methods are used for numerical fatigue life analysis. Stresses derived by the "Theory of Elasticity" are regarded as a relevant criterion in stress–life methods such as

- nominal stress approach,
- structural stress approach and
- local stress approach.

The basis for the first three of the above-mentioned approaches is the so-called S–N curve (Figure 1).

Such a curve is derived by subjecting specimens or components to constant-amplitude (CA) loading at various load levels. Failed specimens are plotted in a double logarithmic diagram, displaying load or stress amplitude versus the number of cycles to failure.

Local elastic–plastic stresses and strains ("true" stresses and strains) are taken into account in the strain–life method, often called the local strain approach.

Further approaches describes fatigue life by growth and coalescence of initial cracks like linear elastic fracture mechanics (LEFM) and elastic–plastic fracture mechanics (EPFM).

The definition of "damage" depends on the approach used. For example, the damage D at a certain load level may be defined as the ratio between the

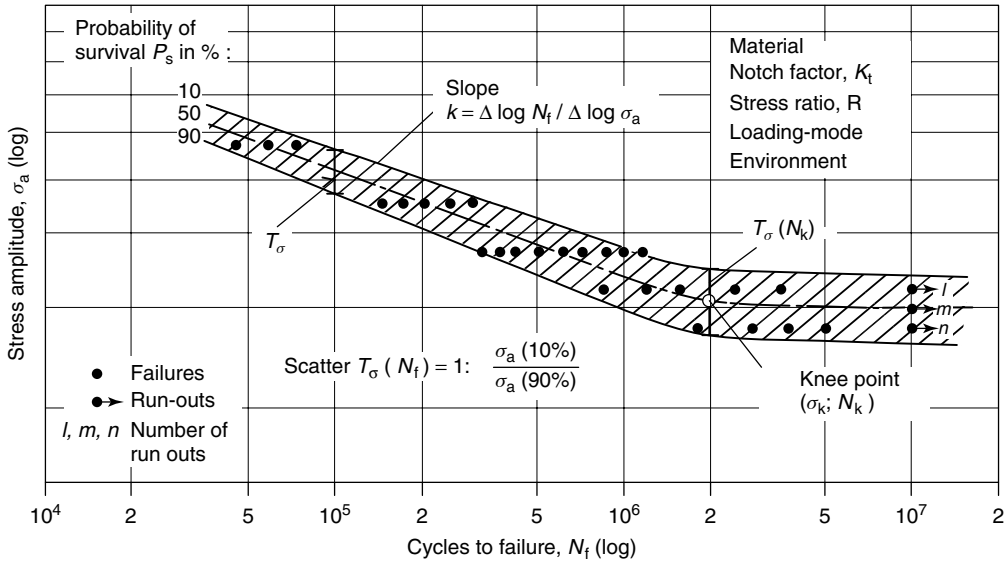


Figure 1. Parameters describing the S–N curve.

number of applied cycles and the number of cycles to failure $D_i = n_i/N_i$ as well as an increase in crack length.

The most suitable approach for a given application depends on the aim of a structural durability analysis and the information that is available. The "local" approaches as well as fracture mechanics typically use FE results.

2 PARAMETERS DETERMINING FATIGUE LIFE

Figure 2 shows the main parameters influencing a structure's fatigue life:

- component's shape
- loading and
- material and manufacturing.

Since the component's shape can be described well using FE methods, the following sections focus on effects due to loading and material.

2.1 Loading

Most often, data acquisition is done as time history recording. Although this type of data contains all the necessary information, it is inconvenient to handle because of its size. Furthermore, a fatigue-related comparison of miscellaneous time histories is difficult. Therefore, other data representations have been developed. They contain only that part of the original information, which is important for the actual analysis. Generally, these representations are called *diagrams or histograms*. Figure 3 shows a time history and the related level crossing and range pair diagrams. The level crossing diagram shows how often the load-time signal passes by a certain load level in one direction (upwards or downwards). The range pair diagram shows how often cycles exceeding a certain load level or amplitude occur within the load sequence.

One of the most important histogram types for fatigue-related analysis is the "rainflow matrix", which contains cycles resulting from a so-called rainflow counting [7–9]. Rainflow counting reveals fatigue-relevant events—the closed hysteresis loops providing amplitude and mean level—in the load-time signal and stores them in an efficient way. A graphical representation of a rainflow matrix is shown in Figure 4.

The rainflow matrix belongs to two-dimensional (2-D) histograms since it contains two pieces of information for every cycle: "from" and "to" level. A 1-D histogram like the "range pair histogram" contains only the amplitude of the cycles. Range pair and level crossing histograms as well as some other 1-D histograms can be easily derived from the rainflow matrix.

For deriving a rainflow matrix, the total load range of the time signal is divided equally into a certain number of "bins" (typically 64 or 100 bins are used). Monotonous parts of the signal are removed and a series of reversal points remain. From this signal, the closed hysteresis loops are identified and stored in the rainflow matrix; small loops may be dropped (hysteresis filtering). In the counting procedure, every closed loop is described by its two anchor points: the load level (bin) where it starts ("from") and the load level (bin) where it ends ("to"). Thus, the "from–to" representation contains information on amplitude, mean, and orientation ("standing" and

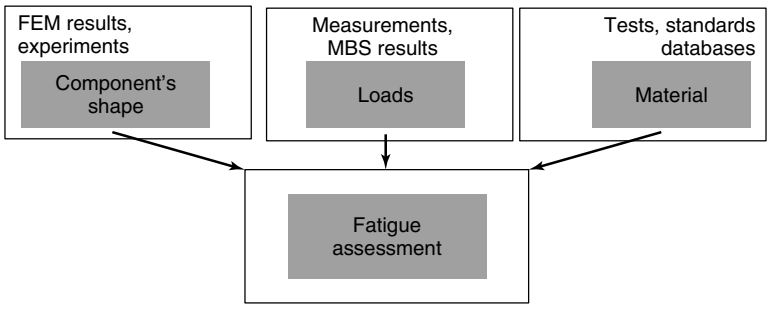


Figure 2. Describing fatigue test results with a S–N curve, important parameters.

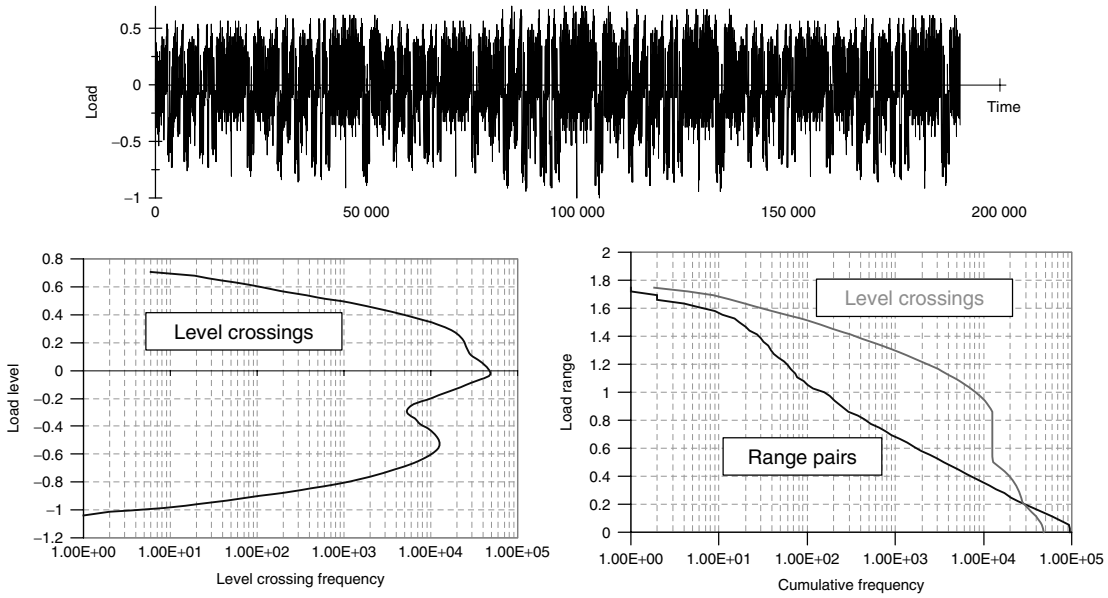


Figure 3. Service loading—Load-time history and representation by level crossing and range pair diagrams.

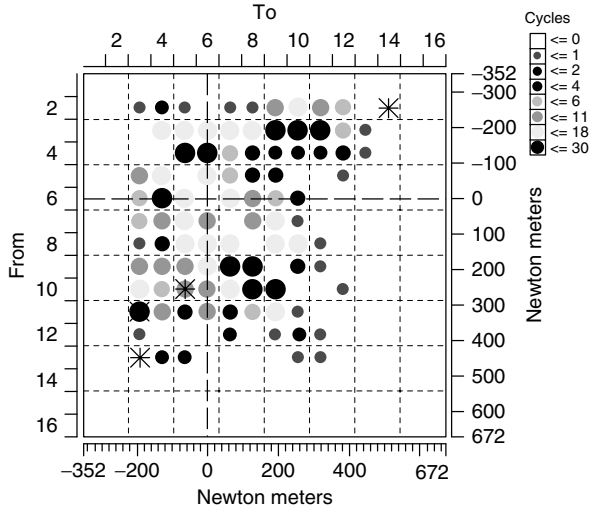


Figure 4. Graphical representation of a rainflow matrix in "from–to" form.

"hanging" inside a bigger hysteresis loop) of the closed hysteresis loops. At the end of the counting, the remaining reversal points that have not formed a closed loop are stored in a so-called residue. The residue may become important in further analysis steps like superposition of rainflow matrices.

The coordinates of the matrix in Figure 4 are the "from" (vertical axis) and "to" (horizontal axis) bins. Additional physical units, e.g., at the centers of bins are displayed. The matrix is divided into two triangular parts by the empty main diagonal. On lines parallel to this main diagonal, cycles with same amplitude but varying mean load are located. The amplitude varies on lines perpendicular to the main diagonal. The bubbles in different gray tones and sizes represent the number of cycles collected in

the corresponding cell. The half-cycles that form the residue are represented by asterisk markers.

Often, load histories (or stress histories derived with respect to a certain location on the component) are analyzed and counted with the rainflow algorithm.

For car body and chassis applications, designing a durability test based on rainflow matrix methods is well accepted. Here, typically time histories reflecting service loading of a vehicle are recorded on proving grounds or public roads. Proving ground tests are the result of many years of experience and contain a mix of test drives (e.g., driving on highway-like as well as rough roads under various vehicle-loading conditions) and especially designed tests.

For durability analyses, axial forces, moments, displacements, or strains are of interest. Data consolidation, e.g., removing spikes and drifts is done in time domain. Transferring load-time histories into (sets of) rainflow matrices reduces the amount of data and simplifies comparisons and further manipulation. Rainflow matrices are edited (e.g., small cycles are removed) and extrapolated to longer measurements or to extreme usage. The final target for a durability analysis consists of a superposition of rainflow matrices [10]. Superposition is, roughly speaking, an addition of the different matrices, which takes the residues into account. The weighting of different events can be accomplished by specifying individual repetition factors. For a rig test, the target in the rainflow domain has to be transformed back in time domain. In case of single channels, this can be achieved by "rainflow reconstruction" [11]. In case of multiaxial loading, the final time histories consist of a number of measured short events (time series). Selecting events and adjusting their repetition factors appropriately allows matching the target.

Using rainflow-based data analysis and synthesis methods, a fatigue test can be setup, which preserves the fatigue-relevant content but leads to a time reduction compared to using the originally measured time histories.

2.2 Material's fatigue properties

Often, CA loading tests are performed to derive the component's S–N curve or material's σ–N curve, respectively. These curves form the basis for the stress–life approach, in which the material behavior is assumed to be linear elastic.

$$\sigma_{\mathrm{a}} = \sigma_{\mathrm{a},k}\left(\frac{N}{N_k}\right)^{-\frac{1}{k}}, \quad \text{for } N \le N_k \tag{1}$$

Parameters are stress amplitude σ_{a}, number of cycles to failure N, and inverse slope k; the index k denotes the knee point of the curve in a double logarithmic diagram. Instead of stresses σ_{a}, loads L_{a} or nominal stresses S_{a} may be used. In the case of a notched rod under axial load F_{ax}, nominal stresses are often defined as $S = F_{\mathrm{ax}}/A$, where A denotes the net area of the cross section.

In the strain–life approach, the cyclic material behavior is described by both the cyclic stress–strain curve (CSSC) (according to Ramberg–Osgood [12]) and the strain–life curve (according to Manson–Coffin–Morrow [13–16]). Typically, strain-controlled tests are carried out on unnotched samples at various strain ranges. Stresses, strains, as well as the associated number of cycles are recorded during each test.

Ramberg and Osgood have suggested a method for describing the CSSC. The total strain amplitude ε_{a} is divided into elastic ($\varepsilon_{\mathrm{a,e}}$) and plastic ($\varepsilon_{\mathrm{a,p}}$) portions and is described by material parameters like Young's modulus E, cyclic strength coefficient K', and cyclic strain hardening exponent n'.

$$\varepsilon_{\mathrm{a}} = \varepsilon_{\mathrm{a,e}} + \varepsilon_{\mathrm{a,p}} = \frac{\sigma_{\mathrm{a}}}{E} + \left(\frac{\sigma_{\mathrm{a}}}{K'}\right)^{\frac{1}{n'}} \tag{2}$$

In the strain–life curve, the endurable strain amplitude σ_{a} is plotted against the number of cycles to crack initiation of technical size N_{i}. The subscript i denotes initiation. Often this failure criterion may be defined by a crack length $2c$ at the surface of $2c \approx 0.5$–1 mm. Plotting strain–life curves for the elastic and the plastic portion on a double logarithmic scale results in straight lines for each, which can be described according to Manson, Coffin, Morrow with Young's modulus E and further material parameters: fatigue ductility coefficient $\varepsilon'_{\mathrm{f}}$, fatigue strength coefficient σ'_{f}, fatigue strength exponent b, and the fatigue ductility exponent c.

$$\varepsilon_{\mathrm{a}} = \varepsilon_{\mathrm{a,e}} + \varepsilon_{\mathrm{a,p}} = \frac{\sigma'_{\mathrm{f}}}{E}\cdot(2N_i)^b + \varepsilon'_{\mathrm{f}}\cdot(2N_i)^c \tag{3}$$

The cyclic parameters of both the Manson–Coffin–Morrow and the Ramberg–Osgood equations are determined using regression analyses of experimental results. The parameters K' and n' can be derived from the data set using the compatibility conditions [12]:

$$K' = \frac{\sigma_f'}{(\varepsilon_f')^{n'}} \tag{4}$$

and

$$n' = \frac{b}{c} \tag{5}$$

Besides deriving cyclic material data from experiments, they can be found in data books or synthetic data may be used, e.g., from Uniform Material Law (UML) proposed by Bäume1 [17].

The stress–strain and strain–life curves are not solely sufficient for fatigue assessment because they do not consider the important effect of mean stresses. A widely spread and accepted parameter to consider mean stresses in the local strain approach is the damage parameter P_{SWT} proposed by Smith, Watson and Topper [18]

$$P_{SWT} = \sqrt{(\sigma_a + \sigma_m)\varepsilon_a E} \tag{6}$$

with stress amplitude σ_a, mean stress σ_m, and strain amplitude ε_a. The endurable damage parameter values are characterized by the P_{SWT}–life curve, which can be derived from strain–life curve

$$P_{SWT}(N) = \sqrt{\sigma_f'^2 \cdot (2N)^{2b} + \sigma_f'^2 \varepsilon_f'^2 E \cdot (2N)^{b+c}} \tag{7}$$

To be able to compare results from stress- and strain-controlled CA tests, even if mean stresses σ_m are slightly different, the damage parameter P_{SWT}—calculated assuming elastic material behavior—is used:

$$P_{SWT,e} = \sqrt{(\sigma_a + \sigma_m)\cdot\varepsilon_a \cdot E} = \sqrt{(\sigma_a + \sigma_m)\cdot\sigma_a} \tag{8}$$

3 FATIGUE STRENGTH ASSESSMENT

A numerical fatigue strength analysis consists of two major steps: the FE analysis of the component behavior when external cyclic load is applied, and a postprocessing step, which interprets the FE results in order to derive the number of load cycles, which the component will endure under a defined loading with a certain probability (Figure 5).

3.1 Determination of local stresses using FEM

Often, the numerical fatigue life assessment is based on FE analyses, assuming the linear behavior of materials and components. The nonlinearity, like nonlinear

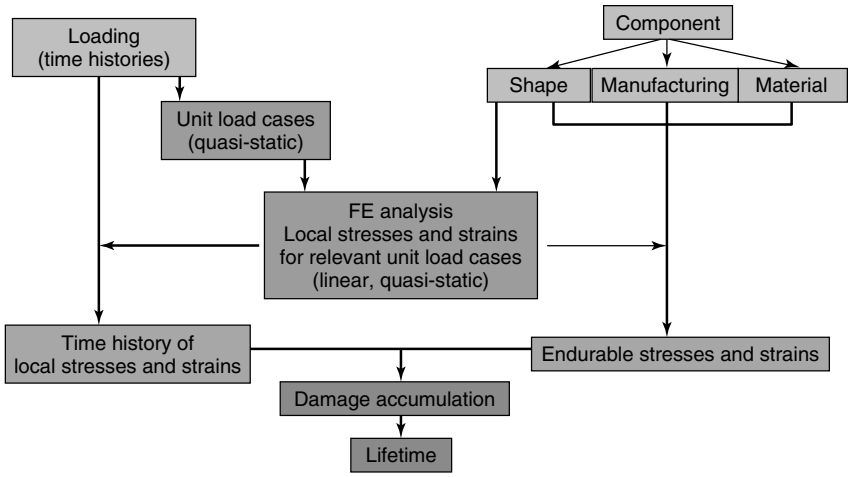


Figure 5. Workflow for fatigue life assessment based on FE results.

material behavior, contact, or large deformations, is neglected. Local stresses derived by the "Theory of Elasticity" are referred to as *elastic* stresses.

For relatively solid components, a static FE analysis can be applied because the lowest eigenfrequency typically lies far above the frequency range of the load-time histories relevant for structural durability. In case a single load $L(t)$ varying in time acts on a component the local "elastic" stresses, σ_e can be computed using the load influence factor c as follows:

$$\sigma_e(t) = c \cdot L(t) = (c \cdot L_{max}) \times \left(\frac{L(t)}{L_{max}}\right) = \sigma_{e,max} \cdot f(t) \tag{9}$$

If multiaxial loading is applied, the basic static load cases (so-called unit load cases) ${}^{s}\sigma_{e,i}$ (hyperscript s denotes static) multiplied with the normalized load-time histories $f_i(t)$ are superposed quasi-statically (Figure 6):

$$\sigma_e(t) = {}^{s}\sigma_{e,1} \cdot f_1(t) + {}^{s}\sigma_{e,2} \cdot f_2(t) + \cdots + {}^{s}\sigma_{e,n} \cdot f_n(t) \tag{10}$$

For thin-walled structures, such an assumption may be invalid. Resonances may be excited with the lowest eigenfrequency lying within the frequency range of the loading. In the numerical analyses, eigenmodes ${}^{m}\sigma_{e,i}$ are determined, starting with the lowest one. In a transient FE analysis, participation factor-time histories $p_i(t)$ are derived on the basis of the given eigenmodes and time histories of all loads acting on the structure. The participation factor-time histories $p_i(t)$ describe the influence of each eigenmode i on the local stresses at a certain point in time t.

The elastic stresses σ_e are computed from the linear combination

$$\sigma_e(t) = {}^{m}\sigma_{e,1} \cdot p_1(t) + {}^{m}\sigma_{e,2} \cdot p_2(t) + \cdots + {}^{m}\sigma_{e,n} \cdot p_n(t) \tag{11}$$

It is also possible to derive "elastic" stresses σ_e based on a combination of basic static modes and eigenmodes, as in the Craig–Bampton approach [19].

$$\sigma_e(t) = {}^{s}\sigma_{e,1} \cdot g_1(t) + {}^{s}\sigma_{e,2} \cdot g_2(t) + \cdots + {}^{s}\sigma_{e,n} \cdot g_n(t) + {}^{m}\sigma_{e,n+1} \cdot g_{n+1}(t) + {}^{m}\sigma_{e,n+2} \cdot g_{n+2}(t) + \cdots + {}^{m}\sigma_{e,n+k} \cdot g_{n+k}(t) \tag{12}$$

The latter approach has the main advantage that typically less modes are needed compared to a solution using eigenmodes only, which—depending on structure and type of loading—may require a high number of modes in order to describe the stresses with sufficient accuracy.

In an FE-based modal analysis, choosing appropriate damping can be challenging. Adjusting damping, based on the results of an experimental modal analysis, helps in improving the numerical analysis. For further information on modal analysis, reference is made to **Modal–Vibration-based Damage Identification**.

Furthermore, the workflow of the fatigue post-processing in the superposition approaches described in equations (11) and (12) is the same as for the quasi-static superposition, equation (10), which

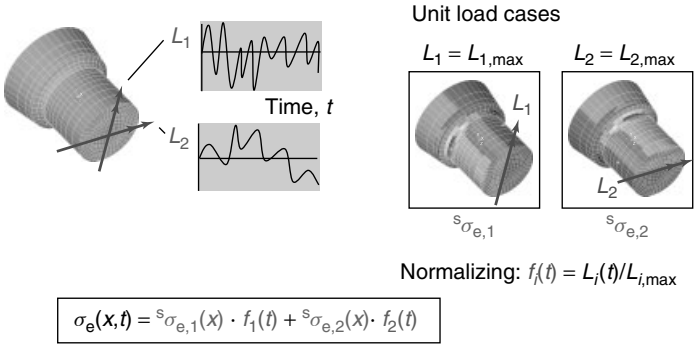


Figure 6. Quasi-static superposition of FE unit load cases (load frequency range below first eigenfrequency).

is supported by commercial fatigue analysis software packages, such as ABAQUS fe-safe, LMS Virtual.Lab Durability, MSC Fatigue, and nCode FE-Fatigue.

Apart from the approaches described in equations (10–12), a transient FE analysis applying all the loads acting on the structure in time domain may be performed. Thus, typically only relatively short events are analyzed due to the effort of calculating all the stresses and strains of a component for each time step. On the other hand, it is advantageous that in a transient analysis even complex contact situations can be handled.

For FE analyses with nonlinear material behavior, the elastic–plastic behavior may be described by the bilinear kinematic hardening rule of Prager–Ziegler [20]. Bilinear stress–strain curves approximate the interpolated Ramberg–Osgood stress–strain curves. More advanced material laws for fatigue analysis were proposed, e.g., by Mróz [21] and Jiang [22, 23].

3.2 Postprocessing of FEM results

For the prediction of the fatigue life of components under a high cycle fatigue (HCF) regime, the "concept of the local fatigue limit" can be applied, in which a purely elastic material behavior is assumed.

Typically in finite life design, elastic–plastic stresses and strains arise at the critical locations of a component. In the more general "local strain approach", elastic and elastic–plastic stresses are computed, and evaluated with regard to the criterion of crack initiation of technical size.

An analysis requires the following input data:

1. A sequence of reversal points of external load or a sequence of local stresses and strains, respectively: assuming that all loads acting on a structure are proportional ($L_i/L_j = \text{const}$) or only a single load is acting (uniaxial loading), a sequence of reversal points of the loading or a rainflow matrix is sufficient. In general, for multiaxial loading, time histories of all loads have to be provided in order to maintain the phase information.
2. Materials data for cyclic loading.
3. Relation between load and local strain.
4. Local macroscopic residual stresses.
5. Selection of a damage parameter P.
6. Modification factor for consideration of the size effect.
7. Modification factor for surface topography, e.g., following Siebel and Gaier [24].

In the following, some of these topics are addressed in more detail:

Materials data for cyclic loading: The material's data for cyclic loading, the stabilized CSSCs, and the strain–life curves can be determined using strain-controlled CA tests on residual-stress-free smooth specimens.

For notched specimens with a homogeneous cyclic material behavior over the cross section, Neuber's rule allows to derive the *relationship between load and local strains* in the notch for many applications with sufficient accuracy.

On the basis of the CSSC,

$$\varepsilon = g(\sigma) \tag{13}$$

Neuber's Rule can be simplified as

$$\sigma \cdot \varepsilon = K_t^2 \cdot S \cdot e \tag{14}$$

where K_t is the stress concentration factor, S is the nominal stress, and e is the nominal strain. The relationship between nominal strain and nominal stress is given by the CSSC $e = g(S)$.

Assuming the elastic behavior of the net section, equation (14) can be further simplified:

$$\sigma \cdot \varepsilon = \frac{\sigma_e^2}{E} \tag{15}$$

Equation (15) can then be directly applied to elastic stresses derived by FE analysis.

The more complex load-notch strain relationship following Seeger [25] has proved to be more accurate, as shown by a very good correlation with the results of the FE analysis.

Size effects: For notched bars with identical cyclic $\sigma-\varepsilon$ behavior of the surface layer and bulk material, the modification factor for consideration of the size effect may be determined according to Siebel–Stieler [26] or the weakest link approach. When using the Smith–Watson–Topper damage parameter P_{SWT} [18], both the modification factors for size and surface

roughness are taken into consideration by shifting the $P-N$ curve in the P direction by the respective factors.

3.3 Surface-strengthened components

Often, residual stresses can be observed at the surface of components, e.g., due to manufacturing processes like turning or grinding. Furthermore, surface strengthening techniques such as shot-peening or gas nitriding are applied in order to improve fatigue strength in the HCF regime [27].

If the local residual stresses are known, they may be considered in the fatigue life assessment using the "thin surface layer model" [28]. This model assumes that the surface layer of a component is of a negligible thickness with regard to its load-bearing capabilities. The thin surface layer model describes the release of initial residual stress through deformation; and, in a generalized form, the cyclic hardening or softening of the material (through using stabilized CSSC). The model does not, however, cover the continuous release of residual stresses arising after 1000 or more load cycles. Under CA loading, the second cycle is already stabilized in the calculation model. The influence of residual stresses upon the deformation behavior of the component as a whole is assumed to be negligible.

The surface layer can have cyclic material data and initial residual stresses that diverge from those of the bulk material. Owing to its negligible thickness, the deformation behavior of the component is not significantly influenced by the layer and the bulk material remains almost free of residual stress. Under these conditions, the total strain of the surface layer when the component is loaded results from the addition of the residual strain ε_{RS} plus the strain ε_B that a component without a surface layer would have. Figure 7 schematically illustrates the determination of local stresses and strains in a component with a thin surface layer of higher strength than the bulk material subjected to compressive residual stresses.

Subsurface crack initiation can be observed under an HCF regime [29], especially with thermochemically surface-strengthened components with low stress gradients at the site of maximum stress and strain.

4 SIZE EFFECTS

In the context of this text, size effect means that—assuming geometrically similar components—the allowable nominal stresses S (respectively the stresses based on the theory of elasticity σ_e at the critical location) decrease with increasing component size. Kloos [30, 31] introduces a systematic, which classifies size effects as

- technological
- stress gradient dependent
- statistical and
- surface treatment related.

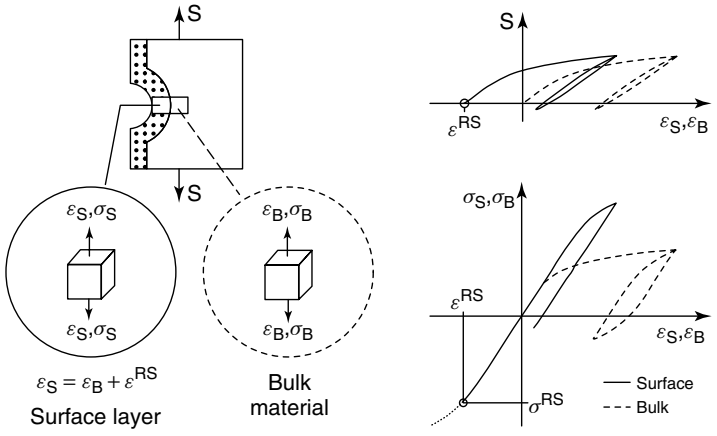


Figure 7. Thin surface layer model: stress-strain behavior in the notch root.

The technological—through use of corresponding component data—and the surface treatment–related size effect can be taken into account to a large extent in the local strain approach. Usually, the transferability of the $P-N$ or $S-N$ curve derived from unnotched specimen to the fatigue critical location of a component has to be ensured. If the highly stressed component area is very small, e.g., in a sharply notched component, the statistical possibility of finding a failure-causing material defect in the highly stressed component area is lower compared to an unnotched specimen with same net section diameter (statistical size effect) and type of loading. A variety of methods exist to take into account the so-called statistical size effect, as well as the structure mechanical size effect. The latter effect describes that stresses derived by the "Theory of Elasticity" are reduced, owing to microscopic and macroscopic yielding in the notch root. In notched components made from mild steels, such yielding leads to true stresses below the elastic stresses even at stress levels close to the knee point of the $S-N$ curve and thus to a better fatigue performance than those that would be expected from the elastic stress level.

Owing to their ease of use, empirically derived factors (so-called size factors) are often applied for the consideration of the size effect. In the HCF regime, e.g., at the knee point of the S–N curve, these size factors n are put in relation to the local fatigue strength of the component (calculated by the "Theory of Elasticity") $K_t \cdot S_\mathrm{k}$ and the material's fatigue strength σ_k (subscript k denotes the knee point):

$$n = \frac{K_t \cdot S_\mathrm{k}}{\sigma_\mathrm{k}} = \frac{\sigma_\mathrm{e,k}}{\sigma_\mathrm{k}} \tag{16}$$

For components with uniform cyclic $\sigma-\varepsilon$ behavior at the fatigue critical location, the size factor can be determined using the approaches proposed by Siebel–Stieler [26]. Similar formulae are given in the (FKM) Forschungskuratorium Maschinenbau Guideline [32], published by the Research Association for the Mechanical Engineering Industry (FKM). Here, the size factor is derived from diagrams or a formula in dependence from the normalized stress gradient χ^* and the material's strength, e.g., the yield strength R_e.

$$n_\chi = f(\chi^*, R_\mathrm{e}) \tag{17}$$

The normalized stress gradient χ^* is defined as follows:

$$\chi^* = \left| \frac{1}{\sigma_\mathrm{e,max}} \cdot \frac{\mathrm{d}}{\mathrm{d}x} \sigma_\mathrm{e} \right| \tag{18}$$

where x denotes a coordinate normal to the surface.

All purely empirical approaches, which describe the size effect on fatigue properties with a size factor based on the normalized stress gradient χ^* only and which neglect the dimensions of both the reference specimen used to derive the material's fatigue strength and the component considered, have a disadvantage. They are somewhat limited to the dimensions and fatigue properties, which are contained in the data set used for their derivation.

Another possibility of describing the size effect is based on the "weakest link" model proposed by Weibull [33–35]. Experiencing the scatter of the fatigue resistance for a constant fatigue life, respectively the scatter of fatigue life for constant load amplitudes, it is assumed that the distribution of internal defect sizes of a material (e.g., length of microscopic cracks) or the local strength corresponds to a Weibull distribution. The largest defect determines the component strength; the subcritical crack growth or coalescence of cracks is of minor importance. Furthermore, it is assumed that the components compared have the same distribution of internal defect sizes (e.g., scatter) and show the same failure mode.

Thereafter, the ratio of the fatigue strength of two components can be determined by comparing the probability to find an internal defect that causes failure in each of the individual components. This probability increases with the size of the highly stressed component area.

5 EXAMPLES

5.1 Uniaxial loading

Fatigue tests were performed on notched, axially loaded specimens. Cyclic material data determined from strain-controlled constant-amplitude loading, residual stress measurements, and FE analyses provided the input data for the calculation.

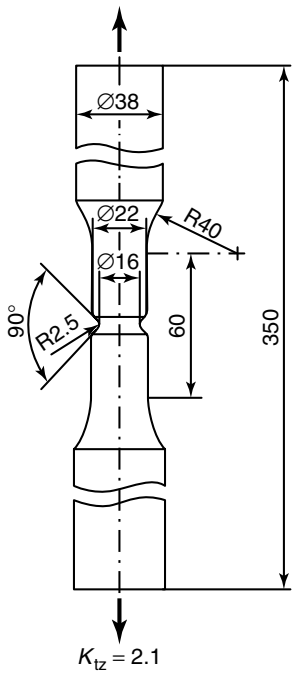


Figure 8. Shape and dimensions of specimen.

The experiments were performed with German grade 42CrMo4 steel (AISI 4040H), quenched and tempered at $T_A = 620\,^{\circ}\mathrm{C}/2.5\,\mathrm{h}$ (ultimate tensile strength $R_m = 1075\,\mathrm{MPa}$, 0.2% offset yield stress $R_{p0.2} = 980\,\mathrm{MPa}$, hardness 330 HV 0.3).

Figure 8 shows the shape of the notched, axially loaded specimen; the axial stress concentration factors is $K_{tz} = 2.1$, and the equivalent stress concentration factor (von Mises criterion) is $K_{tq} = 1.9$.

Tests with constant-amplitude ($R = -1$) and variable-amplitude loading were performed. In the variable-amplitude tests, a random sequence with Gaussian level crossing distribution, a spectrum block size $H_0 = 10^4$ cycles, and a nearly equal number of mean crossings and turning points was used. For all samples (with crack initiation at surface), the criterion of failure was the reaching of a crack length at a surface of $2c = 0.5\,\mathrm{mm}$.

Cyclic material data were determined with strain-controlled CA tests with axially loaded unnotched specimens. The description of the thus-received experimental results for calculation uses relationships following Ramberg–Osgood and Manson–Coffin–Morrow (Figure 9).

Figures 10 and 11 show graphs of the numerically and experimentally determined crack initiation lives of residual-stress-free specimens. The computed curves are shown with solid or dashed lines—these correlate well with the experimental results.

The relation between predicted and experimental results under a Gaussian load sequence corresponds to the general experience made with such comparisons in the literature: If the calculation results fit well with the experimentally determined S–N curve, then the use of the damage parameter according to Vormwald [36] P_J leads to a slightly conservative prediction of the fatigue life curve, while the calculation with the P_{SWT} parameter in conjunction with a P_{SWT}–N curve based on the Manson–Coffin–Morrow relationship and ignoring the endurance limit tends to be somewhat on the unsafe side. Typically, this is adjusted by choosing an allowable damage value for variable amplitude (VA) loading D_{VA} lower than 1.

For analyzing the life of shot-peened specimens with stress concentration factor $K_{tz} = 2.1$, the thin surface layer model is used. As the comparison between the S–N_i curves of untreated and shot-peened samples shows (Figure 12), the influence of residual stress is predominant in the region of low loads, while with higher load levels a release of residual stresses is forced and only the life-reducing influence of surface roughness remains. For the fatigue life curve, the above statements concerning the P_{SWT} parameter apply accordingly.

Applying the stress–life approach based on experimentally determined S–N curves may lead to even better accuracy. Please keep in mind that in the strain–life approach based on cyclic material data, an S–N curve is already a first calculatio result.

5.2 Multiaxial loading

An assessment of a component under multiaxial loading is illustrated using a forged aluminum knuckle of a car's front suspension as an example (Figure 13), [37].

The knuckle is manufactured by forging a wrought aluminum alloy. The material's cyclic properties were derived using the UML [17].

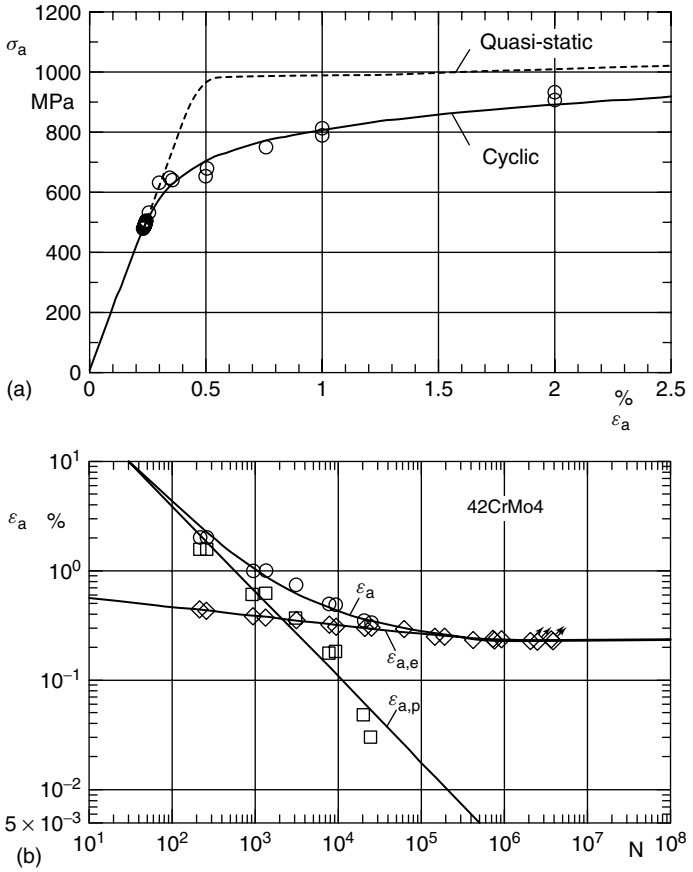


Figure 9. Stabilized CSSC (a) and strain–life (b) curves for steel 42CrMo4.

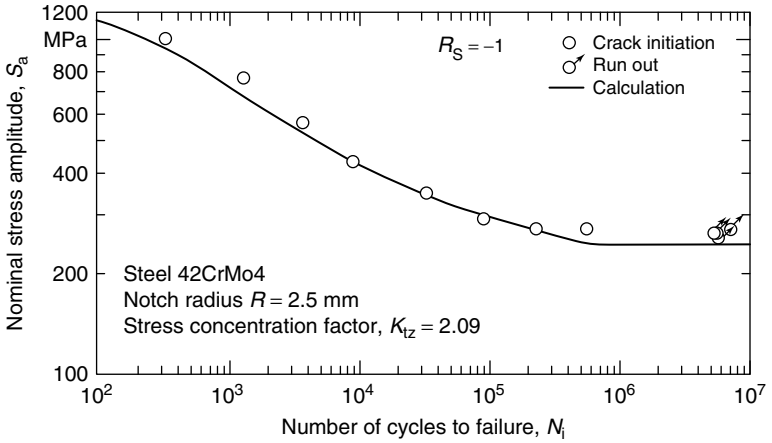


Figure 10. Comparison of numerically and experimentally determined crack initiation life under constant-amplitude loading. $S-N$ curve ($R_S = -1$) of untreated notched specimens.

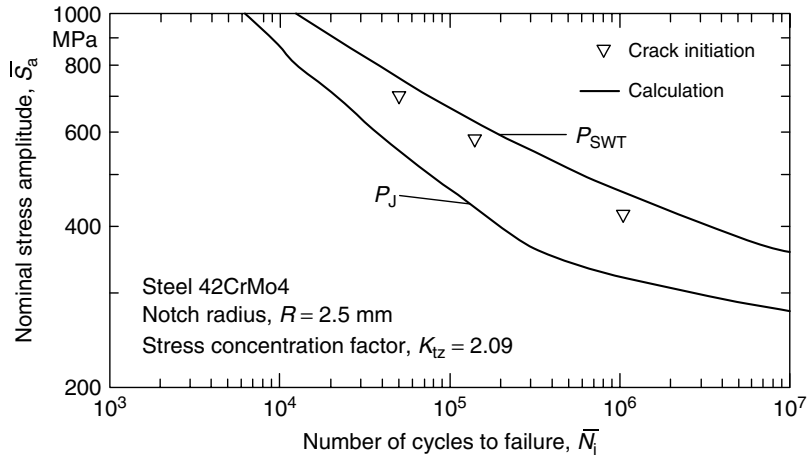


Figure 11. Comparison of numerically and experimentally determined crack initiation life: VA loading (Gaussian load spectrum, ($R_S = -1$) of untreated notched specimens.

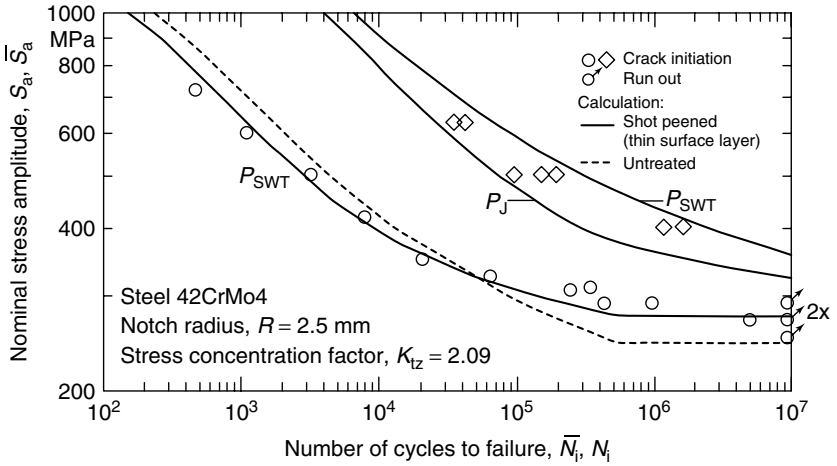


Figure 12. Comparison of numerically and experimentally determined crack initiation life: CA and VA loading ($R_S = -1$) of shot-peened notched specimens (axial residual stresses $\sigma^{R_S} = -700$ MPa).

The standardized load-time history "CARLOS multi" [38] is chosen as a nonproportional load input, which consists of four load channels (longitudinal, lateral, and vertical load vectors acting at the spindle and braking/accelerating forces acting at the tire contact patch) (Figure 14).

The calculation is based on elastic stresses. On the basis of the CAD geometry, an FE mesh with quadratic tetrahedral elements was created and a skinning with membrane elements was applied at the surface. Stresses close to the areas of load application depend strongly on the modeling of this area. Owing to the simplifications made (e.g., rigid element spiders connecting the input points with the solid elements of the component), load input areas are excluded from further analyses. The stiffness of the brake caliper, which is mounted to the knuckle at two points, affects the input of brake loads (and subsequently

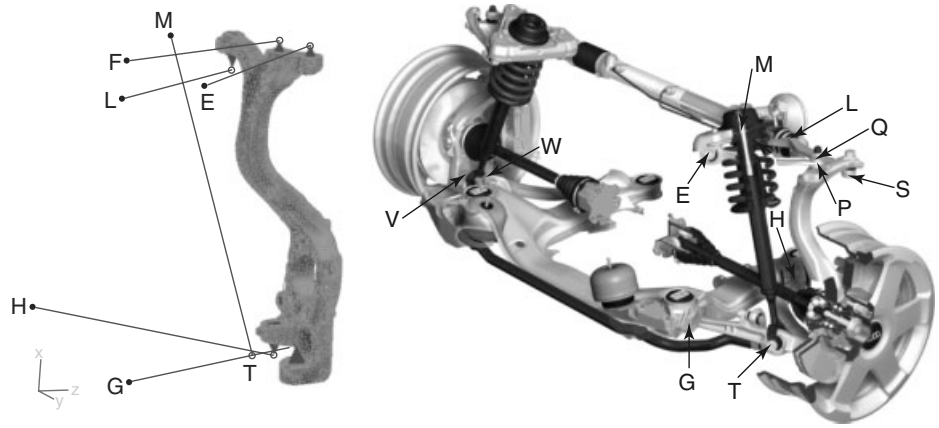


Figure 13. Forged aluminum knuckle and front suspension. [Reproduced from Ref. [37]. © Elsevier, 2004.]

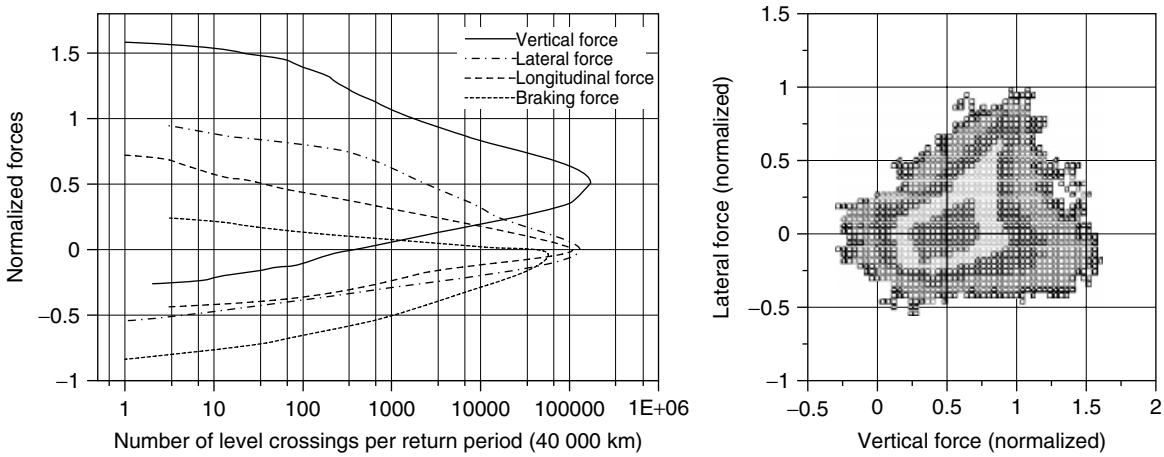


Figure 14. Standardized load-time history CARLOS multi—level crossing distributions and an example for statistical correlations between two channels. [Reproduced from Ref. 38. © Fraunhofer-Institut für Betriebsfestigkeit, 1994.]

the stress distribution in the neighboring area). The caliper is modeled with beam elements representing the appropriate stiffness (Figure 15).

Figure 16 shows the von Mises stress plots for the four unit load cases applying the respective maximum load levels of the corresponding CARLOS multi-load-time histories. For all plots, the same scaling is applied for the stresses. It is obvious that some locations exhibit high stress levels due to several load cases, whereas others exhibit these stress levels only due to a single load case.

For the fatigue life assessment, the critical plane approach is chosen. Local stresses are computed by quasi-static superposition (equation 10). On the basis of the elastic normal stresses in the various planes, the local strain approach is applied [39] in combination with the load–notch strain relation proposed by Seeger–Beste [25], the damage parameter P_{SWT} [18] and the damage accumulation according to Palmgren and Miner. For the sake of simplicity, other influences such as surface roughness, statistical size effects, and changes in local material properties due to the manufacturing process are neglected.

The results of the fatigue life assessment, which has been performed with the software package LMS FALANCS [40], are given as damage plots displaying potentially critical locations (Figure 17).

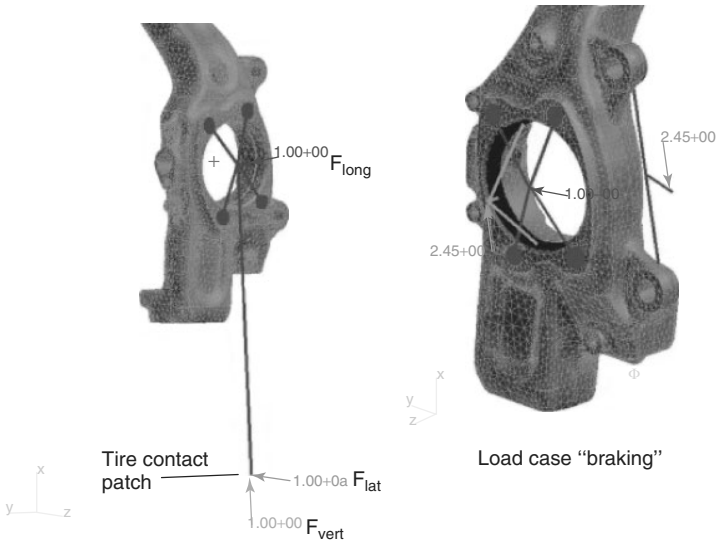


Figure 15. Load input at the knuckle.

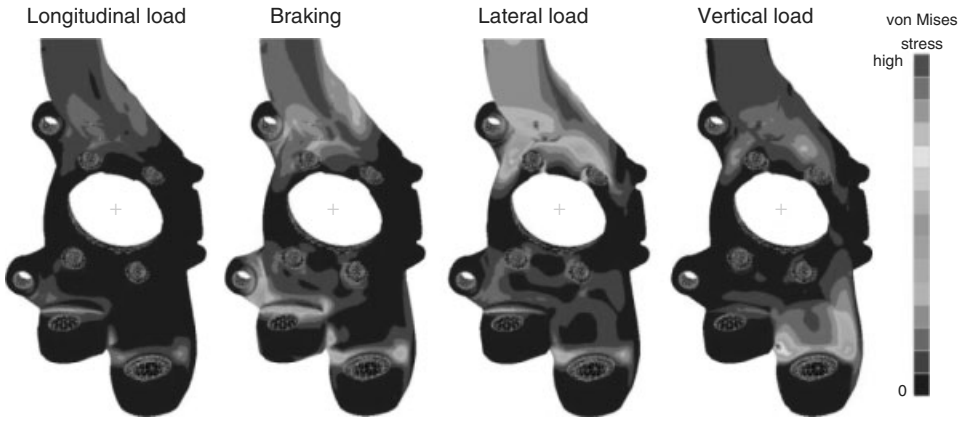


Figure 16. von Mises stresses at the inner side of the knuckle for each unit load case.

Please note that the displayed range of fatigue life covers four decades. Numbers 1 and 2 designate the two most critical locations examined at the knuckle side pointing to the car center.

Figure 18 describes the local biaxial stress state at the two locations. For the evaluation, the maximum principal stresses at the surface were defined as follows: $|\sigma_1| > |\sigma_2|$; $\sigma_3 = 0$. In the histograms, all principle stresses are normalized by the absolute maximum principle stress, which occurs at location 1. The biaxiality ratio is defined as ratio between minimum and maximum principal stress. The behavior at location 1 is strongly influenced by the vertical loading. The local stress state can be described as nearly uniaxial. The direction of the maximum principal stress changes only at very low loads. Location 2 lies in a relatively sharp notch, which is loaded mainly during braking. The constraint caused by the notch leads to a ratio between minimum and maximum principal stress slightly below Poisson's Ratio. With high stress values, no significant variation in the angle of the maximum principal stress can be observed.

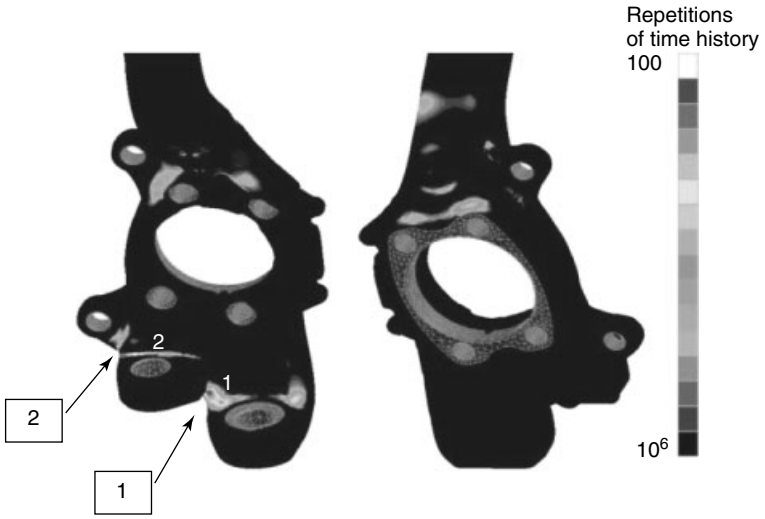


Figure 17. Damage plots.

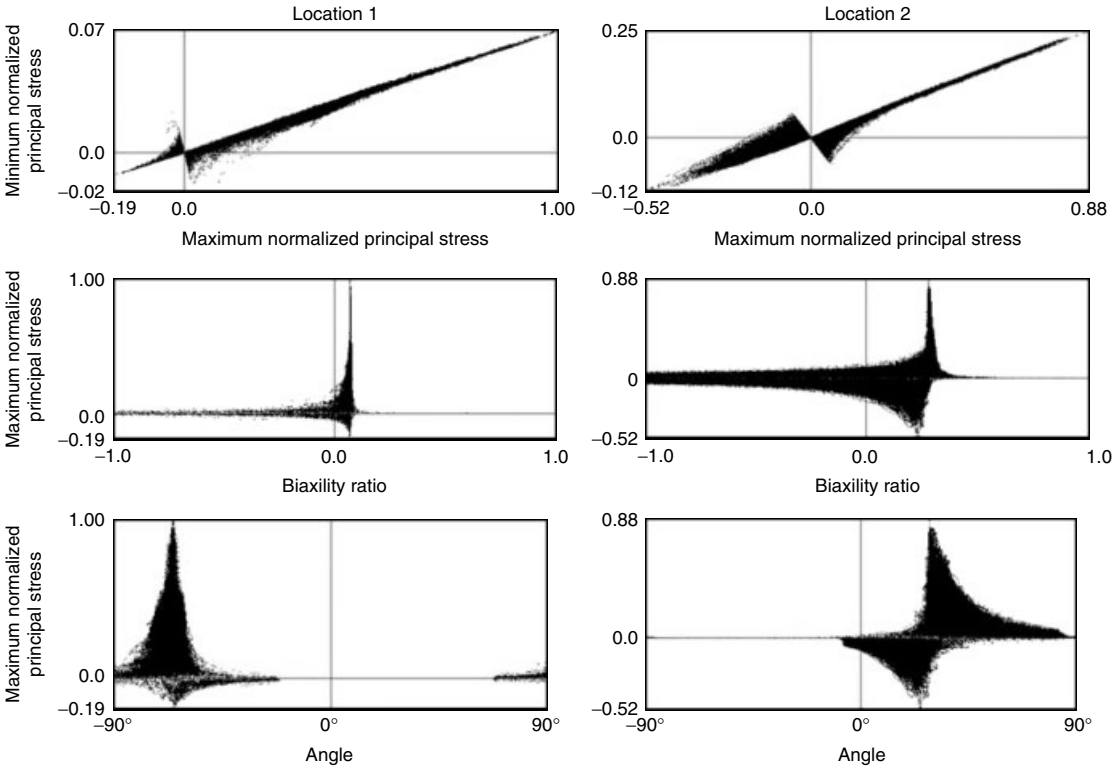


Figure 18. Local elastic stress states at locations 1 and 2 of the aluminum knuckle under loading with CARLOS multi.

In the present example, the structure—a part of multilink car suspension—and the design of the knuckle leads to a separation of the load-transferring functions, thus limiting nonproportional local stress states. The stress state at the locations analyzed turns out to be proportional in spite of the external nonproportional load input, especially with higher loads. For sign-off testing of a single critical location, therefore, a simplified test would be sufficient. If all critical locations on a component or subsystem have to be detected and analyzed, the full set of multiaxial load input has to be taken into account.

The example shows

- how to detect fatigue critical areas, which may be the subject of monitoring and
- how to assess the fatigue damage caused by the loading applied.

Furthermore, knowing the fatigue-relevant locations of a component, actuators and sensors of a structural health monitoring system can be optimally placed on the component in order to detect an occurrence of fatigue cracks.

6 CONCLUSIONS

A broad range of fatigue life assessment methods exists. The most efficient approach depends strongly on the input available.

It has to be pointed out that—compared to relative fatigue strength analyses—a quantitative fatigue life assessment of a component requires profound knowledge of the service loading applied. Load monitoring is important for achieving more reliable input data or to verify assumptions taken in the design phase of structures. It has been shown that numerical fatigue life assessment allows

- detecting fatigue critical areas, which may be the subject of structural health monitoring and
- assessing the fatigue impact of the loading applied.

On the basis of combined load monitoring and fatigue life assessment, an evaluation of the remaining service life is possible.

RELATED ARTICLES

Free and Forced Vibration Models

Civil Infrastructure Load Models for Structural Health Monitoring

Static Damage Phenomena and Models

Damage Evolution Phenomena and Models

Failure Modes of Aerospace Materials

Modal–Vibration-based Damage Identification

Modeling Aspects in Finite Elements

REFERENCES

[1] Dowling NE. *Mechanical Behavior of Materials, Engineering Methods for Deformation, Fracture, and Fatigue*, Third Edition. Pearson Prentice Hall, 2007.

[2] Haibach E. *Structural durability—Methods and data for calculation*. VDI-Verlag: Duesseldorf, 2003.

[3] Bannantine JA, Comer JJ, Handrock JL. *Fundamentals of Metal Fatigue Analysis*. Prentice-Hall: Englewood Cliffs, NJ, 1990.

[4] Stephens RI, Fatemi A, Stephens RR, Fuchs HO. *Metal Fatigue in Engineering*, Second Edition. John Wiley & Sons: New York, 2000.

[5] Sonsino CM. Principles of variable amplitude fatigue design and testing. In *Fatigue Testing and Analysis Under Variable Amplitude Loading Conditions, ASTM STP 1439*. American Society for Testing and Materials: West Conshohocken, PA, 2005; pp. 3–24.

[6] Socie DF, Marquis GB. *Multiaxial Fatigue*. Society of Automotive Engineers: Warrendale, 2000.

[7] Endo T, Mitsunaga K, Nakagawa H. Fatigue of Metals Subjected to Varying Stress—Prediction of Fatigue Lives. *Preliminary Proceedings of the Chugoku-Shikoku District Meeting*, The Japanese Society of Mechanical Engineers, 1967; pp. 41–44.

[8] Murakami Y (ed). *The Rainflow Method in Fatigue*. Butterworth & Heinemann: Oxford, 1992.

[9] Dressler K, Hack M. Fatigue lifetime estimation based on rainflow counted data using the local strain approach. *European Journal of Mechanics A /Solids* 1996 **15**(6):955–968.

[10] Bruder T, Dressler K, Gründer B. Optimal Configuration of Test Schedules. *Proceedings of the JSAE Spring Convention No. 15-00*. Society of Automotive Engineers of Japan: Tokyo, 2000; pp. 8–11.

[11] Dressler K, Hack M, Krüger W. Stochastic reconstruction of loading histories from a rainflow matrix. *Zeitschrift für angewandte Mathematik und Mechanik* 1997 **77**(3):217–226.

[12] Ramberg W, Osgood WR *Description of Stress-Strain Curves by Three Parameters*. Technical Note No. 902. National Advisory Committee for Aeronautics: Washington DC, 1943.

[13] Manson SS. Fatigue: a complex subject—some simple approximation. *Experimental Mechanics* 1965 **5**:193–226.

[14] Manson SS. Experimental support for generalized equation predicting low cycle fatigue. *Transactions of the ASME, Journal of Basic Engineering* 1962 **84**(4):537.

[15] Coffin LF. A study of the effect of cyclic thermal stresses on a ductile metal. *Transactions of the ASME* 1954 **76**:931–950.

[16] Morrow JD. Cyclic plastic strain energy and fatigue of metals. Internal friction, damping, and cyclic plasticity. *ASTM* 1965 45–87.

[17] Bäumel jr. A, Seeger T. *Materials Data for Cyclic Loading*. Supplement 1. Elsevier Science Publishers: Amsterdam, 1990.

[18] Smith KN, Watson P, Topper TH. A stress-strain function for the fatigue of metals. *Journal of Materials, JMLSA* 1970 **5**(4):767–778.

[19] Craig R, Bampton M. Coupling of substructures for dynamic analysis. *AIAA Journal* 1968 **6**(7).

[20] Ziegler H. A modification of Prager's hardening rule. *Quarterly of Applied Mathematics* 1959 **17**:55–65.

[21] Mróz Z. On the description of anisotropic workhardening. *Journal of the Mechanics and Physics of Solids* 1967 **15**:163–175.

[22] Jiang Y, Sehitoglu H. Modeling of cyclic ratchetting plasticity, Part I: development of constitutive relations. *Transactions of the ASME* 1996 **63**:720–725.

[23] Jiang Y, Sehitoglu H. Modeling of cyclic ratchetting plasticity, Part II: comparison of model simulations with experiments. *Transactions of the ASME* 1996 **63**:726–733.

[24] Siebel E, Gaier M. Untersuchungen über den Einfluß der Oberflächenbeschaffenheit auf die Dauerfestigkeit metallischer Bauteile. *VDI-Zeitschrift* 1956 **98**:1715–1723.

[25] Seeger T, Beste A, Amstutz H. In *Proceedings of Fracture 1977*, Taplin DMR (ed). University of Waterloo Press: Waterloo, 1977, Vol. 2, pp. 943–951.

[26] Siebel E, Stieler M. Ungleichförmige Spannungsverteilung bei schwingender Beanspruchung. *VDI-Zeitschrift* 1955 **97**(5):121–126.

[27] Bruder T, Seeger T. Fatigue analysis for surface-strengthened notched specimens using the local strain approach. In *Proceedings of the Sixth International Fatigue Congress: Fatigue 96*, Lütjering G, Nowack H (eds). Pergamon Press: Oxford, New York, Tokyo, 1996, pp. 1327–1332.

[28] Seeger T, Heuler P. *(Ermüdungsverhalten metallischer Werkstoffe)*, Munz D (ed). DGM: Oberursel, 1985, pp. 213–235.

[29] Bruder T, Schön M. Durability analysis of carburized components using a local approach based on elastic stresses. *Materialwissenschaft und Werkstofftechnik* 2001 **32**(4):377–387.

[30] Kloos KH. *Einfluß des Oberflächenzustandes und der Probengröße auf die Schwingfestigkeitseigenschaften*. VDI-Bericht 268. VDI-Verlag, 1976, pp. 63–76.

[31] Kloos KH, Buch A, Zankov D. Pure geometrical size effect in fatigue tests with constant stress amplitude and in programme tests. *Materialwissenschaft und Werkstofftechnik* 1981 **12**(2):40–50.

[32] Wegerdt C, Hanel W, Hänel B, Wirthgen G. *Analytical Strength Assessment, FKM Guideline. 5th revised edition, English version*, Research Association for the Mechanical Engineering Industry: Frankfurt, 2003.

[33] Köhler J. *Statistischer Größeneinfluß im Dauerschwingverhalten ungekerbter und gekerbter metallischer Bauteile*. Ph.D. Thesis, TU München, 1975.

[34] Ziebart W. *Ein Verfahren zur Berechnung des Kerb- und Größeneinflusses bei Schwingbeanspruchung*. Ph. D. Thesis, TU München, 1976.

[35] Böhm J, Heckel K. Die Vorhersage der Dauerschwingfestigkeit unter Berücksichtigung des statistischen Größeneinflusses. *Zeitschrift für Werkstofftechnik—Materials Technology and Testing* 1982 **13**:120–128.

[36] Vormwald M, Seeger T. The consequences of short crack closure on fatigue crack growth under variable amplitude loading. *Fatigue and Fracture of Engineering Materials and Structures* 1991 **14**:205–225.

[37] Bruder T, Heuler P, Klätschke H, Störzel K. Analysis and synthesis of standardized multiaxial load-time histories for structural durability assessment. In *Proceedings Seventh International Conference on Biaxial/Multiaxial Fatigue and Fracture*. Sonsino

CM, Zenner H, Portella PD (eds). DVM: Berlin, 2004, pp. 63–77.

[38] Schütz D, Klätschke H, Heuler P. *Standardized Multiaxial Load Sequences for Car Wheel Suspension Components—Car Loading Standard CARLOS Multi*. Report No. FB-201. Fraunhofer-Institut für Betriebsfestigkeit (LBF): Darmstadt, 1994.

[39] Köttgen VB, Barkey ME, Socie DF. Pseudo stress and pseudo strain based approaches to multiaxial notch analysis. *Fatigue and Fracture of Engineering Materials and Structures* 1995 **18**(9):981–1006.

[40] Barkey ME, Hack M, *et al. LMS FALANCS User Manual*. LMS Durability Technologies GmbH, Kaiserslautern: 2000.

Chapter 45

Modeling for Detection of Degraded Zones in Metallic and Composite Structures

Wiesław Ostachowicz[1,2] and Marek Krawczuk[1,3]

[1] *Institute of Fluid Flow Machinery, PAS, Gdansk, Poland*
[2] *Faculty of Navigation, Gdynia Maritime University, Gdynia, Poland*
[3] *Faculty of Electrical and Control Engineering, Gdansk University of Technology, Gdansk, Poland*

1 INTRODUCTION

Fatigue cracking and delamination are particularly dangerous and are, at the same time, the most common kinds of damages in elements of machines and structures. It is of great importance for the safe operation of the elements of machines and structures to ensure that they are free of any fatigue cracks and delaminations and to determine the extent of fatigue cracks and delaminations in case they are present. Since existing nondestructive methods for the detection of fatigue cracks and delaminations have failed in many practical cases, vibration methods have been continually used for nearly 20 years for the diagnosis of such damage. These methods are based on diagnostic relations between the size and location of failures and changes in some dynamic characteristics of construction elements. In order to establish such relations, and to identify changes of the dynamic characteristics, efficient models that facilitate the assessment of the influence of fatigue cracks and delaminations must be established. In this work, a review of the existing models used for analyses of the influence of fatigue cracks and delaminations on changes in dynamic characteristics of construction elements has been presented.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino ISBN: 978-0-470-05822-0.

2 CONTINUOUS MODELS

Historically, the oldest method, proposed by Hetenyi [1], is applied for modeling fatigue cracks in construction elements made of isotropic materials, in which a crack in a construction element is represented by some additional external equivalent loads (Figure 1a). This method was originally developed to determine the line of static deflection for beams of a nonconstant

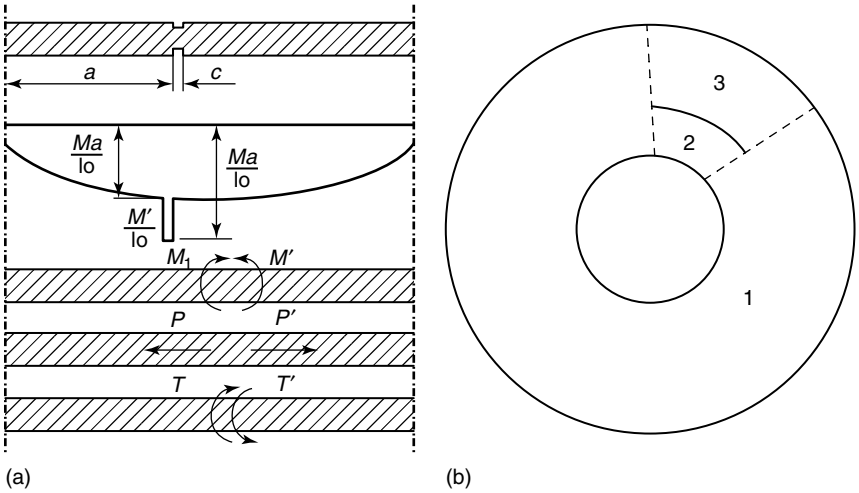


Figure 1. Continuous models of (a) a beam with a transverse fatigue crack (by Thomson [3]) and (b) a thin circular plate with an internal fatigue crack (by Lee [5]).

cross section. Modifications introduced by Kirsmer [2] and Thomson [3] enhanced the method and it is being successfully used for studies of natural frequencies of beams with fatigue cracks. In the following years, models based on the concept of equivalent external loads were further developed by Petroski and Glazik [4]. The principal limitation of these models was the fact that their mathematical description was based on partial differential equations and then they enabled studies and investigation and provided results of sufficiently high accuracy only in the case of construction elements of very simple geometry. Therefore, these models were mainly used for studies of static and dynamic behavior of beams of constant cross sections. On the other hand, the lack of an explicit relation between the size of a fatigue crack and the magnitude of its external equivalent load made it practically inapplicable for the sensitivity analysis of dynamic characteristics of construction elements. Another shortcoming of these models was also the fact that a singular character of the stress and strain fields around the crack tip was omitted. Owing to the above facts, these methods are practically not in use anymore.

Another method used in the analysis of continuous models of construction elements with fatigue cracks or delaminations is the one in which the investigated element is divided into subdomains constrained by its boundaries and the line of a discontinuity (Figure 1b). In order to connect the subdomains, additional boundary conditions are introduced. This approach is applied for isotropic and composite structures. The method was commonly and successfully used to study natural vibrations of thin rectangular plates. Lynn and Kumbasar [6] used the Green's function approach to obtain the Fredholm integral equation of the first kind. Stahl and Keer [7] and Aggarwala and Ariel [8] have solved the eigenvalue problem of simply supported plates by using homogeneous Fredholm integral equations of the second kind. The methods presented in [6–8] were limited to such locations of the crack that enabled the problem to be reduced to a dual series equation. Hirano and Okazaki [9], Neku [10], and Solecki [11] applied fast Fourier transform to the differential equations governing the problem. They obtained a system of integral equations possessing the unknown discontinuities of the deflection and slope across the crack. Lee [5] used a method based on the Rayleigh principle, subsectioning the plates in the determination of the fundamental natural frequency of the circular plate.

The model of delaminated beam, based on the approach described above, is presented in Figure 2. Through-width delamination is parallel to the beam surface located arbitrarily in both the spanwise and thicknesswise directions. The delamination divides the beam into four regions. A Euler beam theory is applied to each region. Apart from the usual

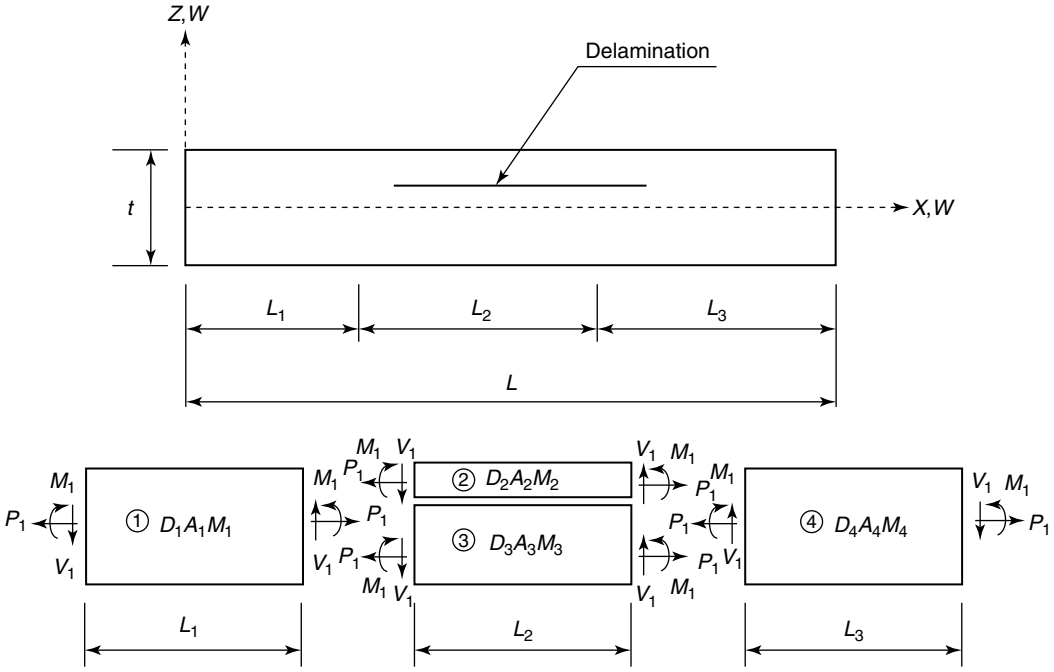


Figure 2. Continuous model of delaminated beam (by Tracy and Pardoen [13]).

conditions of continuity of transverse displacements, slopes, bending moments, and shear forces, two additional conditions were considered (the continuity of axial displacements and forces). This model gives good results for determining the modal parameters of composite laminate. Zou *et al.* [12] presented a review of papers in which models of delaminated beams are described.

The effects of delamination on buckling and postbuckling deformation and delamination growth with various geometrical parameters and loading conditions were studied extensively by Chai *et al.* [14], Botega and Maewal [15], Whitcomb [16], Yin *et al.* [17], and Chen [18]. Natural vibrations of delaminated beams were studied by Ramkumar *et al.* [19] on the basis of the Timoshenko beam theory. The authors, however, did not take into account the effect of coupling of the transverse vibration with the longitudinal wave motion in the upper and lower split layers. Their analytical results predicted significant reduction in the fundamental frequency. Wang *et al.* [20] used the classical beam theory, but, in contrast to Ramkumar *et al.* [19], they considered the coupling effect. With the inclusion of coupling, the calculated fundamental frequency was not appreciably reduced by the presence of a relatively short delamination, and the results were in close agreement with experimental measurements.

Also, results obtained by Christidis and Barr [21] and by Wauer [22] are of great importance in the development of continuous models of construction elements with fatigue cracks. Applying the variational formula of Hu–Washizu, they derived differential equations governing vibrations of beams [21] and turbine blades [22] with fatigue cracks. In order to model changes in their stiffness due to the cracks, they assumed that the stress field along the length of a beam can be approximated by an exponential function. However, limitations of this approach are concerned with the fact that the exponent of the function describing the stress field must be determined experimentally. A singular character of the stress function around the crack tip is also not maintained. It should be noted that such a model does not describe precisely a fatigue crack but rather represents a notch or slot.

The restrictions of the above-mentioned models allow one to analyze the vibration of structures

only in some particular cases. Real engineering constructions are more complex, and the application of other techniques follows naturally.

3 DISCRETE–CONTINUOUS MODELS

Discrete–continuous models as well as discrete models are the most commonly used models to study the dynamic behavior of construction elements with fatigue cracks. In the discrete–continuous models (see Figure 3), a fatigue crack in a construction element is represented by additional springlike elements, compliance of which is calculated according to the laws of fracture. In this manner, a system consisting of an undamaged element and springlike elements modeling the crack is created. The undamaged element and the elements modeling the fatigue crack are connected together by introduction of special boundary conditions. This method can be successfully used for modeling fatigue cracks in one-dimensional construction elements (rods, beams, shafts, columns, and pipes) or in constructions made of such elements (frames and trusses) and plates. In the general case, compliance of springlike elements modeling a fatigue crack in a one-dimensional construction element can be expressed by a 6×6 compliance matrix.

Particular elements of this matrix were studied by many investigators. Liebowitz *et al.* [24] and Okamura *et al.* [25] found the compliance of a springlike element related to longitudinal deformation of a column of a rectangular cross section. Rice and Levy [26] determined the compliance related to the bending and compression modes as well as a coupling term for a thin membrane. Dimarogonas and Massouros [27] derived a formula to calculate compliance of a springlike element related to the torsional mode of a beam of a circular cross section. Anifantis and Dimarogonas [28] showed the form of the compliance matrix of springlike elements modeling a fatigue crack in a beam of a circular cross section except for the term connected to torsion. The full form of this matrix for a beam of a circular cross section can be found in the work of Dimarogonas and Papadopoulos [29] and that of a beam of a rectangular cross section in the work of Krawczuk [30]. The form of the compliance matrix for a beam of a rectangular cross section made of unidirectional composite material was presented in the work of Nikpour and Dimarogonas [31]. The influence of a single transverse fatigue crack on the stability of composite pillars was analyzed by Nikpour [32] who showed that, with increasing slenderness ratio of the columns and depth of the crack, the value of the critical force decreases. Recently, Papadopoulos [33] presented a review of papers in which discrete–continuous models based on strain energy release are described.

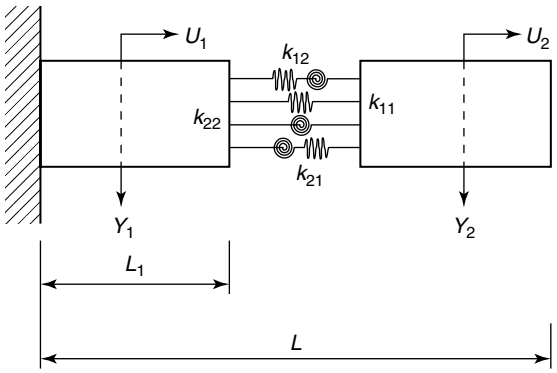


Figure 3. A discrete–continuous model of a shaft (by Papadopoulos and Dimarogonas [23]).

This approach was used in the past in studies of static and dynamic behavior of one-dimensional construction elements. Liebowitz *et al.* [24], Okamura *et al.* [25], and Anifantis and Dimarogonas [34] applied the discrete–continuous model of fatigue damage to investigate the influence of the size and location of fatigue cracks on changes in the critical load and stability of columns made of isotropic material, while Nikpour [32] enhanced these studies by investigation of columns made of composite material. Results on changes in the natural frequencies and mode shapes of beams calculated by the use of this discrete–continuous method can be found in the work of Gudmundson [35], Adams *et al.* [36], Ju *et al.* [37], Springer *et al.* [38], Cuntze and Hajek [39], Papaeconomu and Dimarogonas [40], Liang *et al.* [41], Ostachowicz and Krawczuk [42], Rajab and Al–Sabeeh [43], Rytter *et al.* [44], and Kisa and Arif [45]. However, this method was applied in studies of dynamic behavior of rotors with fatigue cracks most extensively and successfully. The most important studies in this area were carried out by Gash *et al.* [46], Grabowski [47], Ignaki *et al.* [48], Schmied and Kramer [49], Bachschmid *et al.* [50], Dentsoras

and Dimarogonas [51], Papadopoulos and Dimarogonas [23], Wauer [52], Jun *et al.* [53], and Lee *et al.* [54]. This method was also successfully applied in studies of fatigue cracking rings by Dimarogonas [55], as well as in cylindrical shells by Anifantis and Dimarogonas [56].

This method was used by Khadem and Rezaee [57] and Mateev and Boginich [58–60] for the analysis of modal parameters of rectangular cracked plates based on the Kirchoff approach.

It should be noted that, apart from the already mentioned restrictions, this approach is also limited because it is based on partial differential or integral equations. Similar to the method of external equivalent loads, this approach enables one to study, investigate, and obtain results of sufficiently high accuracy only in the case of construction elements of simple geometry like rods, beams, columns, shafts, or rectangular plates with typical boundary conditions.

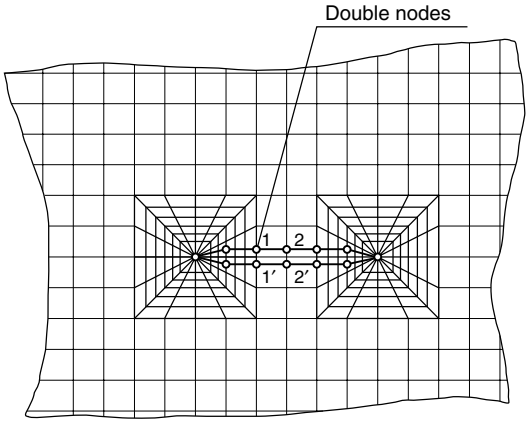


Figure 4. Modeling of a fatigue crack by separation of finite-element nodes and by condensation of the mesh around the crack tip.

4 DISCRETE MODELS

In general, discrete models of fatigue damage are not restricted geometrically, whereas restriction is one of the biggest disadvantages of the continuous or discrete–continuous models. In order to create a discrete model of a construction element with a fatigue crack, the finite-element method is applied most of the times. Other methods such as the boundary element method, graph method, transition matrix method, and the analog method are also used, but these methods are not as popular and commonly used as the finite-element method.

4.1 Finite-element method

The simplest method applied to model construction elements with fatigue damage is based on the use of classical finite elements. In this case, a fatigue crack in the finite element is modeled by reduction of elastic coefficients of the element [61], by reduction of its Young's modulus [62], and by reduction of the cross section area of the element in the crack position [50]. The main disadvantage of these approaches is the fact that the reduced parameters describing a fatigue crack are chosen arbitrarily. Generally, their values are not directly related to the actual size of a crack and therefore a precise study of the influence of the crack depth on changes in dynamic characteristics cannot be made. A singular character of the stress and strain fields around the crack tip is also neglected in these methods.

Another method applied in modeling of fatigue cracks in construction elements based on the finite-element method is the one in which the nodes of finite elements are separated along the crack surface and a condensed mesh of finite elements around the crack tip is used (Figure 4). The condensed mesh of finite elements enables to model properly a singular character of the stress and strain fields around the crack tip. Models based on this approach were successfully applied by Zastrau [63]. This approach gives good results and the existing finite-element software can be effectively used for these studies. Calculating dynamical parameters of construction elements is considerably time consuming.

The methods discussed above can be easily modified by the use of finite elements with singular shape functions around the crack tip [64]. Shen and Pierre have demonstrated that, for an eight-node quadrilateral element, by moving the midside nodes near the crack tip to the quaterside position (Figure 5a), strain singularities are produced. However, for such an element, the singularity conditions prevail only along the edges, and not on an arbitrary ray emanating from the crack tip. This problem was later resolved by collapsing the nodes 1 and 4 of the eight-node quadrilateral element in Figure 5(a), resulting in the

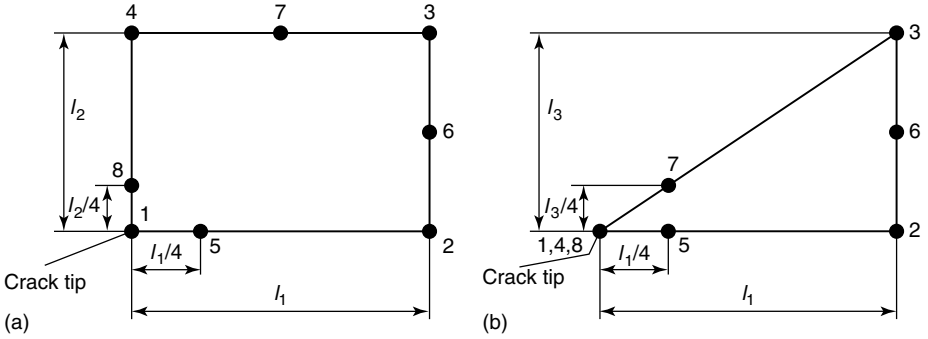


Figure 5. (a) Rectangular finite element with midside nodes at quarter points and (b) triangular element with midside nodes at the quarter points (by Shen and Pierre [64]).

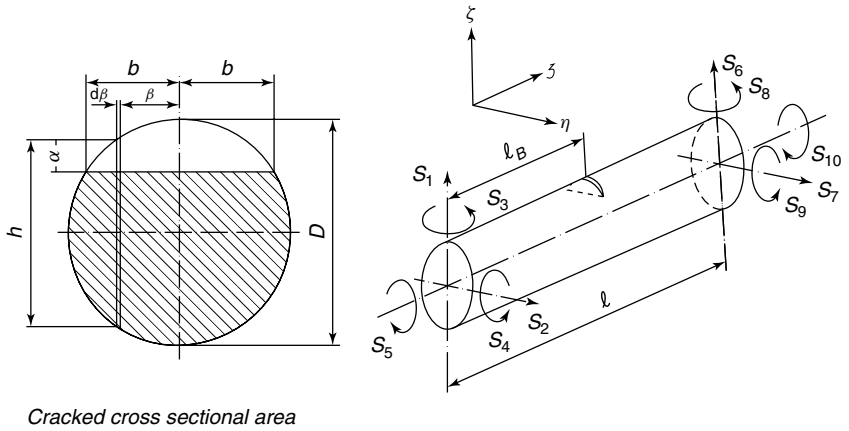


Figure 6. Fragment of shaft with single-sided crack (by Ostachowicz and Krawczuk [70]).

triangular element in Figure 5(b). Then the singularity prevails on all rays emanating from the crack tip. Application of such elements does not require condensation of the finite-element mesh around the crack tip and, consequently, the time required for numerical calculations is significantly shorter.

In the last decade, new finite-element method (FEM)-based models were formulated. Researchers assumed that a failure appears inside the special finite element. A model of a truss finite element with an open one-sided transverse crack was developed by Krawczuk [65]. Models of beam finite elements with fatigue cracks of different types can be found in the work of Haisty and Springer [66], Gounaris and Dimarogonas [67], and Chen and Chen [68]. Krawczuk and Ostachowicz [69] investigated a mathematical, FEM-based model of a beam with a crack, loaded at the end with a constant tensile axial force. The authors assumed that the crack does not propagate and remains open during the beam's vibrations. Admission of a complete opening of the crack in this case was correct because the beam was subjected to the action of a constant axial force.

Ostachowicz and Krawczuk also developed a model of a shaft of constant cross section with a crack. The shaft was modeled by finite elements. The crack was considered to be open. The stiffness matrix for the element with a crack was formulated. The model takes into consideration the torsional–bending interaction in rotor vibration (Figure 6).

The curved beam finite element with a transverse, one-edged, nonpropagating, open crack (Figure 7) was investigated by Krawczuk and Ostachowicz [71]. The authors presented an analysis of the effect of the crack position, and of its location, on the changes in the in-plane natural frequencies and mode shapes of

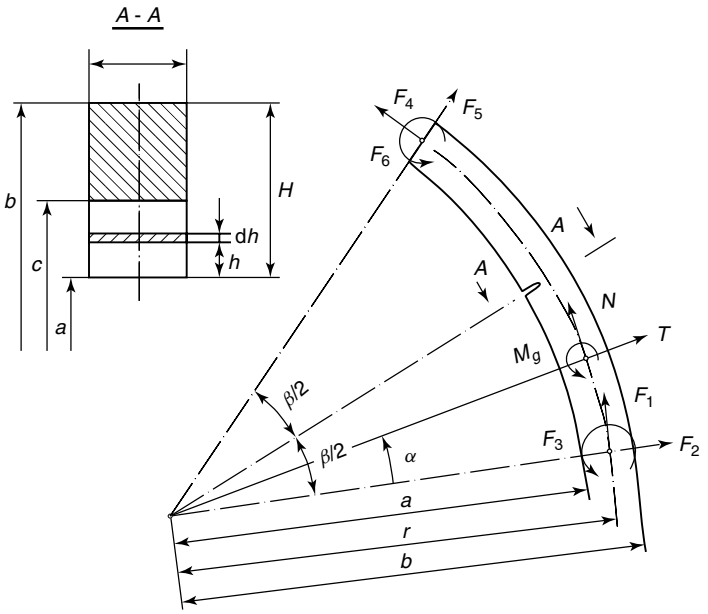


Figure 7. A curved beam finite element with transverse, one-edged, nonpropagating open crack (by Krawczuk and Ostachowicz [71]).

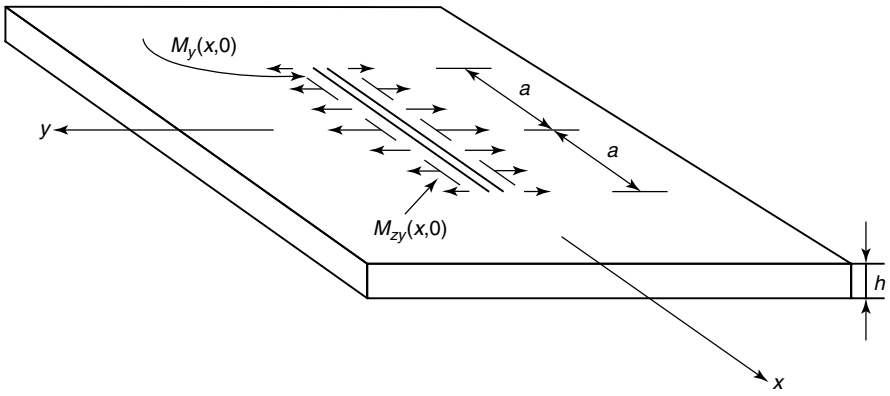


Figure 8. A plate finite element with a fatigue crack (by Qian *et al.* [73]).

the clamped–clamped arch. The authors assumed that the crack only changes the stiffness of the element, with the mass of the element remaining unchanged. The investigated model of the cracked element is restricted to curved beams with a rectangular cross section.

A cracked beam finite element, which is based on elastoplastic fracture mechanics, was formulated by Krawczuk *et al.* [72]. Crack-tip plasticity, at the cracked cross section, was included in the model of the local flexibility. The inertia and stiffness matrices take into account the effect of flexural bending deformation due to the crack presence. They are formulated in closed forms.

Along with one-dimensional models, special models of two- or three-dimensional construction elements (see Figure 8) with fatigue cracks were also investigated. The cracks occurring in a plate can be modeled by the finite-element method in various ways. Plate finite elements with fatigue cracks were used by

Qian *et al.* [73], Krawczuk [74], and Krawczuk and Ostachowicz [75], while a solid finite element with a fatigue crack was developed by Krawczuk and Ostachowicz [76], and a shell element by Krawczuk [77].

Krawczuk and Ostachowicz presented a method of creating the stiffness matrix of a finite plate element with a nonpropagating, internal, open crack (Figure 9). The method is similar to that described by Qian *et el*. [73] but, contrary to their approach, the stiffness matrix of the cracked element was given in a closed form. The additional flexibility matrix was calculated by taking into account the additional elastic stress energy due to the occurrence of the crack in the plate. The method is restricted to cracks whose length is smaller than the dimensions of the element. It has been assumed that the crack changes only the stiffness of the element, whereas the mass of the element remains unchanged.

A method of creating the stiffness matrix of a hexahedral eight-node finite element with a single, nonpropagating, transverse, one-edged crack at half of its length was investigated by Krawczuk and Ostachowicz [76] Figure 10. The crack was modeled by adding an additional flexibility matrix to that of the noncracked element. The terms of the additional matrix were calculated by using an approximated model of the stress intensity factor.

Damage models in composite structures have been studied extensively by many researchers. Krawczuk *et al.* proposed [78] the formulation of a finite

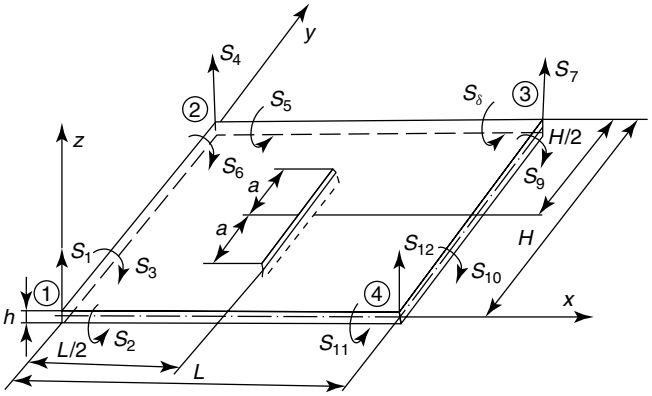


Figure 9. A four-node plate finite element with a crack (by Krawczuk and Ostachowicz [75]).

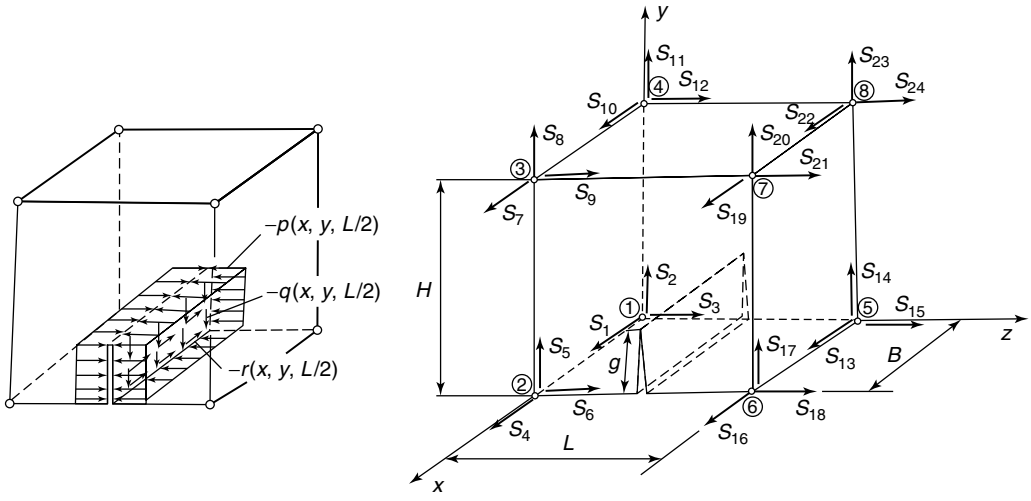


Figure 10. Hexahedral finite element with an open crack (by Krawczuk and Ostachowicz [76]).

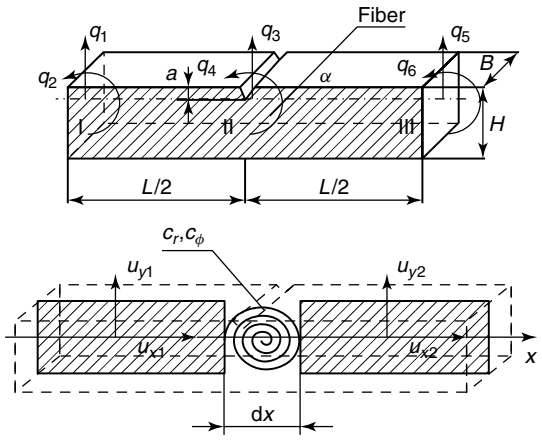


Figure 11. The finite composite beam element with an open crack (by Krawczuk *et al.* [78]).

composite beam element with an open crack. The damaged part of the beam was modeled by a special finite element with a crack (Figure 11), while the undamaged part was substituted by a three-noded beam element. The crack was placed in the middle of the element and remained open. The angle between the fiber and the axis of the element was α. The element had three nodes. Each of them had two degrees of freedom: transverse displacements and rotations. In the paper [78], only the case of flat bending was considered.

Krawczuk *et al.* [79] investigated a model of a layered, delaminated composite beam. The beam was modeled by beam finite elements with three nodes and three degrees of freedom per node (Figure 12). In the delaminated region, additional boundary conditions were applied. It was assumed that the delamination is open (i.e., the contact forces between lower and upper parts are neglected).

The delaminated region is modeled by three finite elements, which are connected at the delamination crack tip where additional boundary conditions are applied. Each element has three nodes with three degrees of freedom, which are axial displacements, transverse displacements, and the independent rotations. In addition to general conditions of beams theory, two different assumptions were used. One was that the extensional and bending stiffnesses are uncoupled.

A model of a finite delaminated plate element was developed by Żak *et al.* [80]. Figures 13 and 14 present a way of modeling the delaminated region in a composite plate with delamination. The delamination is modeled by three-plate finite elements and, to connect them, additional boundary condition are applied at the delamination front. Each finite element has eight nodes with five degrees of freedom.

4.2 Spectral finite-element method

The spectral element method (SEM) is a technique that combines the geometric flexibility of finite

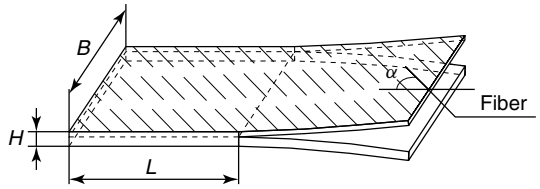


Figure 13. The multilayer composite plate with delamination (by Żak *et al.* [80]).

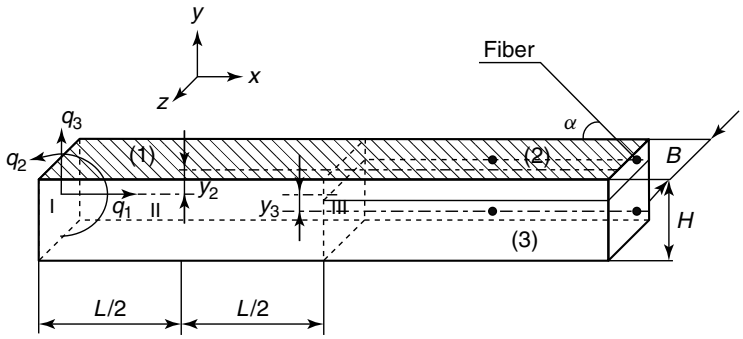


Figure 12. Delamination of a beam modeled by finite elements (by Krawczuk *et al.* [79]).

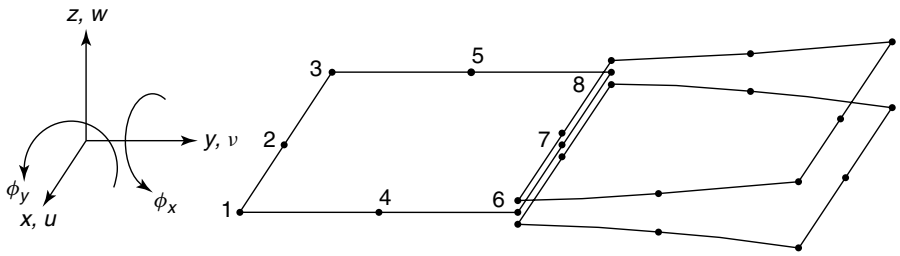


Figure 14. The region of the delamination modeled by finite elements (by Żak *et al.* [80]).

elements with the high accuracy of spectral methods. This method is widely applied for damage detection based on analysis of anomalies in elastic waves propagation. In the literature, two main versions of the method are reported.

The first formulation, called the *forward Fourier transformation (FFT)-based spectral element method*, was proposed by Doyle [81] and developed (especially for damage detection) by Gopalakrishnan *et al.* [82]. This technique is very similar to the technique of the FEM as far as the assembly and the solution of the equation of motion are concerned. Firstly, the excitation signal is transformed into a number of frequency components using the FFT. Next, as a part of a big frequency loop (as opposed to a loop overtime in the conventional finite element (FE) formulation), the dynamic stiffness matrix is generated, transformed, and then a solution is found for each unit impulse at each frequency. This directly yields the frequency response function (FRF) of an analyzed problem. The calculated frequency-domain responses are then transformed back to the time domain using the inverse fast Fourier transformation (IFFT). It proves that this technique is well suited for simple one-dimensional problems but is inefficient when the geometry becomes complex or when two- or three-dimensional problems must be analyzed. This method was applied for analysis of wave propagation in one-dimensional structures with different kinds of damages like cracks [83–85] or delaminations [86–90].

The second formulation of the method, called the *time-domain-based spectral element method*, was proposed by Patera in 1984 [91]. This version is much more versatile for the analysis of elastic wave propagation in structures of complex geometry. The method originates from the use of spectral series for solution of partial differential equations. The

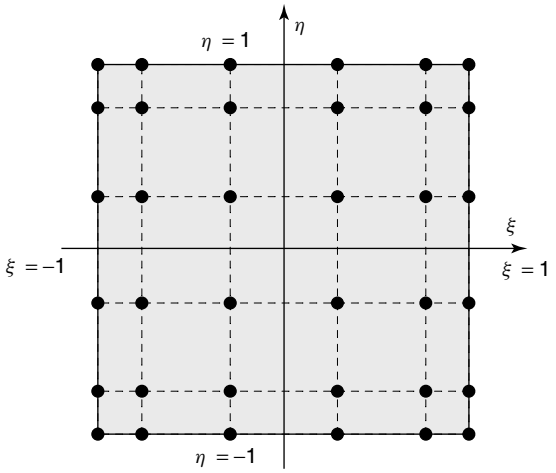


Figure 15. Location of nodes in the spectral finite element (time domain formulation) of the sixth order.

idea of SEM is very similar to the FEM, except for specific approximation. The element interpolation nodes are placed at the zeros of an appropriate family of orthogonal polynomials (Legendre or Chebyshev) Figure 15. A set of local shape functions consisting of Lagrange polynomials, which go through this point, are used. As a consequence, in conjunction with the Gauss–Lobatto–Legendre integration rule, the diagonal mass matrix is obtained. In this way, the cost of numerical calculation is much less than in the case of the classic FE approach. Moreover, the numerical error decreases faster than any power of $1/p$ (spectral convergence), where p is the order of the polynomial expansion. The analysis of the influence of different kinds of damages on elastic wave propagation in one-dimensional structures are presented in papers [92], whereas two-dimensional problems are described in the following works [93, 94].

4.3 Boundary element method

The boundary element method is still not being commonly used in studies of the influence of fatigue damage on dynamic behavior of construction elements. Despite this, this method is being very successfully used in studies of static behavior of construction elements with cracks such as determination of stress intensity factors, strain or stress fields around the crack tip, and similar studies. Also, existing software developed for the boundary element method can be effectively used.

4.4 Transition matrix method

An application of the transition matrix method in modeling and studies of natural frequencies of a beam with a fatigue crack can be found in the work by Sato [95] (Figure 16). This is the only example available in the literature on the use of this method for the investigation of the dynamic behavior of construction elements with fatigue cracks. In the model presented by Sato, only a change of the beam cross section in the place of a crack was considered and therefore, similar to the continuous models, in this model a singular character of the stress and strain fields around the crack tip was not maintained. Thus, mathematical models elaborated by the use of the transition matrix method like the model developed by Sato describe changes in the stiffness of the beam caused by a notch rather than by a real fatigue crack.

4.5 Graph method

The graph method, similar to the boundary element method and the transition matrix method, is not commonly used in studies relating to the influence of fatigue damage on the dynamic behavior of construction elements. This method was used to investigate vibrations of beams with cracks by Ibrahim [96] and Ismail *et al.* [97]. The advantage of this method is that the existing software for the graph method can be very effectively used. An example of a beam with a fatigue crack modeled by the use of the graph method is presented in Figure 17. In this case, the damage is modeled by substitution of the crack by equivalent stiffness. Similar to the continuous models and the transition matrix method, the discussed approach does not take into account a singular character of the stress and strain fields around the crack tip with all the consequences already discussed.

4.6 Analog method

The analog method was applied in modeling of beams and frames with fatigue cracks by Tsyfansky and Beresnevich. [98] and Akgun and Ju [99]. A fatigue crack in this method is modeled by the decrease in the electrical resistance representing the bending stiffness of the beam. Also, in this approach, a singular character of stress and strain fields around the crack tip is neglected. An example of an analog system modeling a cantilever beam with several fatigue cracks is presented in Figure 18.

5 CONCLUSIONS

In this paper, a number of typical models of cracked and delaminated structures have been described. There are many challenges in the development of models of structural stiffness loss due to damage. Typical ones are those that have the ability to identify minor changes of stiffness in composite structures.

Analytical methods to predict changes in the stiffness parameters become dubious with complex structures. The difficulties lie in restrictions described in

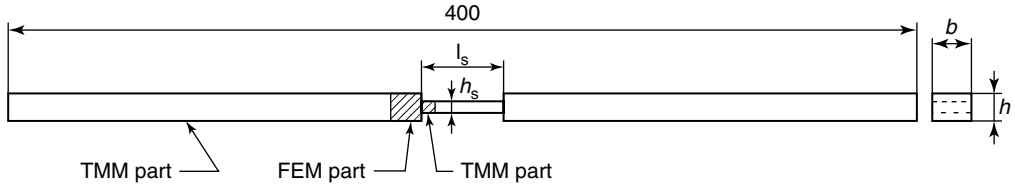


Figure 16. A beam with a fatigue crack modeled by the transition matrix method (by Sato [95]).

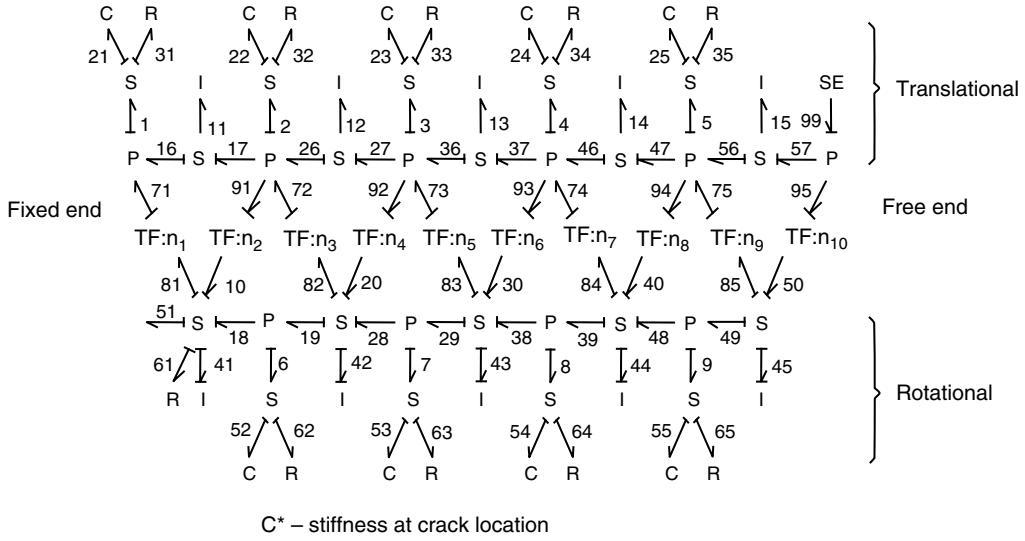


Figure 17. A cantilever beam with a fatigue crack modeled by the graph method (by Ismail *et al.* [97]).

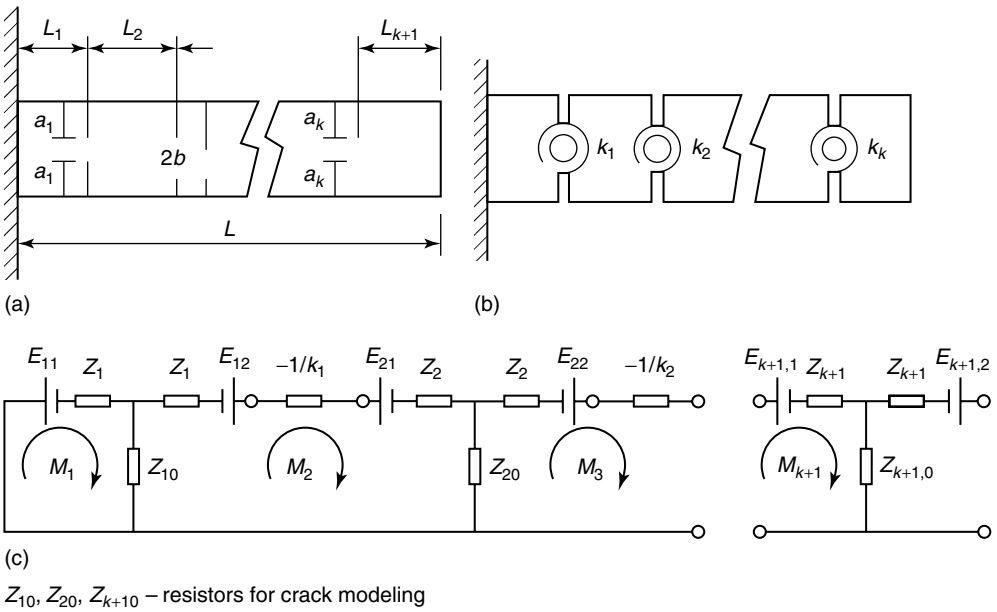


Figure 18. An analog system modeling a cantilever beam with a fatigue crack (by Akgun and Ju [99]).

detail in Section 2. So far, the FEM-based methods have been more realistic for application in engineering constructions.

Laboratory experiments are often conducted to ensure the reality of analytical and numerical models. Therefore, to develop practical and capable detection of stiffness, a large amount of work is still to be done.

Current and future work is concentrated on the integration of various models into a uniform system. Studies are also being carried out on other failures that appear in the composite structure.

RELATED ARTICLES

Free and Forced Vibration Models

Static Damage Phenomena and Models

Damage Evolution Phenomena and Models

Failure Modes of Aerospace Materials

A Simplified Damage Model for SHM Metallic and Composite Structures

Fatigue Life Assessment of Structures

Damage Detection Using Piezoceramic and Magnetostrictive Sensors and Actuators

Damage Measures

Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors

Damage Presence/Growth Monitoring Sensors

Principles of Structural Degradation Monitoring

REFERENCES

[1] Hetenyi M. Deflection of beams of varying cross section. *ASME Journal of Applied Mechanics* 1937 **4**:A49–A52.

[2] Kirmser PG. The effect of discontinuities on the natural frequency of beams. *Proceedings of American Society of Testing and Materials* 1944 **44**:897–900.

[3] Thomson WJ. Vibration of slender bars with discontinuities in stiffness. *ASME Journal of Applied Mechanics* 1949 **17**:361–367.

[4] Petroski HJ, Glazik JL. The response of cracked cylindrical shells. *ASME Journal of Applied Mechanics* 1980 **47**:444–446.

[5] Lee HP. Fundamental frequencies of annular plates with internal cracks. *Computers and Structures* 1992 **43**:1085–1089.

[6] Lynn PP, Kumbasar N. Free Vibration of thin Rectangular Plates having Narrow Cracks with Simply Supported Edges. *10th Midwestern Mechanics Conference*. Fort Collins, 1967; pp. 911–928.

[7] Stahl B, Keer LM. Vibration and stability of cracked rectangular plates. *Journal of Solids and Structures* 1972 **8**:69–91.

[8] Aggarwala BD, Ariel PD. Vibration and bending of a cracked plate. *Engineering Transactions* 1981 **29**:295–310.

[9] Hirano Y, Okazaki K. Vibration of cracked rectangular plates. *Bulletin of the JSME* 1980 **23**:732–740.

[10] Neku K. Free vibration of a simple supported rectangular plate with a straight trough-notch. *Bulletin of the JSME* 1982 **25**:16–23.

[11] Solecki R. Bending vibration of a simple supported rectangular plate with a crack parallel to one edge. *Engineering Fracture Mechanics* 1983 **18**:1111–1118.

[12] Zou Y, Tong L, Steven GP. Vibration-based model-dependent damage (delamination) identification and health monitoring for composite structures. *Journal of Sound and Vibration* 2000 **230**(2):357–378.

[13] Tracy JJ, Pardoen GC. Effect of delamination on the natural frequencies of composite laminates. *Journal of Composite Materials* 1989 **23**:1200–1215.

[14] Chai H, Babcock CD, Knaus WG. One dimensional modelling of failure in plates by delamination buckling. *International Journal of Solids and Structures* 1981 **17**:1069–1083.

[15] Bottega WJ, Maewal A. Delamination buckling and growth in laminates. *Journal of Applied Mechanics* 1983 **50**:184–189.

[16] Whitcomb JD. Parametric analytical study of instability related delamination growth. *Composites Science and Technology* 1986 **25**:19–46.

[17] Yin WL, Sallam SN, Simitses GJ. Ultimate axial load capacity of a delaminated beam-plate. *AIAA Journal* 1986 **24**:123–128.

[18] Chen HP. Shear deformation theory for compressive delamination buckling and growth. *AIAA Journal* 1991 **29**:813–819.

[19] Ramkumar RL, Kulkarni SV, Pipes RB. Free vibration frequencies of a delaminated beam. *34th Annual Technical Conference Proceedings*. Reinforced Composites Institute, Society of Plastics Industry. 1979.

[20] Wang JTS, Liu YY, Gibby JA. Vibration of split beams. *Journal of Sound and Vibration* 1982 **84**:491–502.

[21] Christidis S, Barr AD. One-dimensional theory of cracked Bernoulli–Euler beams. *Journal of Mechanical Sciences* 1984 **26**:639–648.

[22] Wauer J. Vibration of cracked rotating blades. *Machine Vibration* 1992 **1**:126–131.

[23] Papadopoulos CA, Dimarogonas AD. Coupled longitudinal and bending vibrations of a rotating shaft with an open crack. *Journal of Sound and Vibration* 1987 **117**:81–93.

[24] Liebowitz H, Vanderveldt H, Hariss DW. Carrying capacity of notched column. *Journal of Solids and Structures* 1967 **3**:489–500.

[25] Okamura H, Liu WW, Chu CS, Liebowitz H. A cracked column under compression. *Engineering Fracture Mechanics* 1969 **1**:547–564.

[26] Rice JR, Levy N. The part through surface crack in an elastic plate. *ASME Journal of Applied Mechanics* 1972 **39E**:185–194.

[27] Dimarogonas AD, Massouros G. Torsional vibration of a shaft with a circumferential crack. *Engineering Fracture Mechanics* 1980 **15**:439–444.

[28] Anifantis N, Dimarogonas AD. Stability of columns with a single crack subjected to follower and vertical loads. *Journal of Solids and Structures* 1971 **19**:281–291.

[29] Dimarogonas AD, Papadopoulos CA. Vibration of cracked shafts in bending. *Journal of Sound and Vibration* 1983 **91**:583–593.

[30] Krawczuk M. Finite Timoshenko-type beam element with a crack. *Engineering Transactions* 1992 **40**:229–248.

[31] Nikpour K, Dimarogonas AD. Local compliance of cracked composite bodies. *Composites Science and Technology* 1988 **32**:209–223.

[32] Nikpour K. Buckling of cracked composite columns. *Journal of Solids and Structures* 1990 **26**:1383–1386.

[33] Papadopoulos CA. The strain energy release approach for modeling cracks in rotors: a state of the art review. *Mechanical Systems and Signal Processing* 2008 **22**:763–789.

[34] Anifantis N, Dimarogonas AD. Imperfection post buckling analysis of cracked columns. *Engineering Fracture Mechanics* 1983 **18**:693–702.

[35] Gudmundson P. Eigenfrequency changes of structural due to cracks, notches or other geometrical changes. *Journal of the Mechanics and Physics of Solids* 1982 **30**:339–353.

[36] Adams RD, Cawley P, Pye C, Stone BJ. A vibration technique for non-destructively assessing the integrity of structures. *Journal of Mechanical Engineering Science* 1978 **20**:93–100.

[37] Ju FD, Akgun M, Wang ET, Lopez TL. Modal method in diagnosis of fracture damage in simple structure. *Productive Application of Mechanical Vibrations, ASME Publication AMD* 1982 **52**:113–126.

[38] Springer WT, Lawrence KL, Lawley TJ. The effect of a symmetric discontinuity on adjacent in a longitudinally vibrating uniform beam. *Journal of Experimental Mechanics* 1987 **27**:168–173.

[39] Cuntze R, Hajek M. Natural frequencies of a cracked cantilever. *Ingenieur Archiv* 1985 **55**:237–241.

[40] Papaeconomou N, Dimarogonas AD. Vibration of cracked beams. *Computational Mechanics* 1989 **4**:130–137.

[41] Liang RY, Hu J, Choy F. Theoretical study of crack induced eigenfrequency changes on beam structures. *Journal of Engineering Mechanics* 1988 **118**:384–393.

[42] Ostachowicz W, Krawczuk M. Analysis of the effect of cracks on the natural frequencies of a cantilever beam. *Journal of Sound and Vibration* 1991 **150**:191–203.

[43] Rajab MD, Al-Sabeeh A. Vibrational characteristics of cracked shafts. *Journal of Sound and Vibration* 1991 **147**:465–473.

[44] Rytter A, Brincker R, Pilegaard L. Vibration based inspection of civil engineering structures. *Bygningsstatiske Meddelelser* 1991 **62**:79–110.

[45] Kisa M, Arif GM. Free vibration analysis of uniform and stepped cracked beams with circular cross sections. *International Journal of Engineering Science* 2007 **45**:364–380.

[46] Gash R, Person M, Weitz B. Vibrations in rotating machinery. *Dynamic Behaviour of the Laval Rotor with a Cracked Hollow Shaft—A Comparison of Crack Models*. IME: London, 1988, pp. 463–472.

[47] Grabowski B. The vibrational behaviour of a turbine rotor containing a transverse crack. *ASME Journal of Mechanical Design* 1980 **102**:140–146.

[48] Inagaki T, Kaui H, Shiraki K. Transverse vibrations of a general cracked-rotor bearing system. *ASME Journal of Mechanical Design* 1982 **104**:345–355.

[49] Schmied J, Kramer E. Vibrational behaviour of a rotor with a cross-section crack. *Vibrations in Rotating Machinery*. IME: London, 1984, pp. 183–192.

[50] Bachschmid A, Diana G, Pizzigoni B. Vibrations of rotating machinery. *The Influence of Unbalance on Cracked Rotors*. IME: London, 1984, pp. 193–198.

[51] Dentsoras A, Dimarogonas AD. Fatigue crack propagation in resonating structures. *Engineering Fracture Mechanics* 1989 **34**:721–728.

[52] Wauer J. Modelling and formulation of equations of motion for cracked rotating shafts. *Journal of Solids and Structures* 1990 **26**:901–914.

[53] Jun OS, Eun HJ, Earmme YY, Lee CW. Modelling and vibration analysis of a simple rotor with a breathing crack. *Journal of Sound and Vibration* 1992 **155**:273–290.

[54] Lee ChW, Yun JS, Jun OS. Modelling of a simple rotor with a switching crack and its experimental verification. *ASME Journal of Vibration and Acoustics* 1992 **114**:217–225.

[55] Dimarogonas AD. Buckling of rings and tubes with longitudinal cracks. *Mechanics Research Communications* 1981 **8**:179–186.

[56] Anifantis N, Dimarogonas AD. *Identification of Peripheral Cracks in Cylindrical Shells*, ASME–report No. 83—WA/DE—14, ASME, 1983.

[57] Khadem SE, Rezaee M. Introduction of modified comparison functions for vibration analysis of a rectangular cracked plate. *Journal of Sound and Vibration* 2000 **236**:245–258.

[58] Matveev VV, Boginich OE. Vibrodiagnostic parameters of fatigue damage in rectangular plates. Part 1. A procedure of determination of damage parameters. *Strength of Materials* 2004 **36**:549–557.

[59] Matveev VV, Boginich OE. Vibrodiagnostic parameters of fatigue damage in rectangular plates. Part 2. Straight cracks of constant depth. *Strength of Materials* 2005 **37**:30–42.

[60] Matveev VV, Boginich OE. Vibrodiagnostic parameters of fatigue damage in rectangular plates. Part 3. Through the thickness and surface semi-elliptical cracks. *Strength of Materials* 2006 **38**:466–480.

[61] Cawley P, Adams RD. The location of defects in structures from measurements of natural frequencies. *Journal of Strain Analysis* 1979 **14**:49–57.

[62] Yuen MMF. A numerical study of the eigenparameters of damaged cantilever beam. *Journal of Sound and Vibration* 1985 **103**:301–310.

[63] Zastrau B. Vibration of cracked structures. *Archive of Mechanics* 1985 **37**:731–743.

[64] Shen MHH, Pierre C. Natural modes of Bernoulli–Euler beams with symmetric cracks. *Journal of Sound and Vibration* 1990 **138**:115–134.

[65] Krawczuk M. Modelling and identification of cracks in truss constructions. *Finite Elements in Analysis and Design* 1992 **12**:41–50.

[66] Haisty BS, Springer WT. A general beam element for use in damage assessment of complex structures. *ASME Journal of Vibration Acoustics Stress and Reliability in Design* 1988 **110**:389–394.

[67] Gounaris G, Dimarogonas AD. A finite element a cracked prismatic beam for structural analysis. *Computers and Structures* 1988 **28**:309–313.

[68] Chen LW, Chen CL. Vibration and stability of cracked thick rotating blades. *Computers and Structures* 1988 **28**:67–74.

[69] Krawczuk M, Ostachowicz WM. Transverse natural vibrations of a cracked beam loaded with a constant axial force. *ASME Journal of Vibration and Acoustics* 1993 **115**:524–528.

[70] Ostachowicz WM, Krawczuk M. Coupled torsional and bending vibrations of a rotor with an open crack. *Archive of Applied Mechanics* 1992 **62**:191–201.

[71] Krawczuk M, Ostachowicz W. Natural vibrations of a clamped–clamped arch with an open transverse crack. *ASME Journal of Vibration and Acoustics* 1997 **119**:145–151.

[72] Krawczuk M, Żak A, Ostachowicz W. Elastic beam finite element with a transverse elasto-plastic crack. *Finite Elements in Analysis and Design* 2000 **34**(1):61–73.

[73] Qian GL, Gu SN, Jiang JS. A finite element model of cracked plates and application to vibration problems. *Computers and Structures* 1991 **39**:483–487.

[74] Krawczuk M. A rectangular plate finite element with an open crack. *Computers and Structures* 1993 **46**:487–493.

[75] Krawczuk M, Ostachowicz W. A finite plate element for dynamic analysis of a cracked plate. *Computer Methods in Applied Mechanics and Engineering* 1994 **115**:67–78.

[76] Krawczuk M, Ostachowicz W. Hexahedral finite element with an open crack. *Finite Elements in Analysis and Design* 1993 **13**:225–235.

[77] Krawczuk M. Rectangular shell finite element with an open crack. *Finite Elements in Analysis and Design* 1994 **15**:233–253.

[78] Krawczuk M, Ostachowicz W, Żak A. Modal analysis of cracked, unidirectional composite beam. *Composites Part B* 1997 **28B**:641–650.

[79] Krawczuk M, Ostachowicz W, Żak A. Natural vibration frequencies of delaminated composite beams. *Computer Assisted Mechanics and Engineering Sciences* 1996 **3**:233–243.

[80] Żak A, Krawczuk M, Ostachowicz W. Numerical and experimental investigation of free vibration of multilayer delaminated composite beams and plates. *Computational Mechanics* 2000 **26**(3):309–315.

[81] Doyle JF. *Wave Propagation in Structures*. Springer-Verlag: London, 1997.

[82] Gopalakrishnan S, Chakraborty A, Roy Mahapatra D. *Spectral Finite Element Method*. Springer-Verlag: London, 2008.

[83] Krawczuk M, Palacz M, Ostachowicz W. The dynamic analysis of cracked Timoshenko beam by the spectral element method. *Journal of Sound and Vibration* 2004 **264**:1139–1153.

[84] Palacz M, Krawczuk M. Analysis of longitudinal wave propagation in a cracked rod by the spectral element method. *Computers and Structures* 2002 **80**:1809–1816.

[85] Sreekanth K, Roy Mahapatra D, Gopalakrishnan S. A spectral finite element for wave propagation and structural diagnostic analysis in a composite beam with transverse cracks. *Finite Elements in Analysis and Design* 2004 **40**:1729–1751.

[86] Mira M, Gopalakrishnan S. Wavelet based spectral finite element modeling and detection of delamination in composite beams. *Proceedings of the Royal Society of London* 2006 **462**:1721–1740.

[87] Nag A, Roy Mahapatra D, Gopalakrishnan S, Sankar TS. A spectral finite element with embedded delamination for modeling of wave scattering in composite beams. *Composite Science and Technology* 2003 **63**:2187–2200.

[88] Ostachowicz W, Krawczuk M, Palacz M. Detection of delamination in multilayer composite beams. *Key Engineering Materials* 2003 **245-246**:483–490.

[89] Ostachowicz W, Krawczuk M, Cartmell M, Gilchrist M. Wave propagation in delaminated beam. *Journal of Sound and Vibration* 2004 **82**:475–483.

[90] Roy Mahapatra D, Gopalakrishnan S. Spectral finite element analysis of coupled wave propagation in composite beams with multiple delaminations and strip inclusions. *International Journal of Solids and Structures* 2004 **41**:1173–1208.

[91] Patera AT. A spectral element method for fluid dynamics: laminar flow in a channel expansion. *Journal of Computational Physics* 1984 **54**:468–488.

[92] Kudela P, Krawczuk M, Ostachowicz W. Wave propagation modelling in 1D structures using spectral finite elements. *Journal of Sound and Vibration* 2007 **300**:88–100.

[93] Kudela P, Żak A, Krawczuk M, Ostachowicz W. Modelling of wave propagation in composite plates using the time domain spectral element method. *Journal of Sound and Vibration* 2007 **302**:728–745.

[94] Żak A, Krawczuk M, Ostachowicz W. Propagation of in-plane elastic waves in a composite panel. *Finite Elements in Analysis and Design* 2006 **43**:145–154.

[95] Sato H. Free vibration of beams with abrupt changes of cross-section. *Journal of Sound and Vibration* 1983 **89**:59–64.

[96] Ibrahim SR. Incipient failure detection from random decrement time functions. *Journal of Analytical and Experimental Modal Analysis* 1986 **1**:1–8.

[97] Ismail F, Ibrahim A, Martin HR. Identification of fatigue cracks from vibration testing. *Journal of Sound and Vibration* 1990 **140**:305–317.

[98] Tsyfansky SL, Beresnevich VI. Non-linear vibration method for detection of fatigue cracks in aircraft wings. *Journal of Sound and Vibration* 2000 **236**(1):49–60.

[99] Akgun MA, Ju FD. Diagnosis of multiple cracks on a beam structure. *Journal of Analytical and Experimental Modal Analysis* 1987 **2**:149–154.

Chapter 44

A Simplified Damage Model for SHM Metallic and Composite Structures

N. Hu[1], D. R. Mahapatra[2] and Srinivasan Gopalakrishnan[2]

[1] *Department of Aerospace Engineering, Tohoku University, Sendai, Japan*
[2] *Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India*

1 INTRODUCTION

Wave propagation phenomena in solids have received a great deal of attention due to their manifestation in various engineering problems. For instance, the mechanisms of Lamb waves have been rigorously examined over the years for their application to damage identification or structural health monitoring (SHM) (*see* **Ultrasonic Methods**; **Guided-wave Array Methods**; **Modeling of Lamb Waves in Composite Structures**). Applications of various numerical approaches or numerical modeling in this field can be mainly categorized into the following three areas: (i) characterization of Lamb waves, e.g., the dispersion nature; (ii) investigation on Lamb wave scattering by damage; and (iii) direction application of numerically simulated Lamb wave signals in damage identification or damage evaluation (*see* **Modeling of Lamb Waves in Composite Structures**). We review these three application areas one by one as follows.

The nature of dispersion Lamb waves can provide a lot of important information for SHM or damage identification, such as mode selection, nondispersive zone for optimizing the excitation frequency, and wave propagation speed, etc. There are numerous analytical studies on the dispersion nature of Lamb waves, one of them being the well-known transfer matrix approach. They are not reviewed here since our focus is on numerical modeling for complex media. Owing to the complexity of Lamb waves in composites, such as the different wave velocities

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

for different laminate layouts, the finite element method (FEM) has been developed in parallel for this purpose. For example, FEM models that allow particle displacement to vary as cubic polynomials in the z direction (through-thickness direction) and to vary harmonically in the x direction (span direction) were developed to evaluate the dispersion nature of Lamb waves in carbon fiber/resin composite laminates [1, 2]. The models are able to simulate horizontal shear mode (Love wave).

There have also been a number of studies in the second area of Lamb wave scattering by damages in structures, which can provide some fundamental information for damage identification, such as reflection intensity from damage. For instance, for isotropic materials, Karim *et al.* [3] studied Lamb wave scattering from cracks and inclusions in a plate due to a vertical Gaussian beam load using a hybrid FEM and normal function expansion method. Mal and Chang [4] investigated the scattering of Lamb waves from rivet holes and cracks in plates by using hybrid frequency-domain FEM and normal mode expansion (called *global–local FEM technique*) followed by inverse fast Fourier transformation (FFT) to obtain the scattered field in time domain. Hu *et al.* [5] employed a 3-D hybrid FEM to study the Lamb wave scattering caused by a circular hole in metallic plate, and obtained the relationship between the reflection intensity and wave incident angle. In the above studies [3–5], because of the employment of the conventional FEM for whole structures or at least for damaged regions, in general, very fine meshes should be used, especially in the damaged zones. For composite materials, Liu and Achenbach [6] employed the strip element method to model wave propagation in composite laminates containing matrix cracks. Aberg and Gudmundson [7] proposed a higher-order beam theory to analyze the transient waves in laminated beams containing matrix cracks and fiber breakage. The interaction between Lamb waves and delaminations has also been investigated numerically by the conventional 2-D FEM [8] and 3-D FEM [9]. For instance, the detailed position of a delamination in a laminated beam, where the strongest reflection takes place, was identified for S_0 mode [9], which facilitates a more accurate identification of location of the delamination. Further, some recent studies were focused on the improvement of the spectral element method for matrix cracks [10] and delaminations [11, 12] in beam structures. The spectral element method utilizes the exact solution to the strong form of the elastodynamics for finite element (FE) interpolation at discrete frequencies. Computationally, it is a very efficient and powerful method for analyzing the high-frequency responses.

In the third area, tremendous efforts [13–16] have been directed to solve the damage identification problem as an inverse pattern-recognition problem through comparison between numerical analysis results and experimentally captured signals, including artificial neural networks. With the use of numerical models, most of these techniques do not need the experimental *baseline* data of intact structures, although they may be referred to, e.g., as in [14]. Techniques of this kind, which provide detailed information of damage shape or extent, need a careful verification of the effectiveness of the numerical models by comparing them with the experimental data to increase the reliability and accuracy (*see* **Modeling of Lamb Waves in Composite Structures**).

As stated above, owing to the extreme importance of numerical modeling or numerical characterization of elastic waves (e.g., Lamb waves) in the field of damage identification and SHM, over the years, many numerical techniques have been developed for dealing with elastic wave problems. Such methods include: the finite difference method (FDM) [17], the conventional FEM [1, 2, 5, 8, 9, 18–20] and the boundary element method (BEM) [21, 22]. These methods can deal with most complex situations, especially the complex geometry, arbitrary time history of excitation, and the nonlocal properties. However, computational solvability and accuracy based on these approaches deteriorate dramatically as the spatial domain becomes larger and as the time duration of the excitation becomes shorter. For instance, among the versatile computational methods developed till date, the standard *hp*-FEM can deal with most complex situations. However, in the context of the wave propagation problem, the main disadvantage of such *hp*-FEM with polynomial interpolation is that, the higher the frequency, the greater the number of elements required for obtaining accurate results, and, consequently, high computational costs.

Parallel to the above general approaches, there have been quite remarkable progresses on exact or seminumerical methods, which rely upon the

strong form of the boundary value problem (BVP) as opposed to its weak form as in the FEM. For simple 1-D geometry with general boundary conditions and for 2-D or 3-D geometries with periodic or semi-infinite boundary conditions, asymptotic solutions have been employed to develop various numerical methods, namely, the transfer function matrix method [23], dynamic stiffness matrix method [24], space–time FEM [25], strip element method [6], spectral FEM for beams [10–12, 26, 27], and Levy-type plate [28]. In general, these methods that are constructed in the frequency domain need much less memory storage space for necessary data due to a lower level of discretization and application of the exact solution in one direction, which is very effective for the computation of forced wave motion under much higher frequencies. However, it is very difficult to use them for problems with complex geometries. In general, almost all of above exact or seminumerical approaches are built up in the frequency domain and the responses in the time domain should be obtained through inverse FFT.

Another important approach is a spectral or pseudospectral method in the time domain, which may have been first proposed by Patera [29], based on trial functions of the Chebyshev series, for the solution of partial differential equations of laminar flow problems. Despite the terminology, this method in the time domain is completely different from the spectral element methods [10–12, 26–28] in the frequency domain. An important advantage of this method is that the numerical errors decrease more quickly than any power of $1/p$, i.e., the so-called spectral convergence, where p is the order of the polynomial used [30]. At present, the main application fields of this pseudospectral method include: fluid dynamics, heat transfer, acoustics, seismology, and so on. Despite its wide application, the available literature suggests that the use of this pseudospectral method has been quite limited in the fields of elastic wave propagation in 2-D plate structures, e.g., as described only in [31, 32], to the best of the authors' knowledge. The idea of the pseudospectral element method is quite similar to that of FEM except for the specific piecewise interpolation functions such as the orthogonal Chebyshev or the Legendre polynomials that it uses within elements.

From the above brief review of some limited previous studies, it can be seen that successful numerical simulation of elastic wave propagation, especially the wave scattering of structures containing damages, is a key issue for developing highly reliable and applicable SHM techniques or damage identification techniques. Some new numerical approaches are still needed to tackle this problem more efficiently and with greater flexibility. Keeping this in mind, the main objective of this work is to construct a highly accurate element to simulate elastic wave propagation in 2-D problems of metallic and composite materials containing damages. Two different types of interpolation bases are used in two orthogonal directions: (i) the superposed harmonic wave-type solution using wave vector (**k**) from an assumed kinematic theory (e.g., first-order shear deformation theory (FSDT) for the present beam problem) and (ii) the Lagrangian family of interpolation. While dealing with the wave propagation problem in complex geometry, the above description of the displacement field has definite advantage over the exact spectral interpolation. As a consequence of the hybrid interpolation using (i) and (ii), a weak formulation of the BVP becomes necessary. Therefore, we employ a frequency-domain variational approach and derive the frequency-dependent dynamic stiffness matrix, the mass matrix, and the consistent load vector. Compared with many previous studies for wave scattering with the conventional FEM [3–5, 8, 9], by using a coarse mesh, this element can effectively model the region having various damages, i.e., the "interior region", in metallic and composite materials with high flexibility. Ordinary spectral elements are used to model the "exterior regions" or far-field regions for enhancing computational efficiency. Displacement continuity and equilibrium of forces between the "interior region" and the "exterior region" are modeled using the global–local approach [3, 4, 7]. Numerical examples are used to demonstrate the effectiveness of this new element and to study the behavior of wave propagation in cracked or delaminated beams.

2 FIELD INTERPOLATION

2.1 Solution to one-dimensional wave dispersion

In this section, the authors give a brief outline of the steps involved in obtaining the spectral family of

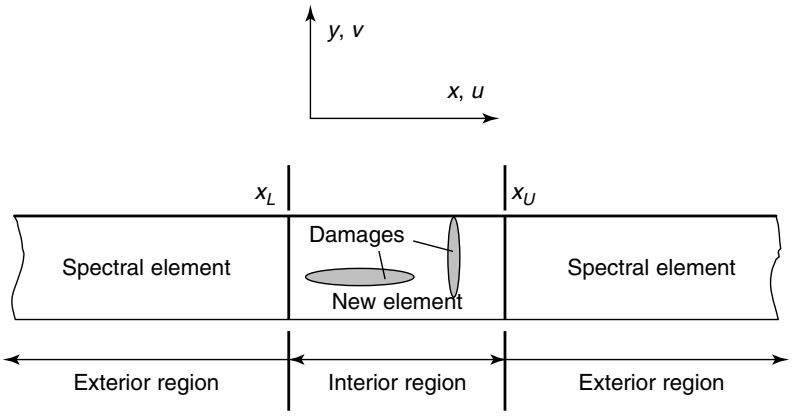


Figure 1. A schematic diagram of the problem geometry.

interpolation functions by solving the characteristic system in terms of the wavenumber (k). These functions are then used to interpolate the displacements along the longitudinal direction (x) of the beam (see Figure 1). In this article, while arriving at the wave dispersion model, the authors apply FSDT for the beam considered as an equivalent single layer. The formulations given by Mahapatra and Gopalakrishnan [27] are employed.

First, cross-sectional stiffness and inertial coefficients of the beam are defined as follows [27]:

$$[A_{kl}, B_{kl}, D_{kl}] = \sum_i \int_{y_i}^{y_{i+1}} \overline{\boldsymbol{Q}}^i_{\boldsymbol{kl}} [1, y, y^2] b_w \, \mathrm{d}y \quad (1a)$$

$$[I_0, I_1, I_2] = \sum_i \int_{y_i}^{y_{i+1}} \rho [1, y, y^2] b_w \, \mathrm{d}y \quad (1b)$$

where y is the thickness coordinate perpendicular to the beam reference plane, b_w the width of the beam, ρ the mass density, and $\overline{\boldsymbol{Q}}^i_{\boldsymbol{kl}}$ the elastic constitutive tensor of the ith layer. In addition, the following parameters are introduced:

$$k_a = \frac{\omega_m}{c_a}, \quad k_b = \frac{\omega_m}{c_b}, \quad k_r = \frac{\omega_m}{c_s} \quad (2a)$$

$$c_a = \sqrt{\frac{A_{11}}{I_0}}, \quad c_b = \sqrt[4]{\frac{D_{11}\omega_m^2}{I_0}}, \quad c_s = \sqrt{\frac{A_{55}}{I_0}} \quad (2b)$$

$$r = \sqrt{\frac{B_{11}^2}{(A_{11}D_{11})}}, \quad s_1 = \omega_m \sqrt{\frac{I_2}{A_{55}}}, \quad s_2 = \sqrt{\frac{I_1^2}{(I_0 I_2)}} \quad (2c)$$

For the coupled axial–flexural shear deformation, without the thickness contractional modes, the wavenumbers (k_j) associated with the individual wave modes at frequency ω_m are determined by solving the following equation:

$$\begin{vmatrix} (k_j^2 - k_a^2) & 0 & \left(\dfrac{s_1 s_2 k_a^2}{k_r} - \dfrac{r k_a k_j^2}{k_r^2}\right) \\ 0 & (k_j^2 - k_r^2) & -ik_j \\ \left(\dfrac{r k_r^2 k_j^2}{k_a k_b^2} - s_1 s_2 k_r\right) & -ik_j & \left(s_1^2 - 1 - \dfrac{k_r^2 k_j^2}{k_b^4}\right) \end{vmatrix} = 0 \quad (3)$$

The characteristic system given by equation (3) can be expressed as follows:

$$a k_j^6 + b k_j^4 + c k_j^2 + d = 0 \quad (4)$$

where

$$a = 1 - r^2 \tag{5a}$$

$$b = \frac{2rs_1s_2k_a k_b^2}{k_r} - (1 - r^2)k_r^2 - \frac{s_1^2 k_b^4}{k_r^2 - k_a^2} \tag{5b}$$

$$c = k_a^2 k_r^2 - 2rs_1s_2k_a k_r k_b^2 - (1 - s_1^2)k_b^4 + \frac{s_1^2(1 - s_2^2)k_a^2 k_b^4}{k_r^2} \tag{5c}$$

$$d = [1 - s_1^2(1 - s_2^2)]k_a^2 k_b^4 \tag{5d}$$

The six roots of equation (4) are given by

$$k_{1,2} = \pm\sqrt{\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{2}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{2}\right)^3}} - \frac{b}{3a}} \tag{6a}$$

$$k_{3,4} = \pm\sqrt{\mu_2\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{2}\right)^3}} + \mu_1\sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{2}\right)^3}} - \frac{b}{3a}} \tag{6b}$$

$$k_{5,6} = \pm\sqrt{\mu_1\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{2}\right)^3}} + \mu_2\sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{2}\right)^3}} - \frac{b}{3a}} \tag{6c}$$

where

$$p = \frac{(27ca^2 - 9ab^2)}{27a^3},$$

$$q = \frac{(27da^2 - 9abc + 2b^2)}{27a^3} \tag{7a}$$

$$\mu_1 = \frac{-1 + i\sqrt{3}}{2}, \quad \mu_2 = \frac{-1 - i\sqrt{3}}{2} \tag{7b}$$

The three pairs of wavenumbers in equation (6a–c) are distinct (except at few frequencies where crossover may occur), since they represent the axial, the flexural, and the shear modes, respectively. The axial mode must strictly have real wavenumbers. The reason can be explained as follows: first assume that the wavenumbers for axial modes are complex. Then they should exist in conjugate form, i.e., $k_1 = a + ib$, $k_2 = a - ib$. Since the bulk wave equations are of second order, even when they are coupled with flexural waves, they propagate with forward and backward wave components. Hence, the associated wave vector, whose amplitude is the wavenumber, must have an opposite sign. The real part of the wave number must have an opposite sign (the imaginary part is attenuation). Therefore, one must have: $\mathrm{Re}[k_1] = \mathrm{Re}[a + ib]$, $\mathrm{Re}[k_2] = \mathrm{Re}[a - ib] = -\mathrm{Re}[k_1] = -\mathrm{Re}[a + ib]$. The last equation is a contradiction. Consequently, the assumption that the bulk wavenumbers are complex is not true.

Depending on the material configuration and geometry, the other modes may have complex wavenumbers. Note that the shear mode starts propagating only for frequencies greater than the cutoff frequency [27], where the term of cutoff frequency means the frequency below which a specified wave fails to propagate in media.

$$\omega_{\text{cutoff}} = \sqrt{\frac{A_{55}}{I_2(1 - s_2^2)}} \tag{8}$$

When $\omega_m < \omega_{\text{cutoff}}$, the shear waves are evanescent in nature. Usually, if there are unsymmetrical damages in beams, the components in the reflected waves contain the axial wave mode. However, in this article, the authors deal with the damages, which are only symmetric with respect to the middle plane of the beam of a symmetric stacking sequence. It is proper to consider flexural and shear waves only.

2.2 Enriched interpolation of the displacement field

After transforming the longitudinal displacement $u(x, y, t)$ and the transverse displacement $v(x, y, t)$

(see Figure 1) from the time domain to the frequency domain, their frequency-domain counterparts are expanded as linear combinations of orthogonal basis functions. The bases for interpolation parallel to the x axis are the four wave modes ψ_i, $i = 1, 2, 3, 4$ (the forward and the backward propagating flexural wave modes, and the forward and the backward propagating/evanescent shear wave modes). The bases for interpolation parallel to the y axis are the Lagrangian family of interpolation functions. Thus, the displacement components of u and v in the frequency domain are expressed as

$$\overline{u}_m(x, y, \omega_m) = \sum_{j=1}^{n} \sum_{l=1}^{NW} N_j(x, y)\psi_l(x)A_j^l \quad (9a)$$

$$\overline{v}_m(x, y, \omega_m) = \sum_{j=}^{n} \sum_{l=1}^{NW} N_j(x, y)\psi_l(x)B_j^l \quad (9b)$$

where A_j^l and B_j^l are the boundary-dependent coefficients for each of the wave modes $l = 1, \ldots, 4; n$ is the number of elemental nodes; NW is the number of used wavenumbers (k_1, k_2, k_3, and k_4); and N_j are the standard FE Lagrangian interpolation functions, i.e., the "old shape functions". In other words, ψ_l are the spectral enrichment functions over N_j. For the present beam problem, they are described as

$$\psi_1 = \mathrm{e}^{-\mathrm{i}[k_1(x-x_L)]}, \quad \psi_2 = \mathrm{e}^{\mathrm{i}[k_2(x_U-x)]} \quad (10a)$$

$$\psi_3 = \mathrm{e}^{\mathrm{i}[k_3(x-x_L)]}, \quad \psi_4 = \mathrm{e}^{-\mathrm{i}[k_4(x_U-x)]} \quad \forall\omega < \omega_{\text{cutoff}} \quad (10b)$$

$$\psi_3 = \mathrm{e}^{-\mathrm{i}[k_3(x-x_L)]}, \quad \psi_4 = \mathrm{e}^{\mathrm{i}[k_4(x_U-x)]} \quad \forall\omega \geq \omega_{\text{cutoff}} \quad (10c)$$

where k_1 and k_2 are the wavenumbers of flexural waves, which are positive real, and negative real numbers, respectively. k_3 and k_4 are wavenumbers of shear waves, which are positive imaginary and negative imaginary numbers when the frequencies are smaller than ω_{cutoff}. x_L and x_U denote the x coordinates of the left and the right cross sections of the interior region, respectively as shown in Figure 1. Generally, it is convenient to define them in the local coordinate system.

The enrichment functions ψ_l in equation (10) occur in pairs, where one member of the pair is the "mirror image" of the other. In fact, the origin of this idea for the displacement field in equation (9) can be traced to a new and general numerical methodology in recent times, namely, the partition of unity finite element method (PUFEM), for solving Helmholtz equations [33–35], where at each of the FE nodes, the potential is expanded into one or multiple discrete series of plane waves. However, no work has been reported for elastic waves in bounded media. The present element is also termed as *PUFEM* in the following content.

To prevent transverse shear locking while applying FSDT, the order of approximation has to be at least quadratic along the y axis. Therefore, 8- or 12-noded isoparametric shape functions are used as the "old shape functions". Rewriting the enriched interpolation functions as $P_{j,l} = N_j\psi_l$, the displacement field can be expressed as

$$\begin{Bmatrix} u(x, y, t) \\ v(x, y, t) \end{Bmatrix} = \sum_{m=1}^{M} \aleph\overline{\mathbf{u}}\mathrm{e}^{-i\omega_m t} \quad (11)$$

where M is the number of sampling points in the frequency domain, and

$$\aleph = \begin{bmatrix} \Re_1 & \mathbf{0} & \Re_2 & \mathbf{0} & \cdots & \Re_n & \mathbf{0} \\ \mathbf{0} & \Re_1 & \mathbf{0} & \Re_2 & \cdots & \mathbf{0} & \Re_n \end{bmatrix} \quad (12)$$

where

$$\Re_j = [\, P_{j,1} \quad P_{j,2} \quad \cdots \quad P_{j,NW} \,] \quad (13)$$

$\overline{\mathbf{u}}$ is a vector consisting of unknown nodal coefficients,

$$\overline{\mathbf{u}} = \{\, \boldsymbol{\alpha}_1 \quad \boldsymbol{\beta}_1 \quad \cdots \quad \boldsymbol{\alpha}_n \quad \boldsymbol{\beta}_n \,\}^T \quad (14)$$

where

$$\boldsymbol{\alpha}_j = \{\, A_j^1 \quad A_j^2 \quad \cdots \quad A_j^{NW} \,\} \quad (15a)$$

$$\boldsymbol{\beta}_j = \{\, B_j^1 \quad B_j^2 \quad \cdots \quad B_j^{NW} \,\} \quad (15b)$$

In equation (15a and 15b), for the different boundary conditions, careful definition of parameters A_j^n and B_j^n is needed. For example, for the energy absorbing boundary condition in the following examples, the parameters in A_j^n and B_j^n corresponding to the backward propagating modes should be set to be zero.

3 WEAK FORMULATION IN FREQUENCY DOMAIN

With the above description of the enriched displacement field, the Hamiltonian for the two-dimensional elastodynamics is given by

$$\Pi_P = \int_{t_1}^{t_2} \left[\frac{1}{2} \iint_\Omega \boldsymbol{\varepsilon}^T \mathbf{D} \boldsymbol{\varepsilon} \, \mathrm{d}x \, \mathrm{d}y - \int_S \mathbf{q}^T \boldsymbol{\delta} \, \mathrm{d}s - \sum_{k=1}^{K} \mathbf{P}_k^T \boldsymbol{\delta}_k - \frac{1}{2} \iint_\Omega \frac{\mathrm{d}\boldsymbol{\delta}^T}{\mathrm{d}t} \boldsymbol{\vartheta} \frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}t} \, \mathrm{d}x \, \mathrm{d}y \right] \mathrm{d}t \tag{16}$$

where t_1 and t_2 are the starting time and ending time, respectively, $\mathbf{D}$ is the elasticity matrix for the plane stress model adopted here, $\boldsymbol{\delta}$ is a vector containing the two displacement components and is defined as

$$\delta = \begin{Bmatrix} u \\ v \end{Bmatrix} \tag{17}$$

The diagonal matrix $\boldsymbol{\vartheta}$ for mass density ρ can be expressed as

$$\boldsymbol{\vartheta} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \tag{18}$$

The terms $\mathbf{P}_k$ and $\mathbf{q}$ in equation (16) are the applied concentrated loads and applied distributed loads, respectively. The strain vector $\boldsymbol{\varepsilon}$ in equation (16) is defined in the following form

$$\boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_{\mathrm{xx}} \\ \varepsilon_{yy} \\ \varepsilon_{\mathrm{xy}} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial u}{\partial x} \\ \dfrac{\partial v}{\partial y} \\ \dfrac{\partial u}{\partial y} + \dfrac{\partial v}{\partial x} \end{Bmatrix} = \sum_{m=1}^{M} \boldsymbol{\Theta} \overline{\mathbf{u}} \mathrm{e}^{-\mathrm{i}\omega_m t} \tag{19}$$

where

$$\boldsymbol{\Theta} = \begin{bmatrix} \dfrac{\partial \Re_1}{\partial x} & \mathbf{0} & \dfrac{\partial \Re_2}{\partial x} & \mathbf{0} & \cdots & \dfrac{\partial \Re_n}{\partial x} & \mathbf{0} \\ \mathbf{0} & \dfrac{\partial \Re_1}{\partial y} & \mathbf{0} & \dfrac{\partial \Re_2}{\partial y} & \cdots & \mathbf{0} & \dfrac{\partial \Re_n}{\partial y} \\ \dfrac{\partial \Re_1}{\partial y} & \dfrac{\partial \Re_1}{\partial x} & \dfrac{\partial \Re_2}{\partial y} & \dfrac{\partial \Re_2}{\partial x} & \cdots & \dfrac{\partial \Re_n}{\partial y} & \dfrac{\partial \Re_n}{\partial x} \end{bmatrix} \tag{20}$$

To obtain the strain–displacement matrix $\boldsymbol{\Theta}$, the derivatives in equation (20) are calculated as follows:

$$\frac{\partial P_{j,l}}{\partial x} = \left[\frac{\partial N_j}{\partial x} + i k_l N_j \right] \psi_l \tag{21a}$$

$$\frac{\partial P_{j,l}}{\partial y} = \left[\frac{\partial N_j}{\partial y} \right] \psi_l \tag{21b}$$

The kinetic energy term in equation (16) can further be expanded as

$$\begin{aligned} \Gamma &= \int_{t_1}^{t_2} \frac{1}{2} \iint_\Omega \frac{\mathrm{d}\boldsymbol{\delta}^T}{\mathrm{d}t} \boldsymbol{\vartheta} \frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}t} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}t \\ &= \frac{1}{2} \iint_\Omega \frac{\mathrm{d}\boldsymbol{\delta}^T}{\mathrm{d}t} \boldsymbol{\vartheta} \boldsymbol{\delta} \, \mathrm{d}x \, \mathrm{d}y \Big|_{t_1}^{t_2} \\ &\quad - \int_{t_1}^{t_2} \frac{1}{2} \iint_\Omega \boldsymbol{\delta}^T \boldsymbol{\vartheta} \frac{\mathrm{d}^2 \boldsymbol{\delta}}{\mathrm{d}t^2} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}t \end{aligned} \tag{22}$$

where

$$\begin{aligned} \frac{\mathrm{d}^2 \boldsymbol{\delta}}{\mathrm{d}t^2} &= \begin{Bmatrix} \dfrac{\mathrm{d}^2 u}{\mathrm{d}t^2} \\ \dfrac{\mathrm{d}^2 v}{\mathrm{d}t^2} \end{Bmatrix} = -\sum_{m=1}^{M} \omega_m^2 \begin{Bmatrix} \overline{u}_m \\ \overline{v}_m \end{Bmatrix} \mathrm{e}^{-\mathrm{i}\omega_m t} \\ &= -\sum_{m=1}^{M} \omega_m^2 \aleph \overline{\mathbf{u}} \mathrm{e}^{-\mathrm{i}\omega_m t} \end{aligned} \tag{23}$$

Subsequently, similar to the displacement field in equation (11), by employing FFT, the distributed load $\mathbf{q}$ and concentrated load $\mathbf{P}_k$ in the time domain are described as

$$\begin{aligned} \mathbf{q}(x, y, t) &= \sum_{m=1}^{M} \overline{\mathbf{q}}(x, y, \omega_m) \mathrm{e}^{-\mathrm{i}\omega_m t}, \\ \mathbf{P}_k(x_k, y_k, t) &= \sum_{m=1}^{M} \overline{\mathbf{P}}_k(x_k, y_k, \omega_m) \mathrm{e}^{-\mathrm{i}\omega_m t} \end{aligned} \tag{24}$$

For the convenience of description, the local elemental variables are transferred into the structural global ones as: $\overline{\mathbf{u}} = \mathrm{T}\overline{\mathbf{U}}$. Finally, equation (16) for an arbitrary element can be rewritten as

$$\Pi_P = \int_{t_1}^{t_2} \sum_{m=1}^{M} \left[\left(\frac{1}{2} \iint_\Omega \overline{\mathbf{U}}^T \mathrm{T}^T \boldsymbol{\Theta}^T \mathbf{D} \boldsymbol{\Theta} \mathrm{T} \overline{\mathbf{U}} \, \mathrm{d}x \, \mathrm{d}y - \int_S \overline{\mathbf{q}}^T \aleph \mathrm{T} \overline{\mathbf{U}} \, \mathrm{d}s - \sum_{k=1}^{K} \overline{\mathbf{P}}_k^T \aleph^k \mathrm{T} \overline{\mathbf{U}} \right. \right.$$

$$-\frac{1}{2}\omega_m^2\iint_\Omega \overline{\mathbf{U}}^T \mathrm{T}^T \aleph^T \boldsymbol{\vartheta} \aleph \mathrm{T} \overline{\mathbf{U}}\,\mathrm{d}x\,\mathrm{d}y\Big)(\mathrm{e}^{-\mathrm{i}\omega_m t})^2\Big]\mathrm{d}t$$
$$-\frac{1}{2}\int\int_\Omega \frac{\mathrm{d}\boldsymbol{\delta}^T}{\mathrm{d}t}\boldsymbol{\vartheta}\boldsymbol{\delta}\,\mathrm{d}x\,\mathrm{d}y\Big|_{t_1}^{t_2} \tag{25}$$

Using a variational principle, the FE equilibrium equation is obtained as

$$(\overline{\mathbf{K}} - \omega_m^2\overline{\mathbf{M}})\overline{\mathbf{U}} = \overline{\mathbf{P}} \tag{26}$$

where the complex matrices are expressed as follows:

$$\overline{\mathbf{K}} = \sum_1^{NE}\int\int_{\Omega_e} \mathrm{T}^T \boldsymbol{\Theta}^T \mathbf{D}\boldsymbol{\Theta}\mathrm{T}\,\mathrm{d}x\,\mathrm{d}y \tag{27a}$$

$$\overline{\mathbf{M}} = \sum_1^{NE}\int\int_{\Omega_e} \mathrm{T}^T \aleph^T \boldsymbol{\vartheta} \aleph \mathrm{T}\,\mathrm{d}x\,\mathrm{d}y \tag{27b}$$

$$\overline{\mathbf{P}} = \sum_1^{NE}\int_S \overline{\mathbf{q}}^T \aleph \mathrm{T}\,\mathrm{d}s - \sum_{k=1}^{K}\overline{\mathbf{P}}_k^T \aleph^k \mathrm{T} \tag{27c}$$

where NE is the number of elements. The main advantage of the above formulation is that a complex distribution in the y axis can be modeled easily by Lagangian FE interpolation, and a long but finite span in the x axis can be modeled accurately using a much smaller number of these new elements compared to the standard Lagrangian FE model. Moreover, the near-exact feature of wave propagation characteristics along the x axis is ensured, which leads to spectral convergence for interpolation parallel to the x axis. Subsequently, the h convergence can be expected for interpolation parallel to the y axis, since the wave modes are only functions of the x axis.

4 NUMERICAL IMPLEMENTATION

In the elemental matrices in equation (27a–c), the integrals encountered are of the form

$$I_{lm} = \int_{-1}^{1}\int_{-1}^{1} f(\xi,\eta)\mathrm{e}^{\mathrm{i}(k_l x)}\mathrm{e}^{\mathrm{i}(k_m x)}|\mathbf{J}|\mathrm{d}\xi\,\mathrm{d}\eta \tag{28}$$

where $\mathbf{J}$ is the Jacobian matrix for isoparametric mapping. The expression $f(\xi,\eta)$ involves the product of the "old shape function", their derivatives, etc. In this section, the evaluation of these integrals is carried out using higher-order Gauss–Legendre integration scheme. At the integration points within an element, x of $\mathrm{e}^{\mathrm{i}(k_l x)}$ and $\mathrm{e}^{\mathrm{i}(k_m x)}$ in equation (28) can be evaluated from the shape functions N_j of the current element. The number of integration points depends on the element nodal spacing with respect to the smallest wavelength ($\lambda_l = 2\pi/k_l$) at a given frequency ω_m. As shown later, only a few of the proposed elements in the spanwise direction of the beam, and hence, a few degrees of freedom can produce sufficient accuracy even for high-frequency excitation if the proper numerical integration scheme is used. Therefore, unlike the Lagrangian FEM, the computational time is mainly consumed at the step of numerical integration while calculating the dynamic stiffness matrix and the mass matrix, but not at the step of FE system solution. Numerical implementation of the proposed element requires efficient integration algorithms, such as frequency-dependent and elemental size-dependent numerical integration schemes. Consider a section with unit length, i.e., $x \in [0, 1]$, the number of cycles of sine and cosine functions within the domain is around $0.1k_1$. Therefore, to integrate one single cycle accurately, at least 30 G integration points are needed. For example, for more accuracy, if 60 G integration points are taken for one cycle of sine and cosine functions, the number of Gauss integration points (NG) for the element i in the x axis is roughly determined as $NG^i = 6L_e^i k_1^i(\omega_m)$, where $k_1^i(\omega_m)$ is the flexural wavenumber of the element i at frequency ω_m, and L_e^i is the length along the x axis of the element i. However, the minimum number of Gauss integration points is set to be 6. Such a consistent choice of frequency-dependent and elemental size-dependent integration scheme can efficiently reduce the computational cost. In the direction parallel to the y axis, the number of Gauss integration points is set to be two for an 8-noded element and three for a 12-noded element, respectively.

Having obtained $\overline{\mathbf{U}}$ containing A_j^l and B_j^l, the displacements in the frequency domain can be calculated using equation (9); therefore, the solution in the time domain is evaluated simply by applying the inverse FFT.

4.1 Assemblage of the interior and the exterior regions using global–local approach

As verified later from the numerical examples, the proposed PUFEM is very efficient for quasi-two-dimensional waves propagation at low, as well as high frequencies. However, at high frequencies, the accuracy is not so high when the forward and the backward propagation exist simultaneously and the propagation distance is long. The main reason is the dissimilar orders in the diagonal elements of the stiffness and the mass matrices, which are contributed by the exponential terms, i.e., the terms determined by wavenumbers k_3 and k_4. For high-frequency waves, the components of stiffness and mass matrices, which involve $\psi_3 = \mathrm{e}^{\mathrm{i}[k_3(x-x_L)]}$ in equation (9) for large x, and $\psi_4 = \mathrm{e}^{-\mathrm{i}[k_4(x_U-x)]}$ in equation (9) for large x_U and small x, are too small compared to the other elements in the matrices. These small elements, especially small diagonal elements of the stiffness and mass matrices, may cause the numerical instability. For instance, to deal with the boundary conditions for propagation in both directions, the forward and backward terms are coupled at the boundary. When x is large at the boundary, the forward terms involving $\psi_3 = \mathrm{e}^{\mathrm{i}[k_3(x-x_L)]}$ are very small. On the other hand, the backward terms involving $\psi_4 = \mathrm{e}^{-\mathrm{i}[k_4(x_U-x)]}$ are well conditioned. The numerical instability is thus inherent.

For the cases of unidirectional propagation of waves, e.g., only forward propagation alone, the elements of the stiffness matrix and the mass matrix involve only the term $\psi_3 = \mathrm{e}^{\mathrm{i}[k_3(x-x_L)]}$. As stated above, although this term has the similar exponential characteristics as x increases, by collocating the diagonal elements corresponding to the exponential decay in the matrix $(\overline{\mathbf{K}} - \omega_m^2 \overline{\mathbf{M}})$ in a decreasing sequence, the numerical instability can be effectively removed. However, this technique is not suited for the cases of the simultaneous propagation of forward and backward waves. For such double directional propagation of high-frequency waves, the PUFEM is only applicable for short traveling distance. In fact, from the authors' numerical experience, the existence of the exponential decay causes many numerical problems. To overcome this problem, the new element needs to be combined with the ordinary spectral element model or any other efficient discretized model of the exterior region. As shown in Figure 1, this new element is employed for the smaller interior region, and the ordinary spectral elements are used to model the exterior region. Another purpose of assembling the ordinary spectral element for the uniform exterior region is to reduce the computational cost to a greater extent.

In the global–local approach, the displacement continuity at the boundary of the two regions must be ensured. With the help of equation (9), and by setting $N_j = 1$, the condition of displacement continuity between the interior and the exterior regions (see Figure 1) at the node j in an arbitrary new element can be expressed as

$$\overline{u}_m(x, y, \omega_m) = \sum_{l=1}^{NW} \psi_l A_j^l = -y\overline{\phi}(x, \omega_m) \quad (29a)$$

$$\overline{v}_m(x, y, \omega_m) = \sum_{l=1}^{NW} \psi_l B_j^l = \overline{w}(x, \omega_m) \quad (29b)$$

where $\overline{\phi}(x, \omega_m)$ and $\overline{w}(x, \omega_m)$ are the rotation and the transverse displacement degrees of freedom in the spectral element on the side of the exterior region (see [27] for details).

There exist three different approaches for enforcing the interface displacement continuity in the context of global–local FE analysis. The first approach is the direct enforcement of the constraints at the interfaces as reported by Gopalakrishnan and Doyle [36]. The second approach is the weak enforcement using Lagrange multiplier as reported by Halliday and Grosh [37]. Finally, the third approach is the weak enforcement based on multipoint constraints as reported by Mahapatra and Gopalakrishnan [12]. While using two dissimilar models, e.g., standard Lagrangian FE model for the interior region and the ordinary spectral element model for the exterior region, direct enforcement of interface constraints [36] appears computationally intensive and requires cross-checking against numerical convergence. Again, from the authors' numerical experience, it is found that the penalty function method is not efficient while using dissimilar models for the exterior and the interior regions, since the choice of penalty parameters is very difficult and the introduction of large penalty parameters may result in numerical instability. In the present global–local approach,

the Lagrange multiplier method is employed to enforce the displacement continuity conditions given in equation (29a and b).

5 NUMERICAL EXAMPLES

5.1 Comparison with the traditional FEM

Consider a one-dimensional (1-D) beam shown in Figure 2, which is subjected to a transverse load $P(t)$ at the free end, i.e., $x = 0$, which is expressed as

$$P(t) = \begin{cases} 0.5\left[1 - \cos\left(\dfrac{2\pi f t}{N}\right)\sin(2\pi f t)\right], & \dfrac{t \le N}{f} \\ 0, & \dfrac{t > N}{f} \end{cases} \tag{30}$$

where f is the central frequency in hertz and N is the number of sinusoidal cycles within a pulse.

To compare with the traditional FEM, the authors consider the low-frequency excitation case, where $f = 50$ Hz and $N = 15$ cycles. The sample duration time T is 4.8 s. The Nyquist frequency is $N_q/2T$, where the number of sampling points is $N_q = 2^k$, $k = 12$, and is used consistently in the several examples that follow. The traditional 2-D eight-noded isoparametric element is employed for comparison. The traditional FEM mesh possesses 84 elements with 2 elements in the thickness direction and 42 elements in the spanwise direction. Only six eight-noded PUFEMs are used with two elements in the thickness direction and three elements in the spanwise direction. The same eight-noded elements are used in all the examples. The beam is made of aluminum with the following properties: $E = 73.0$ GPa, $G = 28.08$ GPa, $\nu = 0.3$, $\rho = 2770\,\mathrm{kg\,m^{-3}}$. The thickness of the beam (h) is equal to 10.0 mm. The span of the beam, for the present example, is taken as 1260.0 mm. The width of the beam is taken as 10.0 mm. The above constants are consistently used in all the examples. The responses at the far ends are assumed to be zero and the reflected wave modes are eliminated. To impose such absorbing boundary conditions, the nodal degrees of freedom corresponding to the backward propagating modes are set to be zero. It means that A_j^2, A_j^4, B_j^2, and B_j^4 in equation (9) are set to be zero at all the nodes. Furthermore, A_j^1, A_j^3, B_j^1, and B_j^3 in equation (9), which are the degrees of freedom corresponding to the forward propagation, are set to be zero at the nodes on the far end. Two measurement points are considered at $x = 0.0$ mm, and at $x = 210.0$ mm from the source of excitation. Comparison of the time histories of deflections at the measurement points is shown in Figure 3. This result reveals that the PUFEM can attain very high accuracy, although the mesh is very coarse. In fact, the present method is almost insensitive to the number of elements when a sufficiently higher-order Gauss integration scheme is employed in the x axis.

5.2 Comparison with the throw-off spectral element

Consider a similar problem as shown in Figure 2. The high-frequency case, where $f = 20$ kHz, and $N = 5$ cycles is investigated. The sample duration time T is 0.012 s. The authors compare the performance of the PUFEM with that of the throw-off spectral element [27], which only tackles the unidirectional wave propagation. It means that in throw-off elements, waves propagate in one direction toward infinity without any reflections. Unlike the example in Section 5.1, the performance of the PUFEM under a much higher excitation frequency is checked here. In this example, the beam is considered to have enough length, i.e., 1200.0 mm. The response at the far end is imposed to be zero and the reflected wave modes are excluded.

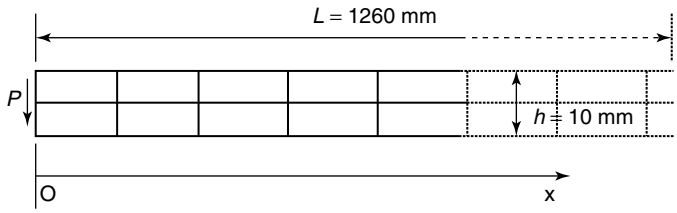


Figure 2. Schematic diagram of a 1-D problem.

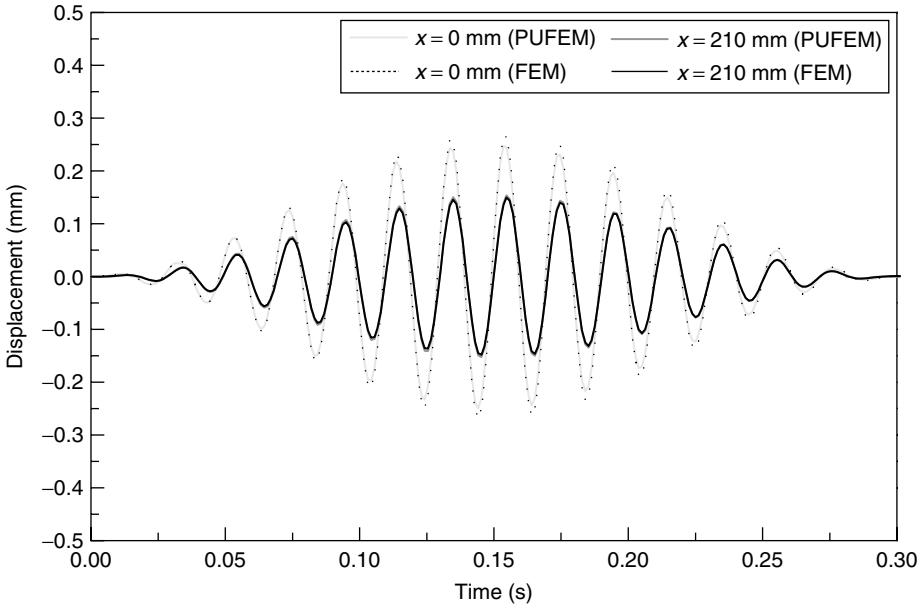


Figure 3. Deflections of the PUFEM and the traditional FEM at two measurement points.

Therefore, the unknown parameters, i.e., A_j^l and B_j^l are dealt in the same way as in the previous example. The present approach employs 16 elements with 2 elements in the thickness direction and 8 elements in the spanwise direction. For high-frequency cases, the amplitude of deflection becomes very small, and here the transverse velocity is selected for plotting. The transverse velocities at two measurement points, i.e., at $x = 0.0$ and 100.0 mm are obtained. Comparison of the time histories of transverse velocity at these measurement points is shown in Figure 4. It may be noted that the result obtained using the PUFEM is in good agreement with those obtained using the spectral throw-off element. The influence of the number of Gauss integration points on the results is investigated. It was found that the integration number determined by $NG^i = 6L_e^i k_1^i(\omega_m)$ can yield the sufficiently converged results.

5.3 Comparison with the ordinary spectral element

In this section, a cantilever beam as shown in Figure 5(a) is considered and two ordinary spectral elements [27] are employed, i.e., one finite and the other throw off. The material properties are the same as in the previous examples, and $f = 20$ kHz, and $N = 5$ cycles are chosen. The length of the beam from the fixed end to the point of application of transverse load is 2000.0 mm. As shown in Figure 5(b), the entire domain in Figure 5(a) is divided into three parts, i.e., two finite ordinary spectral elements with lengths of 990.0 mm, and one interior region with a length of 20 mm, discretized by the PUFEM. Two PUFEMs are employed for the interior region between two normal spectral elements in Figure 5(a), where in the span direction, only one element is used. The comparison of time histories of transverse velocity at the loading point is depicted in Figure 6. This figure shows both the incident and the reflected waves clearly. It can also be noted that both results, which are based on the meshes in Figure 5(a) and (b), are in good agreement. However, there are some small oscillations while using the global–local approach. This phenomenon may be caused by the mismatch between the present elemental dynamic stiffness and that of the ordinary spectral element due to the different assumptions for the displacement field interpolation. The fact is that there is no contraction in the thickness direction considered in the ordinary spectral element model. In fact, by

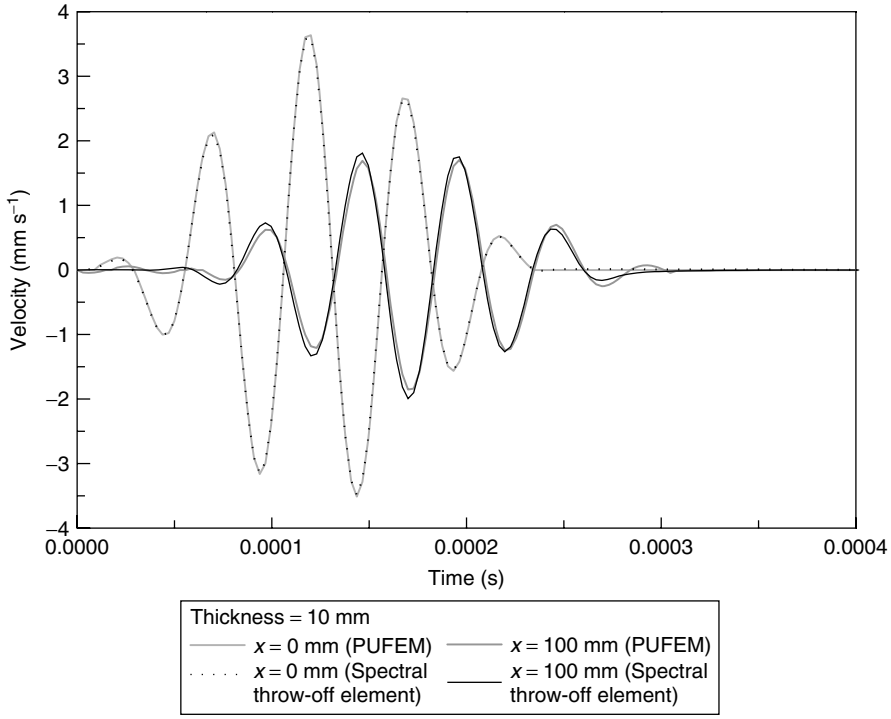


Figure 4. Transverse velocities of the PUFEM and the throw-off spectral element at two measurement points for a beam of thickness of 10 mm.

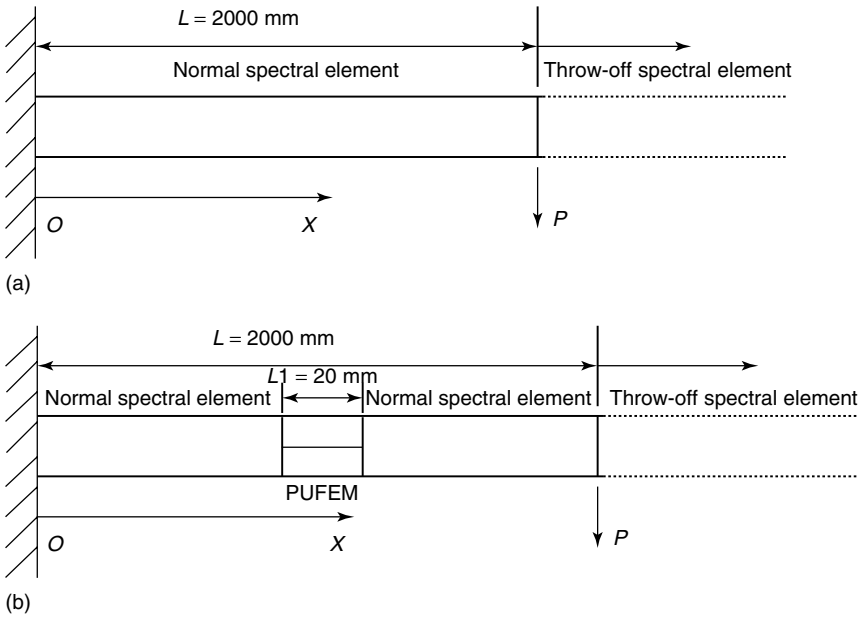


Figure 5. (a) Schematic diagram of a cantilever beam using the spectral element only. (b) Schematic diagram of a cantilever beam using the hybrid approach.

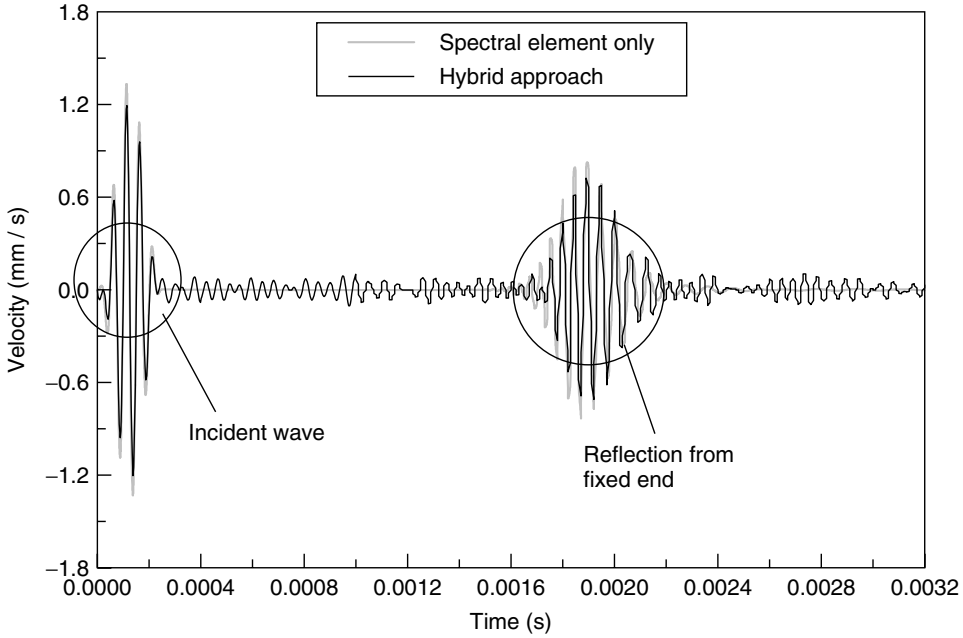


Figure 6. Transverse velocities of the spectral element only and the hybrid approach at the load point.

observing Figure 4, it can also be found that the results of the PUFEM do not completely match with those of the throw-off spectral element. Generally, this mismatch does not cause a serious problem in the low-frequency cases. However, with the increase in the excitation frequency, the effects of mismatch become more obvious. From the authors' numerical experience, increasing the number of the PUFEM along the span and thickness directions to alleviate the effects of this mismatch is not effective.

5.4 A cantilever beam with two symmetric transverse cracks

In this section, another example identical to the previous one is considered, except that there are two symmetric transverse cracks having depth $h1$ as shown in Figure 7. Six PUFEMs are employed to model the interior region. With the PUFEM, the crack geometry can be easily modeled. The placement of the element nodes at the interior region and the crack surfaces are shown in Figure 7. The crack surfaces essentially form the interelement discontinuity. The effect of contact between the two crack surfaces is neglected. The time histories of transverse velocity of the intact and cracked beams ($h1 = 3.0\,\text{mm}$) at the loading point are shown in Figure 8. The first reflection from the cracks can be accurately identified in Figure 8, which arrives at $t = 0.001\,\text{s}$. The first reflection from the fixed end can be identified for both the intact and cracked cases. However, the amplitude of the first reflection from the fixed end in the cracked beam is much lower than that of the intact beam. Furthermore, the second reflection from the fixed end can also be identified for the cracked beam. Here, the second reflection is the combination of (i) part of the first reflection from the fixed end, which is reflected by the cracks back to the fixed end and (ii) the second reflection stated in (i) by the fixed end, which finally arrives at the measurement point. This second reflection does not appear in the case of the intact beam. Moreover, as shown in Figure 9, four transverse cracks are considered. The interior region is 120-mm long, and the distance between the two sets of transverse cracks is 80 mm. Nine PUFEMs are employed for the interior region, and the crack depth $h1 = 3.0\,\text{mm}$. Other conditions are identical to the above example. The transverse velocity at the loading

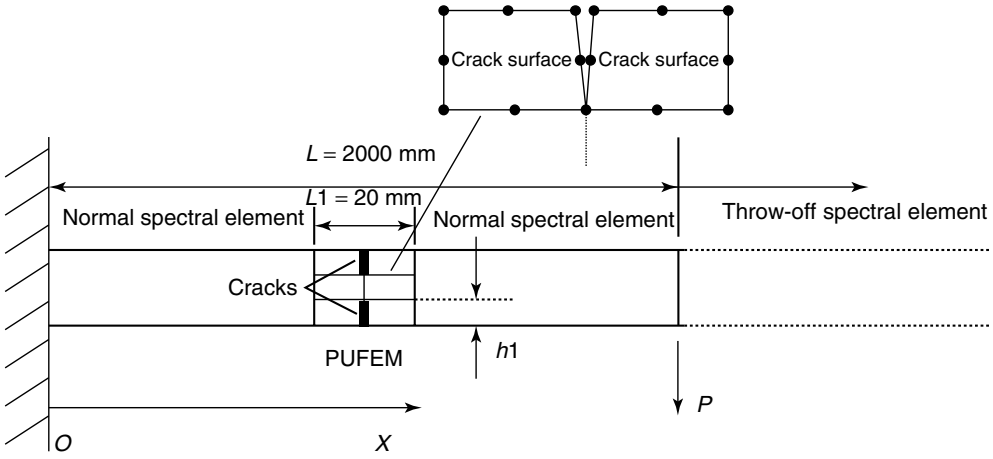


Figure 7. Schematic diagram of a cantilever beam with two symmetric transverse cracks.

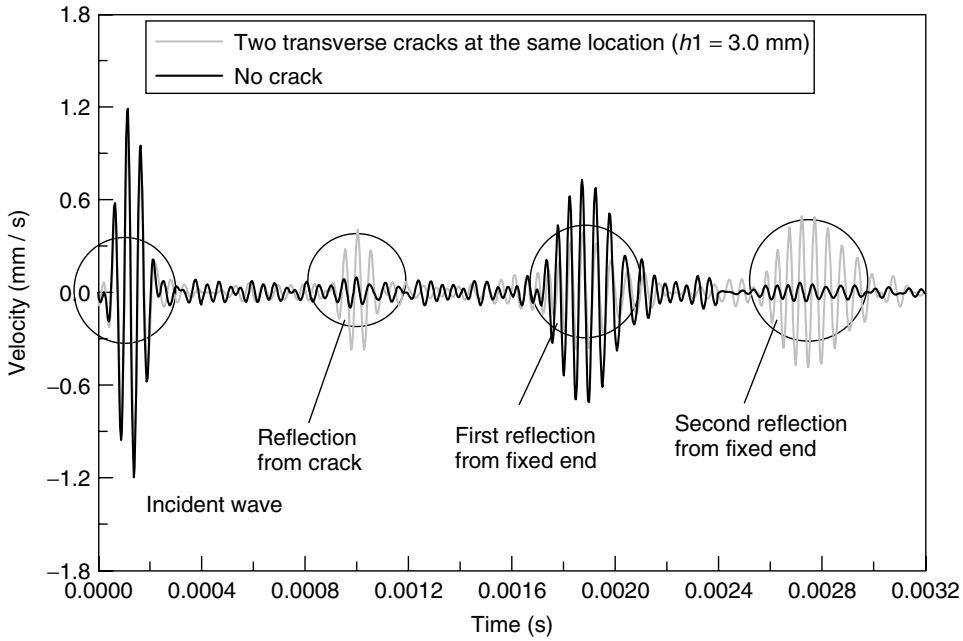


Figure 8. Transverse velocities of intact and cracked beams for $h1 = 3.0$ mm at the load point.

point is shown in Figure 10. It can be seen that the reflections from two sets of cracks overlap due to the short distance between the cracks. Compared with the result of the problem with two transverse cracks in Figure 8, the amplitude of reflection from the four cracks is higher. However, the reflection from the fixed end has decreased significantly.

5.5 A cross-ply cantilever beam with a delamination at midplane

Here, a cross-ply cantilever beam shown in Figure 11 is considered, where a delamination of length of 25 mm is located at the midplane of the beam. We consider six-ply laminates with $[0/90/0]_s$ in

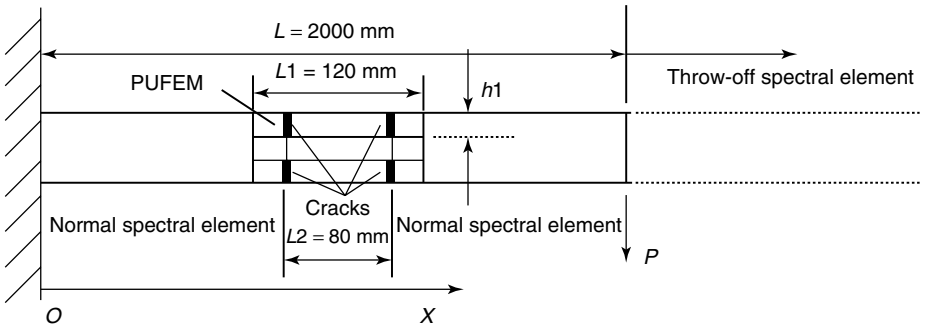


Figure 9. Schematic diagram of a cantilever beam with four transverse cracks.

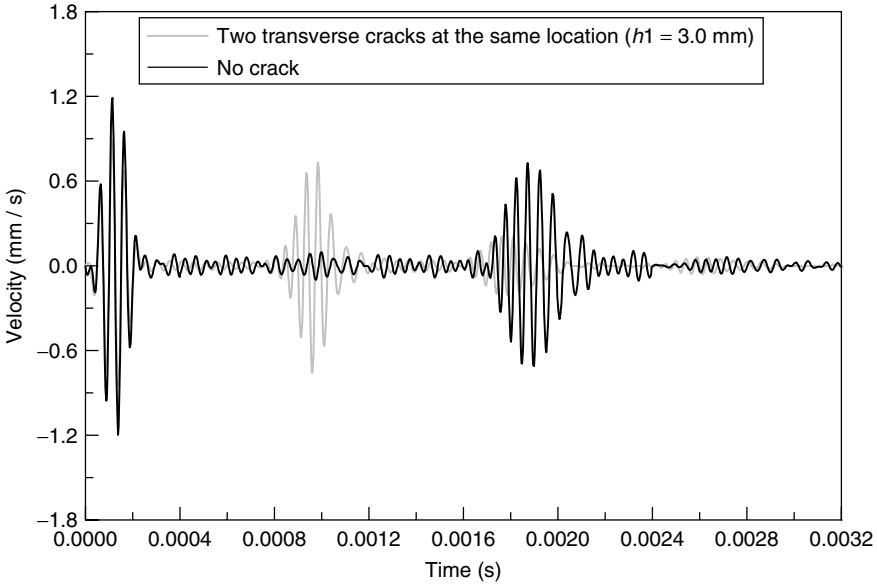


Figure 10. Transverse velocities of intact and cracked beams at the load point (transverse cracks).

stacking sequence with the following material properties of the lamina: $E_{11} = 145.5\,\text{GPa}$, $E_{22} = 9.69\,\text{GPa}$, $G_{12} = G_{13} = 5.97\,\text{GPa}$, $G_{23} = 2.91\,\text{GPa}$, $\nu_{12} = 0.32$, $\rho = 1550\,\text{kg m}^{-3}$. The lamina thickness is 0.5 mm. Six PUFEMs, with three elements in the spanwise direction and two elements in the through-thickness direction, are employed to model the interior region. Other conditions are identical to the above examples. Note that the dispersion relationship in the delamination area is different from that of the intact portion. The effect of contact between the two crack surfaces is neglected. Time histories of transverse velocity at the loading point for the two different crack lengths are shown in Figure 12 for $f = 20\,\text{kHz}$. This figure reveals that the reflection from the delamination can be clearly identified.

6 CONCLUSIONS

This article presents a new global–local approach to model wave propagation in beams with various types of damages. In this approach, the ordinary spectral element method is employed to simulate the behavior of wave propagation in the exterior regions, whereas newly proposed elements are used to model the interior region containing damages in the form of cracks. Formulation of the proposed

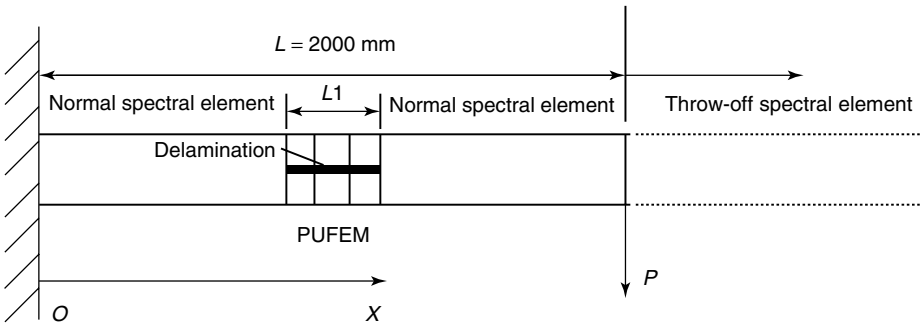


Figure 11. Schematic diagram of a cantilever beam with a delamination at midplane.

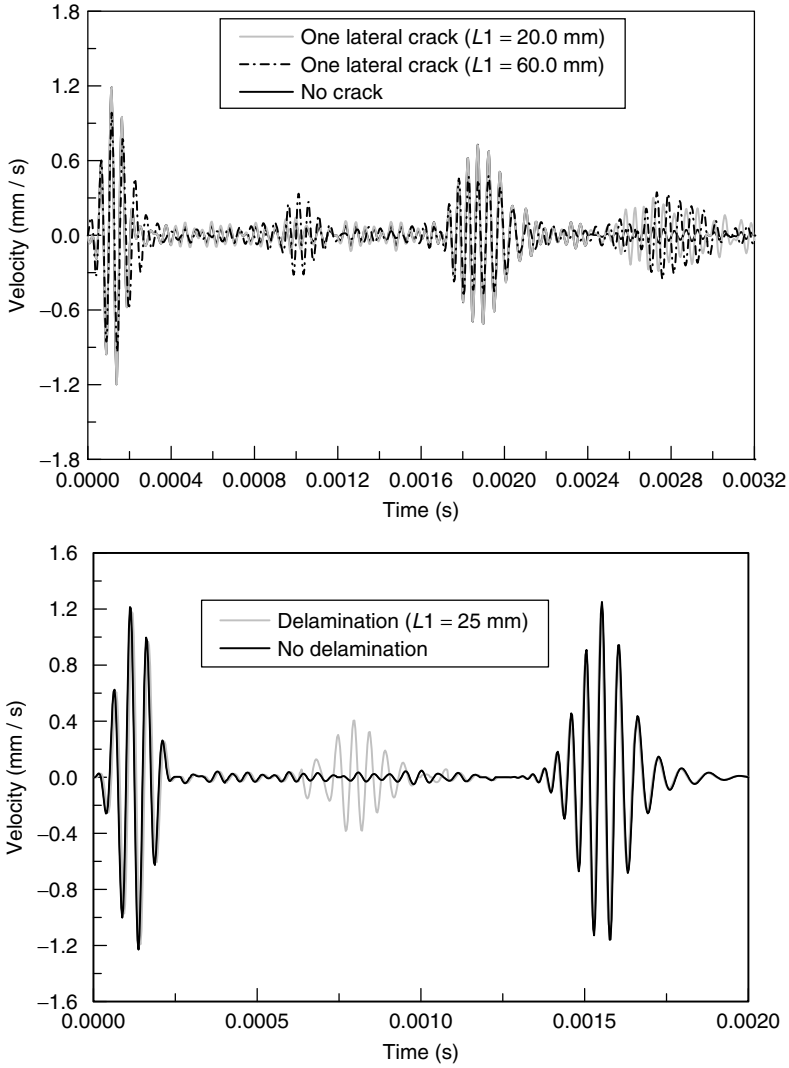


Figure 12. Transverse velocities of intact and delaminated beams at the load point (an in-plane delamination).

element is based on the hybrid interpolation scheme, where the Lagrangian family of interpolation bases is enriched by introducing four wave modes (two flexural wave modes and two shear wave modes) associated with the interpolation along one of the two mutually orthogonal coordinates for the 2-D plane elastodynamic problem. A frequency-domain FE model is then obtained by minimizing the Hamiltonian. The proposed approach essentially balances the advantages of (i) the ordinary spectral element model, which is the direct solution to the strong form, in which the solution is enforced in a pointwise sense, as well as (ii) the weakly formulated FE, where the solution is enforced in a piecewise discrete sense. The advantage of (i) is in the high numerical efficiency while solving transient elastodynamics. The advantage of (ii) is in the efficient handling of complex geometry.

With this balanced approach, accurate analyses can be realized with a very coarse FE mesh. The proposed element is highly efficient and flexible, and hence can be used to model more complex problems. Some numerical examples are shown to illustrate the effectiveness of the 8-noded PUFEM. Wave propagation in beams with various damages in the form of transverse cracks and delaminations in metallic and composite materials has been studied. It is shown from the numerical results that the reflection from the damages can be identified, and the pattern of wave propagation in beams with complex damage configurations can be efficiently studied. The identified reflection from the damage can be used to effectively locate the damage position in SHM.

REFERENCES

[1] Percival WJ, Birt EA. A study of Lamb wave propagation in carbon-fibre composites. *Insight* 1997 **39**:728–735.

[2] Birt EA. Damage detection in carbon-fibre composites using ultrasonic Lamb waves. *Insight* 1998 **40**:335–339.

[3] Karim MR, Awal MA, Kundu T. Elastic wave scattering by cracks and inclusions in plates: in plane case. *International Journal of Solids and Structures* 1992 **29**:2355–2367.

[4] Mal AK, Chang Z. A semi-numerical method for elastic wave scattering calculations. *Geophysical Journal International* 2000 **143**:328–334.

[5] Hu N, Shimomukai T, Fukunaga H, Su Z. Damage identification of metallic structures using A_0 mode of Lamb waves. *Structural Health Monitoring: An International Journal* (in press).

[6] Liu GR, Achenbach JD. A strip element method for stress analysis of anisotropic linearly elastic solids. *Journal of Applied Mechanics* 1994 **61**:270–277.

[7] Aberg M, Gudmundson P. Micromechanical modeling of transient waves from matrix cracking and fiber fracture in laminated beams. *International Journal of Solids and Structures* 2000 **37**:4083–4102.

[8] Guo N, Cawley P. The interaction of Lamb waves with delaminations in composite laminates. *Journal of the Acoustical Society of America* 1993 **94**:2240–2246.

[9] Hu N, Shimomukai T, Yan C, Fukunaga H. Identification of delamination position in cross-ply beams using S_0 Lamb mode. *Composites Science and Technology* 2008 **68**:1548–1554.

[10] Krawczuk M, Palacz M, Ostachowicz W. The dynamic analysis of a cracked Timoshenko beam by the spectral element method. *Journal of Sound and Vibration* 2003 **264**:1139–1153.

[11] Nag A, Mahapatra DR, Gopalakrishnan S, Sankar TS. A spectral finite element with embedded delamination for modeling of wave scattering in composite beams. *Composites Science and Technology* 2003 **63**:2187–2200.

[12] Mahapatra DR, Gopalakrishnan S. Spectral finite element analysis of coupled wave propagation in composite beams with multiple delaminations and strip inclusions. *International Journal of Solids and Structures* 2004 **41**:1173–1208.

[13] Bork U, Challis RE. Artificial neural networks applied to Lamb wave testing of T-form adhered joints. In *Proceedings of the Conference on the Inspection of Structural Composites*, Saffari N (ed). Bentham Press: London, 1994; pp. 127–132.

[14] Su Z, Ye L. Lamb wave propagation-based damage identification for quasi-isotropic CF/EP composite laminates using artificial neural algorithm, Part-I: methodology and database development. *Journal of Intelligent Material Systems and Structures* 2004 **16**:97–111.

[15] Zang C, Friswell MI, Imregun M. Structural damage detection using independent component analysis. *Structural Health Monitoring: An International Journal* 2004 **3**:69–83.

[16] Su Z, Yang C, Pan N, Ye L, Zhou LM. Assessment of delamination in composite beams using shear

horizontal (SH) wave mode. *Composites Science and Technology* 2007 **67**:244–251.

[17] Strickwerda JC. *Finite Difference Schemes and Partial Differential Equations*. Wadsworth-Brooks: Belmont, CA, 1989.

[18] Zienkiewicz OC. *The Finite Element Method*. McGraw-Hill, 1989.

[19] Koshiba M, Karakida S, Suzuki M. Finite element analysis of Lamb waves scattering in an elastic plate waveguide. *IEEE Transactions on Sonics and Ultrasonics* 1984 **31**:18–25.

[20] Alleyne DN, Cawley P. Optimization of Lamb wave inspection techniques. *NDT and E International* 1992 **25**:11–22.

[21] Brebbia CA, Tells JCF, Wrobel LC. *Boundary Elements Techniques*. Springer: Berlin, 1984.

[22] Cho Y, Rose JL. A boundary element solution for mode conversion study of the edge reflection of Lamb waves. *Journal of the Acoustical Society of America* 1996 **99**:2079–2109.

[23] Pao YH, Keh DC, Howard SM. Dynamic response and wave propagation in plane trusses and frames. *AIAA Journal* 1999 **37**:594–603.

[24] Banerjee JR, Williams FW. Exact dynamic stiffness matrix for composite Timoshenko beams with applications. *Journal of Sound and Vibration* 1996 **194**:573–585.

[25] Hou LJ, Peters DA. Application of space-time finite elements to problems of wave propagation. *Journal of Sound and Vibration* 1994 **173**:611–632.

[26] Doyle JF. *Wave Propagation in Structures, Spectral Analysis Using Fast Discrete Fourier Transforms, Second Edition*, Springer-Verlag: New York, 1997.

[27] Mahapatra DR, Gopalakrishnan S. A spectral finite element model for analysis of axial-flexural-shear coupled wave propagation in laminated composite beams. *Composite Structures* 2003 **59**:67–88.

[28] Lee U, Lee J. Spectral-element method for Levy-type plates subject to dynamic loads. *Journal of Engineering Mechanics* 1999 **125**:243–247.

[29] Patera AT. A spectral element method for fluid dynamics: laminar flow in channel expansion. *Journal of Computational Physics* 1984 **54**:468–488.

[30] Pozrikidis C. *Introduction to Finite and Spectral Element Methods using MATLAB®*. Chapman & Hall/CRC: London/Boca Raton, 2005.

[31] Sridhar R, Chakraborty A, Gopalakrishnan S. Wave propagation analysis in anisotropic and inhomogeneous uncracked and cracked structures using pseudospectral finite element method. *International Journal of Solids and Structures* 2006 **43**:4997–5031.

[32] Kudela P, Zak A, Krawczuk M, Ostachowicz W. Modelling of wave propagation in composite plates using the time domain spectral element method. *Journal of Sound and Vibration* 2007 **302**:728–745.

[33] Melenk JM, Babuska I. The partition of unity finite element method. Basic theory and applications. *Computational Methods in Applied Mechanics and Engineering* 1996 **139**:289–314.

[34] Laghrouche O, Bettess P, Astley RJ. Modelling of short wave diffraction problems using approximating systems of plane waves. *International Journal of Numerical Methods in Engineering* 2001 **54**:1501–1533.

[35] Camallo P, Astley RJ. The partition of unity finite element method for short wave acoustic propagation on non-uniform potential flows. *International Journal of Numerical Methods in Engineering* 2006 **65**:425–444.

[36] Gopalakrishnan S, Doyle JF. Spectral super-elements for wave propagation in structures with local non-uniformities. *Computational Methods in Applied Mechanics and Engineering* 1995 **121**:77–90.

[37] Halliday PJ, Grosh K. Dynamic response of complex structural interconnections using hybrid methods. *Journal of Applied Mechanics* 1999 **66**:653–659.

Chapter 47

Damage Detection Using Piezoceramic and Magnetostrictive Sensors and Actuators

Srinivasan Gopalakrishnan

Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

1 INTRODUCTION

In recent years, considerable research has been focused on improving the structural integrity (or increasing the life) of the structure. The structural integrity and the evolution of lightweight structural design form the two most severe design constraints a structure has to satisfy throughout its service life. Structural integrity performance of a structure is affected owing to the presence of minor cracks in the structure that are not visible to the naked eye. Detecting these cracks early in their service life poses a major challenge. An early detection of these cracks and adopting a suitable patch repair scheme will result in prolonging the service life of a structure. The area of crack detection in metallic and composite structure is a well-researched topic world over and large amount of literature is available on various methods of detecting cracks (or delaminations). This article highlights one such method of detecting cracks using embedded/surface-mounted smart magnetostrictive or piezoelectric material patches.

The stringent lightweight design constraints have forced the manufacturers of next generation aerospace and automobile vehicles to look for composites as structural members. They provide numerous opportunities to tailor the strength in the required direction and enable the placement of embedded sensors and actuators at any critical location to monitor the performance of the structure. This facility is not available in conventional metallic structures. Damage tolerant design of such structures is an unexplored area of research. Unlike the design of metallic structures, for composites, this information has not been fully integrated into design extensively. Although matrix cracking, fiber breaking, debonding, etc. are often the initiating failure mechanisms for laminated composites during its manufacture and in service, interlaminar crack, or delamination, is the one that is commonly found vulnerable and can grow, thus reducing the life of a structure. Impact is also a major source of delamination in composites. Traditionally,

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

conventional nondestructive techniques (NDT) such as ultrasonics, fractography, thermography, or tomography have been extensively used to detect the presence of damage. However, these techniques require that the position of damage is known *a priori* and the region that is being inspected is readily accessible [1]. These limitations make them very expensive for damage detection in aircraft or spacecraft structures, which need frequent inspection. Therefore, in-service damage-detection (health monitoring) system is essential, which can constantly monitor the integrity of the structure and also determine the location and extent of damage with built-in sensors. Some of the examples of sensors/actuators embedded in composites for damage detection are reported in [2, 3]. The authors had embedded the piezoelectric material in composites in the form of wafer, fiber, and powder form. Also, the use of shape memory alloy for damage detection is reported in [4]. The use of fiber-optic sensors for damage detection is reported in [5]. This article demonstrates one such built-in system with both collocated and noncollocated magnetostrictive and piezoelectric sensor/actuator combination to detect the presence of damage.

We now review some of the damage-detection methods for structural health monitoring. The most popular approach in structural health monitoring is to use modal parameters such as natural frequencies, damping ratios, and mode shapes to determine the existence of damage in structures. It is based on the principle that the presence of damage will change the natural frequencies of the structure owing to reduction in its stiffness. Monitoring the natural frequencies before and after the damage can confirm its presence. References 6 and 7 show the use of modal test data for evaluating structural integrity of offshore structures. Crawley and Adams [8] used the same technique for composite laminates, wherein they showed that the presence of delamination not only decreases natural frequencies but also increases damping in structures. In all the above studies, variation of mass matrix, mode shapes, uncertainty in structural parameters, or instrumentation accuracy were not considered. West [9] presented what is possibly the first systematic use of mode shapes for damage detection. A finite element (FE) study conducted by Chen and Swamidas [10] on a cracked cantilever plate showed that the higher order modes are least affected by the presence of damage. Reference 11 shows the use of lower order modes for nondestructive damage detection and sizing of cracks in beams. A new method based on the singular value decomposition was proposed by them to detect the structural damage. One of the fundamental difficulties with modal methods is that a small delamination causes very negligible loss of stiffness in composites. As a result, the frequency changes are so small that they are very difficult to measure accurately.

Residual force method (damage force method) is another technique of determining damage in structures. This is done by taking the difference between the measured damaged model properties with the undamaged (baseline) model properties. The advantage of this method is that one need not solve the complete system to determine the damage. References 12, 13 have shown the use of this technique in the modal domain. Schulz *et al.* [14] used the undamaged stiffness and mass matrix, obtained by the FEs, to determine the location of damage. Nag *et al.* [15] modified this method for its use in frequency domain through a spectral element formulation. The main advantage of the spectral formulation is that the system sizes are very small owing to exact mass distribution. Requirement of too many sensor measurements is the main disadvantage of the damage force method.

The recent trend is to use wave-based techniques for health monitoring. Pines [16] showed that damage could be ascertained by looking at the changes in the local wavenumber of a Lamb wave. In [17], a narrow-band modulated pulse, impacted transversely on a laminated composite beam, was used to detect damage. The characteristic of this pulse is that it travels nondispersively even in a dispersive medium. When such a pulse interacts with a crack, it produces an additional reflection from the crack tip owing to impedance mismatch. This reflection is easily recognizable, as it does not change its shape. Knowing the length of travel and speed of the medium, the location of the damage can be assessed. References 15, 18 show the use of spectral element model in damage modeling and assessment. (*See* **Damage Measures** for more details on the methods described in the last few paragraphs.)

Next we consider multifunctional composites with embedded or surface-bonded piezoceramic patch made from lead zirconate titanate (PZT) material. If we strain the structure, the PZT material permittivity

changes, which generates an electric field across the sensor and hence introduces an electrical displacement, which produces a voltage across the sensor, which is called *open circuit voltage* (*OCV*). The structure can, in fact, be strained through another PZT embedded/surface-bonded patch, by passing a voltage in the required poled direction. In the presence of the damage, the OCV changes owing to a different stress state. The difference in the OCVs between the healthy and the damaged structure, which is called *damage-induced voltage* (*DIV*), confirms the presence of damage. This is the principle on which this work is based. The same principle is applied with the help of magnetostrictive material patch, wherein, one has to apply a magnetic field to strain the material, which causes the permeability of the medium to change, which introduces a magnetic field across the sensor. This changes the magnetic flux density, which produces a voltage across the senor, which is the OCV. As stated earlier, the difference in the OCVs of healthy and damaged state will confirm the presence of damage in the structure.

This article is organized as follows. In the next two sections, modeling aspects of the PZT and Terfenol-D embedded/surface-mounted composites are addressed. Modeling aspects in general and FE modeling in particular have already been addressed in **Modeling Aspects in Finite Elements**; **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators**. Hence, only a brief summary is given here. This is followed by some aspects of experimental study undertaken for the analysis of magnetostrictive composites. Next, some numerical examples and results are provided, followed by a summary of the article.

2 DAMAGE DETECTION USING PIEZOELECTRIC SENSORS/ACTUATORS

Damage detection requires a robust mathematical model and a number of measured responses. Both baseline (response from the undamaged structure) and the response from the damaged structure are required for detection purposes. The detection will be based on the voltage generated across the sensor owing to a mechanical load. Voltage changes depending on the stress condition of the structure. Difference in the voltages between the undamaged and the damaged configuration indicates the presence of damages. Voltage signatures are different for different types of damages, which is helpful to characterize the damages. In this section, we use the mathematical model based on FEs, which was explained in detail in **Piezoceramic Materials—Phenomena and Modeling** on the modeling of piezoelectric sensors/actuators. Also, only the final equations are provided for the sake of completeness. The first step in the development of the mathematical model is to write the constitutive model. As explained in **Piezoceramic Materials—Phenomena and Modeling**, the piezoelectric material has two constitutive laws: one is called the *sensing law*, which is extensively used in this section, and the second is the *actuation law*. When the piezoelectric material is embedded in the composites, the stresses would get coupled, resulting in stiffness coupling such as the bending-axial coupling, etc. Assuming that the condition of plane stress exists in the composites, the constitutive law for piezoelectric composites can be written as

$$\begin{Bmatrix} \{\boldsymbol{\sigma}\} \\ \boldsymbol{D}_z \end{Bmatrix} = \begin{bmatrix} [\hat{\boldsymbol{C}}] & -[\hat{\boldsymbol{e}}] \\ [\hat{\boldsymbol{e}}]^T & \hat{\boldsymbol{\mu}} \end{bmatrix} \begin{Bmatrix} \{\boldsymbol{\varepsilon}\} \\ \boldsymbol{E}_z \end{Bmatrix} = \begin{Bmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \sigma_{xz} \\ D_z \end{Bmatrix} = \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} & 0 & -\hat{e}_{31} \\ \hat{C}_{13} & \hat{C}_{33} & 0 & -\hat{e}_{32} \\ 0 & 0 & \hat{C}_{55} & 0 \\ \hat{e}_{31} & \hat{e}_{32} & 0 & \hat{\mu}_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{zz} \\ 2\varepsilon_{xz} \\ E_z \end{Bmatrix} \quad (1)$$

where $\{\boldsymbol{\sigma}\}^{\mathrm{T}} = \{\sigma_{xx} \quad \sigma_{zz} \quad \tau_{xz}\}$ is the stress vector, $\{\boldsymbol{\varepsilon}\}^{\mathrm{T}} = \{\varepsilon_{xx} \quad \varepsilon_{xx} \quad \gamma_{xz}\}$ is the strain vector, $\boldsymbol{D_i}$ is the electrical displacement, and E_z is the electrical field in the out-of-plane direction. $\hat{C}_{ij}$ are the stiffness coefficients and $\hat{\boldsymbol{e}}_{ij}$ are the piezoelectric coefficients. The explicit expression of the stiffness coefficients are provided in **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators** on the modeling of piezoelectric sensors/actuators.

A four-noded smart composite FE with two mechanical degrees of freedom and one electrical degree of freedom was derived in **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and**

Actuators. Here, we use this element as the mathematical model for damage detection. The FE equation is of the form

$$\begin{bmatrix} [\boldsymbol{M}_{uu}] & [\boldsymbol{0}] \\ [\boldsymbol{0}] & [\boldsymbol{0}] \end{bmatrix} \begin{Bmatrix} \{\ddot{\boldsymbol{u}}\}_e \\ \{\ddot{\boldsymbol{E}}_z\}_e \end{Bmatrix} + \begin{bmatrix} [\boldsymbol{K}_{uu}] & [\boldsymbol{K}_{uE}] \\ [\boldsymbol{K}_{uE}]^T & [\boldsymbol{K}_{EE}] \end{bmatrix} \begin{Bmatrix} \{\boldsymbol{u}\}_e \\ \{\boldsymbol{E}_z\}_e \end{Bmatrix} = \begin{Bmatrix} \{\boldsymbol{F}\}_e \\ \{\boldsymbol{q}\}_e \end{Bmatrix} \tag{2}$$

where, $[\boldsymbol{M}_{uu}]$ is the mass matrix, $[\boldsymbol{K}_{uu}]$ is the stiffness matrix corresponding to mechanical degrees of freedom, $[\boldsymbol{K}_{uE}]$ is the stiffness matrix due to electromechanical coupling, and $[\boldsymbol{K}_{EE}]$ is the stiffness matrix due to electrical degrees of freedom alone. Also, the nodal displacement vector at the four nodes is $\{\boldsymbol{u}\}_e = \{u_1\ w_1\ u_2\ w_2\ u_3\ w_3\ u_4\ w_4\}^T$ and the corresponding electric field vector is $\{\boldsymbol{E}_z\}_e = \{E_{z1}\ E_{z2}\ E_{z3}\ E_{z3}\}^T$. Here, $\{\boldsymbol{F}\}_e$ is the elemental nodal vector and $\{\boldsymbol{q}\}_e$ is the elemental charge vector. A detailed expression for these matrices is given in **Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators**.

The solution of equation (2) begins by expanding the equation:

$$\{\boldsymbol{E}_z\} = [\boldsymbol{K}_{EE}]^{-1}\{\boldsymbol{q}\} - [\boldsymbol{K}_{EE}]^{-1}[\boldsymbol{K}_{uE}]^T\{\boldsymbol{u}\} \tag{3}$$

Using the above equation in equation (2) and simplifying, we get

$$[\boldsymbol{M}_{uu}]\{\ddot{\boldsymbol{u}}\} + [\overline{\boldsymbol{K}}_{uu}] = \{\overline{\boldsymbol{F}}\} \tag{4}$$

where

$$[\overline{\boldsymbol{K}}_{uu}] = [\boldsymbol{K}_{uu}] - [\boldsymbol{K}_{uE}][\boldsymbol{K}_{EE}]^{-1}[\boldsymbol{K}_{uE}]^T \tag{5}$$

$$\{\overline{\boldsymbol{F}}\} = \{\boldsymbol{F}\} - [\boldsymbol{K}_{EE}]^{-1}\{\boldsymbol{q}\} \tag{6}$$

First, equation (4) is solved for the applied actuator loading $\{\overline{\boldsymbol{F}}\}$ to obtain the nodal displacement vector $\{\boldsymbol{u}\}$, which is substituted in equation (3), to get the induced electric field in the sensor patches, which is then postprocessed to get the voltage distribution. The approach to obtain the voltage distribution is quite different for the static and dynamic loading cases, which are explained below.

2.1 Postprocessing for sensor output signal

The embedded thin piezoelectric patches electroded on the surfaces are modeled using the above FEs. The present study is to investigate the feasibility and characteristic behavior of these embedded patches for sensing the presence of cracks under static and dynamic loading in metallic or composite structures. While embedding the sensors during the wet lay up process or by any other fabrication technique, residual stresses and residual charge may develop. However, during sensor calibration, such residual charge can be removed by grounding the circuit. The existing residual stresses can be expected to add a bias in the sensor output voltage level without any considerable change in the linear transduction characteristics under transients. Therefore, by assuming no residual stresses and no residual charge on the embedded piezoelectric patches, the sensor equation at the element level can be written as given in equation (3).

2.1.1 Static case

The voltage that develops across each sensor owing to the deformation-induced electric field E_z acting in the out-of-plane direction is given by

$$\overline{V} = \sum E_z \Delta_Z \tag{7}$$

where Δ_Z is the distance between the pair of element nodes across the depth. Corresponding charge Q stored on the surface of this sensor segment is

$$Q = \int D_Z t\, \mathrm{d}x = \int [\hat{e}_{31}\varepsilon_{xx} + \hat{e}_{32}\varepsilon_{zz} + \hat{\mu}_{33}E_z]t\, \mathrm{d}x\Big|_z \tag{8}$$

The above equation is obtained after substituting for electrical displacement from equation (1). The strains can be expressed in terms of nodal displacements using the strain–displacement relationship and the electrical field can be written in terms of nodal electrical field ($E_z = \sum N_i E_{zi}$), where the nodal electrical field is obtained from equation (3). In doing so,

equation (8) becomes

$$Q = \int [\hat{e}]^T [B_u] \{u\}_e t \, dx + \int \hat{\mu}_{33} [N]^T \{E_z\}_e t \, dx \Big|_z \tag{9}$$

In the above equation, $[B_u]$ is the strain–displacement matrix corresponding to normal stress, $\{u\}_e$ is the nodal displacement vector, and $\{E_z\}_e$ is the nodal electrical field vector. Equation (9) is evaluated at each of the piezoelectric sensor locations. Since this sensor segment behaves as a capacitive device with capacitance $C = Q/V$, the stored electric energy $(1/2CV^2)$ can be equated to obtain the equivalent open circuit voltage V_a measurable as the output of the distributed sensing. That is, the equivalent OCV accumulated on a single sensor surface $\Omega = \sum t \, \Delta x$ (having Δx as the length of individual sensor elements) can be computed as

$$V_a = \frac{\sum VQ}{\sum Q} \tag{10}$$

Although the mechanical deformation is under static loading, in actual experiment, the sensor output scaled by a charge amplifier is to be measured as the initial transient after closing a resistive capacitive circuit. This is because of the leakage of the electrostatic charge in the surrounding medium.

2.1.2 Dynamic case

In this case, the developed voltage V and charge Q are both time dependent. Equations (7–9) hold. The current from each sensor element is

$$\dot{Q} = \int [\hat{e}]^T [B_u] \{\dot{u}\}_e t \, dx + \int \hat{\mu}_{33} [N]^T \{\dot{E}_z\}_e t \, dx \Big|_z \tag{11}$$

Unlike the capacitive model in the static case, here the developed electric power $P = V\dot{Q}$ is to be equated to obtain the equivalent open circuit voltage V_a due to distributed sensing over the single sensor surface $\Omega = \sum t \, \Delta x$. That is,

$$V_a = \frac{\sum V\dot{Q}}{\sum \dot{Q}} \tag{12}$$

$$\sum \dot{Q} \neq 0 \tag{13}$$

The above equations for static and dynamic loading cases are used in the examples to follow in the next section to obtain the OCV distribution across the sensors due to the presence of cracks.

2.2 Numerical examples

To study the effectiveness of piezoelectric material patch in predicting the presence of cracks, we consider a double cantilever beam (DCB), wherein, three piezoelectric sensor patches are embedded, two of them above and below the crack and one in the line of the crack as shown in Figure 1. The dimensions of the beam along with the cross section are also shown in the figure. The structure is a unidirectional (0°) graphite/epoxy beam, whose material properties are shown in Table 1.

In the chosen configuration, as shown in Figure 1, the locations of the piezoelectric sensor patches are fixed and the length of the crack (delamination) varies from 25 to 50 mm. For the configuration with delamination length 50 mm, the DCB specimen was discretized with 540 elements (including the sensor patch as shown in the deformed mesh in Figure 2) producing a system size 928×928. The objective here is to see the variation of the OCV as a function of crack length. We now consider voltage distribution under static loading. Static load of 50 N is distributed

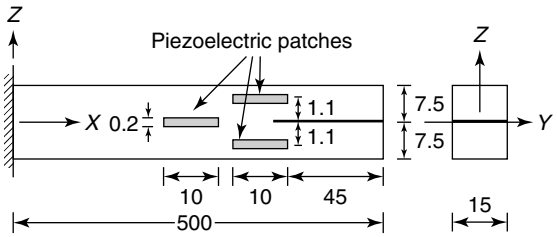


Figure 1. Configuration of a laminated composite double cantilever beam (DCB) with three embedded piezoelectric patches ahead of the advancing delamination tip. All dimensions are in millimeters.

Table 1. Material property of the double cantilever beam specimen used for numerical experiment

Material property	Graphite/epoxy beam	PZT
Young's modulus E_{11} (GPa)	150	63
Young's modulus $E_{22} = E_{33}$ (GPa)	9	63
Shear modulus G_{12} (GPa)	7	24.2
Shear modulus $G_{23} = G_{13}$ (GPa)	25	24.2
Poisson's ratio: $\nu_{12} = \nu_{23} = \nu_{13}$	0.3	0.3
Mass density: ρ (kg m^{-3})	1600	7600
Piezoelectric coefficient (10^{-12} m V^{-1}) d_{31}	—	254
Piezoelectric coefficient (10^{-12} m V^{-1}) d_{32}	—	254
Piezoelectric coefficient (10^{-12} m V^{-1}) d_{33}	—	0
Relative permittivity: $\mu_{11}/\mu_0 = \mu_{22}/\mu_0 = \mu_{33}/\mu_0$	—	1500

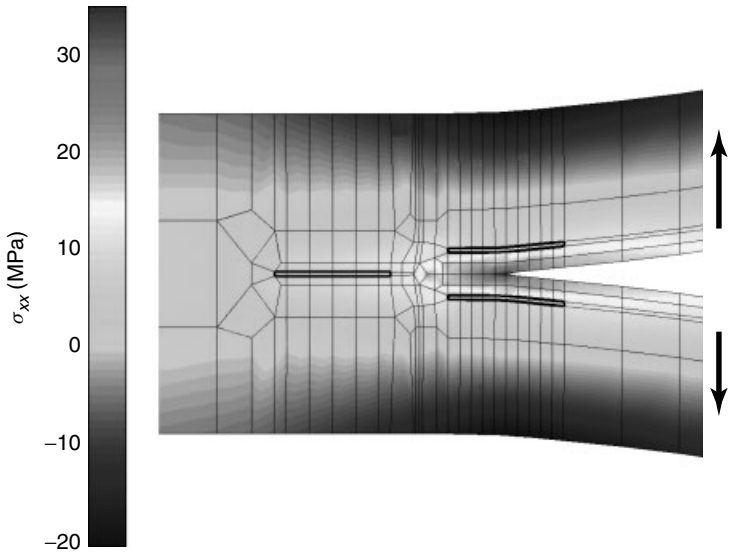


Figure 2. Distribution of stress σ_{xx} under vertical static 50-N loading. Delamination length is 50 mm. Delamination tip is midway between the top and bottom sensor patch.

vertically in the opposite direction at the tip cross section of each of the branches of the DCB specimen (Figure 1). Such a loading will cause what is termed in fracture mechanics as *mode* -1 or *opening mode loading*. Figure 2 shows the distribution of the stresses σ_{xx} for delamination length of 50 mm, where the crack tip is midway between the top and bottom sensor patches. One can clearly see the high stress in the region very close to the crack tip. Locations of the three sensor patches are shown by thick solid lines. Owing to these distributions of stress field, it is important to look into the nature of voltages developed between the parallel electroded surfaces of the front, top, and bottom sensor patches. These are shown in Figure 3(a) and (b), respectively. The locations of the points on the sensor surface at which the voltages are computed are indicated by their x coordinate measured from the fixed end of the DCB specimen. Equation (10) is used to perform the required post-processing. It can be seen from Figure 3(a), for the front sensor patch, that for a smaller delamination length (delamination tip away from the patch), the voltage distribution is almost constant. For higher length of delamination (delamination tip entering in the zone surrounding the sensor group), the voltage at the front portion of the front sensor patch is higher

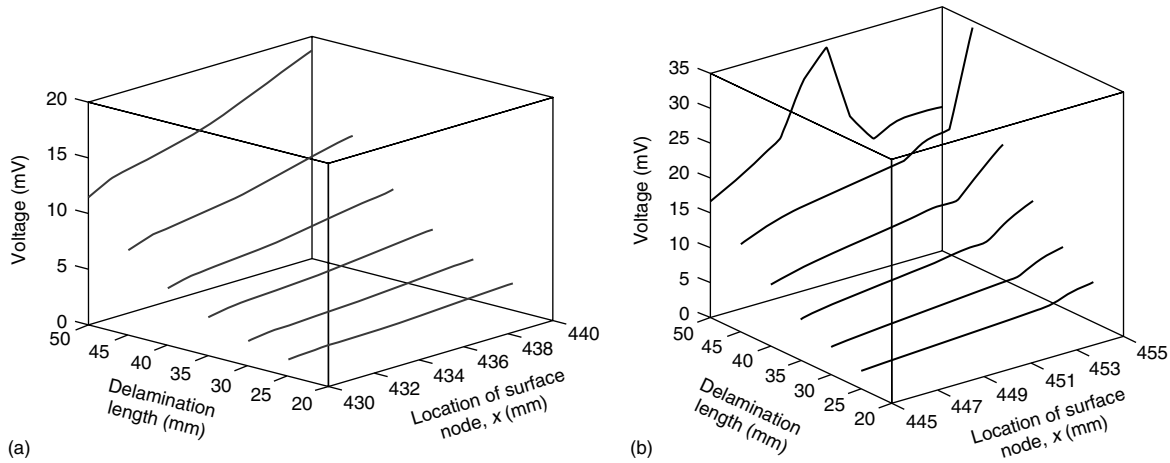


Figure 3. Voltage distribution along x under static 50-N loading for increasing the length of delamination: (a) in the front sensor patch and (b) bottom sensor patch.

than its rear portion. However, the effect of intense crack-tip field approaching toward this front sensor patch could not be captured. This is captured in the bottom and top sensor patches. Figure 3(b) shows the voltage distribution in the bottom sensor patch, where the effect of the crack-tip field entering the midway between the top and bottom sensor patches is clear from the sudden jump in the voltage, first for delamination length of 45 mm and then 50 mm. Owing to symmetry in the material configuration, geometry and loading, the similar feature (but with opposite sign in the voltage) was also captured for the top sensor patch, which is not shown. From these figures, we can clearly see that OCV across the sensor patches can be used as a damage measure to detect the presence of damage.

We now change the loading to a vertical load that is distributed uniformly across the entire cantilever tip. Such a loading, in fracture mechanics terminology, causes shear fracture mode or it is also called the *mode-II loading*. The loading magnitude is kept at 100 N. The OCV distribution in the three piezoelectric sensor patches is shown in Figure 4(a–c), respectively. From the voltage distribution in the front sensor patch in Figure 4(a), it is clear that the dominant stress field σ_{xx} is antisymmetric about the middle plane along x axis. Increase in the voltage due to increasing length of delamination is negligible. From the voltage distribution in the top sensor patch in Figure 4(b), it can be seen that for increasing the length of delamination, the portion of the patch toward the free end of the DCB specimen develops higher voltage compared to the portion nears the front sensor patch, and this variation is almost linear. Slight fluctuation compared to this feature is obtained in the case of the bottom sensor patch, which is shown in Figure 4(c). Interestingly, the antisymmetric stress field σ_{xx}, which is visible from the spanwise antisymmetric voltage distribution in the top as well as bottom sensor patches, gets perturbed as the delamination grows toward the sensor group. Also, the effect of delaminated surface is dominant here in generating the voltage, not the crack-tip field.

These examples have clearly shown the utility of piezoelectric material patches in detecting the presence of damage. Next, we study the voltage distribution under dynamic loading. Here, we study the effects only under shear loading considered previously. In this study, a triangular pulse of 50-μs duration and peak amplitude of 100 N is applied vertically at the cross section of the tip of the DCB specimen (Figure 1). Similar high-frequency transient pulse is typical to many composite structures under impact-type loading, where the damage and structural failure are the most important aspects of structural design and structural health monitoring [18, 19]. First, the time-dependent distribution of the voltage at the spanwise locations on the front, top, and bottom sensor patches are studied. Equations (12) and (13) are used for the required postprocessing. Figure 5(a) shows the

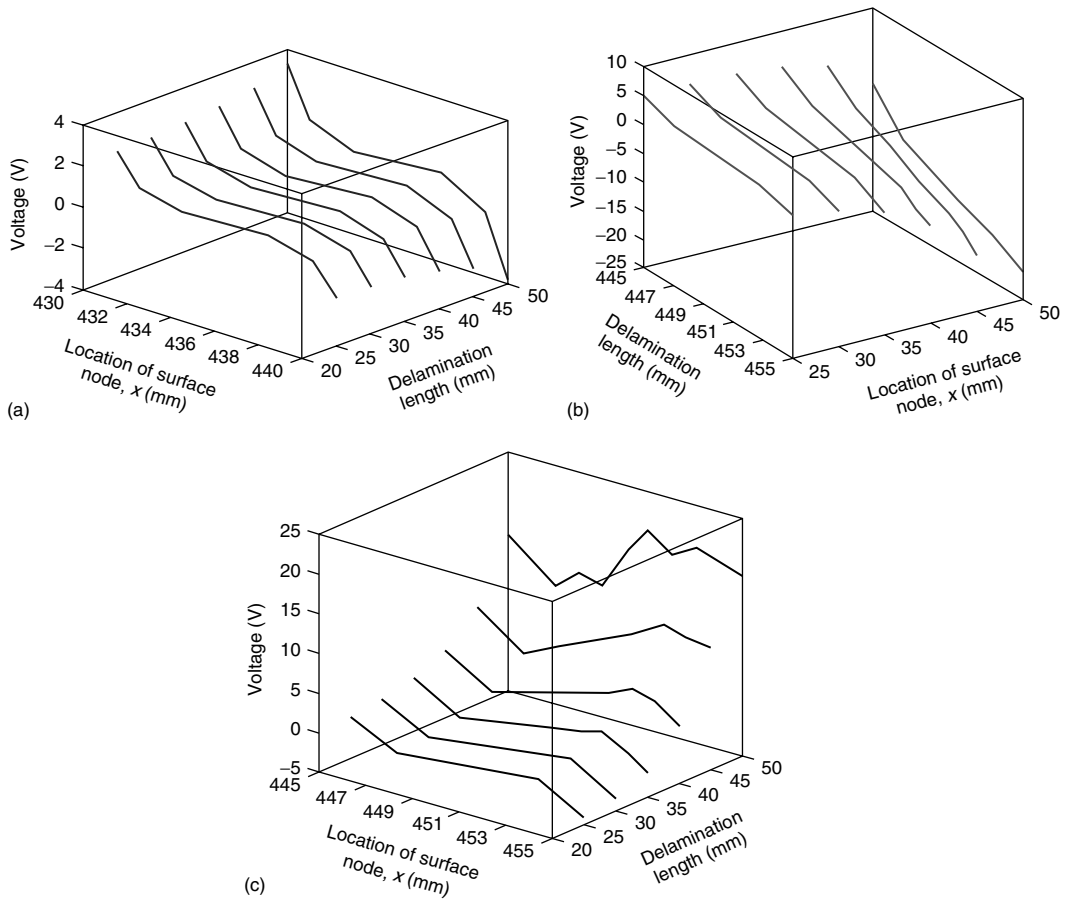


Figure 4. Voltage distribution along x under static 100-N shear loading for the increasing length of delamination: (a) in the front sensor patch, (b) top sensor patch, and (c) bottom sensor patch.

voltage history at the front sensor patch. The points along the x direction on the surface of sensor patches are indicated by the x coordinate measured from the fixed end of the DCB specimen. Since the loading is transient in nature, it produces stress waves traveling with corresponding flexural speed, shear speed, and phase dispersion. Note that the inertia effect is also included, which makes the dynamic behavior different from the quasi-static case. The first peak in each time history, as seen in Figure 4(a), corresponds to the stress waves reaching certain segments of the sensor patch. The second peak in each time history corresponds to the stress waves released from the crack tip. It can be observed that the voltage generated (second peak in the time history) at the front end of the front sensor patch is very high, which is actually the effect of the arrival of the wave at the 50-mm delamination tip (most severe case). In the voltage histories for the top as well as the bottom sensor patches (Figure 5b and c), maximum voltage is generated from the segment confining the delaminated faces behind the crack tip.

In summary, two aspects are very clear from these examples. First, it is shown that there is significant variation in the OCV distribution across the sensor owing to the presence of cracks. These voltage patterns cannot only tell about the presence of cracks (or delamination) but also give a qualitative nature of the damage. Hence, the OCV can indeed be used as a damage measure. The second aspect that is clear is that piezoelectric material can be effectively used as a sensor for damage-detection purposes.

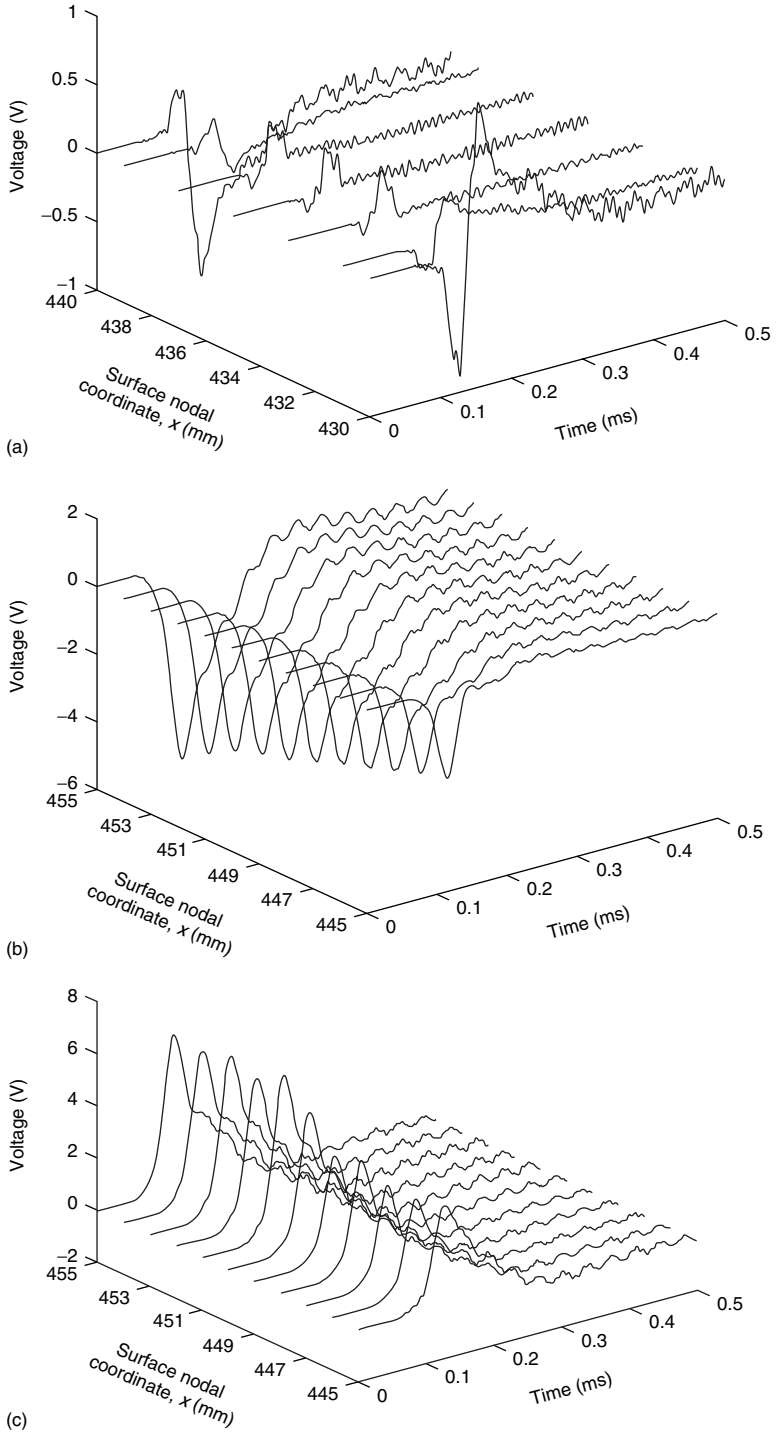


Figure 5. Voltage distribution along x under dynamic 100-N shear loading for increasing the length of delamination: (a) in the front sensor patch, (b) top sensor patch, and (c) bottom sensor patch.

3 DAMAGE DETECTION USING MAGNETOSTRICTIVE SENSORS/ACTUATORS

In this section, a simple FE based on uncoupled constitutive model of magnetostrictive material (Terfenol-D) is considered for modeling of the smart material patches. The results of the model are compared with the experiments. The details of the experiments can be found in [20]. Magnetostrictive material such as "Terfenol-D" was thought of as an actuator material until recently [21]. Compared to PZT material, they offer large block force and large free strain. In addition, they are available in powder form, which makes it possible to embed them in composites. Reference 22 was perhaps the first work reported on the properties of magnetostrictive particle layers that would make them useful for sensing purposes. Reference 23 showed the use of horseshoe coil arrangement for creating the magnetic circuit, which is necessary for damage-sensing application. The present study uses the above concept to demonstrate the effectiveness of magnetostrictive material for damage-sensing purposes. Here, the health monitoring is based on the concept that when an ac current is passed to the actuation coil, it generates mechanical stress (strain). This stress changes not only with the change in delamination location but also with its size. This change in stress produces a change in the magnetic flux, and hence in the voltages across the sensing coil. Measuring the change in the voltage (which is easily measurable quantity compared to strains/stresses) before and after the delamination has taken place determines the condition of the structure.

As mentioned earlier, we use the uncoupled constitutive model of Terfenol-D for the FE analysis purpose and hence assume the magnetic field as the product of coil current and the number of turns of the coil. From the FE point of view, the magnetic field can be converted as a block force and lumped on to the structure. The computation of the block force is first explained.

The magnetostrictive material has the following constitutive law. The first law is the actuation law, while the second is the sensing law.

$$\varepsilon = S^{(H)}\sigma + \mathrm{d}H \tag{14}$$

$$B = \mathrm{d}\sigma + \mu^{(\sigma)}H \tag{15}$$

where, ε is the induced strain, S is the material compliance measured at constant magnetic field H, σ is the induced stress, B is the magnetic flux density, and $\boldsymbol{\mu}$ is the permeability of the medium measured at constant H (*see* **Constitutive Modeling of Magnetostrictive Materials** for more details on the constitutive models for magnetostrictive materials). In the absence of mechanical forces, the strain and induced stress from the actuation law become

$$\varepsilon(t) = \mathrm{d}H(t) \tag{16}$$

$$\sigma(t) = E\varepsilon(t) = E\,\mathrm{d}H(t) \tag{17}$$

where E is the Young's modulus of the material. If A is the cross-sectional area of the smart match, the block force experienced by the smart patch owing to magnetic field intensity of H is given by

$$F(t) = E\,\mathrm{d}AH(t) \tag{18}$$

An actuation current is an alternating current (I) that generates a magnetic flux (B) while passing through one arm of the horseshoe coil and also creates a magnetic field in the layer of smart patch. This current varies sinusoidally with time as

$$I(t) = I_0 \sin \Omega t \tag{19}$$

where I_0 is the applied current amplitude and Ω is the applied current frequency. The magnetic field is related to the current by the expression

$$H(t) = n_{\mathrm{a}} I(t) \tag{20}$$

where n_{a} is the number of turns in the actuation coil. Using equations (19) and (20) in equation (18), the applied force can be written as

$$F(t) = E\,\mathrm{d}AI_0 n_{\mathrm{a}} \sin(\Omega t) \tag{21}$$

The obtained force is applied as concentrated force on the nodes of the patch in the longitudinal direction. This force causes the stress across the sensing coil. From this, the nodal FE stresses σ_{nodal} can be computed. The actual stress is then calculated as

$$\sigma_{\mathrm{total}} = \sigma_{\mathrm{nodal}} + \sigma_{\mathrm{induced}} = \sigma_{\mathrm{nodal}} + E\,\mathrm{d}H \tag{22}$$

The sensing constitutive law (equation 15), in the presence of pure mechanical load becomes

$$B = \mathrm{d}\sigma_{\text{total}} \tag{23}$$

Knowing the stresses from equation (22), and using this in equation (23), the OCV can be computed from the relation

$$V = -n_s A \frac{\mathrm{d}B}{\mathrm{d}t} \tag{24}$$

where, n_s is the number of turns in the sensing coil.

The FE of the beam of size 200 mm × 24 mm × 2.4 mm is discretized using 1920 3-D brick elements. In the longitudinal direction, it is divided into 20 segments, in the lateral direction 6 segments, and in the thickness direction it is divided into 16 segments. The total number nodes in the model is 2499. The magnetostrictive patch is modeled using one 3-D brick element. The complete FE model is shown in Figure 5. The beam is fixed at one end and the equivalent block force is applied in the longitudinal direction of the magnetostrictive patch as shown in Figure 6. Transient dynamic analysis using Newmark time-marching scheme is employed to compute the response histories. Stress histories are computed after postprocessing the displacement histories and subsequently using equation (22). From the stress histories, the OCV is calculated for different times using equation (24).

In order to study the effectiveness of the magnetostrictive patch as a sensor and its ability to sense damage, a cantilever beam structure with both embedded and surface-bonded smart patches is considered. The magnetostrictive patch is bonded (or embedded) in the specimen at a distance of 23 mm

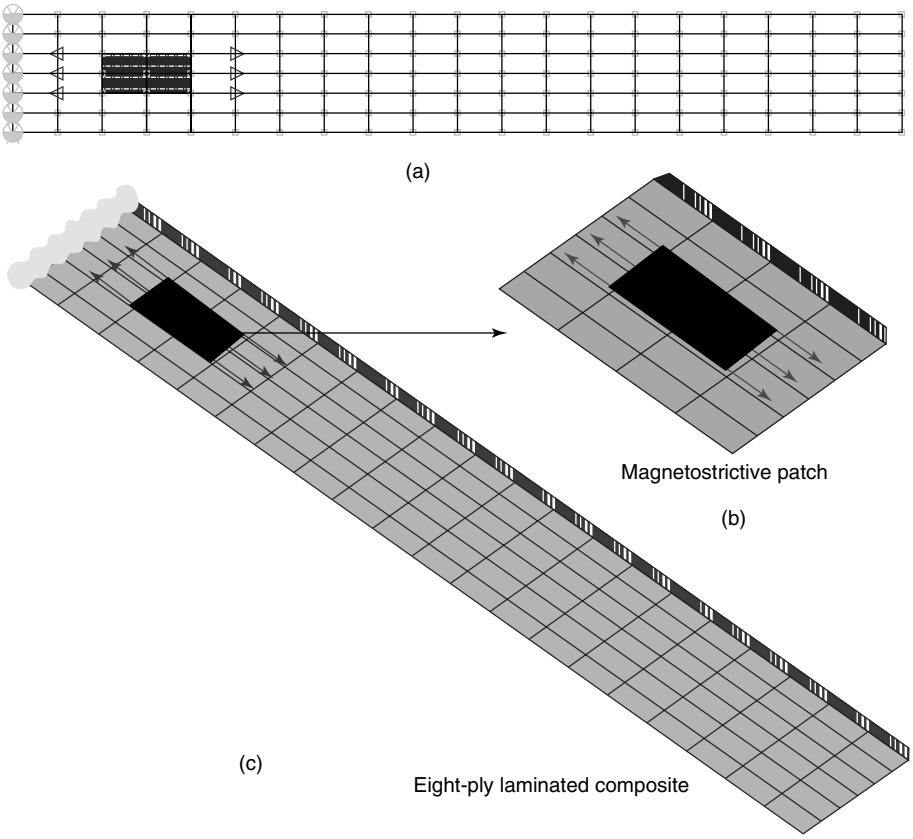


Figure 6. 3-D finite element model for laminated composite beam: (a) plan view of the mesh, (b) enlarged isometric view of the region near the smart patch, and (c) isometric view of the complete model.

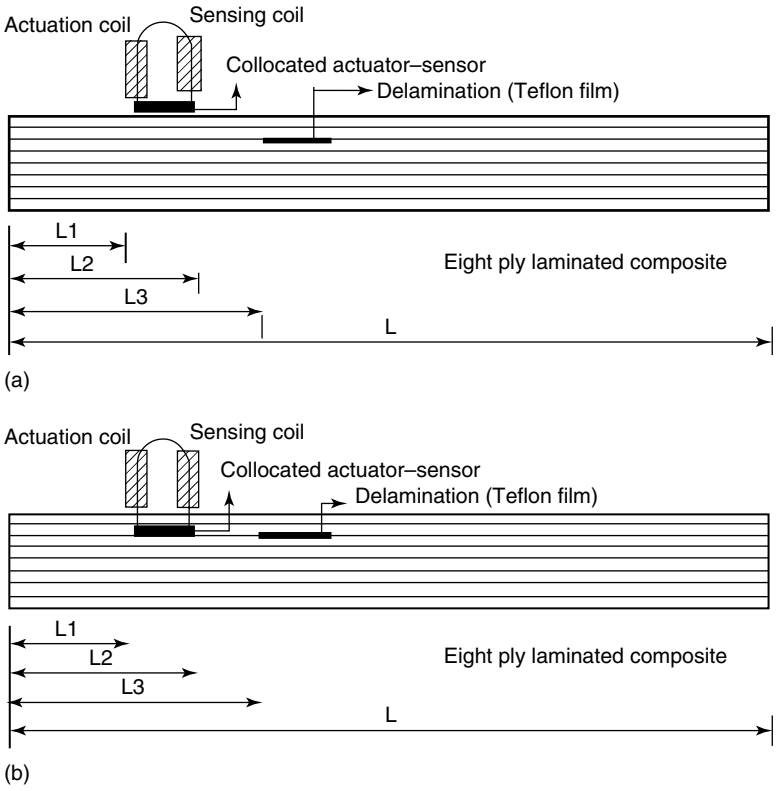


Figure 7. Laminated composite with (a) surface-bonded patch and (b) embedded single patch.

from the root in the longitudinal direction. The delamination length is varied from 24 to 14 mm. The delamination is placed between the second and third layer of an eight-layer unidirectional composite. Figure 7 shows the position of the magnetostrictive patch. The material properties of the laminate and patch are listed in Table 2. It also tells how to induce a magnetic circuit using horseshoe coil to create a magnetic field experimentally. The dynamic excitation is provided by the magnetic field generated by an ac current. This current varies sinusoidally with time and hence the magnetic field as well as the dynamic excitation also varies in a similar manner. The ac current is varied from a frequency range of 100–400 Hz. The current amplitude is varied between 1 and 4 mA. In each case, the OCV is measured and the DIV, which is given by

$$DIV = OCV_{\mathrm{d}} - OCV_{\mathrm{h}} \tag{25}$$

Table 2. Material properties for composite laminate and smart patch used in finite element simulation

Property	Laminated composite	Smart patch
E_{x} (GPa)	53.48	32.7
E_{y} (GPa)	17.92	32.7
E_{z} (GPa)	17.92	32.7
G_{xy} (GPa)	8.92	12.57
G_{yz} (GPa)	8.92	12.57
G_{zx} (GPa)	3.44	12.57
ν_{xy}	0.25	0.3
ν_{yz}	0.25	0.3
ν_{xz}	0.34	0.3
ρ (kg m^{-3})	1500	2330

is computed. Here $\mathrm{OCV_d}$ is the OCV in a delaminated beam and $\mathrm{OCV_h}$ is the OCV in a healthy beam.

First, a beam with surface-mounted patch is analyzed. For this case, delamination of size 24 mm

is introduced between the second and third layer at a distance of 50 mm from the tip of the sensor patch. OCV histories are obtained for varying current amplitudes and frequencies. Figure 8 shows the peak voltage (OCV) as the function of current amplitude for a current frequency of 100 Hz. We see from the figure that OCV increases with the current amplitude and there is excellent agreement of the results with the FE prediction. Figure 9(a) and (b) shows the OCV and DIV time histories for the same delamination configuration at 200-Hz frequency. These figures show that the voltage history exhibits inverted cosine profile. This is due to the presence of time derivative for computing magnetic flux intensity (B)

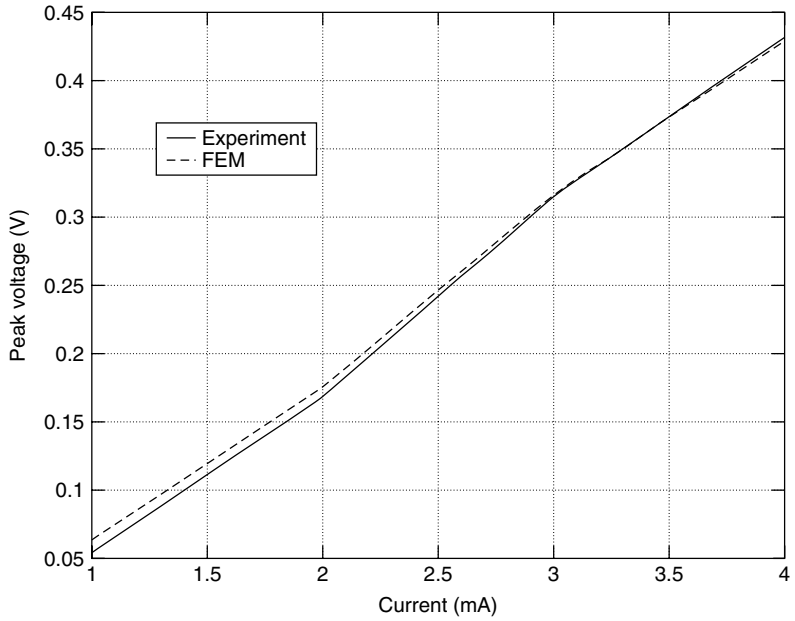


Figure 8. Comparison of experimental and FEM peak open circuit voltages for a beam with surface-bonded patch at 100-Hz frequency.

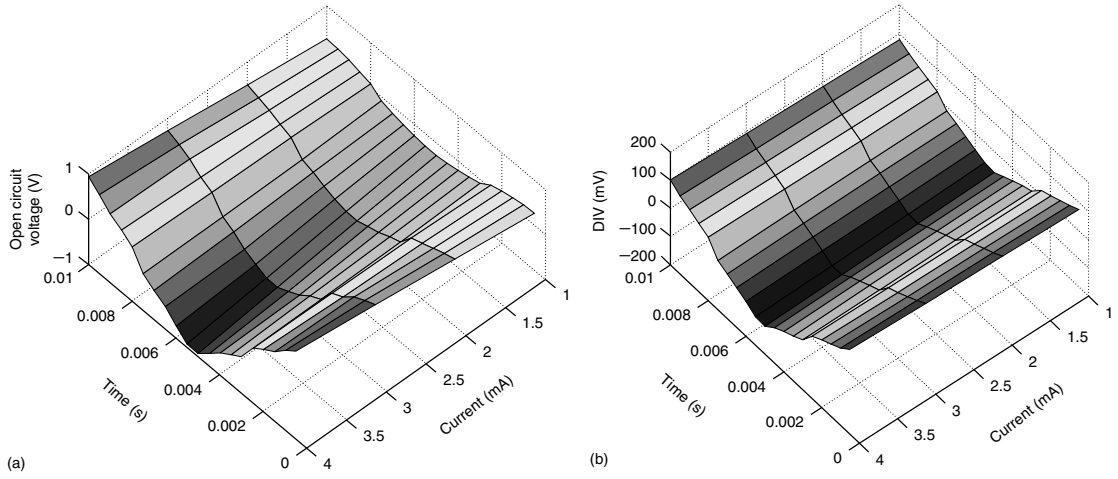


Figure 9. Surface-bonded patch-current frequency at 100 Hz: (a) OCV histories and (b) DIV histories.

Table 3. Damage-induced voltage for surface-bonded single patch with delamination size of 24 mm

Frequency (Hz)	*DIV* (mV) (experiment) [20]				*DIV* (mV) (FEM)			
	1 mA	2 mA	3 mA	4 mA	1 mA	2 mA	3 mA	4 mA
100	74	82	89	98	72	82	87	97
200	78	87	94	103	77	86	93	102
300	84	92	101	109	82	91	99	108
400	89	96	106	112	90	96	104	113

in equation (23). It is seen that the DIV is in the order of millivolts, which is easily measurable using normal instrumentation. Table 3 gives the comparison of DIV obtained from experiments and FE analysis for various frequencies. The table shows that the DIV increases with increase in current amplitudes and current frequencies. These results agree very well with the FE analysis predictions.

In the next example, the smart Terfenol-D patch is embedded in between plies of a composite beam and the OCV and the DIV or computed. In this case, three different sets of experiments are performed. In the first set, the delamination sizes are varied for a fixed location of 50 mm from the smart patch. Three different delamination sizes of 24, 18, and 14 mm, respectively, are considered for the study. These delaminations are again introduced between the second and third layer. OCV histories are obtained for these specimens. Tables 4–6 give the peak DIV values computed from the peak OCV values of

Table 4. Damage-induced voltage for embedded single patch with delamination size of 24 mm

Frequency (Hz)	*DIV* (mV) (experiment) [20]				*DIV* (mV) (FEM)			
	1 mA	2 mA	3 mA	4 mA	1 mA	2 mA	3 mA	4 mA
100	54	67	82	98	52	65	80	95
200	63	77	93	111	62	74	91	107
300	74	89	106	125	73	85	104	120
400	86	102	120	140	86	99	121	138

Table 5. Damage-induced voltage for embedded single patch with delamination size of 18 mm

Frequency (Hz)	*DIV* (mV) (experiment) [20]				*DIV* (mV) (FEM)			
	1 mA	2 mA	3 mA	4 mA	1 mA	2 mA	3 mA	4 mA
100	45	58	72	87	46	57	71	87
200	54	68	82	99	53	66	80	98
300	64	79	94	112	65	78	92	109
400	76	91	107	126	75	89	104	121

Table 6. Damage-induced voltage for embedded single patch with delamination size of 14 mm

Frequency (Hz)	*DIV* (mV) (experiment) [20]				*DIV* (mV) (FEM)			
	1 mA	2 mA	3 mA	4 mA	1 mA	2 mA	3 mA	4 mA
100	32	44	57	71	30	42	54	68
200	41	53	66	83	38	52	62	81
300	50	64	78	95	48	61	76	95
400	62	76	90	108	59	73	91	110

delaminated and healthy beam specimens for different current frequencies and current amplitudes. From the tables, we see that, for the delamination of size 24 mm, the DIV slightly decreases compared to the surface-bonded case, especially for low current frequency. However, at higher current frequencies and current amplitudes, DIV is higher than for the surface-bonded case. Hence, parameters such as the position of the sensor (patch), the current frequencies, and the current amplitudes, all influence the magnitude of DIV. Here again, the experimental results agree well with the FE solutions. When the delamination size is changed to 18 and 14 mm, respectively, we see that the magnitude of DIV drops with the reduction in the delamination size. This is to be expected, since smaller the delamination size, smaller will be

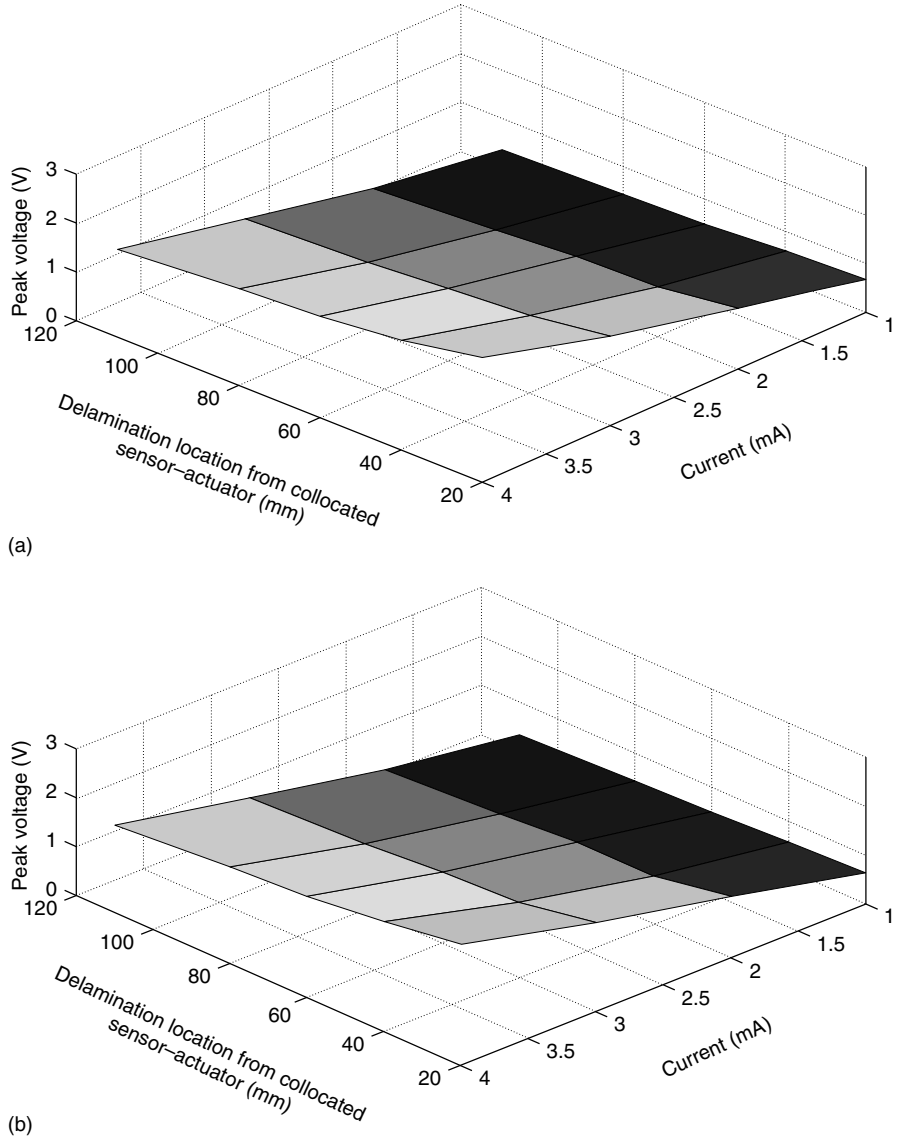


Figure 10. Peak OCV for different delamination locations along the length of a laminated composite with embedded single patch at 100-Hz frequency: (a) experiment and (b) FEM.

the change in the induced stress, and hence smaller will be the DIV. As before, DIV increases with the current frequency and amplitude. FE results, again, match very well with the experiments [20].

In the second set of experiments on a single-patch collocated sensor/actuator configuration, the location of delamination is varied lengthwise with respect to smart patch for a fixed delamination size of 18 mm. OCV histories are obtained for five different locations of 30, 50, 70, 90, and 120 mm from the smart patch. The peak value of OCV is plotted as a function of current amplitude and delamination location at 100-Hz frequency. This is shown in Figure 10. We see from the plots that the peak value of the OCV increases with the current amplitude and, as expected, the maximum value occurs near the sensor location. The parameters that govern the sensitivity of the response are the peak values of DIV very close and far away from the sensor location. From the figures, we see that the peak DIV decreases with the increase in distance between the smart patch and delamination location.

In the last set of experiments on single-patch collocated configuration, the location of delamination of size 18 mm is varied depthwise by introducing it at the successive ply levels starting from the

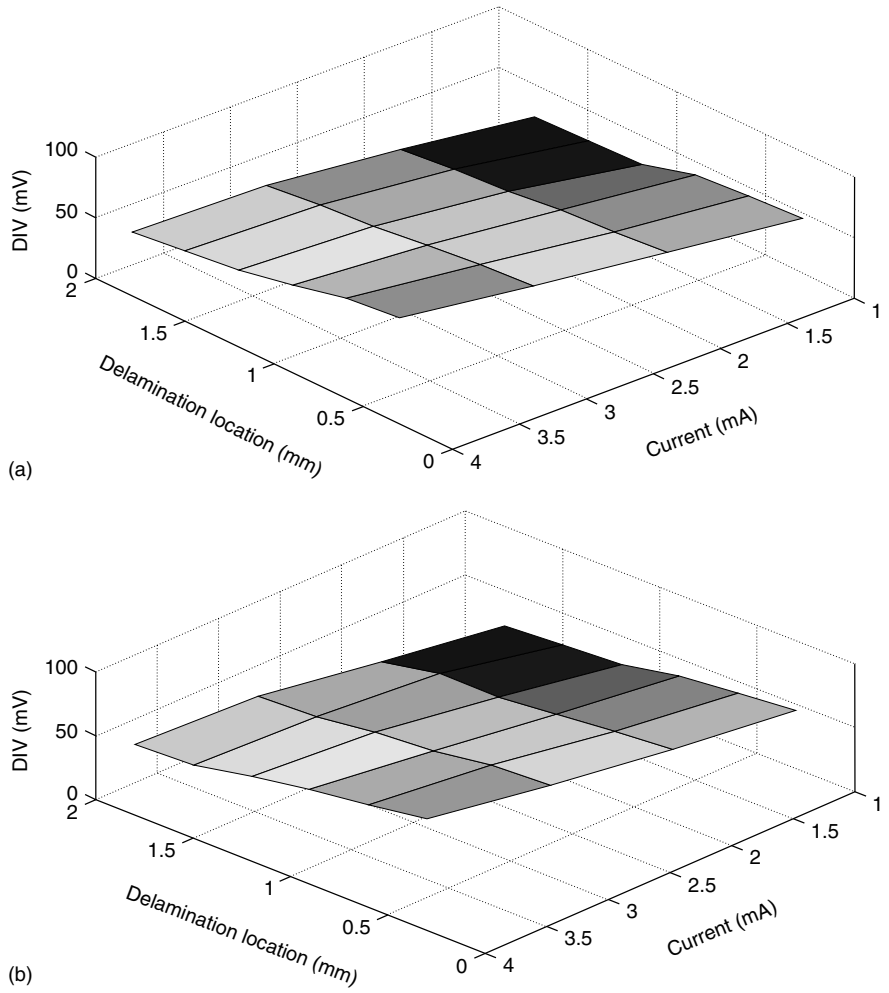


Figure 11. Peak damage-induced voltage for different delamination locations along depth of a laminated composite with embedded single patch at 100 Hz frequency: (a) experiment [20] and (b) FEM.

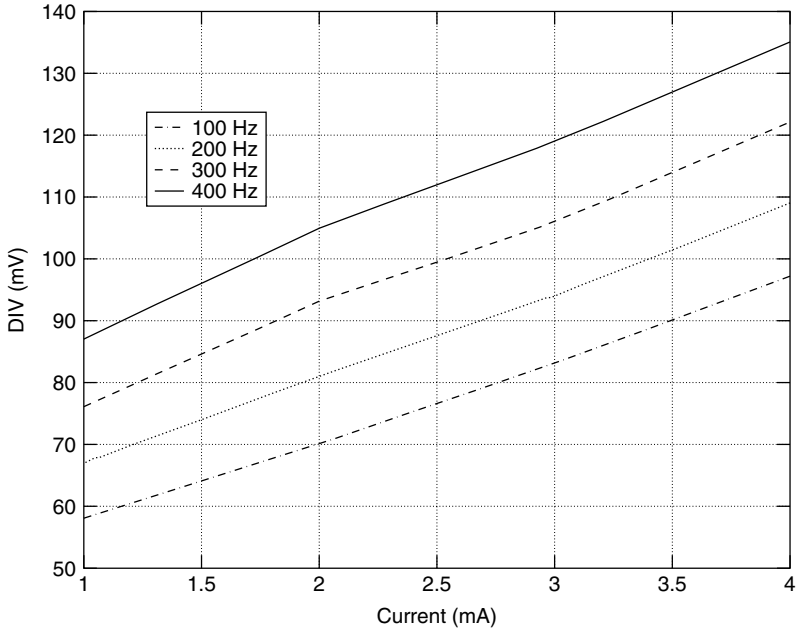


Figure 12. Peak values of damage-induced voltage for delamination of 18 mm located in between collocated sensor and actuator.

second–third layer interface. The lengthwise location of this is at 50 mm from the smart patch. OCV histories are obtained for these configurations. The peak DIV values are plotted as a function of current amplitudes for a current frequency of 100 Hz. This is shown in Figure 11. Higher voltages are predicted when the delamination is very close to the sensor patch (that is, when the delamination is between the second and third layer). When the delamination locations at ply levels are far from the sensor, the induced peak voltage is small and almost constant for all current amplitudes. The next logical step is to introduce the delamination exactly in between the collocated sensor and actuator in the second layer. This is done by making the delamination exactly 18 mm, which is the size of the smart patch in the longitudinal direction. Figure 12 gives the peak DIV values for various current frequencies. These plots indicate a linear behavior of current amplitudes with voltages for all the frequencies. The peak values of DIV change almost 1.5 times when the current amplitude changes from 1 to 4 mA. Hence, it can be concluded that the lengthwise variation of the delamination has more pronounced effect from the sensitivity of the response to the delamination.

4 SUMMARY

This article presents the utility of the smart material patches, which act as both sensors and actuators, to sense the presence of damage in a composite beam. Two different smart materials, namely, the PZT and Terfenol-D, are used in this study. Both these materials require baseline measurement in terms of the OCV across the sensor for predicting the presence of damage. This voltage gets perturbed in the presence of damage. Many numerical experiments are used in this study to demonstrate the concept of damage detection. FEs are used as the mathematical models and some results, especially those with magnetostrictive patches, are compared with the experiments.

REFERENCES

[1] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*, Report LA-13070-MS. Los Alamos National Laboratory, 1996.

[2] Bent AA. *Active Fibre Composite for Structural Actuation*, PhD Thesis. Massachusetts Institute of Technology: Cambridge, MA, 1994.

[3] Shah DK, Chan WS, Joshi SP. Delamination detection and suppression in a composite laminate using piezoceramic layers. *Smart Materials and Structures* 1994 **3**:293–301.

[4] Wang X. Shape memory alloy volume fraction of prestretched shape memory alloy wire-reinforced composites for structural damage repair. *Smart Materials and Structures* 2002 **11**:590–595.

[5] Watkins SE, Sanders GW, Akhavan F, Chandrashekhara K. Modal analysis using fibre optic sensors and neural networks for prediction of composite beam delamination. *Smart Materials and Structures* 2002 **11**:489–495.

[6] Vandiver JK. Detection of structural failure on fixed platforms by measurement of dynamic response. *Proceedings of 7th Annual Offshore Technical Conference*. Houston, TX, 1975; pp. 243–252.

[7] Coppolino RN, Rubin S. Detectability of structural failures in offshore platforms by ambient vibration monitoring. *Proceedings of 12th Annual Offshore Technical Conference*. Houston, TX, 1980; pp. 101–110.

[8] Crawley P, Adams RD. A vibration technique for non-destructive testing of fibre composite structures. *Journal of Composite Materials* 1979 **13**(3):1161–1175.

[9] West WM. Illustration on the use of modal assurance criterion to detect structural changes in an orbiter test specimen. *Proceedings of Air Force Conference on Aircraft Structural Integrity*. St Louis, 1984; pp. 1–6.

[10] Chen Y, Swamidas ASJ. Dynamic characteristics and model parameters of a plate with a small growing surface crack. *Proceedings of the 12th International Modal Analysis Conference*. Honolulu, Hawaii, 1994; pp. 1155–1161.

[11] Ruotolo R, Surface C. Damage assessment of multiple cracked beams: numerical results and experimental validation. *Journal of Sound and Vibration* 1997 **206**:567–588.

[12] Chen JC, Garba JA. On-orbit damage assessment for large space structures. *AIAA Journal* 1988 **26**(9):1119–1126.

[13] Ricles JM, Kosmatka JB. Damage detection in elastic structures using vibratory residual forces and weighted sensitivity. *AIAA Journal* 1992 **30**(9):2310–2316.

[14] Schulz MJ, Naser AS, Pai PF, Chung J. Locating structural damage using frequency response reference functions. *Journal of Intelligent Material Systems and Structures* 1998 **9**:899–905.

[15] Nag A, Roy Mahapatra D, Gopalakrishnan S. Identification of delamination in composite beam using a damaged spectral element. *Structural Health Monitoring* 2002 **1**(1):105–126.

[16] Pines DJ. The use of wave propagation models for structural damage identification. *Proceedings of the International Workshop on Structural Health Monitoring*. Stanford, CA, 1997; pp. 665–677.

[17] Valdes SHD, Soutis C. A structural health monitoring system for laminated composites. *Proceedings of ASME Design Engineering Technical Conference*. Pittsburgh, PA, 9–12 September 2001.

[18] Nag A, Roy Mahapatra D, Gopalakrishnan S, Sankar TS. A spectral finite element with embedded delamination for modelling of wave scattering in composite beams. *Composites Science and Technology* 2003 **63**:2187–2200.

[19] Chakraborty A, Roy Mahapatra D, Gopalakrishnan S. Finite element analysis of free vibration and wave propagation in asymmetric composite beams with structural discontinuities. *Composite Structures* 2001 **55**(1):23–36.

[20] Saida E, Narayan Naik G, Gopalakrishnan S. An experimental investigation of a smart laminated composite beam with magnetostrictive patch for health monitoring applications. *Structural Health Monitoring* 2003 **2**(4):273–292.

[21] Culshaw B. *Smart Structures and Materials*. Artech House: Norwood, 1996.

[22] Krishnamurthy AV, Anjanappa M, Wang Z, Chen X. Sensing of delaminations in composite laminates using embedded magnetostrictive particle layers. *Journal of Intelligent Material Systems and Structures* 1999 **10**:825–835.

[23] Kumar M, Krishnamurty AV. Sensing of delaminations in smart composite laminates using magnetostrictive particle layer and horseshoe coil arrangement. *Journal of Aeronautical Society of India* 2000 **52**:19–25.

Chapter 48
Damage Measures

Massimo Ruzzene
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA

1 INTRODUCTION

The objective of a structural health monitoring (SHM) system is to identify anomalies or damages such as cracks, delaminations, and disbonds in structures. The term identification includes the determination of the existence of damages, their location, and their size as accurately as possible. In the literature, the amount of information that can be obtained regarding a damaged structure is typically classified into five levels: (i) identification of the presence of damage, (ii) determination of the location of the damage, (iii) classification of the type of damage, (iv) quantification of its extent, and (v) estimation of the remaining life of the component under investigation. The definition of an effective measure of damage responds to the requirements of the first four stages of the information, and ideally provides inputs to step number (v). Damage measures proposed in the literature from an SHM perspective are meant to identify and locate the damage, and, in some cases, provide an indication regarding the extent of the damage and its progression.

Many SHM techniques developed over the years are based on the detection of changes in the dynamic behavior of the monitored components. Valuable reviews of the state of the art in dynamics-based SHM can be found in [1–3]. The existing techniques vary on the basis of the type of dynamic response signals used for the analysis and on the features or parameters considered as damage indicators. Such features or parameters must obviously be sensitive to damage and must vary monotonically with the extent of the damage. Dynamics-based inspection techniques are typically classified as vibration-based methods and wave-propagation methods. Vibration-based damage detection techniques typically monitor changes in modal frequencies, changes in measured mode shapes (and their derivatives), and changes in measured flexibility coefficients, while wave-propagation inspections look for reflections and mode conversion phenomena caused by the presence of damage.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Early studies in vibration-based techniques evaluated the influence on the natural frequencies of localized stiffness reductions caused by damage. These investigations have shown that natural frequencies are damage indicators, which generally show low sensitivity, and do not allow the determination of the location of damage. More recent studies have investigated the effects of localized and distributed damage on mode shapes, operational deflection shapes (ODSs), and corresponding curvatures. The detection of small changes in the deformed configuration of the structure can be used to localize damage and potentially estimate its severity. In particular, small variations from an undamaged state can be highlighted by successive spatial differentiations of the deflections, as typically required for estimating curvature modes and associated strain energy. These modal-based methods are very attractive, as they provide information regarding the general state of health of the structure. However, they tend to have limited sensitivity and generally they are not accurate enough to provide detailed information regarding damage type and extent.

The lack in sensitivity and capability of discriminating damage from changes in the operating conditions of modal-based methods can be overcome by applying inspection techniques based on guided ultrasonic waves (GUWs) propagation [3–5]. Guided waves, such as Lamb waves, show sensitivity to a variety of damage types and, as opposed to bulk waves used in traditional ultrasonic techniques, have the ability to travel relatively long distances within the structure under investigation. For this reason, GUWs are particularly suitable for SHM applications, which may employ a built-in sensor/actuator network to interrogate and assess the state of health of the structure. In most applications, GUWs are generated and received by piezoelectric transducers, which can both excite the structure and record its response. As an alternative, an array of transducers and actuators can be distributed over the surface of the structure according to convenient patterns [3, 4, 6, 7]. The fundamentals of this type of operation consist in evaluating the characteristics of the propagation along the wave path between each transducer/receiver pair and detecting reflections associated with damage. The interpretation of the characteristics of the signals and the detection of reflections is, however, complicated by the multimodal and dispersive nature of GUW signals. For this reason, significant efforts are being devoted to improvements in the defect characterization procedures. Advanced signal-processing techniques are being employed to highlight signal features, which are sensitive to the presence of damage and which can be used for its classification and for the estimation of its extent [5].

This article is organized in four sections including this introduction. Section 2 presents an overview of vibration-based damage measures, with specific focus on a few interesting techniques, while Section 3 illustrates research efforts in the area of GUWs-based inspection. Section 4 concludes the article with a short summary.

2 VIBRATION-BASED DAMAGE MEASURES

The analysis of the modal properties of the structure and their variation due to damage has received considerable attention from researchers over the last 30 years. An excellent and thorough review of the research in the area can be found in [1]. Most of the damage measures proposed assume the possibility of comparing current measurements with baseline information regarding the undamaged state of the structure. This is a significant drawback, as it assumes the availability of data on the component under test in its pristine state and also that any measured change is due to damage only, and not due to changing environmental or boundary conditions applied to the component. The techniques presented below are examples of solutions proposed over the years, whose practicality can be significantly improved if coupled with procedures aimed at generating data approximating the undamaged response of the structure. An example of such a procedure, proposed in [8], is presented at the end of this section (Section 2.5).

2.1 Frequency changes

Modal parameters are global properties of a vibrating structure, and their change can indicate the presence of a damage without performing direct measurements at or near the damage site. One of the typical effects of damage is a localized reduction of stiffness, which produces a corresponding change in modal

frequencies. Analytical quantification of the change in modal frequencies related to loss of stiffness in simple beam and plate structures can be found, for example, in [9–11], where perturbation methods are applied to estimate modal properties in the presence of notch-type defects. It is nowadays quite accepted that simple monitoring of frequency changes is not a reliable and sensitive-enough method for early damage detection, and cannot easily provide location information [1]. Noteworthy attempts include the work of Cawley and Adams [12], who investigated frequency shifts due to damage in composite materials. The ratio between shifts at various modes $\Delta\omega_i/\Delta\omega_j$ is used to construct an error term, which allows to estimate the damage location. This technique, which does not account for multiple damage sites, was revisited in [13, 14] using the sensitivity of modal frequency changes in terms of local stiffness and mass changes. The following error function for the ith mode and pth structural member was introduced:

$$e_{ip} = \frac{z_i}{\sum_j z_j} - \frac{F_{ip}}{\sum_j F_{jp}} \tag{1}$$

where z_i is ith term in the array of squared modal frequency changes, which is defined as

$$\{z\} = [F]\{\alpha\} - [G]\{\beta\} \tag{2}$$

with $[F]$ and $[G]$ denoting the changes in elemental stiffness and mass magnitudes, respectively. The member that minimizes the error in equation (1) is determined to be the damaged one, again under the assumption of single defect. The estimation of the frequency change sensitivity is obviously based on the finite element (FE) model of the structure under consideration, which may be problematic in the case of complex structures. Hearn and Testa [15] employed the FE formulation to introduce a damage severity parameter for structural member n representing the stiffness loss of the element, i.e.,

$$[\Delta k_n] = \alpha_n [k_n] \tag{3}$$

which, under the assumption of single damage location, can be directly related to a frequency shift according to the following expression:

$$\Delta\omega_i^2 = \alpha_n \frac{\{\varepsilon_n(\phi_i)\}^{\mathrm{T}}[k_n]\{\varepsilon_n(\phi_i)\}}{\{\phi_i\}^{\mathrm{T}}[M]\{\phi_i\}} \tag{4}$$

where $\{\varepsilon_n(\phi_i)\}$ are the nth member's deformations computed on the basis of the undamaged mode shapes ϕ_i, while $[M]$ and $[k_n]$ describe the mass of the structure and the stiffness of the nth member. This equation directly relates the effect of damage on a specific component n to the corresponding shift in frequency, under the assumption that the evaluation of this direct relation can be performed on the basis of predamage modal information. The hypothesis that damage only produces a change in stiffness $[\delta K]$ is exploited in Richardson and Mannan [16], where the orthogonality properties of the damaged and undamaged structure is used to obtain the following sensitivity equation:

$$\{\phi_i\}^{\mathrm{T}}[\delta K]\{\phi_i\} = (\omega_i^{(d)})^2 - (\omega_i^{(u)})^2 \tag{5}$$

where it is again assumed that damage causes a negligible change in the mode shapes.

2.2 Mode-shape changes

Mode-shape changes are found to be quite sensitive to damage, especially when higher modes are considered, and are able to directly provide damage location information. The problem associated with mode-shape monitoring is clearly related to the need of sufficient spatial measurement resolution, which complicates the experimental procedures. The required measurement resolution can be easily achieved through the use of a scanning laser Doppler vibrometer (SLDV), which has become an important tool for dynamic testing. Alternatively, the number of measurement locations can be reduced if the FE model of the structure under investigation is available for its use in increasing the information on the structure's behavior and for interpolation purposes. Most of the early work on mode-shape analysis consider the modal assurance criterion (MAC) to compare measured, or damaged, modes, with undamaged or numerical ones. For example, West [17] used the MAC to correlate the modes of an undamaged space shuttle orbiter body flap with those after the flap

has been exposed to acoustic loading. Another technique presented in [18] considered a damage measure based on changes in the mode shape and mode-shape slope. Changes were simulated for stiffness reductions in each structural member and compared with measured changes to determine the damage location. Other techniques of various nature exploit comparisons through different types of modal correlation criteria. For example, Fox [19] proposed the concept of an MAC based on measurement points close to a node point for a particular mode ("node line MAC"), Kim *et al.* [20] investigated the use of the partial modal assurance criterion (PMAC) to compare the MAC values of coordinate subsets of the modal vectors, and Ko *et al.* [21] presented a method that uses a combination of modal assurance criterion (COMAC) and sensitivity analysis to detect damage in steel-framed structures. Finally, a damage signature based on the mode shape normalized by the change in the modal frequency of another mode is proposed in [22], as a way to combine frequency shift and changes in mode shapes as damage indicators.

More recent investigations apply the wavelet transform (WT) as a signal-processing tool to highlight the presence of discontinuities in the modal deflections. The spatial information resulting from multipoint measurements are fed to wavelet algorithms to obtain information regarding anomalies related to damage in the deformed shapes. Plots obtained upon the wavelet analysis localize the damage and may be used, after proper calibration, to quantify the extent of damage [23–25]. Figure 1(b) shows, for example, the WT analysis of the fourth mode of the cracked beam shown in Figure 1(a), as presented in [23]. The considered WT of the spatial modal signal clearly locates the damage along the beam length and can be used to monitor its progression through the thickness. Results for plates using a similar approach are presented in [24].

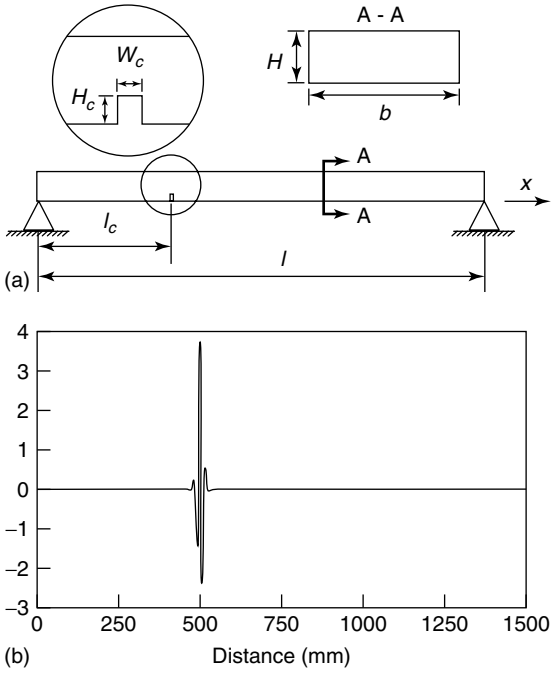


Figure 1. Configuration of cracked beam (a) and result of WT decomposition of fourth mode (b) (from [23]).

2.3 Mode-shape curvature changes

An alternative to using mode shapes to obtain spatial information about structural changes is using mode-shape derivatives, such as curvatures or strain energy distributions obtained from spatial integration of curvature modes. For example, in 1991 Pandey *et al.* [26] demonstrated the use of absolute changes in curvature as good indicators of damage for a beam-type structure. They used a central difference-based numerical differentiation technique to compute the curvature from numerically obtained mode shapes. Their results showed that the absolute changes in curvature were localized around the damaged region and that damage could be quantified by establishing a relationship between the changes in curvature and the extent of damage. The evaluation of changes in the curvature of dynamic deflection shapes as a tool for damage detection and localization were also investigated in [10] and subsequently in [11]. In these studies, the dynamic behavior of beams with notch defects and delaminations was studied analytically and experimentally. Analytical models described the dynamic behavior and the curvature modes of cracked beams through perturbation of the modal parameters of the undamaged beams, so that approximated analytical expressions for the damaged modes were obtained. The analytical studies were supported by experimental investigations performed on simple beam structures. Their results show the potentials of the technique when applied to the first mode of the beam. The limitation to a single mode was

mainly dictated by the limited spatial resolution available in the accelerometers-based experiments. Ho and Ewins [27] formulated a damage index defined as the quotient squared of the corresponding modal curvatures of the undamaged and damaged structure. The damage index was found to be highly susceptible to noise as measurement errors were amplified owing to second-derivative computations based on numerical techniques. They also demonstrated that spatially sparse measurements adversely affect the performance of the damage index. As an alternative, Ho and Ewins [28] investigated the changes in the square of the slope of mode shapes of beams. Oscillations in the slope computations were reduced through polynomial fit of the measured mode shape as compared to the use of finite difference approximation. They also reported that higher derivatives can be more sensitive to damage, but are more subject to numerical errors.

2.4 Damage measure based on energy functional distributions

Other authors have used curvatures for the evaluation of the strain energy distribution over the structure under consideration. This approach has been pursued, for example, by Kim and Stubbs [29] to formulate a damage index based on the comparison of strain energy distributions for damaged and undamaged structures. In [29], and in the subsequent papers by the same authors, the technique is applied to beam structures using mode shapes, ODSs, or time-domain data to obtain information on both damage location and extent. The same technique was extended to plate structures in [30], where accelerometers are used to measure the deflections to be interpolated for successive differentiation. The basis of the technique can be easily illustrated in the case of a beam in bending, for which the strain energy is given by

$$U_i = \frac{1}{2}\int_0^L EI(x)\phi_i^2(x)_{,xx}\,\mathrm{d}x \tag{6}$$

with L and EI denoting the beam's length and flexural stiffness, respectively, while $\phi_{i(x)}$ is the beam's ODS corresponding to its excitation at the ith natural frequency. The beam is subdivided into N regions so that its total strain energy can be expressed as [30] (Figure 2):

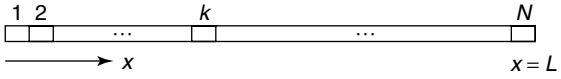


Figure 2. Schematic of beam divisions.

$$U_i = \frac{1}{2}\sum_{k=1}^{N} EI_k \int_{x_k}^{x_{k+1}} \phi_i^2(x)_{,xx}\,\mathrm{d}x$$

$$U_i = \sum_{k=1}^{N} U_{i_k} \tag{7}$$

where it is assumed that the flexural rigidity of the beam over each region is constant. In addition, it is assumed that damage is localized in a single region $k = p$, and that at the damage location,

$$\frac{U_{i_p}}{U_i} \approx \frac{U_{i_p}^*}{U_i^*} \tag{8}$$

so that an estimation of the reduction in stiffness rigidity can be obtained as

$$\frac{EI_p^*}{EI_p} \approx \frac{U_i^*}{U_i}\frac{\displaystyle\int_{x_p}^{x_{p+1}} \phi_i^2(x)_{,xx}\,\mathrm{d}x}{\displaystyle\int_{x_p}^{x_{p+1}} \phi_{i^*}^2(x)_{,xx}\,\mathrm{d}x} = f_{i_p} \tag{9}$$

where the ratio f_{i_p} is the considered damage measure. The damage measure is expected to be equal to 1 over the undamaged regions, and different than one over a damaged region. It is, in general, convenient to combine information obtained from the analysis of several modes (I) and therefore to consider a cumulative damage measure, defined as

$$f_k = \frac{1}{I}\sum_{i=1}^{I} f_{i_k} \tag{10}$$

This cumulative index provides a single piece of information, which combines the results from several ODSs and associated strain energy distributions. ODSs are not affected by damage because their particular location will not contribute, i.e., they will give unit contributions, whereas the index for modes altered by the defect will be combined to provide a robust indication of a defect.

2.5 Interpolation of the measured response and synthesis of undamaged baseline

The damage index as proposed in [30] is based on a comparison between damaged and undamaged strain energy over the considered region of the structure. In practice, however, it may be difficult to have or obtain baseline information from the structures to be analyzed. This, in fact, assumes that either an undamaged specimen or historical data of the same kind as those currently being collected are available. This limitation can be overcome in the presence of high spatial resolution of the measurements, as provided, for example, by SLDVs. The technique introduced in [8] synthesizes undamaged information through spatial decimation of the response. In the proposed approach, the ODSs are measured at several locations over the structure, so that spatial derivatives can be accurately estimated through spline interpolation of the measured data. The ODS $\phi(x, y)$ for a plate structure can, for example, be approximated as

$$\phi(x, y) \cong \sum_{p,q} h_p(x) h_q(y) \Phi_{p,q} \qquad (11)$$

where $\Phi_{p,q}$ defines the value of the ODS measured experimentally at the sensor location p, q, or at a point of an SLDV grid, while $h_p(x), h_q(y)$ are spline basis functions. The curvature estimations can be obtained by taking derivatives of the spline functions, while keeping the nodal or measured values as weighting parameters.

Baseline information can be generated by using a subset of the measurement points. The baseline interpolated deflection can be expressed as

$$\phi^*(x, y) \cong \sum_{r,s} h_r(x) h_s(y) \Phi_{r,s} \qquad (12)$$

where r, s are a subset of the measurement grid points p, q, such that $r < p, s < q$. The resulting undersampling of the data has the purpose of intentionally "missing" any discontinuity or anomaly corresponding to damage, which can generally be detected only through a refined measurement grid. The baseline information can be then differentiated and used for the estimation of the strain energy generically denoted as U^*.

This technique is demonstrated experimentally on a thin aluminum plate with artificially induced damage. The plate measures 35.6 cm × 35.6 cm × 0.10 cm (14″ × 14″ × 0.4″), and it is cantilevered at the base. A piezoceramic disc of 2.79 cm (1.1″) diameter and 0.076 cm (0.030″) thickness is used as an exciter. The plate, with actuator and damage location, is shown in Figure 3. The damage is a 3.6 cm (1.41″)-long, 1.3 mm (0.05″)-wide, and 0.4 mm (0.015″)-deep groove, which was cut in the plate at the location shown. The forced response at the plate's natural frequencies in the excitation range and the corresponding curvatures are evaluated. The modal damage measure shown in Figure 4(a) clearly highlights the location of damage in the plate. Some deviation from unity can be observed at other locations, one of them being located close to the exciter and the other most likely due to measurement and/or numerical noise. This spurious information can be effectively filtered out by considering a cumulative index obtained from the superposition of several modal information. The cumulative index obtained from the superposition of the first five modes of the plate is shown in Figure 4(b), which provides an unambiguous indication about damage presence and location.

The procedure is also demonstrated on the wing skin of an F15 aircraft. The part is a complex structure with rivet holes, thickness variations, and stiffeners manufactured integrally with the skin. The

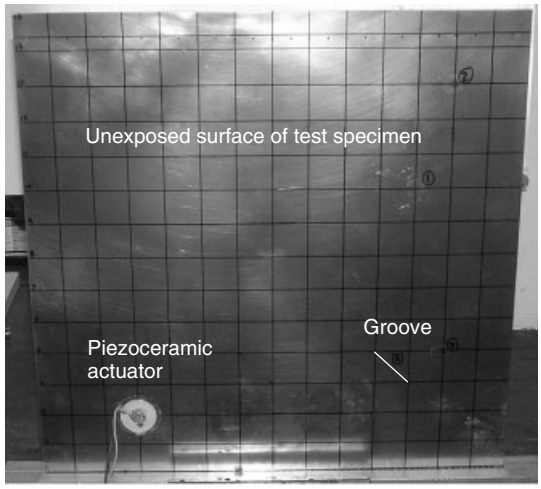


Figure 3. Cantilevered aluminum plate with detail of actuator and damage locations.

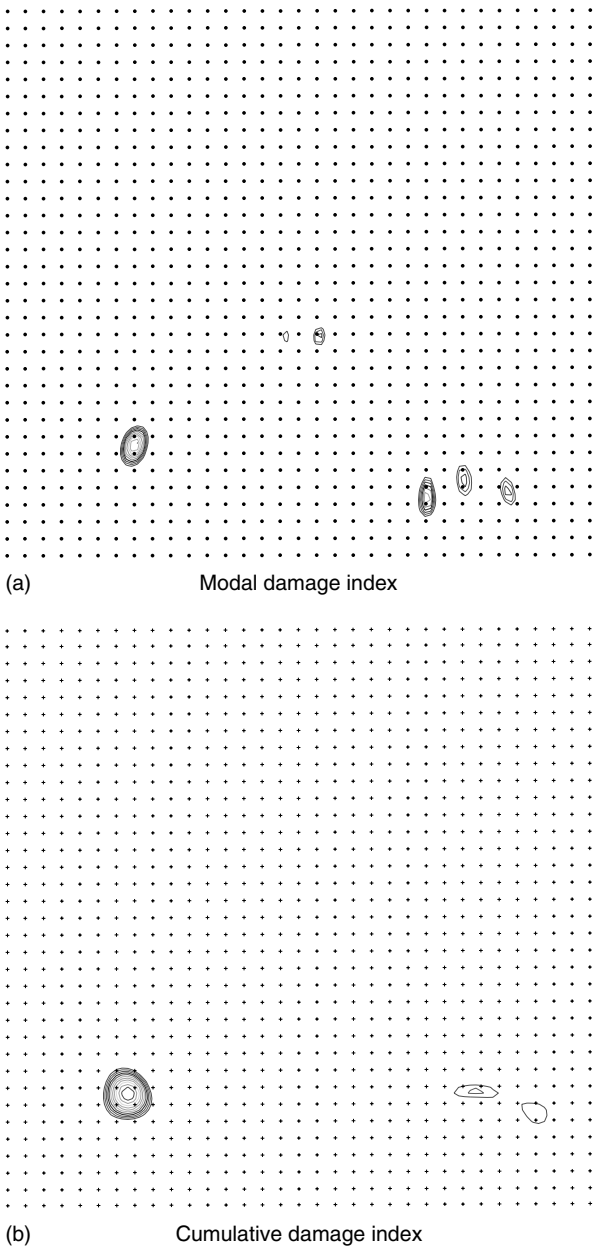


Figure 4. Modal (a) and cumulative (b) damage index.

material of the part is aluminum, and the base thickness is equal to 3.2 mm (1/8″). Figure 5 shows pictures of the front and back of the component to highlight its complexity. In the laboratory setup, the part is hung using elastic bands to fixed rails mounted on the ceiling. Damage is artificially created in a region between two stiffeners. The considered damage is a 3.8 cm (1.5″)-long slit, which causes approximately a 30% reduction in the skin thickness. Modal testing is performed by exciting the skin through a piezoelectric actuator placed at the location shown in Figure 6(a). The deformed configurations

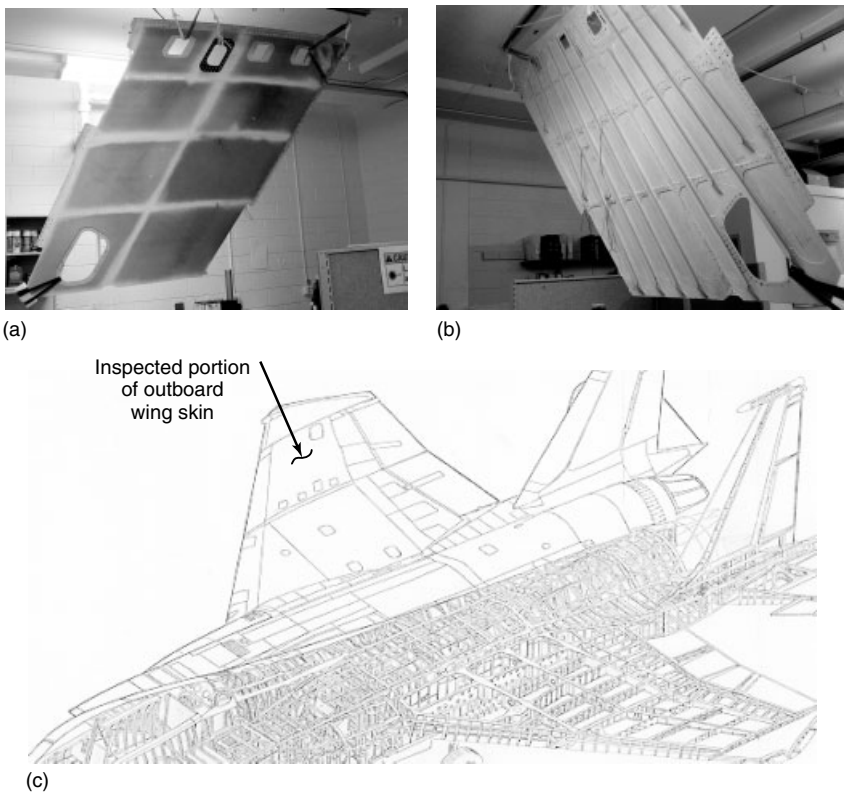


Figure 5. F15 wing skin: front (a) and back (b) views, and location of component on the aircraft (c).

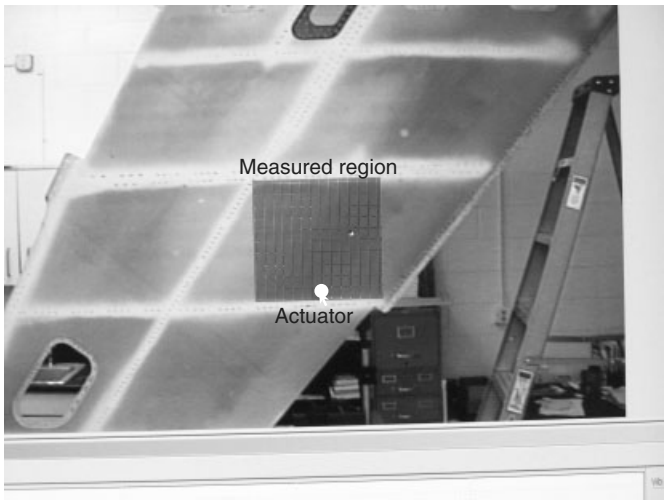
obtained are used to evaluate the energy distribution over the monitored region and the corresponding damage measure, whereby the decimation of the full data set is used to generate baseline information. The damage measure map resulting from the combination of the first 10 modes of the skin is shown in Figure 6(b). The map clearly shows the presence of damage, which is highlighted as a dark region on the white background corresponding to the unit value.

3 GUIDED ULTRASONIC WAVES-BASED DAMAGE MEASURES

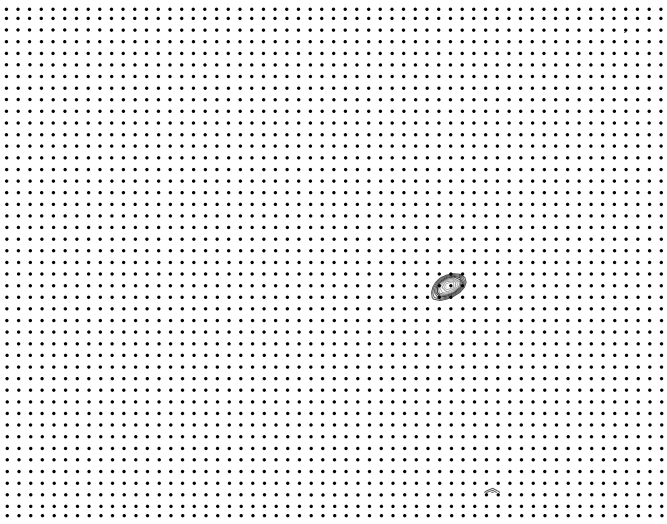
The application of GUWs as inspection tools for SHM is receiving attention by a large number of researchers. The most attractive feature is the ability of GUWs to travel long distances in plate and shell-like structures, which makes the inspection of large areas possible. The complicating factors are the multi-modal and dispersive natures of GUW signals, which require the development of proper interpretation tools. The objective is the extraction of relevant features from the recorded response, which may be related to damage.

3.1 Signal processing

In this regard, proper signal-processing algorithms are essential features of GUW-based SHM techniques. Overviews of signal-processing strategies used for GUW interpretation and damage measure formulation can be found in [3, 5]. Time frequency transforms (TFTs) are well-suited for the analysis, decomposition and denoising of GUW signals, which are typically nonstationary. Short time Fourier transforms (STFT) [31], Wigner–Ville distributions (WVD) [32, 33], and, more recently, the Hilbert–Huang transform (HHT) [34, 35] are examples of



(a)



(b)

Figure 6. Measurement grid considered for modal testing (a) and resulting damage index map (b).

various techniques used to observe the propagation of various modes, to separate reflections from incident waves, and to formulate associated damage measures. The WT in its various forms is a very important and versatile tool, used extensively for denoising, as well as for feature extraction and selection [3]. The discrete wavelet decomposition is applied, for example, in [36] as part of a feature extraction and automatic classification framework developed for GUW inspection of pipes. Reflection measurements are taken to construct a six-dimensional damage index, which is fed to a neural network (NN), tasked to evaluate size and location of a notch damage. The damage index compares features f_i $(i = 1, \ldots, n)$ associated with the reflected signal with the corresponding ones in the signal traveling from transmitter to receiver:

$$DI = \frac{f_i^{(\text{Reflec})}}{f_i^{(\text{Direct})}} \tag{13}$$

Results are presented to show the monotonic variation of the damage index in terms of the defect size, and the effectiveness of the considered features as inputs to the NN for classification purposes. The STFT is applied in [37] to select and isolate the first symmetric mode (S0), known for being particularly sensitive to crack-type damages, and to formulate the following damage index:

$$DI = \left[\frac{\int_{t_i}^{t_f} |S_{\mathrm{sc}}(\omega_0, t)|^2 \, \mathrm{d}t}{\int_{t_i}^{t_f} |S_{\mathrm{b}}(\omega_0, t)|^2 \, \mathrm{d}t} \right]^{\alpha} \tag{14}$$

where S_{sc} is the time varying spectral amplitude of the scattered signal, S_{b} is the corresponding amplitude for a baseline signal, ω_0 is the excitation frequency, t_i and t_f are time bounds, and α is a gain factor set to better match experimental fatigue crack growth trends. The damage measure obtained using built-in piezoceramics showed a good correlation with crack growth observed by visual inspection, and an alternative formulation based on the energy of the A0 mode is suggested for the detection of disbonds.

A network of sensors is used to construct GUW tomograms used to inspect anisotropic composite plates in [38]. The approach accounts for attenuation in the composite material by using the energy of the earliest wave signals as the reconstruction parameter, and by normalizing the wave energy of the defective sample with respect to that of the undamaged one. For quantitative comparisons between tomograms, a parameter β is introduced as the ratio of the considered values in the defect-free region to that of the defective region of a tomogram.

Higher dimensional Fourier transforms (FTs), transforming the signal in the wavenumber/frequency domain, are also used to identify wave modes and to investigate mode conversion phenomena caused by crack-like damages [39]. Examples of numerical investigations of mode conversion phenomena can be found in [40]. Recently, Ruzzene [41] has applied two-dimensional and three-dimensional FTs as tools to decouple incident and reflected waves, and to filter out the incident component from the recorded signal. The experimental application of this concept is enabled by the application of the SLDV, which easily provides the spatial measurement resolution needed to perform FTs in the spatial domain. A damage measure based on this concept is currently in the works. The SLDV was also used by Staszewski *et al.* [42] to investigate maximum amplitudes of low-frequency Lamb waves propagating in plate structures. Amplitude reductions and sudden increases across the defect were considered as indicators of damage.

3.2 Pattern recognition

As in [36], signal-processing tools are often combined with tools that are able to classify the identified signal features and relate them to damage type and extent. The most common technique for pattern recognition is certainly the use of NNs. Examples of their application for GUW SHM can be found in [43], where spectrographic features from Lamb wave signals in the time-frequency domain were used to construct a damage parameters database, through which an NN was then trained for its successive use for the identification of delaminations in quasi-isotropic composite laminates. A multilayer perceptron (MLP) NN is used in [44] as part of a novelty detection method. The technique is applied to a thin, isotropic plate, where GUWs are sequentially transmitted and captured by eight piezoelectric patches bonded on the plate to act both as sensors and actuators. Scattering waveform responses representing normal and damaged conditions are transformed into a set of novelty indices that are fed as inputs to the NN, incorporating the MLP architecture to compute and predict the damage location on the plate.

3.3 Strain energy distribution

The technique presented in Section 2.4, can be easily adapted to transient time signals corresponding to GUW propagation. The strain energy distribution of equation (9) can be rewritten in terms of displacements $w(x, t)$ and corresponding curvatures in the time domain:

$$U_i(t) = \frac{1}{2} \int_0^L EI(x) w_{,xx}(x, t)^2 \, \mathrm{d}x \tag{15}$$

Again, undamaged, or baseline, information can be synthesized through the decimation process described

in Section 2.5, assuming a sufficient number of measurement locations is available. This formulation leads to a time-dependent damage measure, which can be represented as maps that evolve over time as waves propagate within the structure. The advantages of a time-domain damage measure include the possibility of limiting the investigation to a particular time window. In fact, a recognized problem in wave-propagation-based inspections arises when the incoming wave hides the presence of damage and the resulting wave reflections. The analysis of the trailing part of the wave, after the main pulse has decayed, is often rich in information regarding damage location and extent, as damage generally behaves like a secondary wave source. The considered time-domain formulation allows to select the time interval, where structural response and associated energy distribution are most affected by damage. Examples of results obtained through this approach are presented in the following. Wave propagation tests are performed on the F15 wing skin previously considered. A 40 kHz, five-cycle sinusoidal burst excites the structure through a piezoelectric actuator. A snapshot of the recorded response is shown in Figure 7, which demonstrates the complexity of the wavefield as a result of the structural complexity. In particular, it is shown how the stiffeners tend to reflect the wave and cause it to propagate perpendicular to their length. Such a behavior makes the interpretation of the signals through standard time-of-flight estimation particularly difficult, as it would be almost impossible to differentiate multiple reflections from the one associated with the damage. The

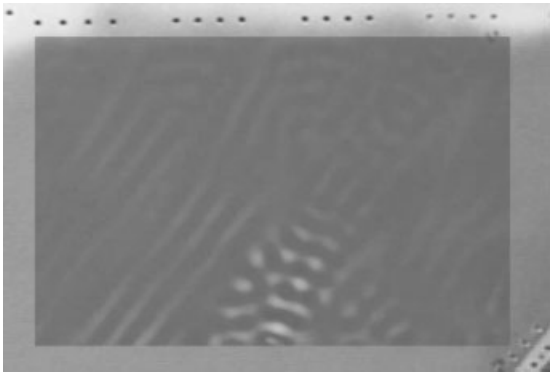


Figure 7. Snapshot of F15 wing skin response resulting from excitation at 40 kHz.

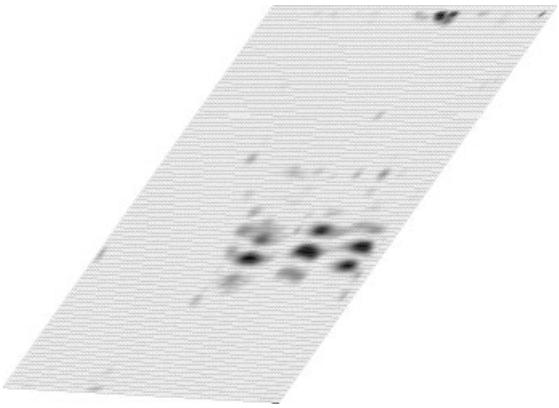


Figure 8. Damage measure evaluated on F15 wing through wave-propagation analysis.

application of the considered damage measure, on the contrary, does not require previous knowledge of the response of the structure, nor does it rely on the availability of a model. It thus represents an attractive solution for the analysis of complex or built-up structures. The damage measure maps obtained on the restricted region, defined by the results of the modal tests described in Section 2.4, are shown in Figure 8, which clearly outline the longitudinal extension and the location of the crack. Results from the experiments conducted on a 1.3 mm (0.050″)-thick metal rib from an F15 aircraft wing are also illustrated. The specimen was removed from service because of the hairline crack in the flange, which can be observed in Figure 9. A piezoceramic actuator was mounted at the location shown in the figure and used to excite elastic waves in the specimen. The excitation is a sinusoidal burst at 35 kHz, which is a frequency that provides the best compromise between short wavelength and wave attenuation through the specimen. The time history corresponding to the propagating elastic wave is recorded at the scan points represented by the dots in Figure 10. The measured displacements are then used as a basis for the interpolation of the response of the scanned region and for the successive formulation of the damage measure in the time domain, which is shown, upon summation over time, in Figure 10. The contours clearly highlight the hairline crack and vividly outline its shape, thus providing its initial characterization.

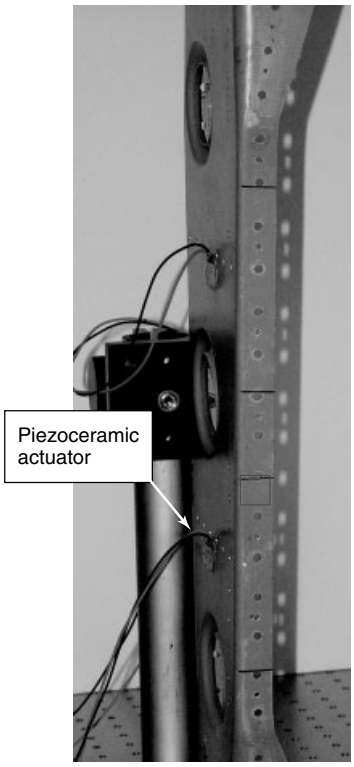
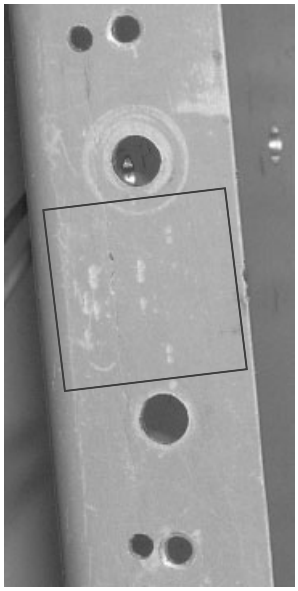


Figure 9. F15 rib.

4 CONCLUSIONS

This article is devoted to vibration-based and wave-propagation-based inspection techniques and to the formulation of damage measures. The overview presents a limited range of solutions proposed in the last 15–20 years, and does not claim to be all-inclusive and exhaustive, due to the wealth of literature available in the field. The presented techniques and the list of references should, however, be sufficient to guide a researcher beginning investigations in this area. The article does not address nonlinear techniques, both in the modal and wave-propagation domains, which, however, are very promising and should be considered as natural extensions of most of the techniques presented earlier. Among the nonlinear techniques, nonlinear acoustics for the evaluation of nonlinear stress–strain parameters as damage precursors and indicators is an area that has received much attention, and that still holds great promises for future research contributions.



(a) Detail of the crack

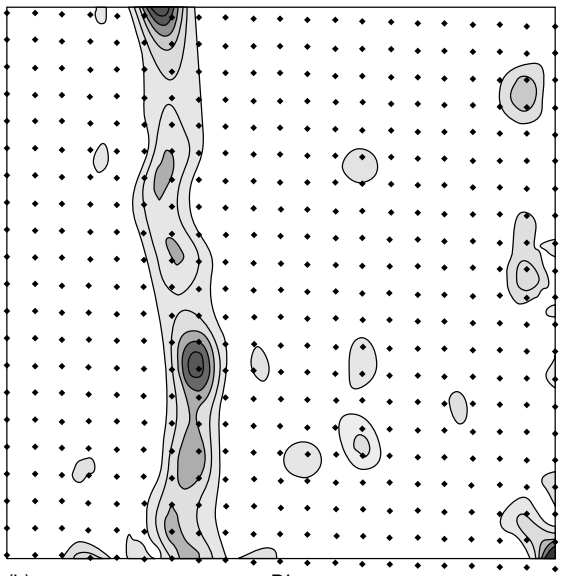


Figure 10. Detail of crack and time-domain damage measure evaluated at 35 kHz.

Future challenges consist in the development of novel and robust interpretation algorithms, as well as in the implementation of novel actuation and sensing strategies. The results presented, in fact, show that detailed spatial measurements provide an enormous

amount of information, which can be exploited in several ways to identify, locate, and quantify damage. Nowadays, such spatial refinement can only be achieved through complex, and often bulky, equipment, or lengthy inspections, and great advantages may be gained if similar detail could be achieved with portable, or *in situ*, devices. The development of such systems could lead to significant advancements in the general area of SHM.

ACKNOWLEDGMENTS

The author wishes the thank Mr V. K. Sharma of Millennium Dynamics Corporation for supplying some of the experimental results shown in this article.

RELATED ARTICLES

Fundamentals of Guided Elastic Waves in Solids

Modal–Vibration-based Damage Identification

Guided-wave Array Methods

Signal Processing for Damage Detection

Novelty Detection

Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators

Full-field Sensing: Three-dimensional Computer Vision and Digital Image Correlation for Noncontacting Shape and Deformation Measurements

Development of an Active Smart Patch for Aircraft Repair

Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge

Integrated Sensor Durability and Reliability

REFERENCES

[1] Doebling SW, Farrar C, Prime MB, Daniel WS. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*, Los Alamos National Laboratory Report LA-13070-MS. Los Alamos National Laboratory, May 1996.

[2] Sohn H, Farrar CR, Francois MH, Devin DS, Daniel WS, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Los Alamos National Laboratory Report LA-13976-MS. Los Alamos National Laboratory, 2003.

[3] Staszewski WJ, Boller C, Tomlinson G. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons, 2004.

[4] Rose JL. A baseline and vision of ultrasonic guided wave inspection potential. *Transactions of the ASME: Journal of Pressure Vessel Technology* 2002 **124**:273–282.

[5] Raghavan A, Cesnik CES. Review of guided-wave structural health monitoring. *The Shock and Vibration Digest* 2007 **39**(2):91–114.

[6] Michaels TE, Michaels JE. Ultrasonic signal processing for structural health monitoring. In *Review of Progress in QNDE*, Thompson DO, Chimenti DE (eds). American Institute of Physics, 2004; Vol. 23.

[7] Giurgiutiu V, Bao J, Zhao W. Piezoelectric wafer active sensor embedded ultrasonics in beams and plates. *Experimental Mechanics* 2003 **43**(4): 428–449.

[8] Sharma V, Ruzzene M, Hanagud S. Damage index estimation in beams and plates using laser vibrometry. *AIAA Journal* 2006 **44**:919–923.

[9] Sharma V, Ruzzene M, Hanagud S. Perturbation analysis of damaged plates. *International Journal of Solids and Structures* 2006 **43**(16):4648–4672.

[10] Luo H, Hanagud S. An integral equation for changes in the structural dynamics characteristics of undamaged structures. *International Journal of Solids and Structures* 1997 **34**(35–36):4557–4579.

[11] Lestari W. *Damage of Composite Structures: Detection Technique, Dynamic Response and Residual Strength*, Ph.D. Thesis. Georgia Institute of Technology, July 2001.

[12] Cawley P, Adams RD. The locations of defects in structures from measurements of natural frequencies. *Journal of Strain Analysis* 1979 **14**(2):49–57.

[13] Stubbs N, Osegueda R. Global non-destructive damage evaluation in solids. *Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis* 1990 **5**(2):67–79.

[14] Stubbs N, Osegueda R. Global damage detection in solids-experimental verification. *Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis* 1979 **5**(2):81–97.

[15] Hearn G, Testa RB. Modal analysis for damage detection in structures. *ASCE Journal of Structural Engineering* 1991 **117**(10):3042–3063.

[16] Richardson MH, Mannan MA. Remote detection and location of structural faults using modal parameters. *Proceedings of the 10th International Modal Analysis Conference*. San Diego, CA, 1992; pp. 502–507.

[17] West WM. Illustration of the use of modal assurance criterion to detect structural changes in an orbiter test specimen. *Proceedings of Air Force Conference on Aircraft Structural Integrity*, 1984; pp. 1–6.

[18] Yuen MMF. A numerical study of the eigenparameters of a damaged cantilever. *Journal of Sound and Vibration* 1985 **103**:301–310.

[19] Fox CHJ. The location of defects in structures: a comparison of the use of natural frequency and mode shape data. *Proceedings of the 10th International Modal Analysis Conference*. San Diego, CA, 1992; pp. 522–528.

[20] Kim J-H, Jeon H-S, Lee C-W. Application of the modal assurance criteria for detecting and locating structural faults. *Proceedings of the 10th International Modal Analysis Conference*. San Diego, CA, 1992; pp. 536–540.

[21] Ko JM, Wong CW, Lam HF. Damage detection in steel framed structures by vibration measurement approach. *Proceedings of the 12th International Modal Analysis Conference*. Honolulu, HI, 1994; pp. 280–286.

[22] Lam HF, Ko JM, Wong CW. Detection of damage location based on sensitivity analysis. *Proceedings of the 13th International Modal Analysis Conference*. Nashville, TN, 1995; pp. 1499–1505.

[23] Zhong S, Oyadiji SO. Crack detection in simply supported beams without baseline modal parameters by stationary wavelet transform. *Mechanical Systems and Signal Processing* 2007 **21**:1853–1884.

[24] Rucka M, Wilde K. Application of continuous wavelet transform in vibration based damage detection method for beams and plates. *Journal of Sound and Vibration* 2006 **297**:536–550.

[25] Chang CC, Chen L-W. Detection of the location and size of cracks in the multiple cracked beam by spatial wavelet based approach. *Mechanical Systems and Signal Processing* 2005 **19**:139–155.

[26] Pandey AK, Biswas M, Samman MM. Damage detection from changes in curvature mode shapes. *Journal of Sound and Vibration* 1991 **145**(2): 321–332.

[27] Ho YK, Ewins DJ. Numerical evaluation of the damage index. *International Workshop on Structural Health Monitoring*. Stanford University, Palo Alto, CA, 2000; pp. 995–1011.

[28] Ho YK, Ewins DJ. On the structural damage identification with mode shapes. *European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 1999; pp. 677–686.

[29] Kim JT, Stubbs N. Crack detection in beam type structures using frequency data. *Journal of Sound and Vibration* 2003 **259**(1):146–160.

[30] Cornwell P, Doebling SW, Farrar CR. Application of the strain energy damage detection method to plate-like structures. *Journal of Sound and Vibration* 1999 **224**(2):359–374.

[31] Prasad SM, Kumar VR, Balasubramaniam K, Krishnamurthy CV. Imaging of defects in composite structures using guided ultrasonics. *Proceedings of the SPIE*. San Diego, CA, 2003; Vol. 5062, pp. 700–703.

[32] Prosser WH, Seale MD, Smith BT. Time–frequency analysis of the dispersion of Lamb modes. *Journal of the Acoustical Society of America* 1999 **105**(5):2669–2676.

[33] Raghavan A, Cesnik CES. Guided-wave signal processing using chirplet matching pursuits and mode correlation for structural health monitoring. *Smart Materials and Structures* 2007 **16**(2):355–366.

[34] Oseguda R, Kreinovich V, Nazarian S, Roldan E. Detection of cracks at rivet holes in thin plates using Lamb wave scanning. *Proceedings of the SPIE*. San Diego, CA, 2003; Vol. 5047, pp. 55–66.

[35] Salvino L, Purekar A, Pines DJ. Health monitoring of 2-D plates using EMD and Hilbert phase. *Proceedings of the 4th International Workshop on Structural Health Monitoring*. Stanford University, Palo Alto, CA, 2005.

[36] Rizzo P, Bartoli I, Marzani A, Lanza di Scalea F. Defect classification in pipes by neural network using multiple guided ultrasonic wave features extracted after wavelet processing. *Transactions of the ASME: Journal of Pressure Vessel Technology* 2005 **127**:294–303.

[37] Ihn JB, Chang FK. Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: I. Diagnostics. *Smart Materials and Structures* 2004 **13**: 609–620.

[38] Prasad SM, Balasubramaniam K, Krishnamurthy CV. Structural health monitoring of composite structures using Lamb wave tomography. *Smart Materials and Structures* 2004 **13**:N73–N79.

[39] Alleyne D, Cawley P. A two-dimensional Fourier transform method for the measurement of propagating multimode signals. *Journal of the Acoustical Society of America* 1991 **89**:1159–1168.

[40] Basri R, Chiu WK. Numerical analysis on the interaction of guided Lamb waves with a local elastic stiffness reduction in quasi-isotropic composite plate structures. *Composite Structures* 2004 **66**:87–99.

[41] Ruzzene M. Frequency/wavenumber filtering for improved damage visualization. In *Review of Progress in QNDE*, Thompson DO, Chimenti DE (eds). American Institute of Physics, 2006; Vol. 25.

[42] Staszewski WJ, Lee BC, Mallet L, Scarpa F. Structural health monitoring using laser vibrometry. Part I and II. *Smart Materials and Structures* 2004 **13**:251–269.

[43] Su Z, Ye L. An intelligent signal processing and pattern recognition technique for defect identification using an active sensor network. *Smart Materials and Structures* 2004 **13**(4):957–969.

[44] Mustapha F, Manson G, Worden K, Pierce SG. Damage location in an isotropic plate using a vector of novelty indices. *Mechanical Systems and Signal Processing* 2006 **21**:1885–1906.

Chapter 49

Modeling of Lamb Waves in Composite Structures

Abir Chakraborty

India Science Laboratory, General Motors R & D, Bangalore, India

1 INTRODUCTION

The methodology of assessing the integrity of structural composites, traditionally based on nondestructive techniques (NDT), has evolved into the generic field of structural health monitoring (*see* **Lamb Wave-based SHM for Laminated Composite Structures**). In NDT, propagation of elastic waves in structures is used to probe for the existence of damage, its location, and its severity. For composite structures, complexity increases owing to the anisotropic and inhomogeneous nature of the constituent materials and the architecture of the composite itself. Further, because of the layered structure of composites, generated waves get reflected and transmitted multiple times resulting in a complex wave pattern and making it difficult to interpret the signal. The boundaries are also responsible for the creation of new kinds of waves. For example, when elastic waves propagate in an isotropic platelike structure, they would experience repeated reflections at the top and bottom surfaces alternately and a new kind of wave—called the *Lamb wave*—is generated from the original primary and secondary waves. Thus, the Lamb wave is a guided wave [1] (*see* **Fundamentals of Guided Elastic Waves in Solids**) propagating in a domain bounded by two parallel traction-free surfaces. The importance of the Lamb wave in NDT applications lies in its ability to inspect large areas at a time by propagating a long distance without attenuation. Hence, the Lamb wave finds immense application in structural health monitoring for ultrasonic inspection of platelike aircraft components, missile cases, pressure vessels, oil tanks, pipelines, etc (*see* **Ultrasonic Methods**).

Classical wave analysis begins with the plane-wave assumption (harmonic dependence of the displacement field on the spatial and temporal variables), i.e.,

$$\mathbf{u}(x, y, z, t) = \mathbf{A}\exp(j(k_x x + k_y y + k_z z - \omega t) = \mathbf{A}\exp(j(\mathbf{k}\cdot\mathbf{x} - \omega t), \quad j^2 = -1 \quad (1)$$

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

where boldface letters denote vectors or matrices. The unknowns in this case, i.e., the wave amplitude vector **A** and the wavenumbers k_x, k_y, and k_z are related to the excitation frequency ω. Substitution of the assumed form of the displacement field in the governing equations

$$\sigma_{mn,\,n} = \rho \ddot{u}_m, \quad \sigma_{mn} = C_{mnpq}\varepsilon_{pq} \tag{2}$$

results in the algebraic form of the governing equation:

$$(\Gamma_{mn} - \rho\omega^2\delta_{mn})A_n = 0, \quad \Gamma_{mn} = C_{mnpq}k_p k_q \tag{3}$$

which is an eigenvalue problem for ω. The characteristic polynomial Γ provides the necessary relation between the wavenumbers and frequency, called *the dispersion relation*. The dispersion relation is also expressed in terms of the frequency and phase speed vector ($\mathbf{c}_\mathrm{p}$)

$$\mathbf{c}_\mathrm{p} = \frac{\omega(\mathbf{k})}{|\mathbf{k}|}\mathbf{k}, \quad \mathbf{k} = (k_x, k_y, k_z) \tag{4}$$

and the frequency and group speed vector ($\mathbf{c}_\mathrm{g}$)

$$\mathbf{c}_\mathrm{g} = \nabla_\mathbf{k}\omega(\mathbf{k}) \tag{5}$$

For isotropic materials, the group speed vector points toward the direction of energy propagation. The group speed is also of higher significance than phase speed as it is related to one of the important parameters of NDT methods, the time of flight. Thus, the first step toward wave analysis is finding out the dispersion relation and the second step is finding out the time-domain response using this relation.

It is to be noted that the governing equation provides only one condition to be satisfied by **k** and ω. However, for the three components of **k** to be completely determined by ω, it is necessary to have three conditions. In the absence of these extra conditions, a few of them are to be assumed, and the rest are expressed in terms of the assumed ones. Thus, for the analysis of three-dimensional (3-D) media of infinite extent, two components of **k** are to be provided, and the third component is solved from the dispersion relation. In this case, the phase and group speed are constant, and the waves are nondispersive in nature. For isotropic material, only two kinds of waves can propagate in a 3-D unbounded media. They are called *dilatational (primary or P-wave)* and *shear (secondary or S-wave) waves*. Further, depending upon the plane of propagation, shear waves are categorized as shear vertical (SV) and shear horizontal (SH) waves. However, for anisotropic material, clear distinction between the P and S waves cannot be made, and they are referred to as *quasi-P* and *quasi-shear* waves.

In the case of singly or doubly bounded media, where the boundary comes into play, the speeds become functions of ω and the waves become dispersive, i.e., they cannot retain their shape while propagating. Historically, the dispersion relation for an elastic isotropic plate with infinite extent in plane strain state was first derived by Lamb [2], thus marking the beginning of a new kind of wave. The presence of boundary imposes another condition to be satisfied by the components of **k** and ω. For the analysis of 2-D media, this extra condition is sufficient to calculate **k** for a given ω. However, for 3-D geometry, one of the wavenumber components should be assumed.

The Lamb wave propagates in doubly bounded media and traction-free boundary conditions are to be imposed at the edges. Imposition of this condition restricts the freedom of choosing wavenumber components and only few admissible wavenumbers exist at a particular frequency. These wavenumbers constitute the fundamental modes of propagation, similar to the fundamental modes of vibration (standing wave). As happens in the latter case, excitation frequency determines modal participation of the Lamb wave modes. Thus, it is necessary to have information about the modes in advance to analyze the wave generated by the actuators.

For isotropic materials, it is possible to distinguish between the symmetric and antisymmetric modes (Figure 1). However, for composite materials, as stated before, this kind of distinction is hard to make. In the case of symmetric laminates, it is possible

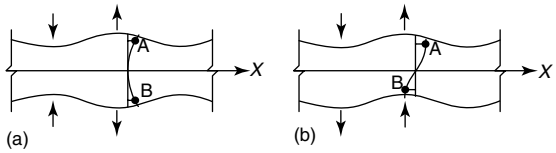


Figure 1. (a) Symmetric Lamb wave propagation and (b) antisymmetric Lamb wave propagation.

to separate the Lamb waves into symmetric and antisymmetric modes. For symmetric modes, one type is designated as quasi-extensional (qS_n), where the dominant component of the polarization vector is along the propagation direction, and the other type is quasi-horizontal shear (qSH_{2n}), where the polarization vector is mainly parallel to the plane of the plate. Similarly, for the antisymmetric types of wave modes, the quasi-flexural (qA_n) and quasi-horizontal shear (qSH_{2n-1}) waves are generated.

In general, there are two theoretical approaches to investigate Lamb waves in composites: one is the exact solution by 3-D elasticity theory, and the other involves approximate solutions by plate theories. For the approach using 3-D elasticity, Nayfeh and Chimenti [3] gave dispersion relations of Lamb waves in a composite lamina. Later, Nayfeh [4] developed the transfer matrix technique to obtain the dispersion curves for Lamb waves in laminates. In this method, the displacements and stresses of one interface are related to the other interface by a system matrix developed from the governing equation and stress displacement relation. This matrix is assembled for more than one layer, and the resulting system needs to be solved after imposing the boundary conditions. A comprehensive review of the theories and applications of Lamb waves is given by Chimenti [5]. Although the exact solutions provide accurate results, the computation for dispersion characteristics of multilayered composites is intensive because of the transcendental nature of the equations. To circumvent this problem, approximate solutions are obtained from higher order laminated plate theories. There are other analytical–numerical approaches for modeling Lamb waves, e.g., on the basis of discrete layer theory and multiple integral transform [6], coupled finite element (FE)–normal mode expansion method [7], or boundary element–normal mode expansion method [8]. The conventional FE method is suitable only for wave propagation simulation, and no information can be obtained for the dispersion relation.

The assumed form of the displacement field indicates an integral transform-based approach for wave-propagation analysis. The spectral finite element method (SFEM) is a confluence of the Fourier transform and conventional FE-based matrix approach, which is an inexpensive way of constructing the Lamb wave modes as well as predicting time-domain signals [9–15]. The SFEM in two dimensions (2-D) is based on simplified geometry like layered media and the method is not vastly different from the transfer-matrix-based method [16–18]. However, the stiffness matrix form, along with the exact representation of the layer, makes it more efficient for modeling multiple layers and for analyzing high-frequency impact loading. The complete solution in SFEM is formulated on the basis of the partial wave technique (PWT). In the PWT-based method of Lamb wave analysis, once the partial waves are found, the wave coefficients are made to satisfy the two nonzero tractions specified at the top and bottom of the layer. In the general case, the formulation is slightly different, as no specific problem-oriented boundary conditions are imposed. Thus a system matrix is established, which relates the tractions at the interface to the interfacial displacements. This generalization enables the use of the system matrix as a finite-element dynamic stiffness matrix, although formulated in the frequency–wavenumber domain. These matrices can be assembled to model different layers of different ply orientation or inhomogeneity, which obviates the necessity for the cumbersome computation associated with multilayer analysis [19]. The only shortcoming of the method is that each spectral layer element (SLE) can accommodate only one fiber angle; thus for different ply stacking sequences, the number of elements will be at least equal to the number of different ply angles in the stacking.

In the next section, the procedure for solving the 2-D elastodynamic equation is outlined for arbitrary load and boundary conditions. Then, the method of obtaining the Lamb wave modes from the 2-D solutions is discussed. Next, different beam theories are evaluated to predict Lamb modes, which are subsequently compared with the 2-D solutions. Finally, the propagation of the lower-order Lamb waves is simulated.

2 SOLUTION OF 2-D ELASTICITY EQUATION

It is assumed that there is no heat conduction in and out of the system, the displacements are small, the material is homogeneous and anisotropic, and the domain is 2-D Euclidean space. The general elastodynamic equation of motion for three dimensions is given by

$$\sigma_{ij,j} = \rho(x_1, x_2, x_3)\ddot{u}_i, \quad \sigma_{ij} = C_{ijkl}(x_1, x_2, x_3)\varepsilon_{kl}$$
$$\varepsilon_{ij} = \frac{(u_{i,j} + u_{j,i})}{2} \tag{6}$$

where comma (,) and dot ($\dot{}$) denote partial differentiation with respect to the spatial variables and time, respectively.

For 2-D model with orthotropic material construction, complexity of the above equation can be further reduced by the following assumptions: the nonzero displacements are $u_1 = u$ and $u_3 = w$ in the direction $x_1 = x$ and $x_3 = z$, respectively (Figure 2). Then the nonzero strains are related to these displacements by

$$\varepsilon_{xx} = u_x, \quad \varepsilon_{zz} = w_z, \quad \varepsilon_{xz} = u_z + w_x \tag{7}$$

The nonzero stresses are then related to the strains by the relation

$$\sigma_{xx} = Q_{11}\varepsilon_{xx} + Q_{13}\varepsilon_{zz}$$
$$\sigma_{zz} = Q_{13}\varepsilon_{xx} + Q_{33}\varepsilon_{zz}, \sigma_{xz} = Q_{55}\varepsilon_{xz} \tag{8}$$

where Q_{ij}'s are the stiffness coefficients, which depend on the ply lay up, its orientation, and the z coordinate of the layer. Substituting equation (8) in equation (6) and imposing the assumptions, the elastodynamic equation for 2-D homogeneous orthotropic media is given by

$$Q_{11}u_{xx} + (Q_{13} + Q_{55})w_{xz} + Q_{55}u_{zz} = \rho\ddot{u}$$
$$Q_{55}w_{xx} + (Q_{13} + Q_{55})u_{xz} + Q_{33}w_{zz} = \rho\ddot{w} \tag{9}$$

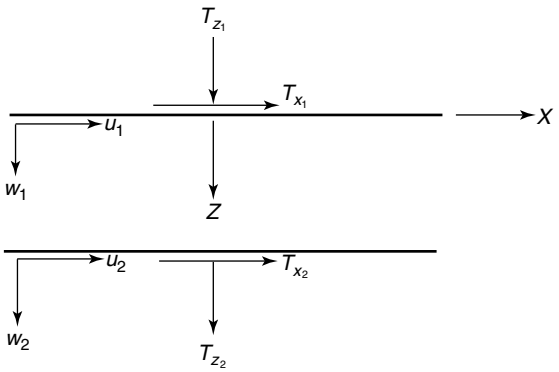


Figure 2. Spectral finite element and the associated degrees of freedom.

We attempt to reduce the governing partial differential equations (PDEs) to a set of ordinary differential equations (ODEs). For this, we need to remove two variables out of the system and introduce two new parameters instead. To achieve this, we take Fourier transform in time and Fourier series in space. With this assumption, the spectral form of the displacement field becomes

$$u(x, z, t) = \sum_{n=1}^{N-1}\sum_{m=1}^{M-1} \hat{u}(z, \eta_m, \omega_n) \begin{Bmatrix} \sin(\eta_m x) \\ \cos(\eta_m x) \end{Bmatrix} \mathrm{e}^{-j\omega_n t} \tag{10}$$

$$w(x, z, t) = \sum_{n=1}^{N-1}\sum_{m=1}^{M-1} \hat{w}(z, \eta_m, \omega_n) \begin{Bmatrix} \cos(\eta_m x) \\ \sin(\eta_m x) \end{Bmatrix} \mathrm{e}^{-j\omega_n t} \tag{11}$$

where ω_n is the discrete angular frequency and η_m is the discrete horizontal wavenumber. As the assumed field suggests, for $M \to \infty$, the model has infinite extent in the positive and negative X direction, although the domain is finite in the Z direction, i.e., it is a layered structure. In particular, the domain can be written as $\Omega = [-\infty, +\infty] \times [0, L]$, where L is the thickness of the layer. The boundaries of any layer are specified by a fixed value of z. The X dependency of the displacement field (sine or cosine) is determined on the basis of the loading pattern. In all subsequent formulation and computation, symmetric load pattern is considered, i.e., $\sin(\eta_m x)$ for u and $\cos(\eta_m x)$ for w. The real computational domain is $\Omega_c = [-X_L/2, +X_L/2] \times [0, L]$, where X_L is the X window length. Discrete values of η_m depend upon the X_L and the number of mode shapes (M) chosen.

This displacement field reduces the governing equations to a set of ODEs as

$$\mathbf{A}\hat{\mathbf{u}}'' + \mathbf{B}\hat{\mathbf{u}}' + \mathbf{C}\hat{\mathbf{u}} = \mathbf{0}, \quad \hat{\mathbf{u}} = \{\hat{u}\ \hat{w}\} \tag{12}$$

where prime denotes differentiation with respect to z. The matrices **A**, **B**, and **C** are

$$\mathbf{A} = \begin{bmatrix} Q_{55} & 0 \\ 0 & Q_{33} \end{bmatrix}$$
$$\mathbf{B} = \begin{bmatrix} 0 & -(Q_{13} + Q_{55})\eta_m \\ (Q_{13} + Q_{55})\eta_m & 0 \end{bmatrix} \tag{13}$$

$$\mathbf{C} = \begin{bmatrix} -\eta_m^2 Q_{11} + \rho\omega_n^2 & 0 \\ 0 & -\eta_m^2 Q_{55} + \rho\omega_n^2 \end{bmatrix} \quad (14)$$

The associated boundary conditions are the specifications of the stresses σ_{zz} and σ_{xz} at the layer interfaces. From equation (8), the stresses are related to the unknowns as

$$\hat{\mathbf{s}} = \mathbf{D}\hat{\mathbf{u}}' + \mathbf{E}\hat{\mathbf{u}},\ \hat{\mathbf{s}} = \{\sigma_{zz}\ \sigma_{xz}\}$$
$$\mathbf{D} = \begin{bmatrix} 0 & Q_{33} \\ Q_{55} & 0 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \eta_m Q_{13} & 0 \\ 0 & -\eta_m Q_{55} \end{bmatrix} \quad (15)$$

The original boundary-value problem reduces to finding $\hat{\mathbf{u}}$, which satisfies equation (12) for all $z \in \Omega_c$ and specification of $\hat{\mathbf{u}}$ or $\hat{\mathbf{s}}$ at $z = 0$ or $z = L$. Once the solution is obtained for different values of z in the frequency–wavenumber domain ($Z - \eta - \omega$ domain, for given values of ω_n and η_m), summation over η_m brings the solution back to the $Z - X - \omega$ domain and inverse Fourier transform brings the solution back to time domain, i.e., $Z - X - t$ domain.

The solution to these ODEs are of the form $u_0 e^{-jkz}$ and $w_0 e^{-jkz}$, which yields the polynomial eigenvalue problem

$$\mathbf{W}\{\mathbf{u}_0\} = \mathbf{0},\ \mathbf{W} = -k^2\mathbf{A} - jk\mathbf{B} + \mathbf{C},\ \{\mathbf{u}_0\} = \{u_0 w_0\} \quad (16)$$

where $\mathbf{W}$ is the wave matrix given by

$$\mathbf{W} = \begin{bmatrix} -k^2 Q_{55} - \eta_m^2 Q_{11} + \rho\omega_n^2 & jk\eta_m(Q_{13} + Q_{55}) \\ -jk\eta_m(Q_{13} + Q_{55}) & -k^2 Q_{33} - \eta_m^2 Q_{55} + \rho\omega_n^2 \end{bmatrix} \quad (17)$$

The singularity condition of $\mathbf{W}$ yields the following equation for determining the spectrum relation:

$$Q_{33}Q_{55}k^4 + \{(Q_{11}Q_{33} - 2Q_{13}Q_{55} - Q_{13}^2)\eta_m^2 - \rho\omega_n^2(Q_{33} + Q_{55})\}k^2 + \{Q_{11}Q_{55}\eta_m^4 - \rho\omega_n^2\eta_m^2(Q_{11} + Q_{55}) + \rho^2\omega_n^4\} = 0 \quad (18)$$

It is to be noted that for each value of η_m and ω_n, there are four values of k, denoted by $k_{lmn}, l = 1, \ldots, 4$, which are obtained by solving equation 18. Explicit solution of the wavenumber k is $k_{lnm} = \pm\sqrt{-b \pm \sqrt{b^2 - 4ac}}/2a$, where a, b, and c are the coefficients of k^4, k^2, and k^0, respectively, in equation 18.

There are certain properties of the wavenumbers that will be explored now. As can be seen from equation 18, for $\eta_m = 0$, the equation is readily solvable to give the roots $\pm\omega\sqrt{\rho/Q_{33}}$ and $\pm\omega\sqrt{\rho/Q_{55}}$. Since none of the ρ, Q_{33}, or Q_{55} can be negative or zero, these roots are always real and linear with ω. When η_m is not zero, k becomes zero for ω satisfying

$$Q_{11}Q_{55}\eta_m^4 - \rho\omega_n^2\eta_m^2(Q_{11} + Q_{55}) + \rho^2\omega_n^4 = 0$$
$$\text{i.e.,}\quad (Q_{11}\eta_m^2 - \rho\omega^2)(Q_{55}\eta_m^2 - \rho\omega^2) = 0$$
$$\text{i.e.,}\quad \omega = \eta_m\sqrt{Q_{11}/\rho}\,,\quad \eta_m\sqrt{Q_{55}/\rho} \quad (19)$$

which are the cutoff frequencies. For frequencies lower than the cutoff frequencies, the roots are imaginary and nonpropagating, and above these frequencies, the roots are real and propagating. For isotropic materials, the cutoff frequencies are given by $c_p\eta$ and $c_s\eta$ [20]. The current expressions for the cutoff frequencies are also reducible to that of isotropic materials if we identify Q_{11} and Q_{55} with $\lambda + 2\mu$ and μ, respectively, where λ and μ are the Lame's parameters. If we identify the qP wave with Q_{33} (or Q_{11}) and the qSV wave with Q_{55}, then, as the cutoff frequencies suggest, for the same value of η_m, it is the qSV wave that becomes propagating first, since $Q_{11} > Q_{55}$. In Figure 3, the wavenumbers are plotted for three different ply angles, 0°, 45°, and 90°. For all the ply angles, Q_{33} and Q_{55} are assumed 9.69 and 4.13 GPa, respectively. For Q_{11} and Q_{13}, the following values are assumed. For 0°, $Q_{11} = 146.3$ GPa and $Q_{13} = 2.98$ GPa, for 45°, $Q_{11} = 44.62$ GPa and $Q_{13} = 1.62$ GPa, and for 90°, $Q_{11} = 9.69$ GPa and $Q_{13} = 2.54$ GPa. In Figure 3, the imaginary part of the wavenumbers is plotted in the horizontal plane and the real part in the vertical plane. Further, the imaginary part of the wavenumbers for 0° and 90° are plotted on the positive side, whereas for 45° it is plotted on the negative side, for distinction. Two different η_m values are considered. The linear variation of the real part of the wavenumbers are for $\eta_m = 0$ and the variation for the rest of

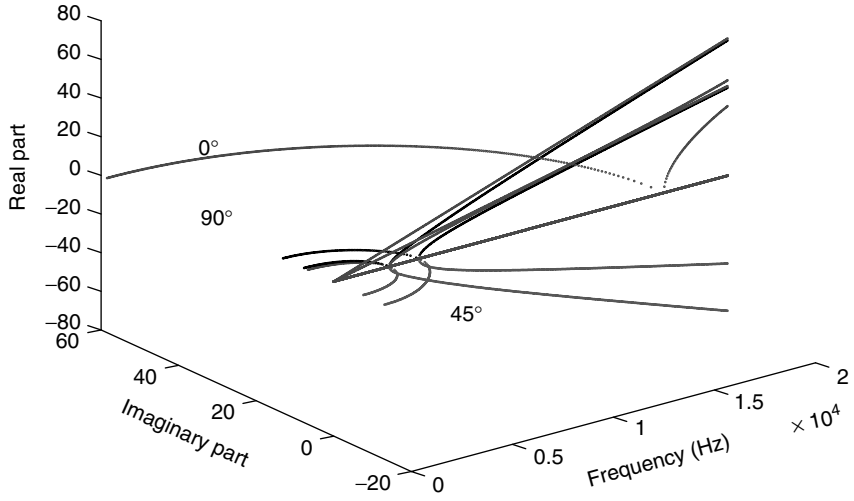


Figure 3. Variation of wavenumber with ω_n ($\eta_m = 10$).

the plots are for $\eta_m = 10$. As discussed previously, the slope of the linear portion depends upon Q_{33} and Q_{55}, and as they are equal for all the ply angles, this part is common for all the ply angles. The difference comes in the imaginary part and the cutoff frequencies. Two different cutoff frequencies are seen in the figure for each ply angle, where the largest value is for the 0° ply angle because of the large Q_{11} value. Further, the shear cutoff frequency is the same for all the ply angles as Q_{55} is the same for all the cases.

Once the required wavenumbers k for which the wave matrix $\mathbf{W}$ is singular are obtained, the solution $\mathbf{u}_0$ at frequency ω_n and wavenumber η_m is

$$u_{nm} = R_{11}C_1 e^{-jk_1 x} + R_{12}C_2 e^{-jk_2 x} + R_{13}C_3 e^{-jk_3 x} + R_{14}C_4 e^{-jk_4 x} \quad (20)$$

$$w_{nm} = R_{21}C_1 e^{-jk_1 x} + R_{22}C_2 e^{-jk_2 x} + R_{23}C_3 e^{-jk_3 x} + R_{24}C_4 e^{-jk_4 x} \quad (21)$$

where R_{ij} are the amplitude coefficients to be determined and they are called *wave amplitudes*. For any wavenumber k_p, the wave amplitudes R_{1p} and R_{2p} satisfy the relation

$$W_{11}R_{1p} + W_{12}R_{2p} = 0, \quad W_{21}R_{1p} + W_{22}R_{2p} = 0 \quad (22)$$

Any of these two relations can be used to express one in terms of the other. Once the four wavenumbers and wave amplitudes are known, the four partial waves can be constructed and the displacement field can be written as a linear combination of the partial waves. Each partial wave is given by

$$\mathbf{a}_i = \begin{Bmatrix} u_i \\ w_i \end{Bmatrix} = \begin{Bmatrix} R_{1i} \\ R_{2i} \end{Bmatrix} e^{-jk_i z} \begin{Bmatrix} \sin(\eta_m x) \\ \cos(\eta_m x) \end{Bmatrix} e^{-j\omega_n t} \quad i = 1 \ldots 4 \quad (23)$$

and the total solution is

$$\mathbf{u} = \sum_{i=1}^{4} C_i \mathbf{a}_i \quad (24)$$

Once the solutions of u and w are obtained in the form of equations (20) and (21) for each value of ω_n and η_m, the same procedure as outlined in the 1-D element formulation is employed to obtain the element dynamic stiffness matrix at ω_n and η_m. Thus, the nodal displacements are related to the unknown constants by

$$\{u_{1nm}\ v_{1nm}\ u_{2nm}\ v_{2nm}\}^{\mathrm{T}} = [\mathbf{T}_{1nm}]\{C_1\ C_2\ C_3\ C_4\}^{\mathrm{T}} \quad (25)$$

i.e.,

$$\{\hat{\mathbf{u}}\}_{nm} = [\mathbf{T}_1]_{nm}\{\mathbf{C}\}_{nm} \quad (26)$$

Using equation 15, nodal tractions are related to the constants by

$$\{\hat{\mathbf{t}}\}_{nm} = [\mathbf{T_2}]_{nm}\{\mathbf{C}\}_{nm}$$
$$\{\hat{\mathbf{t}}\}_{nm} = \{\sigma_{zz1},\ \sigma_{xz1},\ \sigma_{zz2},\ \sigma_{xz2}\} \tag{27}$$

where inhomogeneity of the material is used while evaluating the tractions at the nodes.

Explicit forms of $\mathbf{T}_{2nm}$ and $\mathbf{T}_{1nm}$ are

$$\mathbf{T_1} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ R_{21} & R_{22} & R_{23} & R_{24} \\ R_{11}\mathrm{e}^{(-jk_1L)} & R_{12}\mathrm{e}^{(-jk_2L)} & R_{13}\mathrm{e}^{(+jk_1L)} & R_{14}\mathrm{e}^{(+jk_2L)} \\ R_{21}\mathrm{e}^{(-jk_1L)} & R_{22}\mathrm{e}^{(-jk_2L)} & R_{23}\mathrm{e}^{(+jk_1L)} & R_{24}\mathrm{e}^{(+jk_2L)} \end{bmatrix} \tag{28}$$

$$\begin{aligned} T_2(1, p) &= -Q_{55}(-jR_{1p}k_p - \eta R_{2p}) \\ T_2(2, p) &= jQ_{33}R_{2p}k_p - Q_{13}\eta R_{1p} \\ T_2(3, p) &= Q_{55}(-jR_{1p}k_p - \eta R_{2p})\mathrm{e}^{(-jk_pL)} \\ T_2(4, p) &= \{-jQ_{33}R_{2p}k_p + Q_{13}\eta R_{1p}\}\mathrm{e}^{(-jk_pL)} \end{aligned} \tag{29}$$

where p ranges from 1 to 4.

Thus, the dynamic stiffness matrix becomes

$$[\hat{\mathbf{K}}]_{nm} = [\mathbf{T}_2]_{nm}[\mathbf{T}_1]^{-1}_{nm} \tag{30}$$

which is of size 4×4 and has the parameters ω_n and η_m. This matrix represents the dynamics of an entire layer of any length L at frequency ω_n and horizontal wavenumber η_m. Consequently, this small matrix acts as a substitute for the global stiffness matrix of FE modeling, whose size, depending upon the thickness of the layer, is many orders larger than the size of the layer element stiffness matrix.

2.1 Prescription of boundary conditions

Essential boundary conditions are prescribed in the usual way as is done in FE methods, where the nodal displacements are simply arrested or released, depending upon the nature of the boundary conditions. The applied tractions are to be prescribed at the nodes. It is assumed that the loading function (for symmetric loading) can be written as

$$F(x, z, t) = \delta(z - z_j) \times \left(\sum_{m=1}^{M} a_m \cos(\eta_m x)\right)\left(\sum_{n=0}^{N-1} \hat{f}_n \mathrm{e}^{(-j\omega_n t)}\right) \tag{31}$$

where δ denotes the Dirac delta function, z_j is the Z coordinate of the point where the load is applied, and the z dependency is fixed by suitably choosing the node where the load is prescribed. No variation of load along the Z direction is allowed in this analysis. $\hat{f}_n$ are the Fourier transform coefficients of the time-dependent part of the load, which are computed by fast Fourier transform (FFT), and a_m are the Fourier series coefficients of the x-dependent part of the load.

There are two summations involved in the solution and the two associated windows, one in time T and the other in space X_L. The discrete frequencies ω_n and the discrete horizontal wavenumber η_m are related to these windows by the number of data N and M chosen in each summation, i.e.,

$$\begin{aligned} \omega_n &= 2n\pi/T = 2n\pi/(N\Delta t) \\ \eta_m &= 2(m-1)\pi/X_L = 2(m-1)\pi/(M\Delta x) \end{aligned} \tag{32}$$

where Δt and Δx are the temporal and spatial sample rates, respectively.

2.2 Determination of Lamb wave modes

As defined earlier, the Lamb waves are guided waves (Figure 1), propagating in a free plate, and the two lateral guiding surfaces are traction free. There are two main approaches to the analysis of the Lamb waves. The first one is the method of potentials. In this method, Helmhöltz decomposition of the displacement field is obtained and the governing equations are uncoupled and written in terms of the potentials. Solutions are sought for these potentials,

which contain four arbitrary constants. The displacement field and the stresses are expressed in terms of the potentials, and the imposition of the traction-free upper and lower surfaces generates the necessary condition for finding the unknown constants and the dispersion equation (see [19]). The advantage of this method is that the symmetric and antisymmetric modes can be isolated during formulation (Figure 1a and b). However, the method is applicable only to the isotropic waveguides.

The second approach is based on the PWT, which is discussed below in detail. In the SLE formulation, there are two summations in the solutions. The outer one is over the discrete frequencies and the inner one is over the discrete horizontal wavenumbers. Each partial wave of equation 24 satisfies the governing PDEs (equation 9), and the coefficients C_i, as a whole, satisfy any prescribed boundary conditions. As long as the prescribed natural boundary conditions are nonhomogeneous, no restriction is imposed upon the horizontal wavenumber η and this leads to a double summation solution of the displacement field. However, this is not the case for traction-free boundary conditions on the two surfaces, which are the necessary conditions for generating Lamb waves. The governing discrete equation for finite layer (equation 30), in this case, becomes

$$[\hat{\mathbf{K}}(\eta_m, \omega_n)]_{nm}\{\hat{\mathbf{u}}\}_{nm} = 0 \tag{33}$$

and we are interested in a nontrivial $\mathbf{u}$. Hence, the stiffness matrix $\hat{\mathbf{K}}$ must be singular, i.e., $\det(\hat{\mathbf{K}}(\eta_m, \omega_n)) = 0$, which gives the required relation between η_m and ω_n. Since, ω_n is made to vary independently, the above relation must be solved for η_m to render the stiffness matrix singular, i.e., η_m cannot vary independently. More precisely, for each value of ω_n, there is a set of values of horizontal wavenumber η_m (one for each mode), and for each value of ω_n and η_m, there are four vertical wavenumbers k_{nm}. The difference in this case is in the value of η_m, which is to be solved for, as opposed to its expression in equation 32, and M is the number of Lamb modes considered rather than Fourier modes. Now, for each set of $(\omega_n, \eta_m, k_{nml}), l = 1, \ldots, 4, \hat{\mathbf{K}}$ is singular and $C_l, l = 1, \ldots, 4$ is in the null space of $\hat{\mathbf{K}}$. Now, using equation 24, the total solution can be constructed. Following the normal practice, the traction-free boundary conditions (i.e., $\sigma_{zz}, \sigma_{xz} = 0$) are prescribed at $z = \mp h/2$. Using equation 16, the governing equation for C_i and η_m becomes

$$[\mathbf{W}_2(\eta_m, \omega_n)]\{\mathbf{C}\}_{nm} = \mathbf{0}, \ \ \mathbf{C} = \{C_1, C_2, C_3, C_4\} \tag{34}$$

where $\mathbf{W}_2$ is another form of the stiffness matrix $\hat{\mathbf{K}}$ and given by

$$\begin{aligned}
W_2(1, p) &= (Q_{110}R(1, p)\eta \\
&\quad - jQ_{130}R(2, p)k_p)\mathrm{e}^{jk_p h/2} \\
W_2(2, p) &= (Q_{110}R(1, p)\eta \\
&\quad - jQ_{130}R(2, p)k_p)\mathrm{e}^{-jk_p h/2} \\
W_2(3, p) &= Q_{550}(-R(1, p)k_p + jR(2, p)\eta)\mathrm{e}^{jk_p h/2} \\
W_2(4, p) &= Q_{550}(-R(1, p)k_p + jR(2, p)\eta)\mathrm{e}^{-jk_p h/2}
\end{aligned} \tag{35}$$

The dispersion relation is $\det\{\mathbf{W}_2\} = 0$, which yields $\eta_m(\omega_n)$ and the phase speed for Lamb waves c_{nm} are given by ω_n/η_m. Once the values of η_m are known for the desired number of modes, the elements of $\mathbf{C}_{nm}$ are obtained by the technique of singular-value decomposition as described in [17, 18]. Summing over all the Lamb modes, the solution for each frequency is obtained.

3 LAMB WAVE ANALYSIS BY BEAM/PLATE THEORIES

As stated before, to avoid the complexity of solving the classical elasticity equation, Lamb waves are also constructed from the beam and plate theories. Depending upon the order of the polynomial involved in the assumed displacement field, Lamb wave modes are captured. Generally, the modes are represented well in the low-frequency regime. However, at high frequencies, appearance of the new modes and deviation from the existing modes become too apparent and indicate the boundary of the applicability of beam/plate theories.

In this section, two basic beam theories are considered, namely, the classical or Euler–Bernoulli theory (EBT) and first-order shear deformation or the Timoshenko beam theory (TBT). It is assumed that the boundaries are given by $z = z_1$ and $z = z_2$ and

the wave is propagating in the X direction. Also the Y dimension is considerably low compared to the X dimension. The displacement field in terms of X displacement, U, and Z displacement, W, for each of the beam theories is as follows:

$$\text{EBT:} \quad U(x,z,t) = u(x,t) - z\partial W(x,t)/\partial x$$
$$W(x,z,t) = w(x,t) \tag{36}$$
$$\text{TBT:} \quad U(x,z,t) = u(x,t) - z\phi(x,t)$$
$$W(x,z,t) = w(x,t) \tag{37}$$

Assuming plane-wave solution for all the unknowns, i.e.,

$$\{u, w, \phi\}(x,t) = \{\hat{u}, \hat{w}, \hat{\phi}\}\exp(j(kx - \omega t)) \tag{38}$$

the algebraic form of the governing equation takes the familiar form

$$\mathbf{W}\{\mathbf{u}_0\} = \{\mathbf{0}\} \tag{39}$$

Imposing the condition of nontrivial solutions, i.e., $\det(\mathbf{W}) = 0$, generates the dispersion relation, i.e., the $k - \omega$ relation. Appealing to the definition of phase (ω/k) and group speeds $(\mathrm{d}\omega/\mathrm{d}k)$ the Lamb wave modes can be obtained. This method is relatively simple, as only polynomial equations are involved (owing to the assumed polynomial form of the displacement field). The number of Lamb wave modes captured depends upon the number of unknowns in each theory. For example, in EBT, the first symmetric (u) and antisymmetric (w) modes participate, whereas, in the TBT, the second antisymmetric mode is present along with the previously mentioned modes.

It is worth realizing that the governing equations considered in each case are devoid of any distributed load. Thus the top and bottom surfaces of the beam are stress free, which is exactly the condition of Lamb wave propagation. This is why the dispersion relation of the beam/plate generates the dispersion relation of the Lamb wave. To extend this idea further, if we are interested in finding the Lamb wave modes for a pipeline system, the method worth trying would be to compute the dispersion relation for shell structure, rather than directly working with the axi-symmetric 3-D elasticity solution.

4 PROPAGATION OF LAMB WAVE

First, the Lamb wave modes are computed by directly solving the 2-D elasticity equation following the procedure outlined in Section 2.2. For this purpose, AS4/3502 composite lamina with the following material properties are considered: $E_1 = 144.5\,\text{GPa}$, $E_{2,3} = 9.63\,\text{GPa}$, $\nu_{13,12} = 0.3$, $\nu_{23} = 0.02$, $G_{23,13,12} = 4.12\,\text{GPa}$, and $\rho = 1389\,\text{kg}\,\text{m}^{-3}$. An angle-ply lamina of 2-*mm* thickness is considered for the Lamb wave propagation study. Analysis is performed for three different fiber directions, 0°, 45°, and 90°.

The dispersion relation (relation between $c_\text{p} = \omega/\eta$ and ω) is usually left in the form of a determinant equal to zero because of its complexity. Hence, solution for this kind of implicit equation requires special treatment. The solution, in particular, is multivalued, unbounded, and complex (although the real part is of interest). One way to solve these equations is to appeal to the strategies of nonlinear optimization, which are based on nonlinear least-square methods. There are several choices of algorithms, like the trust-region dogleg method, the Gauss–Newton method with a line search, or a Levenberg-Marquardt method with line search. Here, the MATLAB function *f solve* is used, and the default option for medium-scale optimization—the trust-region dogleg method—is adopted, which is a variant of the Powell's dogleg method [21].

Apart from the choice of algorithm, there are other subtle issues in root capturing for the solution of wavenumbers, as the solutions are complicated in nature. Moreover, except for the first one or two modes, all the other roots escape to infinity at low frequency. For isotropic materials, these cutoff frequencies are known *a priori*. However, no expressions can be found for anisotropic materials and in most cases the modes (solutions) should be tracked backward, i.e., from the high-frequency to the low-frequency region. In general, two strategies are essential to capture all the modes within a given frequency band. Initially, the whole region should be scanned for different values of the initial guess, where the initial guess should remain constant for the whole range of frequency. This sweeping opens up all the modes in that region, although they are not completely traced. Subsequently, each individual mode should be followed to the end of the domain

or to a preset high value of the solution. For this case, the initial guess should be changed for each frequency to the solution of the previous frequency step. Also, sometimes it is necessary to reduce the frequency step in the vicinity of high gradient of the modes. Once the Lamb modes are generated, they are fed back into the frequency loop to produce the frequency-domain solution of the Lamb wave propagation, which through inverse Fourier transform produces the time-domain signal. As the Lamb modes are generated first, they need to be stored separately. To this end, data are collected from the generated modes at several discrete points in the whole range of frequency. Next, a cubic spline interpolation is performed for a very fine frequency step within the same range. While generating the time-domain data, the interpolation is performed from these finely graded data to get the phase speed (hence, η).

To get the time history of the propagating Lamb waves, a modulated pulse of 200-kHz center frequency is applied at one end of an infinite plate, and the X and Z velocities are measured for a propagating distance of $320h$, where h is the thickness of the plate. While studying the time-domain representation, the thickness of the plate is taken as 10 mm, which amounts to a frequency-thickness value of 2. This increased thickness is taken because, for this value, at least three modes are excited in all the cases, as shown by their respective dispersion curves (Figures 4, 8, and 11).

In all the dispersion plots of the Lamb modes, the abscissa is given in terms of frequency times the thickness. Figure 4 shows the first 10 Lamb modes for fiber angle 0°. As is seen in the figure, the first antisymmetric mode (mode 1) converges to a value of $1719\,\mathrm{m\,s^{-1}}$ in a range of 1 MHz-mm, where all the other modes converge. In analogy to the isotropic case, this is the velocity of the Rayleigh surface waves in 0° fiber laminae. The first symmetric mode (mode 2) starts above $10\,000\,\mathrm{m\,s^{-1}}$ and drops suddenly at around 1.3 MHz-mm to converge to $1719\,\mathrm{m\,s^{-1}}$, before which it has a fairly constant value. All the other higher order modes escape to infinity at various points in the frequency range. Also, the symmetric and the antisymmetric pair of each mode escape at almost the same frequency.

For the 0° ply stacking, Lamb modes are computed by the beam theories and compared with the previous solutions in Figure 5. As the figure suggests, EBT fails to predict mode 1, although it predicts mode 2

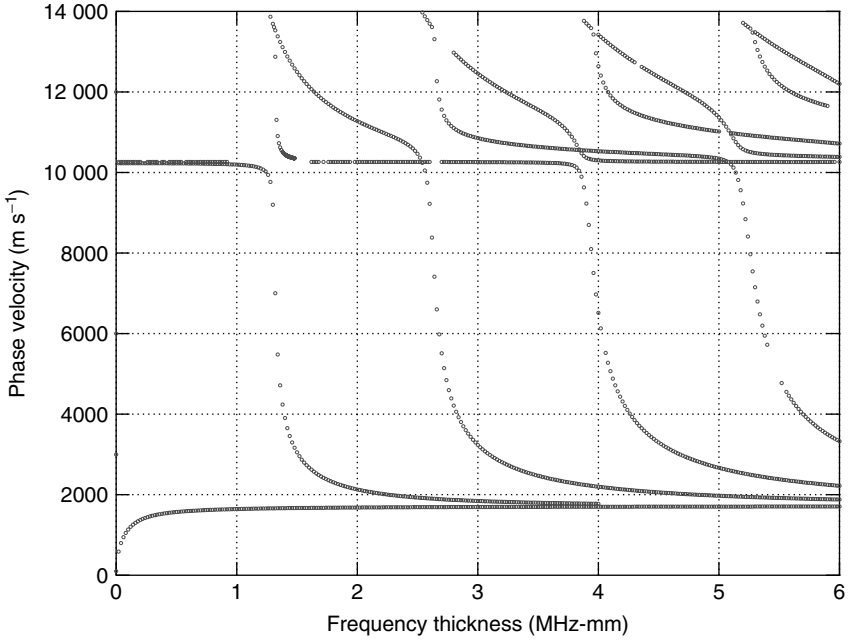


Figure 4. Lamb wave modes for 0° ply angle.

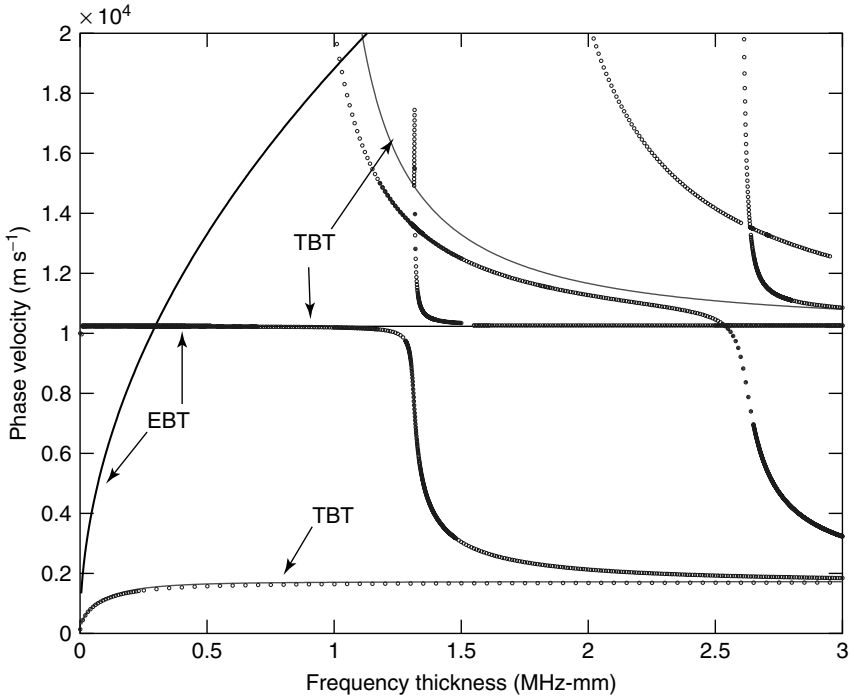


Figure 5. Lamb wave modes for 0° ply angle: comparison with beam theories.

accurately. On the other hand, TBT predicts mode 1 and 2 accurately but approximates the second antisymmetric mode. This prediction can be improved if a correction factor is introduced in the shear modulus. Thus, depending upon the order of the beam theory, different Lamb wave modes are captured.

Propagation of these modes are plotted in Figure 6 and 7 for the first three modes (a_0, s_0, and a_1), here referred to as modes 1, 2, and 3 respectively. The Z velocity history is plotted in Figure 6, whereas the X velocity history is plotted in Figure 7. The figures readily show the different propagating modes, each corresponding to one blob. It is to be noted that the wave propagation velocity is given by the group speed (and not the phase speed). Hence, Figure 4 cannot help us to predict the appearances of different modes. However, as Figure 6 and 7 suggest, mode 2 has a lower group speed than mode 1, and mode 3 has a group speed much higher than both modes 1 and 2. One difference in the $\dot{u}$ and $\dot{w}$ history can be observed. For $\dot{u}$, the higher mode generates velocity of comparatively lesser magnitude, whereas, for $\dot{w}$, the magnitude is the highest.

Next the fiber angle is changed to 45°, and the Lamb modes are plotted in Figure 8. Here, the phase velocity of mode 1 (a_0) is lower than the previous values for 0° (1690 m s^{-1}). Also, the initial phase velocity of mode 2 (s_0) has come down to less than 6000 m s^{-1} in comparison to its 0° counterpart (10 000 m s^{-1}). Further, the cutoff frequencies of all the higher modes are smaller compared to the previous case. Also, there are considerable differences in these cutoff frequencies for each pair of symmetric and antisymmetric modes, which is absent in the 0° case. Further, the number of modes is increased to 11 from 10 in the previous case. The time-domain representations of the propagating waves are shown in Figures 9 and 10. In this case, however, the second mode has higher group velocity than the first mode, and the third mode has the highest group speed.

Finally, the fiber angle is changed to 90°, and the resulting dispersion relation is plotted in Figure 11. The shifting of the modes to the left of the figure continues as the number of modes is increased to 12. Further, the first symmetric mode has come down to 2600 m s^{-1} and the first antisymmetric mode is

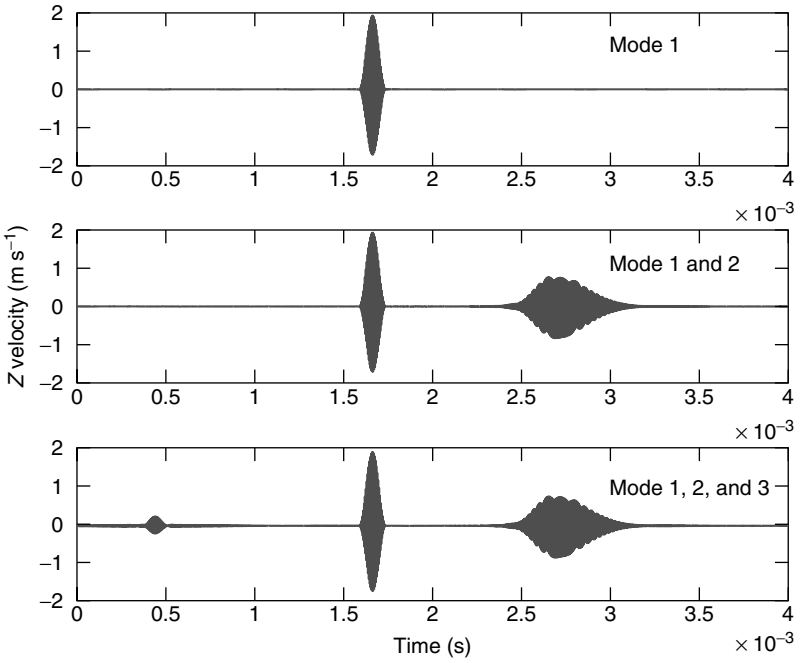


Figure 6. Lamb wave propagation for 0° ply angle, transverse mode, $L = 320h$.

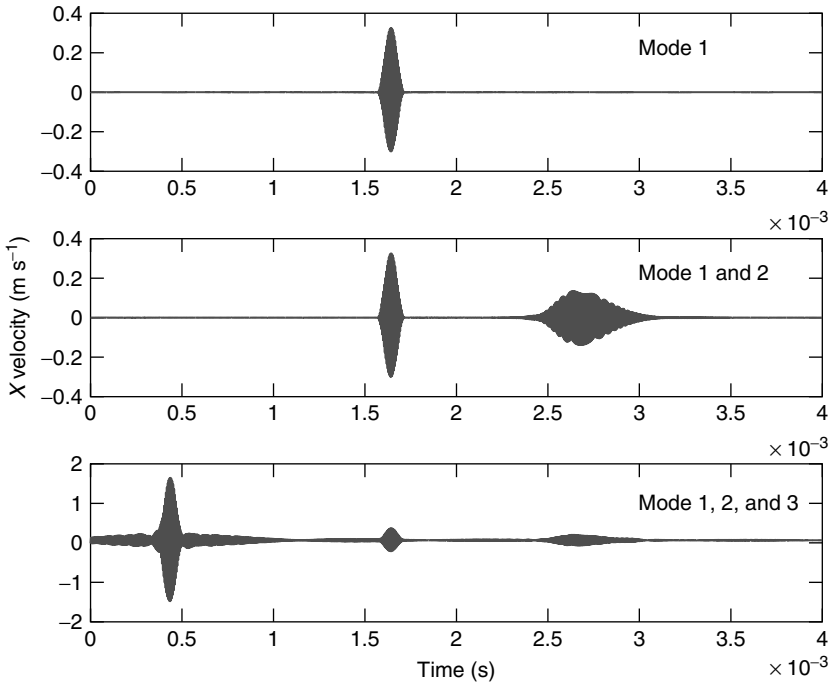


Figure 7. Lamb wave propagation for 0° ply angle, longitudinal mode, $L = 320h$.

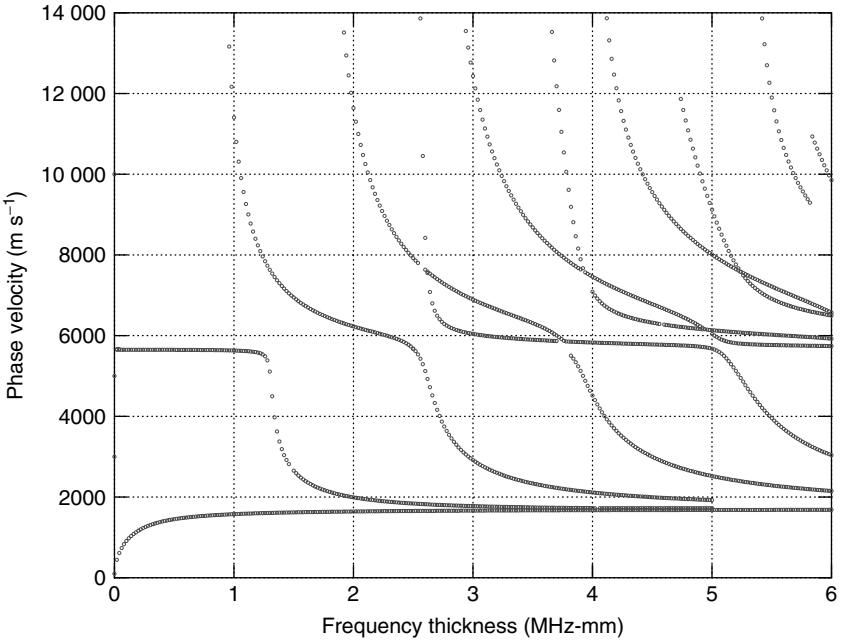


Figure 8. Lamb wave modes for 45° ply angle.

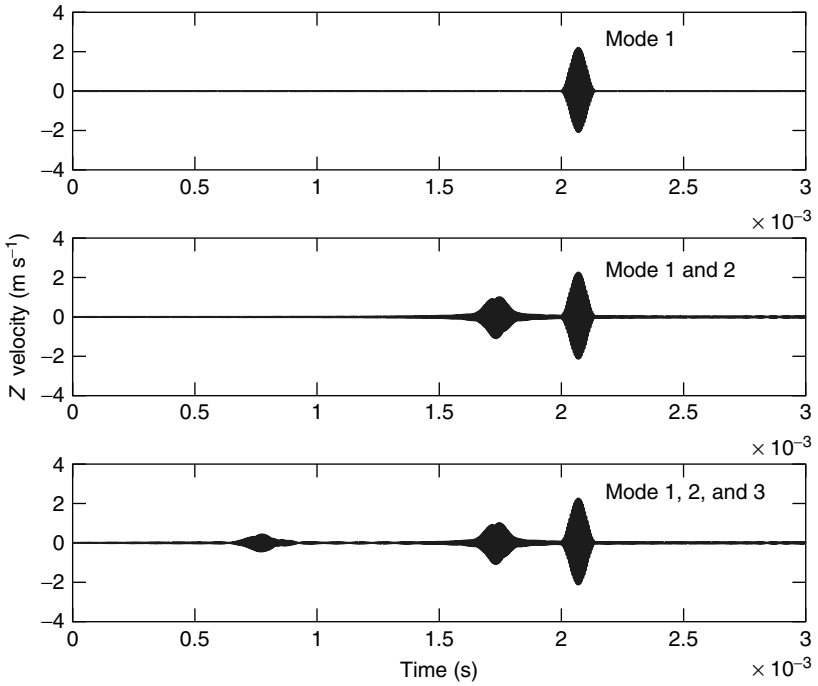


Figure 9. Lamb wave propagation for 45° ply angle, transverse mode, $L = 320h$.

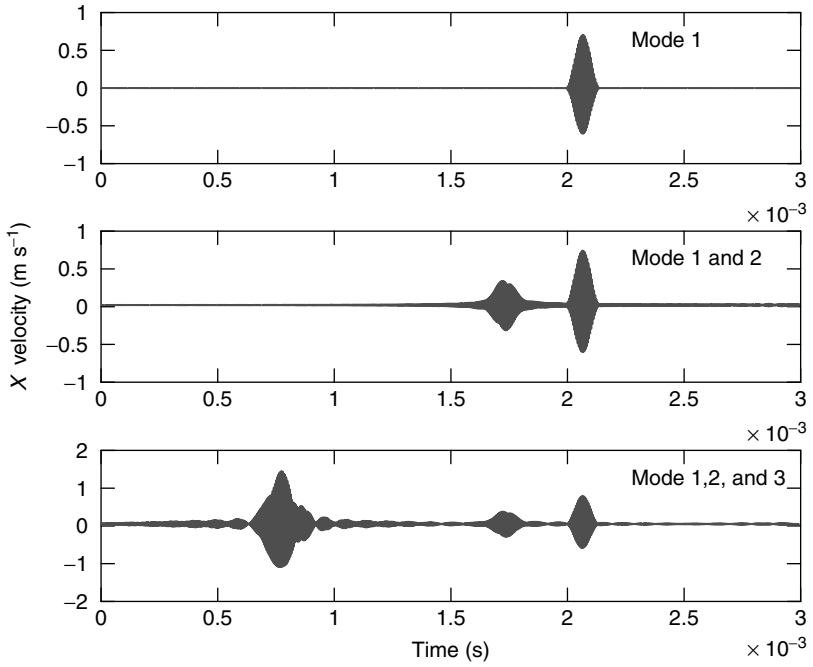


Figure 10. Lamb wave propagation for 45° ply angle, longitudinal mode, $L = 320h$.

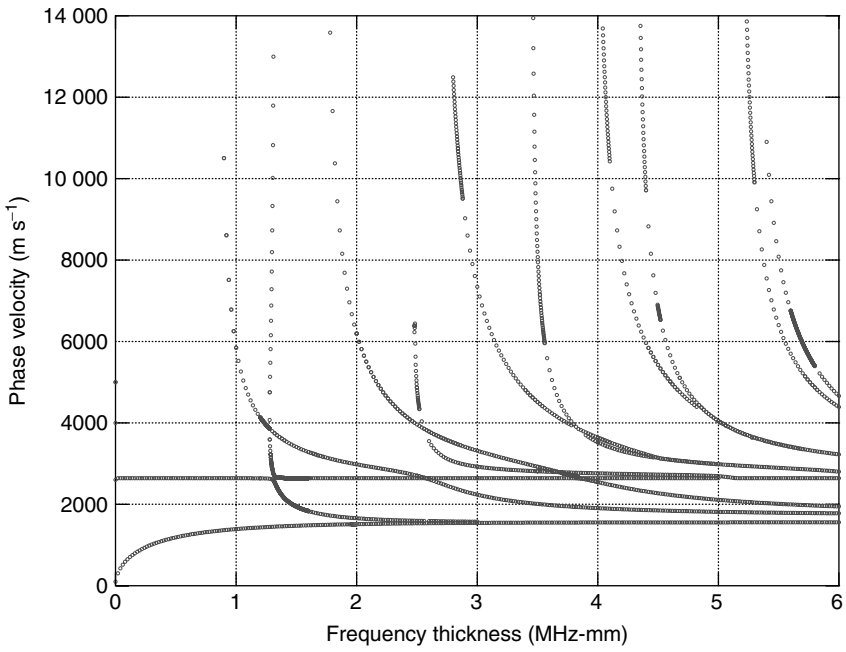


Figure 11. Lamb wave modes for 90° ply angle.

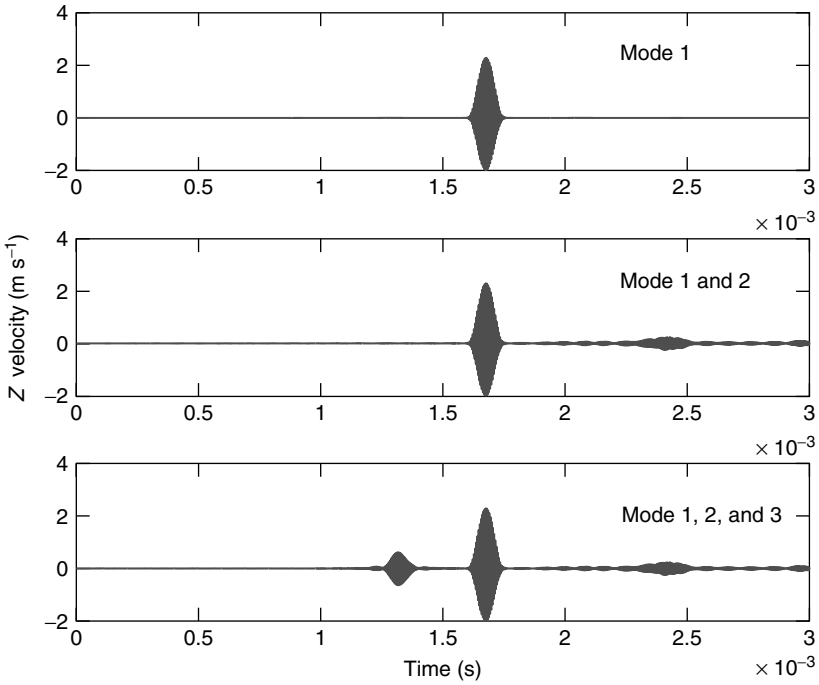


Figure 12. Lamb wave propagation for 90° ply angle, transverse mode, $L = 320h$.

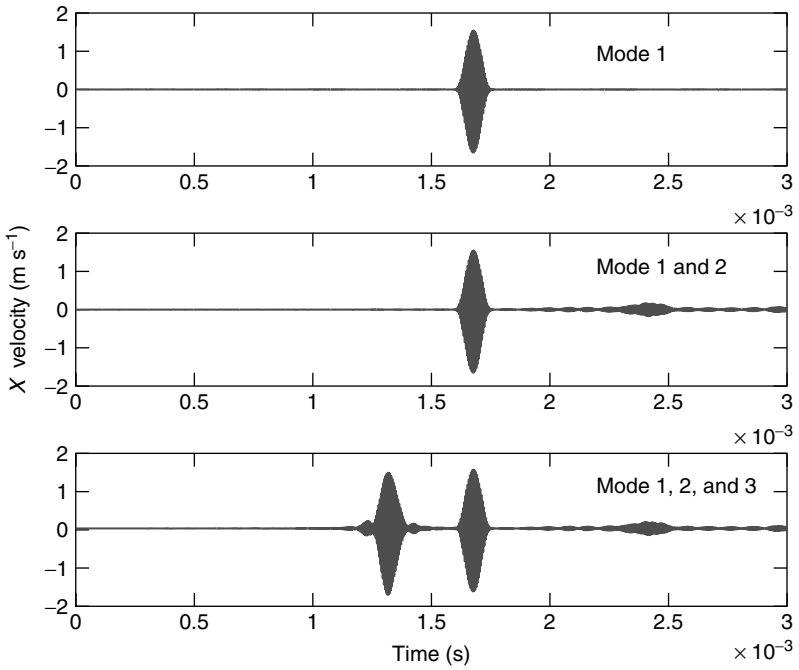


Figure 13. Lamb Wave propagation for 90° ply angle, longitudinal mode, $L = 320h$.

reduced to a converged speed of $1510\,\mathrm{m\,s^{-1}}$. For these modes, the propagating Lamb wave is plotted in Figures 12 and 13 for $\dot{u}$ and $\dot{w}$, respectively. As the figures suggest, mode 2 again has lower group speed compared to mode 1, and mode 3 has higher speed than both modes 1 and 2. However, the difference between the mode 3 group speed and mode 2 group speed is not very high as opposed to the previous cases.

5 CONCLUSION

A general procedure is presented to solve the 2-D elastodynamic equation for anisotropic media and the method of obtaining the Lamb wave modes as a special case. The solution is compared with the beam solutions and approximations involved in various theories are demonstrated. The complexities involved in obtaining the Lamb wave modes from 2-D equations and various strategies to overcome them are discussed in detail. The study of Lamb wave propagation reveals some important effects of ply angle. It is observed that the increase in the ply angle increases the number of active modes within a defined frequency range and reduces the cutoff frequencies and the phase speeds of a particular mode. Moreover, the difference in the cutoff frequencies increases with an increase in the ply angle.

REFERENCES

[1] Viktorov I. *Rayleigh and Lamb Waves*. Plenum Press: New York, 1967.

[2] Lamb H. On waves in an elastic plate. *Proceedings of the Royal Society of London, Series A* 1917 **93**:114–128.

[3] Nayfeh A, Chimenti D. Free wave propagation in plates of general anisotropic media. *Journal of Applied Mechanics* 1989 **56**(4):881–886.

[4] Nayfeh A. The general problem of elastic wave propagation in multilayered anisotropic media. *Journal of the Acoustical Society of America* 1991 **89**(4):1521–1531.

[5] Chimenti D. Guided waves in plates and their use in materials characterization. *Applied Mechanics Reviews* 1997 **50**(5):247–284.

[6] Veidt M, Liub T, Kitipornchai S. Modelling of Lamb waves in composite laminated plates excited by interdigital transducers. *NDT and E International* 2002 **35**(7):437–447.

[7] Moulin E, Assaad J, Delebarre C, Grondel S, Balageas D. Modeling of integrated Lamb waves generation systems using a coupled finite element normal mode expansion method. *Ultrasonics* 2000 **38**:522–526.

[8] Zhao G, Rose J. Boundary element modeling for defect characterization potential in a wave guide. *International Journal of Solids and Structures* 2003 **40**(11):2645–2658.

[9] Doyle J. *Wave Propagation in Structures*. Springer: New York, 1997.

[10] Gopalakrishnan S, Martin M, Doyle J. A matrix methodology for spectral analysis of wave propagation in multiple connected Timoshenko beam. *Journal of Sound and Vibration* 1992 **158**:11–24.

[11] Gopalakrishnan S, Doyle J. Wave propagation in connected waveguides of varying cross-sections. *Journal of Sound and Vibration* 1994 **175**(3): 347–363.

[12] Mahapatra DR, Gopalakrishnan S, Shankar T. Spectral-element-based solution for wave propagation analysis of multiply connected unsymmetric laminated composite beams. *Journal of Sound and Vibration* 2000 **237**(5):819–836.

[13] Mahapatra DR, Gopalakrishnan S, Shankar T. A spectral finite element model for analysis of axial-flexural-shear coupled wave propagation in laminated composite beams. *Composite Structures* 2003 **59**(1):67–88.

[14] Mahapatra DR, Gopalakrishnan S. A spectral finite element for analysis of wave propagation in uniform composite tubes. *Journal of Sound and Vibration* 2003 **268**(3):429–463.

[15] Chakraborty A, Gopalakrishnan S. Various numerical techniques for analysis of longitudinal wave propagation in inhomogeneous one-dimensional waveguides. *Acta Mechanica* 2003 **194**:1–27.

[16] Rizzi S, Doyle J. A spectral element approach to wave motion in layered solids. *Journal of Vibration and Acoustics* 1992 **114**:569–577.

[17] Chakraborty A, Gopalakrishnan S. A spectrally formulated plate element for wave propagation analysis in anisotropic material. *Computer Methods in Applied Mechanics and Engineering* 2005 **194**(42–44):4425–4446.

[18] Chakraborty A, Gopalakrishnan S. A spectrally formulated finite element for wave propagation analysis in layered composite media. *International*

Journal of Solids and Structures 2004 **41**(18–19): 5155–5183.

[19] Rose J. *Ultrasonic Waves in Solid Media*. Cambridge University Press, 1999.

[20] Rizzi S. *A Spectral Analysis Approach to Wave Propagation in Layered Solids*, Ph.D. Thesis. Purdue University: West Lafayette, IN, 1989.

[21] Powell M. A Fortran subroutine for solving systems of nonlinear algebraic equations. In *Numerical Methods for Nonlinear Algebraic Equations*, Robinowitz P (ed), Gordon & Breach, New York 1970; Chapter 7, pp. 115–161.

Chapter 50

Probabilistic Approaches to Sensor Layout Design, Data Processing, and Damage Detection

Sankaran Mahadevan, Xiaomo Jiang and Robert F. Guratzsch
Vanderbilt University, Nashville, TN, USA

1 INTRODUCTION

This article focuses on uncertainties in sensor performance and data and on methods to address these uncertainties and to improve the reliability of structural diagnosis. Although substantial research has been conducted in the past decade [1–17], accurate real-time structural health monitoring (SHM) is still a particularly challenging problem because of several important issues related to sensors: (i) sensitivity of damage detection to sensor configuration, (ii) presence of noise in the sensor data, (iii) incomplete (missing) nature of the sensor data, and (iv) variability in sensor data. This article presents probabilistic methodologies to address these issues.

The development of an effective SHM system for condition assessment depends on two important factors: sensing technology and the associated signal analysis and interpretation algorithm. Many sensor technologies (e.g., strain gauges, thermocouples, and accelerometers) are available for use of health monitoring and damage diagnosis [2, 18–20]. These traditional sensors are embedded in or attached to a structure at selected locations and are used to measure the dynamic response of the structure with the purpose of monitoring the structural integrity and performance. Recently, several new sensor technologies, such as fiber-optic sensors [21], remote wireless or noncontact sensing technologies [22], and active and passive ultrasonic sensing methods [23], have been developed for SHM. More recently, Blackshire and coworkers [24–26] developed and applied

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

surface-bonded piezoelectric transducers in experimental settings for aerospace materials.

Upon instrumentation of a structural system with sensors, an enormous amount of response data can be quickly collected. Information reliability is a critical issue in using the sensor data for damage detection. The data collected by sensors contain *imperfections*, in the form of imprecision (errors, fuzziness, etc.), incoherence (conflicting information), and uncertainty (partial knowledge, randomness, etc.), and could also be affected by sensor performance degradation. Yet, such data do contain a wealth of useful information that needs to be mined and interpreted to identify the structural condition. A critical challenge in SHM is how to deal with the data imperfection and extract useful information in support of decision making. The focus of this article is to develop probabilistic methodologies for sensor layout design and sensor data analysis that explicitly address imperfectly sensed vibration data and information uncertainty and maximize the probability of successfully identifying the structural damage.

Instrumentation of sensors in structures is expensive, and it is infeasible to instrument the entire structure. Therefore, it is desirable to develop an optimal sensor configuration that can minimize the cost, while maximizing the reliability of the SHM system. However, sensed data from real structures is nondeterministic owing to natural variability (described through stochastic or random processes) and uncertainties (due to lack of knowledge) in the inspection setup, measurement conditions, and different measured quantities of interest. Sensor configuration is optimal if, in addition to minimizing cost, it accounts for both sensor reliability and uncertainties in measurement conditions. Ignoring uncertainty and reliability issues may result in an expensive and ineffective SHM system design that adversely impacts the real-time condition assessment or damage detection after the data is collected. Therefore, one of the contributions of this article is to present a probabilistic analysis method to optimize sensor placement.

After instrumentation of a structural system using sensors, SHM requires analysis of the sensed data to detect changes in the global or local conditions of the structure. This is still a challenging problem owing to (i) the complicated nonlinear behavior of structural systems under natural hazards or even ambient effects (e.g., wind), (ii) the incomplete (missing outputs and/or immeasurable inputs) and noise-contaminated nature of the sensed data, and (iii) the uncertainty and variability in the sensed data. While significant advances are being made in innovative hardware technologies such as sensors and communication and measurement tools, equally important is the development of advanced data processing and damage assessment techniques to take advantage of the improvements in hardware. This article pursues a probabilistic data processing methodology (wavelets, dynamical neural network (NN), fuzzy clustering, and Bayesian statistics) to achieve this objective.

In general, damage detection is realized by comparing the sensor data collected from the real structural system with measurements from the healthy or undamaged structure, or against predictions by a trained model of the healthy structure. The existing damage detection techniques can be divided into two categories: vibration-based and wave-propagation-based methods. The vibration-based technique detects the damage-induced changes in the dynamic response of the entire structure under external excitations, in either time or frequency domains. Features are usually extracted from the time-series measurement or modal analysis of the structure for the detection purpose. Time-series analysis-based pattern recognition techniques, such as autoregressive and autoregressive with exogenous inputs prediction models, have been widely applied for health monitoring and damage detection of various structural systems [7, 27, 28]. Recently, modal analysis-based techniques have also been developed for detecting loose bolts in space operation vehicles [8, 29].

In the wave propagation approach, an impact or excitation is applied to a continuum structure, the structure generates dispersive waves, and changes in the wave propagation parameters (i.e., reflection and transmission coefficients, and wave travel times) are used to detect damage [30–35]. The wave propagation parameters are more sensitive to structural damage than mode shapes and frequencies in the classical vibration-based methods, and therefore have good potential for damage detection in a continuum structure.

Sensor data, however, always contain noise. A dilemma is that it is not possible to know with any measure of certainty whether and how much the measured data are corrupted by noise. If denoising

techniques are indiscriminately applied to data with very little noise, then useful information may be removed from the data, leading to erroneous results [36]. Therefore, it becomes an especially challenging problem to effectively denoise the sensor data and improve the accuracy of structural damage detection. Recently, Jiang and Mahadevan [17, 37] developed a Bayesian discrete wavelet packet transform (DWPT) denoising technique to investigate the effect of noise on the accuracy of structural system identification and damage detection. A comparative study has demonstrated that the proposed approach outperforms the existing denoising methods [37].

There are two different modeling approaches for structural damage detection: parametric and nonparametric methods. Unlike the parametric or physically based method, the system model in the nonparametric or nonphysically based approach does not represent any physical quantity directly, but it is trained to approximate a physical structure and predict its response. As such, the approach does not require complete measurements of the structural response. In addition, a nonlinear autoregressive moving-average approach is commonly used in the nonparametric methods for mapping the input–output relationship. Thus, the nonparametric approach can effectively represent a nonlinear structural system to address its nonlinear behavior. Nonparametric methods, however, usually do not perform damage isolation or quantification. Parametric methods may be further utilized to isolate the damage for health monitoring of a structural system.

Both parametric and nonparametric damage detection approaches need to account for the nondeterministic nature of sensor data and the modeling error of the analytical model. In addition to the uncertainty in the analytical or predictive model, sensor data always contain uncertainty resulting from the inspection setup, measurement conditions, and different measured quantities of interest. Ignoring the uncertainties may result in an erroneous decision making in the structural condition assessment after the data is collected. Recently, several researchers have pursued Bayesian probabilistic methods for structural damage detection and health monitoring, and addressed data uncertainty and modeling errors [38–41].

In the following sections, probabilistic methodologies are presented to address the aforementioned challenges. The proposed methodologies have been grouped into two categories and presented. In the first part, a probabilistic analysis approach is presented to address the issue of optimum layout design of sensor arrays of SHM systems under uncertainty. The sensor layout optimization combines probabilistic FEA, damage detection algorithms, and reliability-based optimization techniques. It is illustrated for the optimum sensor layout design for a thermal protection system (TPS) panel of a space operations vehicle (SOV), using FEA under transient mechanical and thermal loads, and uncertainty quantification techniques. The finite element model is validated with sensor data, accounting for uncertainties in experimental measurements and model predictions.

In the second part, a Bayesian wavelet probabilistic methodology is presented to address the noisy, incomplete, and uncertain characteristics of the sensor data. A Bayesian DWPT-based denoising approach is employed to perform data cleansing prior to damage detection. A nonparametric system identification method is applied to predict dynamic responses of the structure subjected to external excitations. Bayesian hypothesis testing is developed to assess the difference between the sensor data and the model prediction. The Bayesian assessment metric is treated as a random variable, and its probability density function (pdf) is constructed using Monte Carlo simulation to incorporate possible uncertainties. The method is validated using the sensor data collected from a five-story test steel frame.

2 SENSOR LAYOUT DESIGN

This section presents a probabilistic methodology for maximizing the reliability of damage detection by designing the locations of SHM system sensors. This includes the following steps: (i) probabilistic FEA, (ii) damage detection, and (iii) sensor placement optimization (SPO). The methodology is illustrated for layout design of sensor arrays for a TPS panel [42].

Several studies have investigated SPO during recent years. Hiramoto *et al.* [43] and Abdullah *et al.* [44] have addressed the need to place actuators in an optimal way to control the behavior of dynamic structures. Hiramoto *et al.* use the explicit solution of the algebraic Riccati equation to determine the optimal actuator placement, whereas Abdullah *et al.*

utilize genetic algorithms (GAs) to solve the optimization. GAs have also been employed to search for optimal locations of actuators in active vibration control [45–48].

Related more closely to SPO of SHM systems of next generation flight vehicles, Li *et al.* [49] proposed an algorithm that aims to identify modal frequencies and mode shapes best, as well as increase the signal-to-noise ratio. However, it is not shown that a sensor array that best identifies modal frequencies and mode shapes optimizes more traditional SHM performance measures such as the probability of correct classification. Gao and Rose [50, 51] define a probabilistic SPO approach, where a probabilistic damage detection model that describes detection probabilities over a confident monitoring region with radius R is defined for each sensor of a given sensor set. The entire effectiveness of the sensor network is then assumed to be the joint effect of all sensors as estimated at a point by the union probability of all sensors. A covariance matrix adaptation evolution strategy is used to search the decision variable domain. Difficulties arise in defining the probabilistic damage detection models and sources for uncertainty are not identified specifically. A similar SPO framework that addresses imprecise detection probabilities, as well as uncertain terrain properties, is proposed by Dhillon *et al.* [52]; Parker and Frazier [53] address SPO for SHM based on the concept of observability from the fields of dynamic systems theory and engineering design optimization. The technique uses a dynamic model of the structure in question to obtain performance measures with respect to damage detection and localization; however, it does not include uncertainty.

To the authors' best knowledge, the issues associated with SPO under uncertainty for SHM systems due to the spatial and temporal stochastic variability of material, geometric, and loading parameters have not been sufficiently addressed. The methodology presented in this article includes the stochastic nature of the model input parameters to perform a probabilistic FEA to derive the stochastic characteristics of the model outputs, which are used with appropriate damage detection algorithms to estimate probabilistic performance measures of a given sensor layout. Single-objective and multiobjective functions that use the probabilistic performance measures individually and in combination are considered.

2.1 Sensor placement optimization method

2.1.1 Probabilistic analysis

Structural model parameters such as distributed loads, and material and geometric properties, have temporal and spatial variability and cannot be expressed as single random variables, but must be represented as random processes and random fields [54]. Thus random process/field modeling is a key step in probabilistic FEA. Several methods, such as Karhunen–Loeve expansion [55], the Pierson–Moskowitz wave spectra [56], Sakamoto's polynomial chaos decomposition [57], and Shinozuka's Gaussian stochastic process formulation [58], have been used to simulate Gaussian random processes. Random field realizations can be used to simulate component thickness, material moduli, and spatially distributed loads such as thermal and pressure loading. Representing spatially or temporally distributed model inputs through discretized random process/field realizations allows the inclusion of their uncertainty in >FEAs.

Once the model input parameters are randomly generated via the discretization of random processes/fields and applied as inputs to FEA models, repeated simulations of the FEA at each realization are used to generate statistical and/or sensitivity information on model outputs at each possible sensor location. For practical purposes, each node of the FEA model may represent a possible sensor location.

2.1.2 Damage detection algorithms

Damage detection and location identification algorithms of a continuum structure include wavelet-based approaches [59], two-stage modal frequency analysis [60], and methods for eddy-current-based damage detection [61]. Property matrix updating, nonlinear response analysis, and damage detection using NNs are all methods used to manipulate the information gathered by SHM systems for decision making. However, most structural damage detection methods and algorithms found in the literature examine the changes in the measured structural vibration response and analyze the modal frequencies, mode shapes, and flexibility/stiffness coefficients of the structure [62]. This can be achieved either

actively or passively, where active damage detection algorithms use the system response to an auxiliary excitation and passive methodologies use only the responses to operational vibrations. A comprehensive review of the state-of-the-art damage detection and location identification algorithms is provided in Doebling *et al.* [62].

The probabilistic FEA in the previous section quantifies the statistics of the model outputs at all possible sensor locations. Additional analysis is needed to estimate the probability of correctly identifying the structural state of a component for a given sensor layout, x (i.e., $P(CD) = P(\text{correct structural classification}|\text{sensor layout } x)$). This can be accomplished via any appropriate diagnostics signal analysis procedure (i.e., damage detection algorithm). The signal analysis procedure employed in this study follows the general concepts of Duda *et al.* [63] and utilizes the feature extraction and state classification methodologies defined by DeSimio *et al.* [64]. Repeated analyses using different realizations of the random inputs to healthy and damaged structural FEA models and their respective state classification construct a classification matrix from which several performance measures of the given sensor layout can be estimated. Further details of such a procedure are given in Section 2.2.3.

2.1.3 Sensor placement optimization

The SPO problem can be generalized as "given a set of n candidate locations, find a locations, where $a \ll n$, which provide the best possible performance" [65] in damage detection. Studies by Padula [65–67] and Raich and Liszkai [68] have examined the problems and issues involved with SPO. Integer and combinatorial optimization methods have been used to optimize the placement of actuators for vibration control and noise attenuation. In addition, GAs for the optimization of sensor layouts [68] have been proposed. Multivariate stochastic approximation using simultaneous perturbation gradient approximation allows for the inclusion of noise in function evaluations or experimental measurements and has been shown to be efficient for large-dimensional problems [69].

An approach to SPO that includes uncertainty is to employ Snobfit (stable noisy optimization by branch and fit) [70], an optimization scheme that is designed for bound-constrained optimization of noisy objective functions, which are costly to evaluate because of their computational or experimental complexity. The major advantage of using Snobfit is that the algorithm does not require a previously determined set of candidate sensor locations, but rather considers the following optimization problem.

$$\begin{aligned} &\min \ f(x) \\ &s.t. \ x \in [u, v] \end{aligned} \tag{1}$$

where x is continuous and $[u, v]$ is a bounded box in $\Re^n$ with a nonempty interior [70].

The underlying idea of the optimization formulation is to identify a sensor layout, x, that will maximize some performance measure, such as the probability of correctly classifying the structure as either healthy or damaged (i.e., classifying the structure as healthy when it is indeed healthy and as damaged when it is damaged). Here x represents a vector containing the coordinates of the SHM sensors for a given layout. From the reliability analysis described above and a diagnostics signal analysis procedure, a performance measure such as $P(CD)$ is known. This allows the optimization formulation given in equation (1) to be utilized, where $f(x) = -P(CD)$ and $[u, v]$ are the geometric constraints on x given by the physical dimensions of the structure.

2.2 Illustrative example

The proposed methodology is implemented using the following example problem for illustrative purposes. The structure under consideration is a simplified TPS component that is described in detail in [71], and shown in Figure 1. The test article consists of a heat-resistant, 6.35 mm-thick aluminum plate, held in place via four 6.35 mm-diameter bolts located 12.7 mm. from the edges of the plate.

2.2.1 Structural simulation and model validation

The structure under consideration is modeled using the commercial finite element software ANSYS [72]. A portion of the finite element model is shown in Figure 2. Four-noded shell elements (Shell63) and two-noded spring elements (Combin14) are utilized to model the aluminum plate and bolted boundary

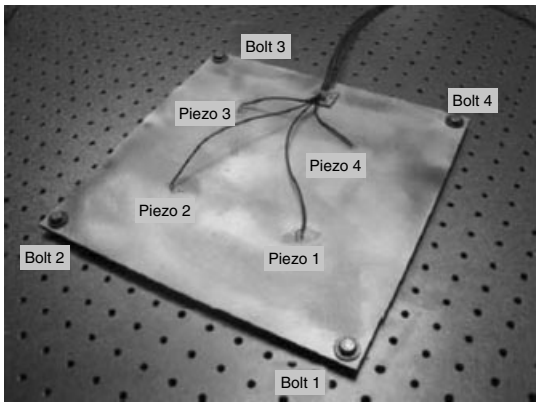


Figure 1. Experimental setup of TPS test article showing bolts and piezoelectric transducer placement [71].

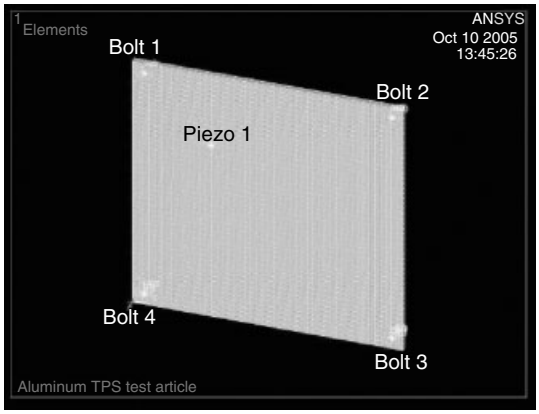


Figure 2. Finite element model of TPS component.

conditions. Approximately 3300 nodes and 2800 elements comprise the 19 836 degree of freedom (DOF) models. In Figure 2, the four points located near the corners of the plate simulate the bolted boundary conditions via 48 spring elements per bolt with varying stiffness coefficients (depending on which structural state the model simulates), while the point near the center of the upper left quadrant of the plate simulates the piezoelectric actuator. The analysis is transient and includes a dynamic mechanical load consisting of a sinusoidal frequency sweep, exciting the structure from 0 to 1500 Hz in approximately 2.0 s. This excitation represents the auxiliary input used with active damage detection algorithms. Owing to the high frequency of the excitation function, a mode superposition (MSP) transient analysis was used to evaluate the FEA model simulations. MSP analysis sums factored mode shapes obtained from a modal analysis to calculate the dynamic response [72].

2.2.2 Probabilistic FEA

In the current example, plate thickness, Young's modulus, Poison's ratio, and density are modeled as Gaussian random fields with independent but equal correlation structures along orthogonal axes. A two-dimensional stochastic process was generated for these model inputs using the spectral representation as defined in equation (1) via Shinozuka's formulation [73] and the Wiener–Khinchine relations [74]. The Gaussian random field $g_o(x_1, x_2, \phi)$ can be simulated by the following series as N_1 and N_2 approach infinity:

$$g_o(x_1, x_2, \phi) = 2 \sum_{k_1=1}^{N_1-1} \sum_{k_2=1}^{N_2-1} \left[\sqrt{S(\omega_{k_1}) S(\omega_{k_2}) \Delta\omega_1 \Delta\omega_2} \times \cos(\omega_{k_1} x_1 + \omega_{k_2} x_2 + \phi_{k_1,k_2}) \right] \quad (2)$$

where $\Delta\omega_i = \omega_{u_i}/N_i$ and $\omega_{k_i} = k_i \Delta\omega_i$, for $i = 1, 2$. Here ω_{u_i} is the upper cutoff frequency beyond which $S(\omega_{k_i})$ is considered to be zero. $S(\omega_{k_i})$ is the two-sided power spectral density (PSD) function of the random field in the i direction and ϕ_{k_1,k_2} an array containing the independent random phase angles uniformly distributed between 0 and 2π. N_i defines the number of terms to be included in the dual summation in the i direction. The random fields in this article use the following PSD functions: $S(\omega_{k_i}) = 1/4\sigma_i^2 b_i^3 \omega_{k_i}^2 \cdot \exp(-b_i \omega_{k_i})$ for $i = 1, 2$. Here σ_i is the standard deviation of the stochastic process in the i direction and b_i its corresponding "correlation distance".

For the random fields considered as FEA inputs to models of the test article, $b_1 = b_2 = 3$ and $\sigma_1 = \sigma_2 = 1$, where the magnitude of $g_o(x_1, x_2, \phi)$ is scaled after the test to match the mean and coefficient of variation (COV) of the random field to be simulated. $\omega_{u_1} = \omega_{u_2} = 5\pi$, while $N_1 = N_2 = 35$. Table 1 lists the mean and COV used for each of the random fields simulated with equation (2).

Temperature uncertainty was included as a random variable uniformly distributed between 65 and 75 °F.

Table 1. Mean and COV values used for random field simulation

	Panel thickness	Young's modulus	Poison's ratio	Density
Mean	6.24 mm	6.72E04 MPa	0.3	7.169 kg m^{-3}
COV	0.02	0.02	0.02	0.02

The following temperature effect model was constructed via a quadratic regression analysis of data by Kohavi and Provost [75]:

$$F(t) = (-1.151525 \times 10^{-6})t^2 + (2.75775 \times 10^{-5})t + 1.00067 \quad (3)$$

where $F(t)$ is a scale factor for Young's modulus and t is the plate temperature in degrees Fahrenheit.

Repeatedly executing deterministic FEAs using realizations of the model inputs provides data for statistical analysis of the model responses. For the example at hand, 500 simulations using 500 realizations of the random inputs were executed; 100 simulations of the healthy model, 100 simulations of the model damaged at bolt 1, 100 simulations of the model damaged at bold 2, and so on, where a damaged bolt refers to a bolt at 25% nominal torque (damage was simulated analytically by altering the stiffness constants of the spring elements surrounding each bolt location). These five sets of simulations and their corresponding response statistics are used for damage detection.

2.2.3 Damage detection and state classification

Figure 3 shows a typical sensor layout, where sensor location 1 is the point of input excitation and is stationary, while sensor locations 2, 3, and 4 are the points of sensing and are variable. Also shown in Figure 3 are the locations of the four bolts that hold the test structure in place and are the locations of fastener damage. The hatched areas in Figure 3 are regions where it is infeasible to place SHM sensors.

From the pool of simulation output of the probabilistic FEA consisting of temporal displacement data, an equivalenced von Mises stress for each possible sensor location is calculated by ANSYS 2004 [72].

$$\sigma_{vM} = \sqrt{\sigma_x^2 - \sigma_x\sigma_y + \sigma_y^2 + 3\tau_{xy}^2} \quad (4)$$

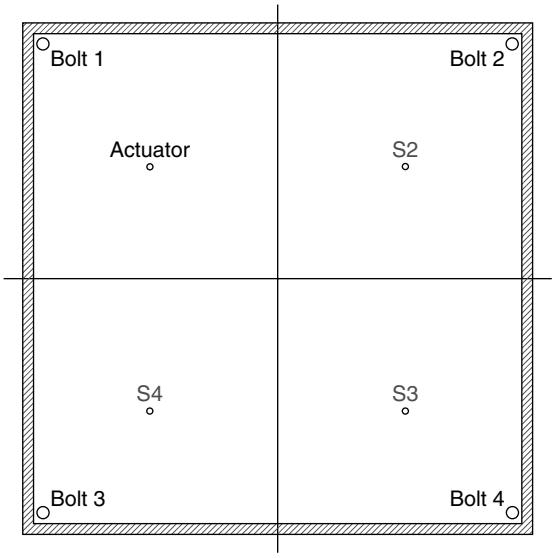


Figure 3. TPS test plate with typical sensor layout, actuator, and fastener damage locations.

where σ_x, σ_y, and τ_{xy} are the in-plane stress components, which are estimated from the displacement records of the four nearest neighboring nodes. Plane stress conditions are assumed.

From the estimated von Mises stress records at sensor locations S2, S3, and S4, a set of features is extracted. Features are characteristics unique to a signal generated under a given set of parameters. The set of features utilized for this example problem is based in the frequency domain and is extracted via the well-known Welch method [76, 77] from the PSDs of the signals. This damage detection algorithm is then applied to testing data, which consists of the second 50 simulations of each structural state. This yields a classification matrix corresponding to a given sensor layout, from which several performance measures may be estimated. Refer to Guratzsch and Mahadevan [42] for details about the feature extraction. A sample classification matrix is shown in Table 2.

Using the information contained in the classification matrix one can estimate several probabilistic performance measures of a given sensor layout, such as the probability of false alarm (type I error), the probability of missed detection (type II error), the probability of correct detection (accuracy), and the probability of misdetection (1-accuracy) [75]. P(false alarm) is defined as the likelihood that the

Table 2. Sample classification matrix for a given sensor layout

		Classified states				
		Damaged 1	Damaged 2	Damaged 3	Damaged 4	Healthy
True states	Damaged 1	1	21	0	0	1
	Damaged 2	78	98	0	2	0
	Damaged 3	0	1	91	7	1
	Damaged 4	0	10	0	89	1
	Healthy	0	11	0	1	88

damage detection algorithm classifies a healthy structure as damaged. P(missed detection) is the probability that the damage detection method classifies a damaged structure as healthy. Accuracy is measured via P(correct detection), which is defined as the probability that the damage detection method will classify a given structure correctly into its proper structural state (i.e., P(classify structure as $\omega_i \mid$ structural state is ω_i)). The compliment of P(correct detection) is P(misdetection). These probabilities can be used to evaluate a given sensor array [42]. The performance measures are expressed as follows:

$$P(\text{false alarm}) = \frac{\text{Sum of first four elements in row 5 of CM}}{\text{Sum of all elements in row 5 of CM}} = P(\text{type I}) \quad (5)$$

$$P(\text{missed detection}) = \frac{\text{Sum of first four elements column 5 of CM}}{\text{Sum of all elements in column 5 of CM}} = P(\text{type II}) \quad (6)$$

$$P(\text{correct detection}) = \frac{\text{Sum of diagonal elements of CM}}{\text{Sum of all elements of CM}} = P(CD) \quad (7)$$

$$P(\text{misdetection}) = 1 - P(\text{correct detection}) \quad (8)$$

Evaluating equations (5)–(8) for the classification matrix shown in Table 2 yields the following results: P(false alarm) = 0.12; P(missed detection) = 0.033; P(correct detection) = 0.89; and P(misdetection) = 0.11. As an example, the performance measures for the five randomly selected sensor layouts shown in Figure 4 are shown in Table 3.

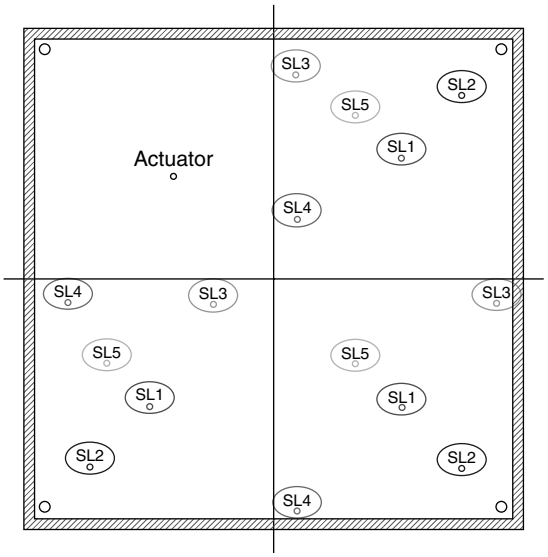


Figure 4. Five randomly selected sensor layouts.

Table 3. Performance measures corresponding to randomly selected sensor layouts of Figure 4

Sensor layout	$P(CD)$	P(type I)	P(type II)
SL1	0.916	0.12	0.0075
SL2	0.860	0.08	0.0050
SL3	0.866	0.12	0
SL4	0.894	0.11	0.0025
SL5	0.872	0.10	0.0075

2.2.4 Sensor placement optimization

The software package Snobfit [70], programmed in Matlab [78], is used to solve the optimization formulation given by equation (1) iteratively. Table 4

Table 4. Results: optimal sensor arrays corresponding to different objective functions

Objective function $f(y) =$	N_{ite}	N_{obj}	Optimal solution coordinates for sensors (in.)			E	Corresponding performance measures		
			S2	S3	S4		$P(CD)$	P(type I)	P(type II)
$-P(CD)$	71	258	8.75, 6.75	6.0, 3.5	3.5, 0.75	0.0104	0.944	0.13	0.0075
P(type I)	12	58	7.0, 8.5	11.73, 0.27	5.75, 1.25	0.0426	0.916	0.01	0
P(type II)	n/a	n/a	7.0, 8.5	11.73, 0.27	5.75, 1.25	n/a	0.916	0.01	0
$-0.5P(CD) + 0.25P$(type I) $+ 0.25P$(type II)	55	268	6.75, 8.75	11.60, 0.40	5.75, 1.25	0.0208	0.932	0.03	0.0025
$0.5(1 - P(CD)) + 0.25P$(type I) $+ 5.0P$(type II)	44	196	7.0, 8.5	11.73, 0.27	5.75, 1.25	0.0375	0.916	0.01	0

Units conversion for coordinates of sensors S2, S3 and S4: 1 in. = 25.4 mm

summarizes the optimization results, where N_{ite} is the number of Snobfit iterations, N_{obj} is the number of objective function evaluations, and E is the measure of accuracy of the quadratic model at the optimal solution as estimated by Snobfit. The coordinates given are with respect to the bottom left corner of the plate (Figure 3).

From Table 4, it can be concluded that although the solution varies for different objective functions, the optimal sensor arrays corresponding to objective functions 2 through 5 are nearly identical. Additionally, it was observed during Snobfit's iterations that the optimal solutions to objective functions 2 through 5 were robust and insensitive to small changes in the independent variables (i.e., shifting sensors S2, S3, and/or S4 by less than 0.25 in. in any direction, did not significantly alter the performance measures). However, the solution to the first objective function was very sensitive with respect to small changes in the independent variables (i.e., shifting sensors S2, S3, and/or S4 by less than 0.25 in. in any direction, significantly degraded the performance measures).

3 DENOISING SENSOR DATA

Suppose that the sensor data is contaminated by an additive white Gaussian noise, $\varepsilon(t_i)$, as follows:

$$f(t_i) = g(t_i) + \varepsilon(t_i) \quad i = 1, 2, \ldots, N \qquad (9)$$

where $f(t_i)$ and $g(t_i)$ represent the noisy and denoised sensor data, respectively. The noise term $\varepsilon(t_i) \sim N(0, \sigma^2)$ is a vector of independent, identically distributed (i.i.d.) errors with mean zero and variance σ^2. In the DWPT decomposition of the time series, the noise item $\varepsilon(t_i)$ in equation (9) is also resolved into a series of corresponding noise ε_{jk}. Then Bayesian hypothesis testing is applied to the DWPT coefficients. The goodness of fit of the denoised time series is measured by its signal-to-noise ratio (SNR). It is defined as the ratio of the sum of square of the signals to that of the noise. The thresholded coefficients are then used to reconstruct the denoised time series.

Figure 5 shows the Bayesian DWPT denoising approach done in five steps (identified by numbers 1–5 in Figure 5):

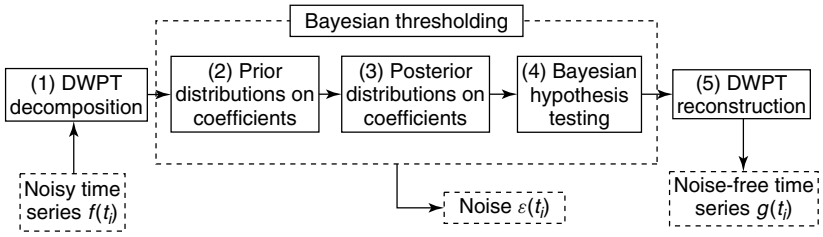


Figure 5. Bayesian discrete wavelet packet transform (DWPT) denoising procedure for time series.

1. The measured data is resolved into a series of wavelet coefficients by a DWPT decomposition.
2. Noninformative prior distributions are imposed on the decomposed coefficients.
3. Posterior distributions of the coefficients are derived based on Bayes theorem.
4. The noise is removed from the coefficients through a Bayesian hypothesis testing.
5. The denoised data is reconstructed through an inverse wavelet transform.

Refer to Jiang *et al.* [37] for the details of the Bayesian DWPT denoising method.

This denoising approach has been demonstrated to outperform the conventional wavelet-based and translation-invariant soft thresholding methods [37]. It has the following three features that are advantageous in structural damage detection. First, it avoids the subjective selection of the threshold required in conventional wavelet-based denoising methods. Second, it can address the uncertainty of noise effectively through Bayesian hypothesis testing. Third, it provides more accurate data denoising, particularly for the sensor data even with low noise, due to the adroit integration of the discrete wavelet packets decomposition and Bayesian hypothesis testing.

4 DAMAGE DETECTION USING INCOMPLETE DATA

Nonparametric system identification is used in this section for damage detection using incomplete data. The nonparametric dynamic fuzzy wavelet neural network (WNN) model developed by Adeli and Jiang [79] is used in this research for nonparametric system identification of structures to generate model predictions. The general dynamic input–output mapping in the model is

$$\hat{y}_i = \sum_{k=1}^{M} w_k \sum_{j=1}^{D} \varphi\left(\frac{X_{ij} - c_{kj}}{a_{kj}}\right) + \sum_{j}^{D} b_j X_{ij} + d$$
$$i = 1, \ldots, N, a \in \Re, \varphi(.) \in L^2(\Re) \quad (10)$$

where $\varphi(.)$ is the nonorthogonal Mexican hat wavelet function; X_{ij} is the jth value in the ith input vector, X_i, and c_{kj} is the jth value in the kth cluster of the multidimensional input vector obtained using the fuzzy C-means clustering approach [80]. The parameter D is the input dimension or the size of the input vector in nonlinear autoregressive moving average with exogenous inputs (NARMAX) approach [81]. The parameter M is the number of wavelets, which is also equal to the number of the fuzzy clusters as well as the number of wavelet nodes used in the WNN model. The parameters $a_{kj} \neq 0$ denote the frequency (or scale) corresponding to the multidimensional input vector; w_k represents the kth wavelet coefficient linking the hidden node to the output; b_j is the weight of the link of the jth input to the output; d is a bias term, and $\Re$ is the set of real numbers. The parameter N is the number of input vectors.

Unlike conventional NN models such as backpropagation NN, the fuzzy WNN model is a dynamic NN that preserves the time sequence of the input vectors and memorizes the past of the time-series data. The model is based on the integration of multiple paradigms including chaos theory (based on nonlinear dynamics theory), wavelets (a signal processing method), and two complementary soft computing methods, i.e., fuzzy logic and NNs. It has been demonstrated to provide more accurate nonlinear approximation than the conventional NN [79].

Note that the fuzzy WNN model is trained using the data collected from a healthy structure subjected to a low-level excitation, using the adaptive Levenberg–Marquardt least squares (LM-LS) algorithm [16, 17, 37, 82]. Therefore, the trained model represents a structural system without any damage such that the predicted outputs should represent the structural responses without any damage as well. The trained model is then used to predict dynamic responses of the structure with unknown conditions under a high-level excitation. Both sensor data and model predictions under the same excitation are used to assess the health status of the structure based on the Bayes metric described in Section 5.

5 DAMAGE DETECTION UNDER UNCERTAINTY

5.1 Bayes metric

Structural damage evaluation involves comparing the model prediction (based on original healthy condition

of a structure) with the experimental result (representing current condition of the structure). When the outputs predicted from the model are compared with the sensor data (under some input excitation), the difference between them will be used to detect the structural damage. The relative root mean square (rrms) error has been widely used to measure structural damage [5, 83, 84]. It is customary to assume that structural damage has occurred when the rrms error exceeds a predefined threshold level obtained by trial and error (for example, 0.6). However, the rrms error method is not always accurate because two other sources also contribute to this error: (i) training of the model to approximate the structural properties and (ii) sensor data, which are imperfect and contain measurement noise. As such, the rrms error method may be ineffective for structural damage detection when imperfection and uncertainty in the data are considered.

In this article, Bayesian hypothesis testing is pursued as a quantitative measure to detect the structural damage based on the denoised sensor data. Let y_{true} be the (usually unknown) true response of the original healthy structure under the input excitation being considered at present, y_{exp} the sensed structural response under the current (may be unknown) condition, and y_{pred} the prediction output representing structural responses under healthy condition. (All three quantities—y_{true}, y_{exp}, and y_{pred}—are under the same input excitation being used for damage detection). Within the context of binary hypothesis testing, consider two hypotheses H_0 and H_1. The point null hypothesis (H_0: $y_{\text{exp}} = y_{\text{pred}}$) indicates that the structure is healthy. The alternative hypothesis (H_1: $y_{\text{exp}} \neq y_{\text{pred}}$) indicates that the structure is damaged. Their prior probabilities of acceptance are denoted by $\pi_0 = Pr(H_0)$ and $\pi_1 = Pr(H_1)$. On the basis of Bayes theorem, the relative posterior probabilities of the two hypotheses are obtained as

$$\frac{Pr(H_0|\text{data})}{Pr(H_1|\text{data})} = \left[\frac{Pr(\text{data}|H_0)}{Pr(\text{data}|H_1)}\right]\left[\frac{Pr(H_0)}{Pr(H_1)}\right] \quad (11)$$

The term in the first set of square brackets on the right hand side of equation (11) is referred to as the "Bayes factor" [85]. Thus, the Bayes factor is the ratio of probability of observing the data given the null hypothesis (i.e., structure is healthy) to the probability of observing the data given the alternate hypothesis (i.e., structure is damaged), obtained as follows [86]:

$$B(y_{\text{pred}}) = \frac{P(\text{data}|H_0: y_{\text{exp}} = y_{\text{pred}})}{P(\text{data}|H_1: y_{\text{exp}} \neq y_{\text{pred}})} \quad (12)$$

Mahadevan and Rebba [87] showed that, under practical conditions, the Bayes factor simply becomes the ratio of posterior to prior pdfs of the structural response at $y_{\text{exp}} = y_{\text{pred}}$ (Figure 6):

$$B(y_{\text{pred}}) = \frac{h_2}{h_1} = \left.\frac{f(y_{\text{pred}}|\text{data})}{f(y_{\text{pred}})}\right|_{y_{\text{exp}}=y_{\text{pred}}} \quad (13)$$

When $B(y_{\text{pred}})$ is greater than unity, the sensor data is said to favor the null hypothesis. However, if $B(y_{\text{pred}}) < 1$, it may be inferred that the data supports the alternative hypothesis that the structure is damaged. Since $B(y_{\text{pred}})$ is nonnegative, the value of $B(y_{\text{pred}})$ is often converted into the logarithm scale for the convenience of comparison among a large range of values (i.e., $b_{01} = \ln[B(y_{\text{pred}})]$, where ln[.] is a natural logarithm operator with a basis of e). As such, a positive value of b_{01} indicates that the structure is healthy (i.e., accepting H_0) while a negative value indicates that the structure is damaged (i.e., rejecting H_0). Kass and Raftery [86] suggest interpreting b_{01} between 0 and 1 as weak evidence in favor of H_0, between 3 and 5 as strong evidence, and $b_{01} > 5$ as very strong evidence. Negative b_{01} of the same magnitude is said to favor H_1 by the same amount. A key difference from the rrms metric is that the Bayes-factor metric explicitly accounts for the uncertainty in sensor data during the computation of the posterior pdf [87].

Assume $\mathbf{y}_{\text{exp}} = \{y_{1,\text{exp}}, y_{2,\text{exp}}, \ldots, y_{n,\text{exp}}\}$ and $\mathbf{y}_{\text{pred}} = \{y_{1,\text{pred}}, y_{2,\text{pred}}, \ldots, y_{n,\text{pred}}\}$ to be n samples of

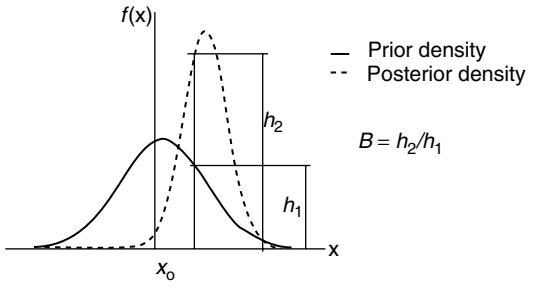


Figure 6. Bayes-factor metric B for damage detection.

sensor data and model predictions, respectively. Let $e_i = y_{i\,\mathrm{exp}} - y_{i\mathrm{pred}}$ represent the difference between the ith experimental data and the ith model prediction, and $\varepsilon_{\mathrm{obs}} = \{e_1, e_2, \ldots, e_n\}$ represent n values of the error, which is usually assumed to follow a normal distribution $N(\mu, \sigma^2)$ with σ known in this study (estimated using the data set $\boldsymbol{\varepsilon}_{\mathrm{obs}}$). Here we do not know the distribution of the difference *a priori*, so we assume Gaussian as an initial guess, and then do a Bayesian update. The damage detection problem becomes testing $H_0: \mu = 0$ versus $H_1: \mu \neq 0$ with $\mu|H_1 \sim N(\rho, \tau^2)$, in which ρ and τ are two parameters of the prior density of μ under the alternative hypothesis (denoted by $f(\mu|H_1)$). If no information on $f(\mu|H_1)$ is available, the parameters $\rho = 0$ and $\tau^2 = \sigma^2$ are suggested by Migon and Gamerman [88]. This selection assumes that the amount of information in the prior is equal to that in the observation, which is consistent with the Fisher information-based method [86].

Note that H_0 becomes a simple hypothesis with $f(\mathrm{data}|H_0) = f(\boldsymbol{\varepsilon}_{\mathrm{obs}})$, which is referred to as the marginal likelihood of H_0 given data $\boldsymbol{\varepsilon}_{\mathrm{obs}}$. Using Bayes theorem, Jiang and Mahadevan [17] explicitly derived an expression to calculate the Bayes factor in the logarithm scale as follows:

$$b_{01} = \ln[B(\mathbf{y}_{\mathrm{pred}})] = \frac{1}{2}\ln\left(\frac{n\tau^2 + \sigma^2}{\sigma^2}\right) + \frac{n}{2}\left[\frac{(\overline{\varepsilon}_{\mathrm{obs}} - \rho)^2}{n\tau^2 + \sigma^2} - \frac{\overline{\varepsilon}_{\mathrm{obs}}^2}{\sigma^2}\right] \quad (14)$$

Thus, a value of b larger than 0 indicates that the error data $\boldsymbol{\varepsilon}_{\mathrm{obs}}$ are judged to support H_0 (i.e., the structure is healthy). Otherwise, the error data $\boldsymbol{\varepsilon}_{\mathrm{obs}}$ are judged to reject H_0 (i.e., the structure is damaged).

5.2 Confidence assessment

On the basis of the Bayes factor calculated using equation (14), quantitative assessment of confidence in the structural damage detection result may be done in two ways: deterministic and stochastic. In the deterministic context, given a set of sensed data $\mathbf{y}_{\mathrm{exp}}$ and model prediction $\mathbf{y}_{\mathrm{pred}}$, the likelihood ratio in the logarithmic scale is calculated using equation (14). The uncertainties in data and prediction are not considered. Thus, the Bayesian measure of evidence that the structure is healthy may be quantified by the posterior probability of the null hypothesis $Pr(H_0|\mathrm{data})$, denoted by λ. Using the Bayes theorem, the value of λ can be derived:

$$\lambda = P(H_0|\mathbf{y}_{\mathrm{exp}}) = \frac{B(\mathbf{y}_{\mathrm{pred}})P(H_0)}{P(H_1) + B(\mathbf{y}_{\mathrm{pred}})P(H_0)} \quad (15)$$

Equation (15) is used to quantify the confidence in accepting the null hypothesis (i.e., the structure is healthy) based on the specific experimental data and prediction output. Before conducting experiments, it is usually assumed that $P(H_0) = P(H_1) = 0.5$ owing to the absence of any prior knowledge about the two hypotheses. In that case, equation (15) is simplified as $\lambda = B(\mathbf{y}_{\mathrm{pred}})/[1 + B(\mathbf{y}_{\mathrm{pred}})]$, and $B(\mathbf{y}_{\mathrm{pred}}) \to 0$ indicates 0% confidence in accepting the null hypothesis, and $B(\mathbf{y}_{\mathrm{pred}}) \to \infty$ indicates 100% confidence.

In the stochastic context, the Bayes factor, $B(y_{\mathrm{pred}})$, is treated as a random variable to consider the uncertainties in both experimental data and model prediction. Given the pdf of $\boldsymbol{\varepsilon}_{\mathrm{obs}}$, the M sets of mean ($\overline{\varepsilon}$) and sum of squared error (SSE) (s) values are created by sampling m error data using Monte Carlo simulation technique ($m = 5000$ and $M = 10\,000$ are taken in the example presented in this section). The resulting M b_{01} values by equation (14) are used to construct its density distribution function. As such, the probability of the event that the structure is healthy (i.e., accepting the null hypothesis) can be obtained by using the pdf or estimated by finding the proportion of $b_{01} > 0$, i.e., $\gamma = Pr(b_{01} > 0)$. Thus, the stochastic approach directly provides a quantitative probability of damage.

6 ILLUSTRATIVE EXAMPLE

As an illustrative example, the sensor data used in the numerical implementation are collected from a test five-story steel frame. The proposed probabilistic methodology has also been illustrated by Jiang and Mahadevan [17] using the sensor data collected from a 38-story concrete building test model. All sensor data in this example are denoised using the Bayesian DWPT denoising method. A three-level DWPT decomposition is found adequate for signal denoising using the proposed denoising approach

[37]. Structural damage evaluation is conducted on the test structure using both original and denoised data, using the proposed probabilistic evaluation method.

6.1 Problem description

This structure was tested at the National Center for Research on Earthquake Engineering in Taiwan [5]. It is a 3-m long, 2-m wide, and 6.5-m high steel frame (Figure 7a). This test frame was subjected to five different levels of excitations, all derived from an original Kobe earthquake acceleration time history, that is, the acceleration amplitudes at 20, 32, 40, 52, and 60% of the original Kobe time history amplitude. These excitations are denoted as Kobe1, Kobe2, Kobe3, Kobe4, and Kobe5, respectively. Acceleration responses were measured using acceleration sensors at the four corners of each floor in two horizontal directions (x and y directions in Figure 7a) over a period of 25 s at increments of 0.001 s. Thus, around 25 000 output data points are recorded in a single test. All acceleration response data are normalized to the gravity acceleration (g) to improve computational efficiency in the structural system identification. As an example, Figure 7(b) shows the acceleration responses at the first, second, and third floors of the test frame under Kobe1 earthquake.

In this example, only the acceleration response data at the first, second, and third floors in the x direction (Figure 7a) are used as both training and testing data sets. The dynamic fuzzy WNN model is constructed using the measured input–output data under the lower Kobe1 earthquake and then trained using the adaptive LM-LS algorithm, as described by Adeli and Jiang [79]. The trained model is then used to represent the structural system and predict the acceleration responses at the second floor of the test frame under the other four stronger earthquakes, i.e., Kobe2, Kobe3, Kobe4, and Kobe5.

6.2 Damage detection

All measured data are denoised using the proposed Bayesian DWPT approach. The first 10 000 points of the pairs of original and denoised experimental data obtained at Kobe1 earthquake are used to train the model, one set at a time, resulting in two trained models. Each trained model is tested using the remaining 15 000 points of the denoised data and the original experimental data. Furthermore, each trained model is used to predict the original and the denoised acceleration responses under four different levels of excitations (Kobe2 to Kobe5). As such, 10 different sets of system identification results are produced, including the two testing cases. The rrms errors and b_{01} values between the measured and predicted output are computed for all 10 test cases and shown in Table 5. The 10 corresponding probabilities in accepting the null hypothesis are calculated using equation (15) and also shown in Table 5.

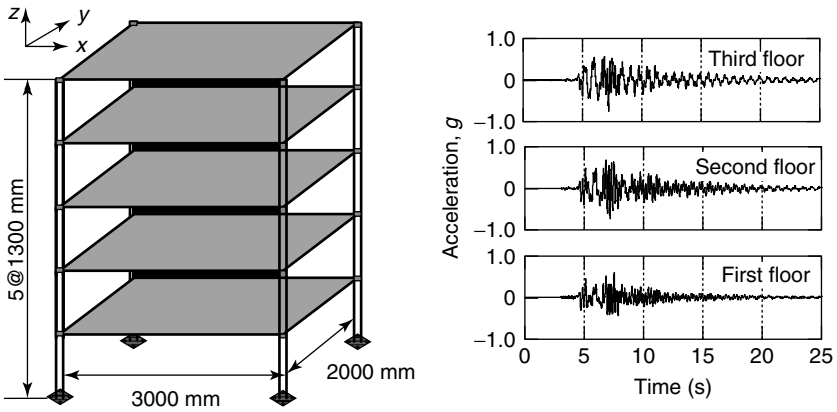


Figure 7. Five-story test steel frame and its acceleration responses under Kobe1. (a) Test frame structure and (b) sensed data under Kobe4 earthquake.

Table 5. Detection results at the second floor of a five-story frame in terms of rrms, b_{01}, and γ metrics

Excitation		Kobe1	Kobe2	Kobe3	Kobe4	Kobe5
rrms	Original	0.12	0.233	0.250	0.207	0.199
	Denoised	0.06	0.150	0.162	0.157	0.181
b_{01}	Original	2.757	2.438	−0.765	−5.469	−5.073
	Denoised	3.167	1.196	−3.917	−11.076	−9.944
λ (%)	Original	94.0	92.0	31.8	0.4	0.6
	Denoised	96.0	76.8	2.0	0.0	0.0
γ (%)	Original	91.8	82.6	40.9	8.6	9.7
	Denoised	94.5	65.8	14.7	0.6	1.4

6.2.1 *Deterministic case*

A difference between the prediction output and the sensor data is used as an indication of the structural damage. Assuming a damage threshold $\varepsilon = 0.15$, two observations are made from the results. First, all rrms values obtained using four sets of noisy data (Kobe2 to Kobe5 in Table 5) are larger than ε, implying that the frame structure is damaged under all four scenarios. However, when the original data is denoised and then used for structural system identification and damage evaluation, the rrms values of only the last three scenarios (Kobe3 to Kobe5) are larger than ε, implying that the structure is damaged under only the three scenarios. The experimental observations described by Hung *et al.* [5] demonstrate that there exists distinguishable damage or inelastic structural behavior under only the strong earthquake Kobe3 to Kobe5. Therefore, it is reasonably concluded that the noise in the sensed data adversely affects the detection results. Second, the damage detection results obtained using the Bayesian metric have further indicated that the Bayesian data denoising approach followed in this study improves the accuracy of detection results significantly.

As an example, Figure 8 shows the comparison of detection results using the original and denoised acceleration response data of the second floor subjected to Kobe4 earthquake. Figure 8(a) shows the comparison result in the time domain, while Figure 8(b) shows the comparison in the frequency domain. The PSDs are used in Figure 8(b) because the difference between noisy and denoised response data cannot be identified visually in the time-series plots (Figure 8a). The PSD of every time series is calculated through fast Fourier transform with the length 1024. Refer to Stoica and Moses [89] for details of spectral analysis. It is observed that the difference of PSDs between the original data and the denoised data is significant in the frequency domain.

6.2.2 *Stochastic case*

Using the statistics of the error data (i.e., its mean and variance) for various scenarios, the value of $\overline{\varepsilon}_{\text{obs}}$

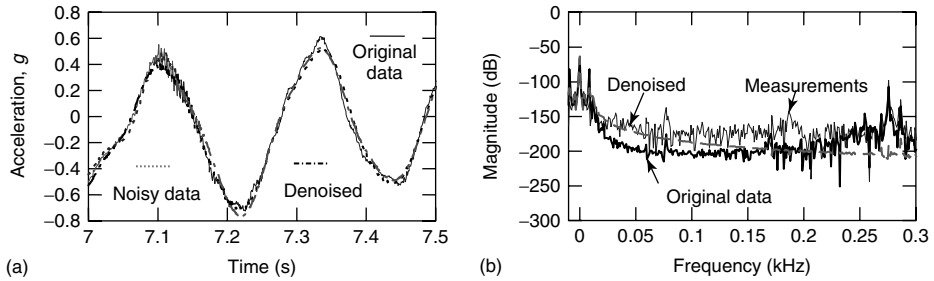


Figure 8. Comparison of detection results using original and denoised acceleration response data under Kobe4. (a) Time domain and (b) power spectral densities.

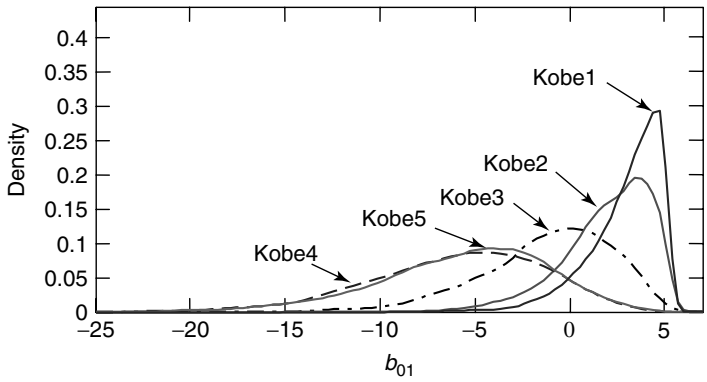


Figure 9. Simulated results of $\gamma = \Pr(b_{01} > 0)$ for various excitation levels in five-story frame: the densities are kernel density estimates based on samples of size 10 000 from the respective densities in the denoised case.

and σ^2 are simulated by sampling 5000 data points using Monte Carlo technique. In each scenario, a b_{01} value is obtained using equation (14) for the simulated data, and 10 000 simulations are run to yield a pdf of b_{01}. The overall probabilities of accepting the event that the structure is healthy (γ), i.e., accepting the null hypothesis, for all 10 scenarios are approximated by finding the proportion of $b_{01} > 0$ from the simulated results and summarized in Table 5. It is observed that the probabilistic assessment results lead to the same conclusion as obtained from the deterministic evaluation. As an example, Figure 9 shows the pdfs of b_{01} for five denoised cases (Kobe1 to Kobe5).

7 CONCLUDING REMARKS

In this article, probabilistic methodologies are presented to address several important and challenging issues in SHM that are related to sensors. The methodologies include a probabilistic FEA approach to optimize sensor placement, a wavelet-based Bayesian approach to cleanse sensor data, a nonparametric model based on NN and fuzzy clustering to perform structural system identification using incompletely sensed data, and a Bayesian hypothesis testing approach to assess the structural status, considering uncertainties in both model prediction and experimental observations. The proposed methodologies are investigated for application in two different disciplines, namely, aerospace (TPS panel) and civil (five-story frame structure) engineering.

Further work is required in regard to validating the proposed probabilistic methodologies for more complicated applications. For example, in sensor layout optimization, additional investigation is necessary to determine the optimum number of sensors. Instead of the fixed number of sensors considered in this article, it is more reasonable to assume a variable number of sensors to maximize the reliability of damage detection. As the number of sensors distributed across the structure increases, the estimated SHM performance measures might be expected to improve. However, owing to weight penalties associated with additional sensors, as well as complexity constraints with respect to the amount of data acquired by the sensing system that requires processing (and therefore processing power), the number of sensors applied to the structure must be minimum.

Future work needs to account for sensor damage and reliability under various environmental and operating conditions, both in sensor layout design and damage detection. A bond graph-based parametric approach appears to be a promising approach to distinguish between structural and sensor damage [90, 91]. Using the bond graph, a temporal causal graph can be derived to represent the causal relations among the structural variables and parameters. The possible causes (either structural damage or sensor fault) leading to the system damage can be isolated by a damage signature matrix. After isolation, the quantification of the amount of damage (parameter estimation) can be done using analysis of only the isolated substructure containing the

damage, leading to computational efficiency. Preliminary results with structural frames are promising; further work is needed for continuum structures in combination with FEA.

ACKNOWLEDGMENTS

The research described in this article is partly sponsored by the United States Air Force Research Laboratory (project monitor: Mark Derriso) through subcontracts to Anteon Corporation and General Dynamics Information Technology. The authors gratefully acknowledge this support. The authors also gratefully acknowledge valuable discussions and help from Dr Steven Olson at University of Dayton Research Institute and Dr Martin DeSimio at Alliant Techsystems, Inc., with respect to the TPS panel example.

REFERENCES

[1] Paté-Cornell E, Fischbeck PS. Probabilistic risk analysis and risk-based priority scale for tiles of the space shuttle. *Reliability Engineering and System Safety* 1993 **40**(3):221–238.

[2] Housner GW, *et al.* Structural control: past, present, and future. *ASCE Journal of Engineering Mechanics* 1997 **123**(9):897–971.

[3] Masri SF, Smyth AW, Chassiakos AG, Caughey TK, Hunter NF. Application of neural networks for detection of changes in nonlinear systems. *ASCE Journal of Engineering Mechanics* 2000 **126**(7):666–676.

[4] Park G, Cudney HH, Inman DJ. An integrated health monitoring technique using structural impedance sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**(6):448–455.

[5] Hung S-L, Huang C-S, Wen C-M, Hsu Y-C. Nonparametric identification of a building structure from experimental data using wavelet neural network. *Computer-Aided Civil and Infrastructure Engineering* 2003 **18**(5):358–370.

[6] Farrar CR, *et al. Damage Prognosis: Current Status and Future Needs*, Technical Report LA–14051–MS. Los Alamos National Laboratory, Los Alamos, NM, 2003.

[7] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Technical Report LA–13976—MS. Los Alamos National Laboratory, Los Alamos, NM, 2003.

[8] Derriso MM, Braisted W, Rosenstengel J, DeSimio M. The structural health monitoring of a mechanically attached thermal protection system. *Journal of the Minerals, Metals and Materials* 2004 **59**(3):36–39.

[9] Peairs DM, Park G, Inman DI. Improving accessibility of the impedance-based structural health monitoring method. *Journal of Intelligent Material Systems and Structures* 2004 **15**(2):129–139.

[10] Peairs DM, Park G, Inman DI. Practical issues of activating self-repairing bolted joints. *Smart Materials and Structures* 2004 **13**(6):1414–1423.

[11] Yuen KV, Au SK, Beck JL. Two-stage structural health monitoring approach for phase I benchmark studies. *ASCE Journal of Engineering Mechanics* 2004 **130**(1):16–33.

[12] Yang J, Chang F-K. Verification of a built-in health monitoring system for bolted thermal protection panels. *12th SPIE Annual International Symposium on Smart Materials and Structures*. San Diego, CA, 2005.

[13] Yang J, Chang F-K. Detection of bolt loosening in C-C composite thermal protection panels: I diagnostic principle. *Smart Materials and Structures* 2006 **15**(2):581–590.

[14] Yang J, Chang F-K. Detection of bolt loosening in C-C composite thermal protection panels: II experimental verification. *Smart Materials and Structures* 2006 **15**(2):591–599.

[15] Ni YQ, Zhou XT, Ko JM. Experimental investigation of seismic damage identification using PCA-compressed frequency response functions and neural networks. *Journal of Sound and Vibration* 2006 **290**(1–2):242–263.

[16] Jiang X, Adeli H. Psuedospectra, MUSIC, and dynamic wavelet neural network for damage detection of high-rise buildings. *International Journal of Numerical Methods in Engineering* 2007 **71**(5):606–629.

[17] Jiang X, Mahadevan S. Bayesian wavelet methodology for structural damage detection. *Structural Control and Health Monitoring* 2007, DOI: 10.1002/stc.230. in press.

[18] Aktan AE, Farhey DN, Helmicki AJ, Brown DL, Hunt VJ, Lee KL, Levi A. Structural identification for condition assessment: experimental arts.

ASCE Journal of Structural Engineering 1997 **123**(12):1674–1684.

[19] Catbas FN, Aktan AE. Condition and damage assessment: issues and some promising indices. *ASCE Journal of Structural Engineering* 2002 **128**(8):1026–1038.

[20] Chong KP, Carino NJ, Washer G. Health monitoring of civil infrastructures. *Smart Materials and Structures* 2003 **12**(3):483–493.

[21] Wood KH, Brown TL, Wu MC, Gause CB. Fiber optic sensors for cure/health monitoring of composite materials. *Proceeding of 3rd International Workshop on Structural Health Monitoring*. Stanford, CA, 2001.

[22] Prosser WH, Brown TL, Woodard SE, Fleming GA, Cooper EG. Sensor technology for integrated vehicle health management of aerospace vehicles. *Proceedings of 29th Annual Review of Progress in Quantitative Nondestructive Evaluation*. Bellingham, WA, 2002.

[23] Staszewski WJ, Boller C, Grondel S, Biemans C, O'Brien E, Delebarre C, Tomlinson GR. Damage detection using stress and ultrasonic waves. In *Health Monitoring of Aerospace Structures*, Staszewski WJ, Boller C, Tomlinson GR (eds). John Wiley & Sons: Cambridge, MA, 2004.

[24] Blackshire JL, Giurgiutiu V, Cooney A, Doane J. Characterization of sensor performance and durability for structural health monitoring systems. *Proceedings of SPIE 12th International Symposium on Smart Structures and Materials*. San Diego, CA, 2005.

[25] Blackshire JL, Cooney A. Characterization of bonded piezoelectric sensor performance and durability in simulated aircraft environments. *Proceedings of 32nd Review of Progress in Quantitative NDE*. Brunswick, ME, 2005.

[26] Blackshire JL, Cooney A. Evaluation and improvement in sensor performance and durability for structural health monitoring systems. *Proceedings of SPIE's 13th International Symposium on Smart Structures and Materials*. San Diego, CA, 2006.

[27] Sohn H, Farrar CR, Hunter NF, Worden K. Structural health monitoring using statistical pattern recognition techniques. *ASME Journal of Dynamic Systems, Measurement, and Control* 2001 **123**(4): 706–711.

[28] Nichols JM, Nichols CJ, Todd MD, Seaver M, Trickey ST, Virgin LN. Use of data-driven phase space models in assessing the strength of a bolted connection in a composite beam. *Smart Materials and Structures* 2004 **13**(2):241–250.

[29] Hundhausen RJ, Adams DE, Derriso M, Kukuchek P, Alloway R. Loads, damage identification and NDE/SHM data fusion in standoff thermal protection systems using passive vibration-based methods. *2nd European Workshop on Structural Health Monitoring*. Munich, 2004; pp. 959–966.

[30] Pines DJ. The use of wave propagation models for structural damage identification. *Structural Health Monitoring, Current Status and Perspectives*. Stanford University: Palo Alto, CA, 1997, pp. 665–677.

[31] Lovell P, Pines D. Damage assessment in a bolted lap joint. *Proceedings of SPIE on Smart Structures and Materials 1999: Smart Systems for Bridges, Structures, and Highways*, Newport Beach, California, USA, 1998; Vol. 3325, pp. 112–126.

[32] Purekar A, Lakshmanan A, Pines D. Detecting delamination in composite rotorcraft flexbeams using the local wave response. *Proceedings of SPIE on Smart Structures and Integrated Systems*, Newport Beach, California, USA, 1998; Vol. 3329, pp. 523–535.

[33] Alves CJS, Ribeiro PMC. Crack detection using spherical incident waves and near-field measurements. In *Boundary Elements XXI*, Brebbia CA, Power H (eds). WIT Press, 1999, pp. 355–364.

[34] Krawczuk M, Palacz M, Ostachowicz W. Wave propagation in plate structures for crack detection. *Finite Elements in Analysis and Design* 2004 **40**(9–10):991–1004.

[35] Bartoli I, Marzani A, Lanza di Scalea F, Viola E. Modeling wave propagation in damped waveguides of arbitrary cross-section. *Journal of Sound and Vibration* 2006 **295**(3–5):685–707.

[36] Huang S, Mahadevan S. Multivariate model validation using PCA and similarity factors. *Proceedings of PMC. 04*. Albuquerque, NM, 2004.

[37] Jiang X, Mahadevan S, Adeli H. Bayesian wavelet packet denoising for structural system identification. *Structural Control and Health Monitoring* 2007 **14**(2):333–356.

[38] Sohn H, Law KH. A Bayesian probabilistic approach for structure damage detection. *Earthquake Engineering and Structural Dynamics* 1997 **26**(12): 1259–1281.

[39] Vanik MW, Beck JL, Au SK. Bayesian probabilistic approach to structural health monitoring.

ASCE Journal of Engineering Mechanics 2000 **126**(7):738–745.

[40] Ching J, Beck JL. Bayesian analysis of the phase II IASC–ASCE structural health monitoring experimental benchmark data. *ASCE Journal of Engineering Mechanics* 2004 **130**(10):1233–1244.

[41] Katafygiotis LS, Lam HF, Mickleborough N. Application of a statistical model updating approach on phase *i* of the IASC–ASCE structural health monitoring benchmark study. *ASCE Journal of Engineering Mechanics* 2004 **130**(special issue): 34–48.

[42] Guratzsch RF, Mahadevan S. Structural health monitoring sensor placement optimization under uncertainty. *AIAA Journal* 2008, in press.

[43] Hiramoto K, Doki H, Obinata G. Optimal sensor/actuator placement for active vibration control using explicit solution of algebraic Riccati equation. *Journal of Sound and Vibration* 2000 **229**(5): 1057–1075.

[44] Abdullah M, Richardson A, Hanif J. Placement of sensor/actuators on civil structures using genetic algorithms. *Earthquake Engineering and Structural Dynamics* 2001 **30**(8):1167–1184.

[45] Simpson MT, Hansen CH. Use of genetic algorithms to optimize vibration actuator placement or active control of harmonic interior noise in cylinder with floor structure. *Noise Control Engineering Journal* 1996 **44**(4):169–184.

[46] Yan YJ, Yam LH. Optimal design of number and locations of actuators in active vibration control of a space truss. *Smart Materials and Structures* 2002 **11**(4):496–503.

[47] Demetriou MA. Integrated actuator-sensor placement and hybrid controller design of flexible structures under worst case spatiotemporal disturbance variations. *Journal of Intelligent Material Systems and Structures* 2004 **15**(12):901–921.

[48] Peng F. Actuator placement optimization and adaptive vibration control of plate smart structures. *Journal of Intelligent Material Systems and Structures* 2005 **16**(3):263–271.

[49] Li D, Li H, Fritzen CP. A new sensor placement algorithm in structural health monitoring. *Proceedings of 3rd European Workshop on Structural Health Monitoring*. Granada, 5–7 July 2006.

[50] Gao H, Rose JL. Ultrasonic sensor placement optimization in structural health monitoring using evolutionary strategy. *Proceedings to AIP Conference on Quantitative Nondestructive Evaluation*, Portland, Oregon, USA, 6 March 2006; Vol. 820, pp. 1687–1693.

[51] Gao H, Rose JL. Sensor placement optimization in structural health monitoring using genetic and evolutionary algorithms. *Proceedings of SPIE Smart Structures and Materials*. San Diego, CA, 26 February 2006.

[52] Dhillon SS, Chakrabarty K, Iyengar SS. Sensor placement for grid coverage under imprecise detection. *Proceedings of International Conference on Information Fusion*. Annapolis, MD, 7–11 July 2002.

[53] Parker DL, Frazier WG. Experimental validation of optimal sensor placement algorithms for structural health monitoring. *Proceedings of 3rd European Workshop on Structural Health Monitoring*. Granada, 5–7 July 2006.

[54] Haldar A, Mahadevan S. *Reliability Assessment Using Stochastic Finite Element Analysis*. John Wiley & Sons: New York, 2000.

[55] Huang S. *Simulation of Random Processes Using Karhunen-Loeve Expansion*, Ph.D. Dissertation. National University of Singapore, 2001.

[56] Tedesco JW, McDougal WG, Ross CA. *Structural Dynamics: Theory and Application*. Addison Wesley Longman: Menlo Park, CA, 1999.

[57] Sakamoto S, Ghanem R. Polynomial chaos decomposition for the simulation of non-Gaussian nonstationary stochastic processes. *ASCE Journal of Engineering Mechanics* 2002 **128**(2):190–201.

[58] Deodatis G, Micaletti RC. Simulation of highly skewed non-Gaussian stochastic processes. *ASCE Journal of Engineering Mechanics* 2001 **127**(12): 1284–1295.

[59] Corbin M, Hera A, Hou Z. Location damage regions using wavelet approach. *Proceedings of the 14th Engineering Mechanics Conference*. Austin, TX, 21–24 May 2000.

[60] Au SK, Yuen KV, Beck JL. Two-stage system identification results for benchmark structure. *Proceedings of the 14th ASCE Engineering Mechanics Conference*, CD ROM. Austin, TX, 2000.

[61] Banks HT, Joyner ML, Bincheski B, Winfree WP. Real time computational algorithms for eddy-current-based damage detection. *Inverse Problems* 2002 **18**(3):795–823.

[62] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of Structures and Mechanical Systems from Changes in*

their Vibration Characteristics: a Literature Review. Los Alamos National Laboratory: Los Alamos, NM, 1996.

[63] Duda RO, Hart PE, Stork DG. *Pattern Classification*, *Second Edition*. John Wiley & Sons: New York, 2001.

[64] DeSimio M, Miller I, Derriso M, Brown K, Baker M. Structural health monitoring experiments with a canonical element of an aerospace vehicle. *Proceedings of 2003 IEEE Aerospace Conference*. Big Sky, MT, 8–15 March 2003.

[65] Padula LS, Kincaid RK. *Optimization Strategies for Sensor and Actuator Placement*, Report NASA/TM-1999-209126. NASA Langley Research Center, April 1999.

[66] Padula LS, Kincaid RK. *Aerospace Applications of Integer and Combinatorial Optimization*, NASA TM-110210. NASA Langley Research Center, October 1995.

[67] Padula LS, Kincaid RK. Optimal sensor/actuator locations for active structural acoustic control. *Proceedings of 39th AIAA/ASME/ASCE/AHS/ASC Structures, Dynamics and Materials Conference*. Long Beach, CA, 20–23 April 1998.

[68] Raich AM, Liszkai TR. Multi-objective genetic algorithm methodology for optimizing sensor layouts to enhance structural damage detection. *Proceedings of 4th International Workshop on Structural Health Monitoring*. Stanford, CA, 15–17 September 2003.

[69] Spall JC. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins APL Technical Digest* 1998 **19**(4):482–492.

[70] Huyer W, Neumaier A. *SNOBFIT – Stable Noisy Optimization by Branch and Fit*, http://www.mat.univie.ac.at/~neum/software/snobfit/ (date accessed 25 May 2004).

[71] Olson S, DeSimio M, Derriso M. Fastener damage estimation in a square aluminum plate. *Structural Health Monitoring* 2006 **5**(2):173–183.

[72] ANSYS. *ANSYS Release 9.0 Documentation*, 2004.

[73] Shinozuka M, Deodatis G. Simulation of stochastic processes by spectral representation. *Applied Mechanics Review* 1991 **44**(4):191–204.

[74] VanMarcke E. *Random Fields: Analysis and Synthesis*. The MIT Press: Cambridge, MA, 1983.

[75] Kohavi R, Provost F. Glossary of terms. *Machine Learning* 1998 **30**(2–3):271–274.

[76] Welch PD. The use of fast Fourier transform for the estimates of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 1967 **AU-15**:70–73.

[77] Oppenheim AV, Schafer RW. *Digital Signal Processing*. Prentice-Hall: Englewood Cliffs, NJ, 1975.

[78] MathWorks. *MATLAB Version 7.0.4.365 (R14) Service Pack 2*, 29 January 2005.

[79] Adeli H, Jiang X. Dynamic fuzzy wavelet neural network model for structural system identification. *ASCE Journal of Structural Engineering* 2006 **132**(1):102–111.

[80] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press: New York, 1981.

[81] Chatfield C. *The Analysis of Time Series: an Introduction*, *Sixth Edition*. Chapman & Hall/CRC Press: Boca Raton, FL, 2004.

[82] Jiang X, Adeli H. Dynamic wavelet neural network for nonlinear system identification of high-rising building. *Computer-Aided Civil and Infrastructure Engineering* 2005 **20**(4):316–330.

[83] Masri SF, Nakamura M, Chassiakos AG, Caughey TK. Neural network approach to the detection in structural parameters. *Journal of Engineering Mechanics, ASCE* 1996 **122**(4):350–360.

[84] Wu ZS, Xu B, Yokoyama K. Decentralized parametric damage detection based on neural networks. *Computer-Aided Civil and Infrastructure Engineering* 2002 **17**(3):175–184.

[85] Jeffreys H. *Theory of Probability*, *Third Edition*. Oxford University Press: London, 1961.

[86] Kass R, Raftery A. Bayes factors. *Journal of the American Statistical Association* 1995 **90**(430): 773–795.

[87] Mahadevan S, Rebba R. Validation of reliability computational models using Bayes networks. *Reliability Engineering and System Safety* 2005 **87**(2): 223–232.

[88] Migon HS, Gamerman D. *Statistical Inference: an Integrated Approach*. Arnold: London, 1999.

[89] Stoica P, Moses RL. *Introduction to Spectral Analysis*. Prentice-Hall: Upper Saddle River, NJ, 1997.

[90] Biswas G, Mahadevan S. A hierarchical model-based approach to system health management. *Proceedings, IEEE Aerospace Conference*. Big Sky, MT, 2007.

[91] Moustafa A, Daigle M, Roychoudhury I, Shantz C, Biswas G, Mahadevan M, Koutsoukos X. Fault diagnosis of civil engineering structures using the bond graph approach. *Proceedings, 7th International Conference on Diagnosis (DX07)*. Nashville, TN, 2007; pp. 146–153.

FURTHER READING

Karnoop DC, Margolis DL, Rosenberg RC. *System Dynamics: Modeling and Simulation of Mechatronic Systems*. John Wiley & Sons: New York, 2000.

Chapter 51

Development of Fuzzy Rules for Damage Detection and Location

Ranjan Ganguli

Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India

1 INTRODUCTION

Structural damage leads to changes in some measurable system properties. In structural health monitoring (SHM), we try to detect the damage from such measurable system behavior. The problem of damage detection becomes complicated because of the presence of noise in the measured data and errors in modeling. We can visualize that structural damage causes patterns in the measurement space, and our objective is to device algorithms that can predict the location and size of damage from these patterns (*see* **Statistical Pattern Recognition**). In recent years, methods such as those based on neural networks (*see* **Artificial Neural Networks**), genetic algorithms (GAs), and system identification have been used to solve damage detection problems. An alternative method for structural damage detection, which has several advantages over the neural network and GA approaches, is based on fuzzy logic and is discussed in this article.

As mentioned by Boller in a recent article [1], there is no need for a health monitoring system to locate damage to within a few millimeters. The cost and effort involved in predicting damage to a high level of accuracy can be prohibitive. In addition, because of measurement, model, and signal-processing inaccuracies, a health monitoring system that claims to predict damage with great accuracy is likely to give false alarms and lose the faith of maintenance personnel using it. A better idea is to roughly locate the damage to within about 1 m using a health monitoring system and then to use standard non-destructive evaluation (NDE) methods for a closer analysis of the damage area [1]. Such an approach is well suited to the use of fuzzy logic, which gives linguistic outputs such as "large damage at root of beam" instead of numerical values.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

For successful SHM, we need to solve the inverse problem relating the change in measurements between the damaged and undamaged structure to the presence, location, and size of the structural damage (*see* **Damage Presence/Growth Monitoring Sensors**). The inverse problem is complicated by incomplete information (not all states of the system are available) and uncertainty in modeling, measurement, and signal processing. As mentioned before, the inverse problem for structural damage detection is typically solved by computational intelligence methods like neural network [2–7] and GA [8–10]. GAs used directly for damage detection problems can have some limitations. For example, GA has to run each time the structure is evaluated to obtain the measured dynamic characteristics. This can be very time consuming and may not be suitable for on-line applications. In contrast, neural networks are computationally fast once they have been trained. However, neural networks have the reputation of being black boxes that are difficult to understand.

Fuzzy systems allow for easier understanding because they are expressed in terms of linguistic variables [11], which are expressed in terms of a set of rules. In many ways, a set of rules obtained from observation and experience is what human experts also use for monitoring the health of many structures and for making decisions. Fuzzy systems have a built-in fuzzification process at the front end that accounts for uncertainty and does not need to be trained on several cycles of noisy data like neural networks to account for uncertainty [12]. It is well known that feedforward neural networks are universal function approximators [13]. Recently, it has been proved that classical feedforward neural networks can be approximated to an arbitrary degree of accuracy by a fuzzy logic system (FLS), without having to go through the laborious training process needed by a neural network [14]. Therefore, fuzzy systems share the universal approximation characteristics with neural networks and can perform any function approximation or pattern-recognition task that a neural network can perform.

Fuzzy logic is a useful and efficient method for modeling complex systems where uncertainty and imprecision can be important, and can be used to solve the damage detection problem. Fuzzy systems address uncertainty directly by using linguistic reasoning, which is more robust to uncertainty than pure numerical reasoning. Very few researchers have used fuzzy logic for structural damage detection, primarily because of the lack of familiarity with fuzzy logic among most structural engineers. Some researchers have used fuzzy logic to improve neural network approaches for damage detection. For example, Ramu and Johnson [15] and Nyongesa *et al.* [16] used fuzzy logic for improving the performance of neural networks for damage detection of composite structures. Ramu and Johnson created a fuzzy neural network to detect, classify, and estimate the extent of damage in a composite structure from its vibration response. They mention that fuzzy representation was found to be the most efficient means of treating uncertainties. Nyongesa *et al.* [16] developed an automated system for feature recognition from shearograms of composites impacted by damage. The use of shearograms for impact damage detection in composites is difficult because the fringe patterns for damage such as a delamination is not clear. The pattern-recognition problem is noisy and has significant uncertainty. The use of a neural network pattern classifier along with a fuzzy logic inference was found to improve damage detection accuracy.

A generalized methodology for structural damage detection using fuzzy logic is presented by Sawyer and Rao [17]. They point out that the development of smart structures with built-in damage detection systems are well suited to automated reasoning tools such as fuzzy logic. Fuzzy associations between structural damage and observable system response were generated using finite element simulations. They considered static and dynamic response of structures to determine their health state. Damage was modeled using the principles of continuum damage mechanics by appropriately reducing the stiffness at the damage location. They found that the fuzzy approach is better suited to tolerate noise and uncertainty than system-identification-type methods. Ganguli [18] demonstrated the use of fuzzy logic for damage detection and isolation in presence of noise for a BO-105 hingeless helicopter rotor. Fuzzy rules were generated from numerical data obtained from finite element simulations. Frequencies were used for damage detection, and it was found that a gross estimate of damage location and size could be obtained from the first four frequencies. He showed that the fuzzy system performed very well even in the presence of noisy data. He also pointed out that with the advent of

embedded and surface-mounted piezoelectric sensors and actuators in structures, operational health monitoring of structures from structural response data using FLSs has become possible.

Sazonov *et al.* [19] developed a fuzzy logic expert system to automate the process of damage detection from strain energy mode shapes. They pointed out that strain energy mode shapes require curvature information and are therefore sensitive to measurement noise, which gets amplified by the double differentiation of displacement needed to get curvature. The peaks in the strain energy mode shapes are the damage indicators, and false peaks can result with noisy data. They developed a set of fuzzy rules from a finite element model of a simple beam. The fuzzy system was then developed to mimic a human expert and very accurate damage detection results were obtained, using mode shapes acquired by impact testing and/or noncontact laser vibrometer methods.

The traditional fuzzy system is an effective tool for damage detection. However, their rules need to be developed and fine-tuned by the user on the basis of the available data. For example, the rules need to be regenerated for different structures and for each time that the set of measurements or the fault changes. The process of rule development is therefore time consuming, not very accurate, and certainly not optimal. Therefore, to make the rule generation and tuning process automatic, GA is used in advanced soft computing techniques, known as the *genetic fuzzy system* [20, 21]. These systems combine the approximate reasoning capability of fuzzy systems with the learning capabilities of GAs. Recently Pawar and Ganguli [12] used a GA for automating the process of rule generation for a fuzzy system for application to structural damage detection. They demonstrated the algorithm for a beam-type structure and a helicopter rotor blade modeled using the finite element method. Such a system represents an alternative to the neural network and GA-based methods for structural damage detection problem.

We see that the use of fuzzy logic in SHM has been rather limited, compared to other soft computing methods. In this article, an introduction is given for developing a fuzzy system using some simple problems. The article is organized as follows. First, a brief description of fuzzy logic is given. Then an FLS is formulated for structural damage detection. The example problem used here is damage detection in a cantilever beam, using frequencies. Next, a method of using GA to automate the development of fuzzy rules is presented. This genetic fuzzy system is illustrated using a composite thin-walled beam with matrix cracking as the damage and frequencies as measurements. Finally, some concluding remarks are made.

2 FUZZY LOGIC SYSTEM

Before using fuzzy logic for damage detection, we introduce the FLS as a function approximation and decision making tool. An FLS is a nonlinear mapping of an input feature vector into a scalar output [22, 23]. Fuzzy set theory and fuzzy logic provide the framework for the nonlinear mapping. FLSs have been widely used in engineering applications because of the flexibility they offer designers and their ability to handle uncertainty. Only a brief introduction to FLSs is given here. Several books provide a detailed introduction to fuzzy logic [22, 24–26].

A typical multi-input single-output (MISO) FLS performs a mapping from $V \in R^m$ to $W \in R$ using four basic components: rules, fuzzifier, inference engine, and defuzzifier. Here, $f : V \in R^m \rightarrow W \in R$ where $V = V_1 \times V_2 \times \cdots \times V_n \in R^m$ is the input space and $W \in R$ is the output space. A typical FLS is shown schematically in Figure 1. Once the rules driving the FLS have been fixed, the FLS can be expressed as a mapping of inputs to outputs.

Rules can come from experts or can be obtained from numerical data. Expert knowledge is typically transformed into fuzzy rules by interviewing

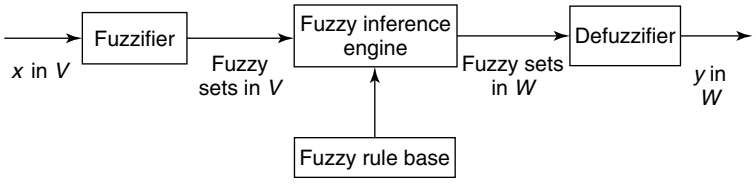


Figure 1. Schematic representation of a fuzzy logic system.

knowledge workers about the problem. Rules from numerical data can be obtained by qualitative analysis of the data obtained by parametric study, or by more formal methods such as GAs as discussed later in this article. It is also possible to combine rules obtained from numerical data with those obtained from experts. Whatever be the approach used to create the rules, they are expressed as a collection of IF-THEN statements such as "IF u_1 is HIGH, and u_2 is LOW, THEN v is LOW". To formulate such a rule we need an understanding of the following:

1. Linguistic variables versus numerical values of a variable. For example, a height of 190 cm is a numerical value that can be labeled with the linguistic variable "Tall". Similarly, a frequency change of 5% due to damage can be labeled as "High".
2. Quantifying linguistic variables. For example, u_1 may have a finite number of linguistic terms associated with it, such as NEGLIGIBLE, SMALL, MEDIUM, HIGH, and VERY HIGH, which is done using fuzzy membership functions.
3. Logical connections between linguistic variables (for example, AND, OR, etc.).
4. Implications such as "IF A THEN B". We also need to understand how to combine more than one rule.

The fuzzifier operates on numerical data and maps crisp input numbers into fuzzy sets. It is needed to activate rules that are expressed in terms of linguistic variables. The process of transferring numerical data into linguistic variables through the fuzzifier is a direct approach to handling uncertainty built-in at the front end of the fuzzy system. An inference engine of the FLS maps fuzzy sets to fuzzy sets and determines the way in which the fuzzy sets are combined. In several applications, crisp numbers are needed as an output of the FLS. In those cases, a defuzzifier is used to calculate crisp values from fuzzy values. The key terms and elements of an FLS are defined next.

Fuzzy sets

A fuzzy set F is defined on a universe of discourse U and is characterized by a degree of membership $\mu(x)$, which can take on values between 0 and 1. Here U represents the set of all possible values that can be taken by the number x. A fuzzy set generalizes the concept of an ordinary set whose membership function only takes two values, zero and unity.

Linguistic variables

A linguistic variable u is used to represent the numerical value x, where x is an element of U. A linguistic variable is usually decomposed into a set of terms $T(u)$, which cover its universe of discourse.

Membership functions

The most commonly used shapes for membership functions $\mu(x)$ are triangular, trapezoidal, piecewise linear, or Gaussian. The designer of the FLS selects the type of membership function used. The Gaussian functions overlap to some degree across all fuzzy sets, and therefore allow all fuzzy rules to fire simultaneously. Therefore, they smooth out the output signal [27]. There is no theoretical requirement that membership functions overlap. However, one of the major strengths of fuzzy logic is that membership functions can overlap. FLS systems are robust because decisions are distributed over more than one input class. For convenience, membership functions are normalized to one, so they take values between 0 and 1, and thus define the fuzzy set.

Inference engine

Rules for the fuzzy system can be expressed as:

$$R_i : \textbf{IF } x_1 \text{ is } F_1 \textbf{ AND } x_2 \text{ is } F_2 \text{ AND } \ldots \ x_m \text{ is } F_m$$
$$\textbf{THEN } y = C_i, \quad i = 1, 2, 3, \ldots, M$$

where m and M are the number of input variables and rules, x_i and y are the input and output variables, and $F_i \in V_i$ and $C_i \in W$ are fuzzy sets characterized by membership functions $\mu_{F_i}(x)$ and $\mu_{C_i}(x)$, respectively. Each rule can be viewed as a fuzzy implication $F_{12\cdots m} = F_1 \times F_2 \times \cdots F_m \rightarrow C_i$, which is a fuzzy set in $V \times W = V_1 \times V_2 \times \cdots V_m \times W$ with membership function given by

$$\mu_{R_i}(x, y) = \mu_{F_1}(x_1) * \mu_{F_2}(x) * \cdots * \mu_{F_m}(x_m) \times \mu_{C_i}(y) \qquad (1)$$

where $*$ is the T-norm with $x = [x_1 x_2 \cdots x_m] \in V$ and $y \in W$. This sort of rule covers many applications. The algebraic product is one of the most widely used T-norms in applications, and leads to a product inference engine.

Defuzzification

Popular defuzzification methods include maximum matching and centroid defuzzification [28]. While centroid defuzzification is widely used for fuzzy control problems where a crisp output value is needed, maximum matching is often used for pattern matching problems where we need to know the output class only. In structural damage detection, it is often better to leave the damage location, size, and other information in linguistic terms. For example, "large damage at root" is more useful and credible than numerical values. In fact, a linguistic output is useful for creating maintenance alerts and suggests prognostic action.

Suppose there are K fuzzy rules and, among them, K_j rules ($j = 1, 2, \ldots, L$ and L is the number of classes) produce class C_j. Let D_p^i be the measurements of how the pth pattern matched the antecedent conditions (**IF** part) of the ith rule, which is given by the product of membership grades of the pattern in the regions that the ith rule occupies.

$$D_p^i = \prod_{i=1}^{m} \mu_{li} \tag{2}$$

where m is the number of inputs and μ_{li} is the degree of membership of measurement l in the fuzzy regions that the ith rule occupies. Let $D_p^{\max}(C_j)$ be the maximum matching degree of the rules (rules $j_l, l = 1, 2 \ldots, K_j$) generating class C_j.

$$D_p^{\max}(C_j) = \max_{l=1}^{K_j} D_p^{j_l} \tag{3}$$

then the system will output class C_{j^*}, provided that

$$D_p^{\max}(C_{j^*}) = \max_j D_p^{\max}(C_j) \tag{4}$$

If there are two or more classes that achieve the maximum matching degree, we select the class that has the largest number of fired fuzzy rules (a fired rule has a matching degree greater than zero) [29].

3 FUZZY LOGIC SYSTEM FOR DAMAGE DETECTION

The description of the FLS given above may appear abstract and mathematical. We illustrate the FLS with a simple problem of damage detection in a cantilever beam using natural frequencies (*see* **Modal–Vibration-based Damage Identification**). A schematic representation of the problem requirement is shown in Figure 2. Damage is modeled using a reduction in element stiffness and a percentage damage parameter D is defined such that

$$D = 100\frac{E^{\text{undamaged}} - E^{\text{damaged}}}{E^{\text{undamaged}}} \tag{5}$$

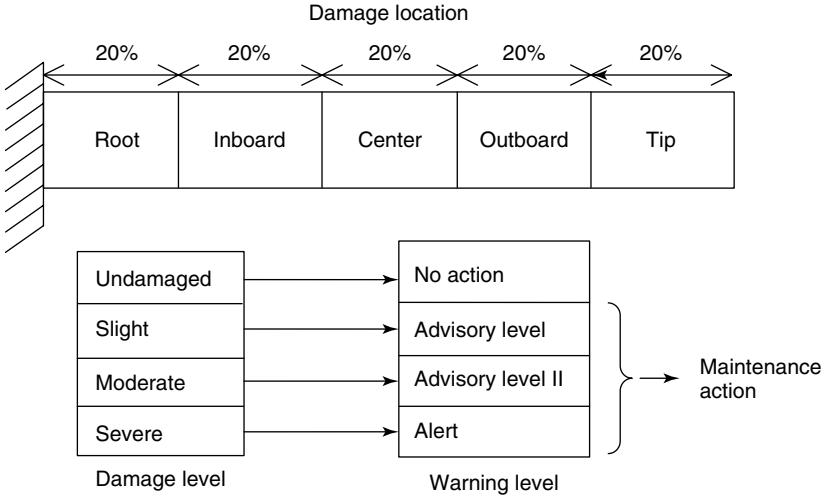


Figure 2. Schematic representation of structural damage detection in a cantilever beam.

where E is the Young's modulus of the material. The beam is 4.94-m long and has a mass per unit length of $6.46\,\mathrm{kg\,m^{-1}}$ and a flexural modulus of $103\,955.3\,\mathrm{N\,m^2}$ [18].

The beam is divided into five uniform segments of equal lengths. These segments are labeled as "root", "inboard", "center", "outboard", and "tip", as shown in Figure 2. The structural damage in each segment is modeled by stiffness reductions (D) of 5, 15, and 25% and these three damage sizes are classified as "slight damage", "moderate damage", and "severe damage", respectively. Damage sizes below "slight damage" are classified as undamaged. A damage size larger than "severe damage" is classified as "catastrophic damage".

From a practical implementation viewpoint, given a slight damage, the warning issued by the diagnostic system is classified as a "level I advisory". For a moderate damage, the warning issued is a "level II advisory" and for severe damage, the maintenance action is an "alert". No warning is issued for an undamaged beam. This form of classification of the structural damage allows for development of a user-friendly decision system that can be deployed in a handheld electronic device or field computer. With the requirements of the damage detection system defined, we proceed to formulate the fuzzy logic behind it.

3.1 Input and output

Let the beam be divided into five damage locations as shown in Figure 2. We want to detect damage in this beam using the first four frequencies. Define measurement deviations from the undamaged beam frequencies as measurement "deltas" given as

$$\Delta\omega = 100\frac{\omega^{(\mathrm{u})} - \omega^{(\mathrm{d})}}{\omega^{(\mathrm{u})}} \tag{6}$$

Here $\omega^{(\mathrm{u})}$ and $\omega^{(\mathrm{d})}$ correspond to the undamaged and damaged frequency. Inputs to the FLS developed for damage detection are then measurement deltas and outputs are structural damage. We have four measurements represented by z and five fault locations represented by x. The objective is to find a functional mapping between z and x, which can be represented as

$$\mathrm{x} = F(\mathrm{z}) \tag{7}$$

where

$$\mathrm{x} = \{\text{Root, Inboard, Center, Outboard, Tip}\}^{\mathrm{T}} \tag{8}$$

$$\mathrm{z} = \{\Delta\omega_1, \Delta\omega_2, \Delta\omega_3, \Delta\omega_4\}^{\mathrm{T}} \tag{9}$$

Each frequency measurement delta has uncertainty.

3.2 Fuzzification

The structural damage locations are crisp numbers. For example, "root" ranges from 0 to 20% of the blade, "inboard" from 20 to 40%, "center" from 40 to 60%, "outboard" from 60 to 80%, and "tip" from 80 to 100%, as shown in Figure 2. To get a degree of resolution of the extent of damage, each of these damage locations is allowed several levels of damage and split into linguistic variables. For example, consider "root" as a linguistic variable. Then it can be decomposed into a set of terms

$$\begin{aligned} T(\text{root}) = \{&\text{Undamaged, slight damage,}\\ &\text{moderate damage, severe damage,}\\ &\text{catastrophic damage}\} \end{aligned} \tag{10}$$

where each term in $T(\text{root})$ is characterized by a fuzzy set in the universe of discourse $U(\text{root}) = \{0, 30\}$, in terms of damage variable D. The other structural damage variables are fuzzified in a similar manner.

The measurement deltas $\Delta\omega_1$, $\Delta\omega_2$, $\Delta\omega_3$, and $\Delta\omega_4$ are also treated as fuzzy variables. To get a high degree of resolution, they are further split into linguistic variables. For example, consider $\Delta\omega_1$ as a linguistic variable. It can be decomposed into a set of terms

$$\begin{aligned} T(\Delta\omega_1) = \{&\text{Negligible, very low, low,}\\ &\text{low medium, medium, medium high,}\\ &\text{high, very high, very very high}\} \end{aligned} \tag{11}$$

where each term in $T(\Delta\omega_1)$ is characterized by a fuzzy set in the universe of discourse $U(\Delta\omega_1) = \{0, 9\}$. The other three measurement deltas are defined using the same set of terms as $\Delta\omega_1$. Measurement deltas larger than those covered by the universe of discourse will represent an extensive structural damage indicative of a catastrophic failure.

Fuzzy sets with Gaussian membership functions are used for the input variables. These fuzzy sets can be defined as follows:

$$\mu(x) = \mathrm{e}^{-0.5((x-m)/\sigma)^2} \tag{12}$$

where m is the midpoint of the fuzzy set and σ is the uncertainty (standard deviation) associated with the variable. Table 1 gives the linguistic measure associated with each fuzzy set and midpoint of the set for each measurement delta. The midpoints are selected to span the region ranging from an undamaged beam (all measurement deltas are zero) to one with significant damage. The standard deviations for $\Delta\omega$ are 0.35%, and are selected to provide for enough intersection between the fuzzy sets so as to optimize accuracy of detection. Note that we have developed the fuzzy sets manually using a simple uniform discretization in the measurement variables.

3.3 Rules and fault isolation

Rules for the fuzzy system are obtained by fuzzification of the numerical values obtained from the finite element analysis using the following procedure [23, 30]:

A set of four measurement deltas corresponding to a given structural fault is input to the FLS and the degree of membership of the elements of $\Delta\omega_1$, $\Delta\omega_2$, $\Delta\omega_3$, and $\Delta\omega_4$ are obtained. Therefore, each measurement has nine degrees of membership based on the linguistic measures in Table 1.

Table 1. Gaussian fuzzy sets for frequencies of a cantilever beam

Linguistic measure	Symbol	Midpoints of $\Delta\omega$
Negligible	N	0
Very low	VL	1
Low	L	2
Low medium	LM	3
Medium	M	4
Medium high	MH	5
High	H	6
Very high	VH	7
Very very high	VVH	8

1. Each measurement delta is then assigned to the fuzzy set with the maximum degree of membership.
2. One rule is obtained for each fault by relating the measurement deltas with maximum degree of membership to a fault.

These rules are tabulated in Table 2 for the frequencies. The linguistic symbols used in this table are defined in Table 1. These rules can be read as follows for the "moderate damage at root" fault:

IF $\Delta\omega_1$ is medium high **AND** $\Delta\omega_2$ is low **AND** $\Delta\omega_3$ is very low **AND** $\Delta\omega_4$ is low **THEN** moderate damage at root.

The rules for the other faults can be similarly interpreted. These rules provide a knowledge base and represent how a human engineer would interpret data to isolate structural damage using frequency shifts. For any given input set of measurement deltas, the fuzzy rules are applied for a given measurement, we have degrees of membership for each fault. For fault isolation, we are interested in the most likely fault. The fault with the highest degree of membership is selected as the most likely fault.

A close inspection of Table 2 shows that each of the rules is independent of the other. This is useful for preventing confusion between the rules. The level of discretization in the measurement deltas in terms of the selection of midpoints and standard deviations must be sufficiently high to make the rules independent, for best performance of the FLS. Also note that the set of rules have a built-in damage detection function, since if the first rule fires, the beam is undamaged. Firing of any of the other 15 rules indicates a damaged beam. The problem of location and identification of the damage size is handled through the 15 rules shown in Table 2 after Rule 1. Thus, the three critical functions of damage detection, location, and identification can be handled by a set of simple fuzzy rules. As shown in Figure 2, these rules can also be used to give maintenance alarms.

3.4 Accessing the performance of a fuzzy logic system

If an FLS has been developed properly, it will give excellent results for noise-free numerical data.

Table 2. Rules for fuzzy system for damage detection in a cantilever beam

Rule no.	Damage	Frequency measurement deltas			
		$\Delta\omega_1$	$\Delta\omega_2$	$\Delta\omega_3$	$\Delta\omega_4$
1	Undamaged	N	N	N	N
2	Slight damage at root	VL	VL	N	N
3	Slight damage at inboard	VL	N	VL	N
4	Slight damage at center	N	VL	N	VL
5	Slight damage at outboard	N	VL	VL	N
6	Slight damage at tip	N	N	N	VL
7	Moderate damage at root	MH	L	VL	L
8	Moderate damage at inboard	LM	VL	L	VL
9	Moderate damage at center	VL	LM	VL	L
10	Moderate damage at outboard	N	L	LM	L
11	Moderate damage at tip	N	N	VL	L
12	Severe damage at root	VVH	M	LM	LM
13	Severe damage at inboard	MH	VL	M	LM
14	Severe damage at center	L	H	VL	M
15	Severe damage at outboard	N	M	MH	LM
16	Severe damage at tip	N	N	VL	LM

In such cases of ideal data, the fuzzy rules are similar to a classical expert system. However, the strength of the fuzzy system comes out when numerical data is contaminated with noise, which is the case for most damage indicators used in structural damage detection. The FLS is tested using noise contaminated simulated data. Given a computed frequency measurement delta $\Delta\omega$, a random number $u = [-1, 1]$ and a noise level α, the noisy simulated measurement delta is given as $\Delta\omega_{\text{noisy}} = \Delta\omega(1 + u\alpha)$. We observe that noise addition to $\Delta\omega$ accounts for both model and measurement uncertainty. Five thousand noisy data points are generated for each seeded fault, and the percentage success rate for the fuzzy system in isolating a fault is calculated. Table 3 shows the average success rate of damage detection for increasing levels of noise. It is clear that the damage detection accuracy is 100% for a noise level of up to 15% in the data, and shows a gradual reduction after that.

The example problem discussed above clearly shows the ability of the FLS to detect damage accurately from noisy data. For most structural damage detection problems, such a fuzzy system can be developed manually by engineers. However, as the complexity of the problem increases in terms of number of measurements, damage levels, and damage locations, it becomes difficult to develop the fuzzy rules manually. Each time a fuzzy discretization is selected and rules obtained, the performance of the FLS needs to be evaluated. If the performance is not good, one has to adjust the fuzzy sets until a good performance is achieved. The process of adjusting the fuzzy sets involves changing the midpoints and standard deviations of Gaussian sets, if they are being used, calculating the new rules for the fuzzy sets selected, and then evaluating them for ideal and noisy data. This process involves trial and error, and becomes very cumbersome with increasing number of fuzzy sets and rules. In the next section, we

Table 3. Average percent success rate in fault detection using frequencies at different noise levels

Damage level	$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.25$	$\alpha = 0.3$
Undamaged	100	100	100	100	100	100	100
Slight	100	100	100	100	99	97.24	93.94
Moderate	100	100	100	100	99.33	97.84	96.51
Severe	100	100	100	100	99.7	98.24	96.32

discuss an approach for automating the rule development process for fuzzy systems by using optimization methods such as the GA.

4 GENETIC FUZZY SYSTEM

In this section, we formulate the genetic fuzzy system for detection of crack density and location of matrix cracks in a thin-walled hollow circular composite beam, shown in Figure 3. Inputs to the genetic fuzzy system are measurement deltas and outputs are crack density and locations of cracks (*see* **Lamb Wave-based SHM for Laminated Composite Structures**). The objective is to find the mapping between eight frequency measurement deltas and 10 locations of matrix cracking.

Composite cross sections used in many applications are thin-walled and are produced with various cross-sectional dimensions and ply orientations. Here, a hollow circular cross section with uniform diameter and thickness along the length is considered. The wall consists of $[\pm\theta_m/90_n]_s$ family of composite laminates. The ply orientation angle is with respect to the longitudinal direction (along the length of the beam). The dynamic analysis is done by a one dimensional (1-D) finite element analysis. For this

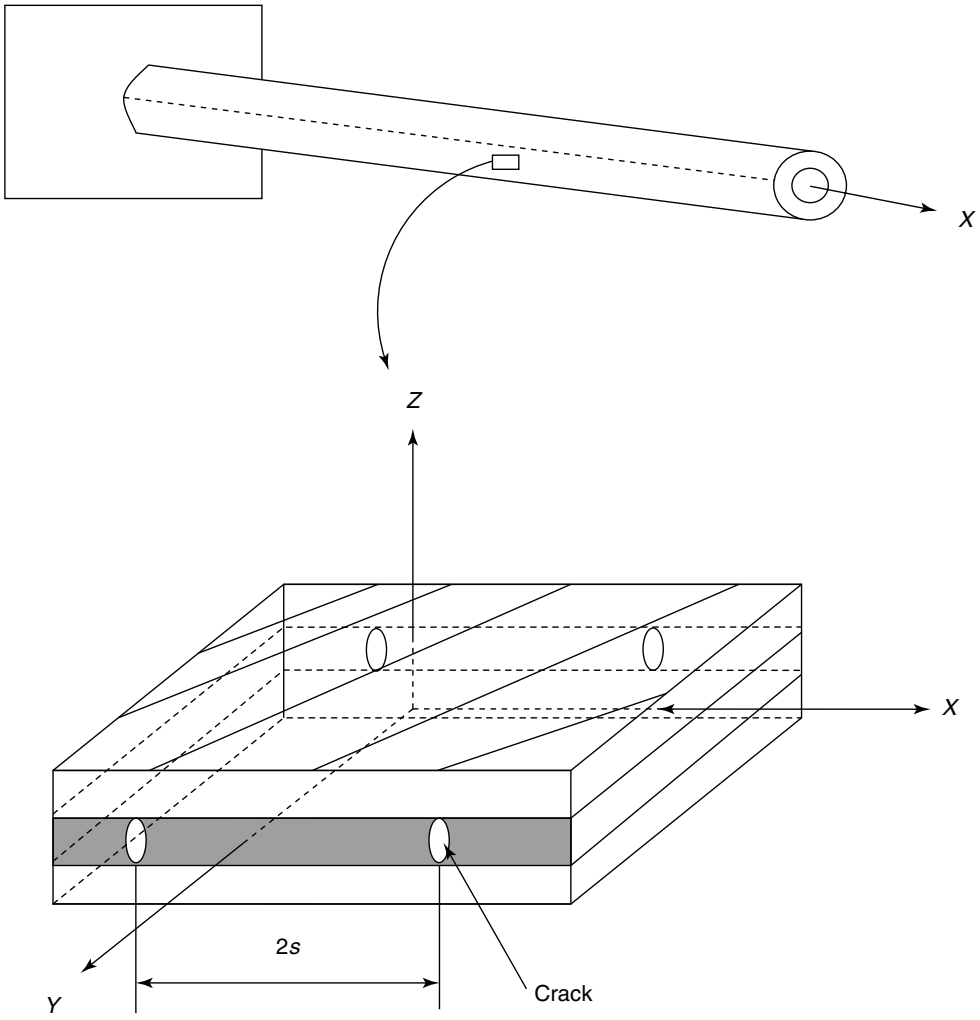


Figure 3. A hollow circular cantilever beam showing a small cross section of the wall with matrix cracks.

1-D model the effective elastic modulus is required, which accounts for all composite properties including the effect of matrix cracking. The effective elastic modulus is formulated considering the small cross section of the beam as a 2-D composite plate.

The beam length is 18 m and the outer diameter of the circular cross section is 600 mm. These properties are taken from Polyzois *et al.* [31] Material properties corresponding to glass epoxy are $E_1 = 48\,\text{GPa}$, $E_2 = 13.30\,\text{GPa}$, $\nu_{12} = 0.235$, and $G_{12} = 5.17\,\text{GPa}$ where the subscript 1 denotes the fiber direction and the subscript 2 denotes the transverse direction. The other properties are material density $\rho = 1.94\,\text{g cm}^{-3}$ and the wall contains 50 plies each 0.22-mm thick. The total thickness of the beam wall cross section is 11 mm. The natural frequencies of the beam are compared with Polyzois *et al.* [31]. Detailed calculations of measurement deltas (change in frequencies) for a thin-walled hollow circular composite beam structure are explained in [32].

In this fuzzy system, location of matrix cracking L_{10} ranges from 0 to 10% of the beam from fixed end, L_{20} from 10 to 20%, and so on, up to L_{100} from 90 to 100%. To get a degree of resolution of the extent of matrix cracking, each of these matrix cracking locations is allowed several levels of damage level and split into linguistic variables. These classifications are based on the numerical results obtained for matrix cracking. The matrix crack density can broadly be classified as "undamaged" for zero crack density, as "slight damage" for 1–5 crack density, as "moderate damage" for 5–10, as "severe damage" for 10–20 and as "very severe damage" for more than 20. Crack density, in this study, refers to the number of cracks per 10 cm.

The measurement deltas $\Delta\omega_1, \Delta\omega_2, \ldots, \Delta\omega_8$ are also treated as fuzzy variables. Fuzzy sets with Gaussian membership functions are used to define these input variables. Change in frequency (measurement delta) is calculated by using finite element simulations for a combination of 10 different locations and four different levels of damages (undamaged, slight damage, moderate damage, and severe damage). Crack densities greater than 20 represent very severe damage and are excluded when formulating the fuzzy system.

Rules for the fuzzy system are obtained by fuzzification of the numerical values obtained from finite element analysis. The fuzzy sets corresponding to $\Delta\omega_1, \Delta\omega_2, \ldots, \Delta\omega_8$ are generated by taking the $\Delta\omega$ values obtained by the finite element method (FEM) solution as midpoints of the membership function corresponding to a location of matrix cracking and damage level. This strategy for selecting the midpoint ensures that the maximum degree of membership ($\mu = 1$) for each fuzzy set occurs at the values of $\Delta\omega$ since the Gaussian function is highest at the midpoint. The standard deviation of the Gaussian membership functions are calculated using GA for optimization of the success rate as discussed next.

Typical resolution for the natural frequencies of a lightly damped mechanical structure is 0.1% [33]. Therefore, we consider a noise level of 0.1 to simulate data used for developing the genetic fuzzy system. Noise is added to the finite element simulations. Here the noise level $\alpha = 0.1$ corresponds to normal-quality data, $\alpha = 0.05$ to good-quality data, and $\alpha = 0.15$ to low-quality data.

By generating noisy frequency deltas and testing the fuzzy system for a known damage, we can define a success rate as

$$S_R = \frac{N_c}{N} \times 100 \tag{13}$$

Here N_c is the number of correct classifications among N number of noisy training samples. Since the midpoints of the fuzzy sets are tuned using FEM, the success rate is a function of the standard deviations of the Gaussian functions for the fuzzy system, i.e.,

$$S_R = S_R(\sigma_{ij}) \tag{14}$$

The design of the fuzzy system can be posed as an optimization problem and can be written in standard form [12] as

Maximize

$$S_R(\sigma_{ij})$$

subject to the move limits

$$\sigma^{\min} \le \sigma_{ij} \le \sigma^{\max}, \qquad i = 1, 2, \ldots, M \quad j = 1, 2, \ldots, P \tag{15}$$

The use of formal optimization to design the fuzzy system leads to an optimal diagnostic system, which provides the best results for the given structure, measurement set, and noise level in the data. GAs provide useful optimization tools since they can find

global optima and do not need gradient information that is required by most traditional optimization algorithms. Details of GAs can be found in [33].

The schematic representation of the FLS and the genetic fuzzy system are shown in Figures 1 and 4, respectively. Figure 1 shows the components of a classical FLS where data is fuzzified, the rules are evaluated, and the defuzzification is used to determine the final output of the system. For the current problem, inputs are measurement deltas ($\Delta\omega$) and the outputs are the damage presence, location, and size. Figure 4 shows how the FLS in Figure 1 is trained. Data from the finite element simulation is used to obtain the midpoints of the fuzzy system to maximize the success rate (S_R) by using the standard deviations as design variables. The outputs of the genetic fuzzy system are damage presence, location, and size, which are the same as those for the fuzzy system in Figure 1. However, the addition of learning using GA allows the genetic fuzzy system to optimize its rule base.

In this way, the "undamaged" level of damage is represented by one rule and the three damage levels "slight", "moderate", and "severe" at 10 locations are represented by a total of 10 rules each. Therefore, the complete matrix crack detection system can be represented by 31 rules. For example "slight" damage at 0–10% part of the beam can be written as "slight L_{10}". The rules are detailed and complex due to the large number of measurements and locations, and are given in Table 4 in terms of the midpoints and standard deviations of the fuzzy sets.

For the GA, the population size, crossover probability, mutation probability, and maximum number of generations are 20, 0.8, 0.05, and 30, respectively. These parameters are obtained by numerical experiments. Midpoints and standard deviations are two data sets required for the formulation of the genetic fuzzy system for damage detection of structures. These midpoints change for different structures and damage combinations since they are dependent on geometrical and material properties of the structure and on the damage model. Since this method is based on model-based diagnostics, an FEM of the structure needs to be developed. Once such an FEM is developed, GA is used with this model to design the fuzzy system. Once the genetic fuzzy system is developed, it can detect and isolate faults from measured frequencies. Table 5 shows the average success rate of the fuzzy system for different levels of noise in the data for $[\pm\theta_8/90_9]$ types of composites. Here, a training noise level of 0.1 was used. It can be seen that the fuzzy system is robust to quite high levels of noise in the data.

An interesting feature of the genetic fuzzy system is that though damage location accuracy can decline

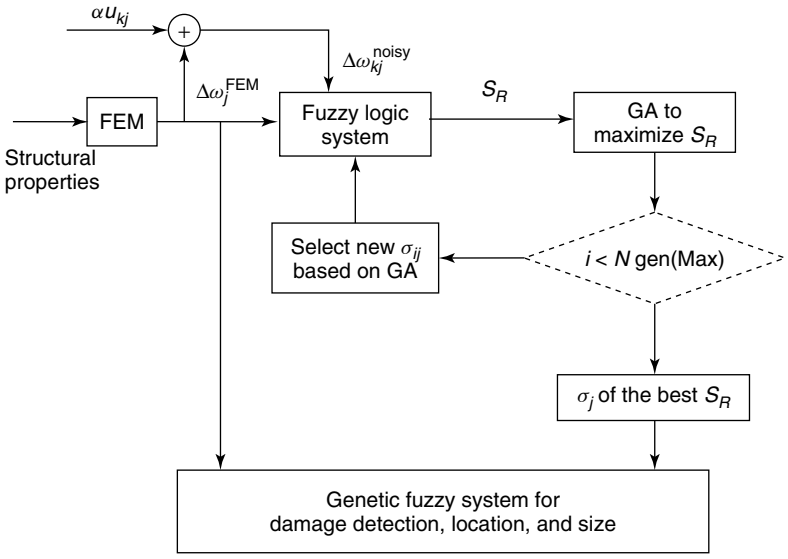


Figure 4. Schematic representation of a genetic fuzzy logic system.

Table 4. Rules for genetic fuzzy system (mean values, with standard deviations in brackets)

No.	Rule	$\Delta\omega_1$	$\Delta\omega_2$	$\Delta\omega_3$	$\Delta\omega_4$	$\Delta\omega_5$	$\Delta\omega_6$	$\Delta\omega_7$	$\Delta\omega_8$
1	Undamaged	0.00 (0.33)	0.00 (0.34)	0.00 (0.32)	0.00 (0.35)	0.00 (0.31)	0.00 (0.32)	0.00 (0.31)	0.00 (0.33)
2	Slight L_{10}	1.11 (0.34)	1.06 (0.33)	1.02 (0.34)	0.99 (0.33)	0.96 (0.32)	0.94 (0.31)	0.92 (0.34)	0.92 (0.34)
3	Slight L_{20}	1.03 (0.30)	0.85 (0.32)	0.80 (0.33)	0.82 (0.34)	0.88 (0.30)	0.93 (0.33)	0.94 (0.34)	0.92 (0.34)
4	Slight L_{30}	0.96 (0.32)	0.79 (0.30)	0.88 (0.31)	0.94 (0.31)	0.87 (0.32)	0.81 (0.31)	0.88 (0.32)	0.96 (0.31)
5	Slight L_{40}	0.90 (0.34)	0.85 (0.34)	0.94 (0.33)	0.82 (0.34)	0.87 (0.32)	0.96 (0.35)	0.85 (0.34)	0.87 (0.31)
6	Slight L_{50}	0.86 (0.30)	0.93 (0.30)	0.84 (0.34)	0.88 (0.33)	0.93 (0.31)	0.82 (0.30)	0.96 (0.33)	0.84 (0.30)
7	Slight L_{60}	0.82 (0.31)	0.97 (0.32)	0.80 (0.31)	0.95 (0.30)	0.82 (0.31)	0.94 (0.31)	0.86 (0.33)	0.93 (0.34)
8	Slight L_{70}	0.81 (0.31)	0.94 (0.34)	0.92 (0.34)	0.81 (0.34)	0.96 (0.32)	0.86 (0.33)	0.86 (0.30)	0.96 (0.31)
9	Slight L_{80}	0.79 (0.31)	0.87 (0.32)	0.98 (0.30)	0.91 (0.32)	0.81 (0.31)	0.89 (0.33)	0.96 (0.33)	0.87 (0.31)
10	Slight L_{90}	0.79 (0.33)	0.81 (0.30)	0.89 (0.33)	0.97 (0.33)	0.98 (0.34)	0.91 (0.31)	0.84 (0.32)	0.84 (0.32)
11	Slight L_{100}	0.79 (0.33)	0.79 (0.31)	0.80 (0.30)	0.82 (0.32)	0.85 (0.31)	0.89 (0.34)	0.93 (0.32)	0.96 (0.34)
12	Moderate L_{10}	2.13 (0.34)	2.03 (0.31)	1.95 (0.33)	1.89 (0.35)	1.84 (0.31)	1.79 (0.33)	1.77 (0.33)	1.75 (0.35)
13	Moderate L_{20}	1.97 (0.34)	1.63 (0.32)	1.53 (0.30)	1.58 (0.31)	1.69 (0.33)	1.78 (0.30)	1.80 (0.30)	1.75 (0.32)
14	Moderate L_{30}	1.83 (0.33)	1.52 (0.34)	1.68 (0.33)	1.80 (0.32)	1.67 (0.34)	1.56 (0.32)	1.68 (0.31)	1.84 (0.32)
15	Moderate L_{40}	1.72 (0.34)	1.62 (0.31)	1.80 (0.31)	1.57 (0.31)	1.66 (0.33)	1.84 (0.30)	1.62 (0.30)	1.67 (0.33)
16	Moderate L_{50}	1.64 (0.31)	1.78 (0.34)	1.61 (0.30)	1.68 (0.33)	1.77 (0.32)	1.58 (0.31)	1.84 (0.31)	1.61 (0.30)
17	Moderate L_{60}	1.58 (0.32)	1.86 (0.34)	1.54 (0.32)	1.82 (0.32)	1.58 (0.31)	1.79 (0.33)	1.64 (0.34)	1.77 (0.31)
18	Moderate L_{70}	1.55 (0.35)	1.80 (0.31)	1.77 (0.31)	1.54 (0.30)	1.83 (0.32)	1.64 (0.32)	1.66 (0.34)	1.84 (0.32)
19	Moderate L_{80}	1.53 (0.33)	1.67 (0.34)	1.89 (0.34)	1.74 (0.31)	1.54 (0.32)	1.71 (0.33)	1.84 (0.34)	1.67 (0.31)
20	Moderate L_{90}	1.52 (0.30)	1.56 (0.32)	1.70 (0.34)	1.85 (0.30)	1.88 (0.31)	1.75 (0.31)	1.61 (0.33)	1.61 (0.31)
21	Moderate L_{100}	1.52 (0.30)	1.52 (0.33)	1.53 (0.34)	1.57 (0.30)	1.62 (0.30)	1.70 (0.33)	1.77 (0.31)	1.84 (0.31)
22	Severe L_{10}	3.67 (0.32)	3.50 (0.34)	3.36 (0.31)	3.25 (0.34)	3.16 (0.32)	3.10 (0.34)	3.06 (0.31)	3.04 (0.31)
23	Severe L_{20}	3.40 (0.33)	2.83 (0.34)	2.65 (0.33)	2.73 (0.33)	2.92 (0.30)	3.08 (0.32)	3.10 (0.34)	3.02 (0.30)
24	Severe L_{30}	3.17 (0.32)	2.64 (0.31)	2.91 (0.30)	3.11 (0.31)	2.89 (0.30)	2.70 (0.33)	2.90 (0.34)	3.17 (0.31)
25	Severe L_{40}	2.98 (0.32)	2.81 (0.32)	3.10 (0.31)	2.72 (0.31)	2.87 (0.34)	3.18 (0.32)	2.82 (0.30)	2.89 (0.33)
26	Severe L_{50}	2.84 (0.30)	3.08 (0.32)	2.80 (0.30)	2.90 (0.32)	3.07 (0.30)	2.73 (0.34)	3.18 (0.32)	2.79 (0.33)
27	Severe L_{60}	2.74 (0.30)	3.21 (0.33)	2.67 (0.32)	3.14 (0.34)	2.73 (0.31)	3.10 (0.34)	2.83 (0.31)	3.06 (0.31)
28	Severe L_{70}	2.68 (0.30)	3.11 (0.30)	3.06 (0.33)	2.68 (0.33)	3.16 (0.32)	2.84 (0.32)	2.88 (0.34)	3.17 (0.34)
29	Severe L_{80}	2.65 (0.33)	2.89 (0.34)	3.26 (0.34)	3.01 (0.30)	2.68 (0.30)	2.96 (0.30)	3.17 (0.34)	2.89 (0.33)
30	Severe L_{90}	2.64 (0.31)	2.70 (0.31)	2.94 (0.30)	3.20 (0.32)	3.24 (0.30)	3.02 (0.34)	2.79 (0.33)	2.79 (0.30)
31	Severe L_{100}	2.63 (0.32)	2.64 (0.32)	2.66 (0.33)	2.72 (0.33)	2.82 (0.31)	2.94 (0.32)	3.07 (0.33)	3.18 (0.31)

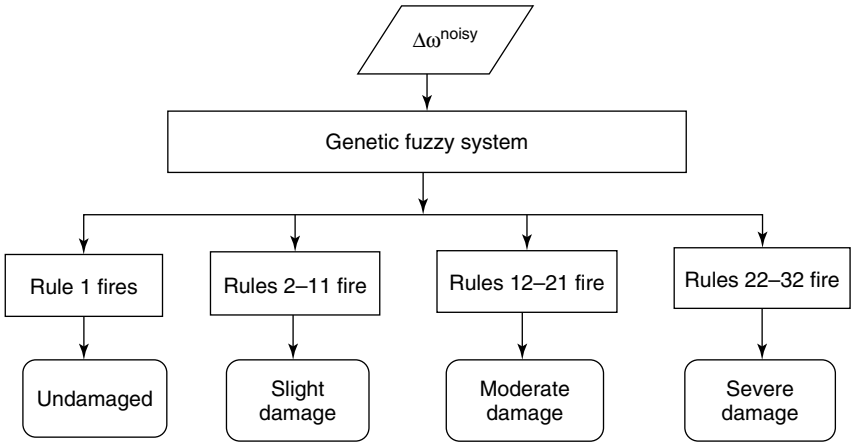


Figure 5. Damage detection and identification using genetic fuzzy system.

Table 5. Average success rate of genetic fuzzy system at different noise levels

Ply angle	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.15$
0	100	97.31	89.40
30	100	99.80	96.27
60	100	100	100

at higher noise levels, the estimate of damage detection remains good for high levels of damage [32]. Figure 5 shows how the rules in Table 4 can be used to detect whether damage is present and also to identify the size of damage.

5 CONCLUDING REMARKS

Fuzzy logic provides a powerful approach for solving the inverse problem of structural damage detection from noisy data. The use of linguistic rules and a built-in fuzzifier at the front end of the fuzzy system provides an elegant approach to handling uncertainty in decision making. Simple fuzzy systems can be easily developed for most damage detection problems, using a method to relate the most likely situations to rules. As the number of measurements, damage locations, and number of damage types increases, the genetic fuzzy system can be considered. We also note that it is also possible to combine fuzzy rules generated from data to those generated from expert knowledge, a feature that is not possible with the neural network approach.

Fuzzy systems have enormous potential in SHM as they can handle uncertainty in measurements and models. They can be used for addressing uncertainty in finite element models due to variations in material and geometric properties. However, fuzzy systems suffer from the "curse of dimensionality" as the number of measurements and damage levels increase since the number of fuzzy rules can become very large. In general, it is advantageous to use fuzzy systems for SHM where the number of measurements and damage levels is relatively small.

ACKNOWLEDGMENTS

I acknowledge my former PhD student Prashant Pawar for his work on the genetic fuzzy system and project assistant Arun Kumar for help in preparing the manuscript.

REFERENCES

[1] Boller C. Next generation structural health monitoring and its integration into aircraft design. *International Journal of Systems Science* 2000 **31**(11):1333–1349.

[2] Okafor AC, Dutta A. Optimal ultrasonic pulse repetition rate for damage detection in plates using neural networks. *NDT and E International* 2001 **34**(7):469–481.

[3] Xu YG, Liu GR, Wu ZP, Huang XM. Adaptive multilayer perception networks for detection of

cracks in anisotropic laminated plates. *International Journal of Solids and Structures* 2001 **38**(32–33): 5625–5645.

[4] Bar HN, Bhat MR, Murthy CRL. Identification of failure modes in GFRP using PVDF sensors: ANN approach. *Composite Structures* 2003 **65**:231–237.

[5] Choi SW, Song EJ, Hahn HT. Prediction of fatigue damage growth in notched composite laminates using an artificial neural network. *Composites Science and Technology* 2003 **63**(5):661–675.

[6] Reddy RRK, Ganguli R. Structural damage detection in a helicopter rotor blade using radial basis function neural networks. *Smart Structures and Materials* 2003 **12**:232–241.

[7] Zhang Z, Friedrich K. Artificial neural networks applied to polymer composites: a review. *Composite Science and Technology* 2003 **63**(14):2029–2044.

[8] Sherratt PJ, Panni DC, Nurse AD. Damage assessment of composite structures using inverse analysis and genetic algorithms. *Damage Assessment of Structures* 2001 **204**(2):409–418.

[9] Nag A, Mahapatra DR, Gopalakrishnan S. Identification of delamination in composite beams using spectral estimation and genetic algorithm. *Smart Materials and Structures* 2002 **11**(6):899–908.

[10] Xu YG, Liu GR, Wu ZP. Damage detection for composite plates using lamb waves and projection genetic algorithm. *AIAA Journal* 2002 **40**(9): 1860–1866.

[11] Zadeh L. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems* 1996 **4**(2):103–111.

[12] Pawar PM, Ganguli R. Genetic fuzzy system for damage detection in beams and helicopter rotor blades. *Computer Methods in Applied Mechanics and Engineering* 2003 **192**:2031–2057.

[13] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are uniform approximators. *Neural Networks* 1989 **2**(3):359–366.

[14] Hong XL, Chen PCL. The equivalence between fuzzy logic systems and feedforward neural networks. *IEEE Transactions on Neural Networks* 2000 **11**(2):356–365.

[15] Ramu SA, Johnson VT. Damage assessment of composite structures using fuzzy logic integrated neural-network approach. *Computers and Structures* 1995 **57**(3):491–502.

[16] Nyongesa HO, Otenio AW, Rosin PL. Neural fuzzy analysis of delaminated composites from shearography imaging. *Composite Structures* 2001 **54**(2–3): 313–318.

[17] Sawyer JP, Rao SS. Structural damage detection and identification using fuzzy logic. *AIAA Journal* 2001 **38**(12):2328–2335.

[18] Ganguli R. A fuzzy logic system for ground based structural health monitoring of a helicopter using Model Data. *Journal of Intelligent Materials Systems and Structures* 2001 **13**(6):397–408.

[19] Sazonov ES, Klinkhachorn P, Gangarao HVS, Halabe UB. Fuzzy logic expert system for automated damage detection from changes in strain energy mode shapes. *Nondestructive Testing and Evaluation* 2001 **18**(1):1–20.

[20] Cordon O, Herrera F, Villar P. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base. *IEEE Transactions on Fuzzy Systems* 2001 **94**(4):667–674.

[21] Cordon O, Gomide F, Herrera F, Hoffmann F, Magdalena L. Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems* 2004 **141**:5–31.

[22] Kosko B. *Fuzzy Engineering*. Prentice Hall: 1997.

[23] Wang LX, Mendel JM. Generating fuzzy rules by learning from examples. *IEEE Transactions on Fuzzy Systems* 1992 **22**(6):103–111.

[24] Ross TJ. *Fuzzy Logic with Engineering Applications*. John Wiley & Sons: 2004.

[25] Harris J. *Fuzzy Logic Applications in Engineering Science*. Springer: 2005.

[26] Mukaidono M, Kikuchi H. *Fuzzy Logic for Beginners*. World Scientific: Singapore, 2001.

[27] Mengali G. The use of fuzzy logic in adaptive flight control systems. *The Aeronautical Journal* 2000 **104**(1031):31–37.

[28] Chi Z, Yan H, Pham T. *Fuzzy Algorithms: with Applications to Image Processing and Pattern Recognition*. World Scientific: Singapore, 1998.

[29] Chi Z, Yan H. ID3 derived fuzzy rules and optimized defuzzification for handwritten character recognition. *IEEE Transactions on Fuzzy Systems* 1996 **4**(1):24–31.

[30] Abe S, Lin MS. A method for fuzzy rules extraction directly from numerical data and its application to pattern recognition. *IEEE Transactions on Fuzzy Systems* 1995 **3**(1):18–28.

[31] Polyzois D, Raftoyiannis G, Ibrahim S. Finite elements for the dynamic analysis of tapered composite poles. *Composite Structures* 1998 **43**: 25–34.

[32] Pawar PM, Ganguli R. Matrix crack detection in thin-walled composite beam using genetic fuzzy system. *Journal of Intelligent Materials Systems and Structures* 2005 **16**(5):395–409.

[33] Goldberg D. *Genetic Algorithms in Search, Optimization and Machine Learning*. Pearson Education: 2001.

FURTHER READING

Friswell MI, Penny JET. Is damage detection using vibration measurements practical? In *EUROMECH 65 International Workshop: DAMAS 97; Structural Damage Assessment Using Advanced Signal Processing Procedures*. Sheffield.

Chapter 53
Operational Loads Sensors

William F. Ranson and Reginald I. Vachon
Direct Measurements, Inc., Atlanta, GA, USA

1 INTRODUCTION

Load sensors for structural health monitoring (SHM) measure force, torque, pressure, and strain. The data are translated into information to determine the effect of loads and/or vibration on operating dynamic systems and static infrastructures. Sensors used to make these measurements are accelerometers, frequency devices, or strain gauges. The theory of operation of each is fundamental. Nonetheless, there are a number of manifestations of theory for each by commercial companies and research laboratories providing accelerometer, frequency, and strain gauge devices. Some of these devices are described and the general theory for the major types of accelerometers and strain gauges are presented.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

The transition from usage monitoring to individual component damage tracking is the challenge for these types of sensors. The areas of regime recognition, loads, strain prediction, and component serialization need to improve for enhanced structural health assessment. Of these three requirements, loads and strain prediction along with serialization represent the direct application for SHM.

The current developments of sensors of this type are to overcome the limitations of complex instrumentation, long wiring lengths, and in rotating components of the elimination of slip rings. Strain sensor development is focused on wireless technology, serialization, and energy harvesting for power. The accelerometers used in health usage monitoring systems (HUMS) such as the Navy V-22 and the Army Apache and Blackhawk helicopters are incorporating regime recognition in the onboard data acquisition systems.

2 ACCELEROMETERS

2.1 Overview

Accelerometers are one of the basically three vibration measuring instruments, which measure displacements, vibrations, and accelerations as a result of harmonic motion of a base relative to a mass system. The most common types are the displacement and acceleration measurement instruments. They consist

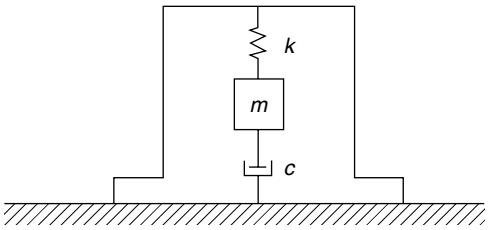


Figure 1. Schematic of an accelerometer.

of a case containing a spring mass damper system of the type shown in Figure 1 and a device measuring the displacement of the mass relative to the case. The mass is usually constrained to move along a given axis. Measurement of the mass relative to the case is measured electrically. When the frequency of the base is low relative to the spring mass damper of the system, the instrument is proportional to the acceleration of the case.

2.2 Theory

A single-degree-of-freedom mechanical model with viscous damping as shown in Figure 1 can be used to describe the mechanical behavior of instruments to measure force, pressure, and motion. Accelerometers are a special case of this type of instrument where a component with frequencies far below the natural frequency of the transducer, the relative motion between the base and the transducer, is proportional to the base acceleration. Figure 1 is a schematic that represents the vibration of a base by $y(t) = Y \sin \omega t$. From Newton's second law of motion [1], the equation of motion for the seismic mass is obtained. Thus

$$m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} + c\frac{\mathrm{d}z}{\mathrm{d}t} + kz = m\omega^2 Y \sin \omega t \qquad (1)$$

where m is the mass of the seismic body, k is the linear spring constant, and c is the viscous damping constant, $z = x - y$ is the relative displacement between seismic mass and the base, and $z, \mathrm{d}z/\mathrm{d}t, \mathrm{d}^2 z/\mathrm{d}t^2$ are the relative displacement, velocity, and acceleration.

A steady-state solution is assumed in the form $z = Z \sin(\omega t - \phi)$ and $Z = m\omega^2 Y$.

The solution yields

$$Z = \frac{Y\left(\omega/\omega_n\right)^2}{\sqrt{\left(1 - \left(\omega/\omega_n\right)^2\right)^2 - \left(2\zeta\left(\omega/\omega_n\right)\right)^2}} \qquad (2)$$

$$\tan\phi = \frac{2\zeta\left(\omega/\omega_n\right)}{1 - \left(\omega/\omega_n\right)} \qquad (3)$$

where $\zeta = c/c_c$ is the viscous damping ratio and ω_n is the seismic mass natural frequency.

Accelerometers have a small frequency ratio $\omega/\omega_n \ll 1$ and $\zeta \ll 1$ (small damping); then

$$Z = \frac{\omega^2 Y}{\omega_m^2} = \frac{\text{Base acceleration}}{\omega_n^2} \qquad (4)$$

The crystal accelerometer behaves like an underdamped, spring mass system with a single degree of freedom. Equation (1), a classical second-order differential equation, can be used to describe the behavior of the electromechanical crystal accelerometer system.

Piezoelectric crystals subjected to a force exhibit an electrical charge. Slicing a quartz crystal, for example, relative to an xyz axis results in a sensor producing a charge based on the direction of the applied force. This fact is used to design accelerometers used in dynamic systems as well as for quasi-static measuring systems using a quartz crystal with special signal conditioning. Examples can be given for quasi-static measurements over periods of minutes and even hours.

Piezoelectric crystal sensors are of high impedance or low impedance. Low-impedance sensors have a built-in charge-to-voltage converter and require an external power source to power the device and to separate the output voltage from the bias dc voltage. These low-impedance systems are tailored to a particular application. High-impedance units require a charge amplifier or external impedance converter for charge-to-voltage conversion. The high-impedance sensors are more versatile than the low-impedance devices.

Time constant and drift are two terms associated with the use of charge amplifiers with piezoelectric devices. The time constant is the discharge time of an ac-coupled circuit. One time constant is the period of time the input decays by 63–37% of its original value.

The product of the time constant resistor R_t and range capacitor C_t yields the time constant. Figure 2 depicts a high-gain inverting voltage amplifier typically used with crystal sensors.

The amplifier yields an output voltage that can be expressed by

$$V = -\frac{Q}{C_\mathrm{r}}\left[\frac{1}{\left[1+\frac{1}{A}C_\mathrm{r}(C_t+C_\mathrm{r}+C_\mathrm{c})\right]}\right] \quad (5)$$

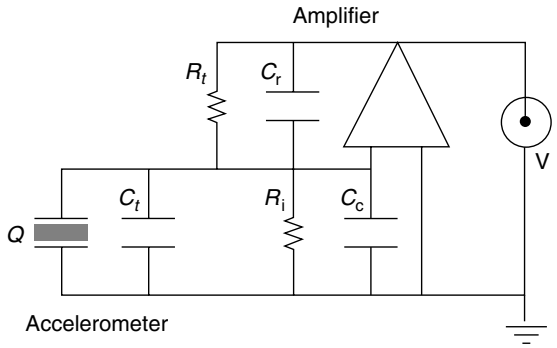


Figure 2. Typical accelerometer circuit.

where C_t is the sensor capacitance, R_i is the cable and sensor resistance, C_c is the cable capacitance, C_r is the range capacitor, R_t is the time constant, A is the open-loop gain, and Q is the charge generated.

When the open-loop gain is high, the output voltage is a function of the input charge and range capacitance.

$$V = -\frac{Q}{C_\mathrm{r}} \quad (6)$$

High-impedance crystals coupled with an amplifier with a high open-loop gain produce a usable output voltage. Quartz-based piezoelectric sensors are widely used because they operate up to temperatures of 498.9 °C, and stresses of up to approximately 137 895 Pa, ultrahigh insulation resistance of 1014 Ω, which permits low frequency (1 Hz), negligible hysteresis, high rigidity, and high linearity.

There are a variety of accelerometers. Listing includes capacitive spring mass based; electromechanical servo; null balance; resonance; magnetic induction; optical; surface acoustic wave; laser; bulk micromachined piezoresistive; bulk micromachined capacitive; capacitive spring mass based; piezofilm; piezoelectric; shear mode accelerometer; thermal

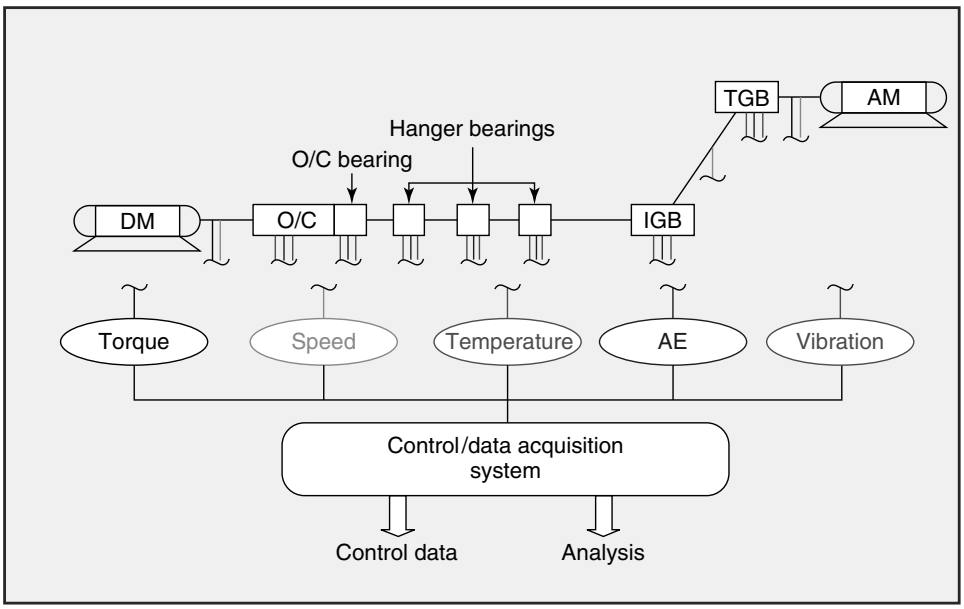


Figure 3. Schematic of the test stand (courtesy of Dr Abdel Bayoumi, Department of Mechanical Engineering, University of South Carolina). [Reproduced with permission from Dr. Abdel Bayoumi.]

Figure 4. Photograph of the test stand (courtesy of Dr Abdel Bayoumi, Department of Mechanical Engineering, University of South Carolina). [Reproduced with permission from © Dr. Abdel Bayoumi.]

(submicrometer CMOS process); and modally tuned impact hammers.

2.3 Example of operational load sensors for SHM

An excellent example of operational load sensors is the University of South Carolina (USC) drivetrain test stand laboratory. The test stands (see Figures 3 and 4) in this facility are capable of testing drivetrain components (bearings, gearboxes, swashplates, oil coolers, and shafts) of AH-64, UH-60, and ARH-70 aircraft. Additionally, they are able to provide up to 150% of aircraft power at full speed for the components under testing. These stands are also capable of handling shaft misalignment requirements while remaining safe. Other stands at USC include AH-64, ARH-70, CH-47, and UH-60 hydraulic pump stands, as well as AH-64, ARH-70, CH-47, and UH-60 main rotor swashplate bearing assembly stands. All test stands utilize several data acquisition systems, including current in-flight monitoring systems such as the HUMS (Multi Signal Processing Unit (MSPU) and/or Integrated Mechanical Diagnostics (IMD)-HUMS), as well as a specialized laboratory data acquisition system capable of recording torque, speed, temperature, vibration, and acoustic emissions. All test stands are controlled on the basis of monitored data measures of torque, speed, and temperature, which are collected every 2 s. All vibration and acoustic emissions data are collected every 2 min. Operational load sensors of this type have been developed for rotary wing aircraft. References 2–5 describe applications in maneuver regime recognition and SHM using drivetrain, gearbox, engine, rotor track, and balance accelerometers as part of the HUMS onboard system.

3 STRAIN GAUGES AND STRAIN MEASUREMENT

3.1 Overview

Strain gauges measure the relative displacements of a small straight line segment between usually undeformed and deformed configurations of a body subjected to loads. The electrical resistance strain gauge is based on this principle and was first established by Lord Kelvin in 1856 and accounted for more than 80% of the stress analysis applications in the 1980s. Also, electrical resistance strain gauges are widely used as sensors in transducers that have been developed to measure load, torque, and pressure. Historically, strain measurement for operational SHM has relied on electrical resistance strain gauge applications. In recent years, optical-based

techniques have been developed to measure strains. Some of these very useful laboratory technologies [6, 7] include digital image correlation, interferometry, photoelasticity, holographic interferometry, and fiber-optic strain gauges.

The concentration here is the electrical resistance strain gauge and one optical technique suitable for in-the-field measurement and verification of engineered residual strains intended to enhance fatigue life for operating systems such as fastener holes in aircraft structures. The theory of each is discussed and examples are given.

3.2 Electrical resistance strain gauge

The characteristics for electrical resistance gauges are listed in [6, 8] based on the premise that they are usually the most economical. Some of the gauge characteristics are as follows:

1. The gauge constant should be stable with respect to both temperature and time.
2. Strain measurement accuracy of $\pm 1\,\mu\mathrm{m\,m^{-1}}$ (μin./in.) over a range of $\pm 5\%$ strain.
3. The gauge length and width should be small.
4. The inertia of the gauge should be minimal for the measurement of dynamic strains.
5. The response or output of the gauge should be linear over the entire strain range of the gauge.

These characteristics are not only criteria but also constraints predicated on the limitations of the electrical resistance gauge. They also provide a basis for comparison and contrasting of other strain gauges such as inductance, capacitance, and optical gauges.

3.3 Theory of electrical resistance strain gauge

The electrical resistance strain gauge uses the analog of the change in resistance to resistance ratio to translate the change in resistance of the wire gauge as it undergoes strain into strain readings. The strain sensitivity (S_A) is termed the *gauge factor*. This is expressed by

$$S_A = \frac{dR/R}{dl/l} \tag{7}$$

where dR is the change in resistance, R is the initial resistance, dl is the length change of a small line segment, and l is the initial length of a small line segment.

Equation (7) can be now expressed in terms of the strain:

$$\varepsilon_g = S_A \frac{dR}{R} \tag{8}$$

where $\varepsilon_g = dl/l$ is the gauge strain from the measured resistance change.

Theoretically, the change in the resistance of the gauge is a function of surface deformations. Real considerations that affect the change in resistance include the adhesive used to affix the gauge to the surface, stability of the surface, temperature, and the material. Also, axial strain measurement is not sufficient to achieve complete analysis. A combination of gauges is required to yield orthogonal strains and associated shear strains. The electrical resistance strain gauge has wide use and utility. There are a variety of gauge configurations and wire filaments used to achieve stability, temperature sensitivity, and strain characteristics of the sensor. Many strain gauge filament materials degrade over time and are temperature sensitive. Thus, long-term continuous applications require attention to drift and temperature compensation. Selecting a gauge is based on the gauge characteristics that include gauge factor, thermal coefficient of resistively, resistance, and temperature coefficient of gauge factor. There are several manufacturers of gauges and the wide variety of gauges can be appreciated by reviewing available products.

3.4 Example of operational strain gauge application

The need exists to improve the accuracy of load/strain prediction by directly measuring data in flight. Microstrain has demonstrated a sensor that relies on piezoelectric energy harvesters capable of measuring strain, acceleration, and temperature. This type of sensor is a powerful combination of real-time sampling rates up to 4 KHz and onboard data storage. Microstrain sensors are configured to be efficient when using energy to sample, record, and communicate over a bidirectional radio link. Microstrain has

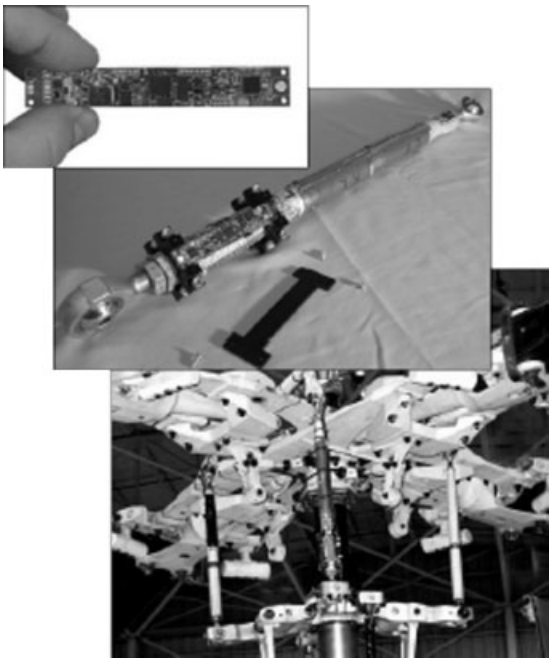


Figure 5. Wireless strain gauge on a Bell pitch link.

successfully demonstrated this technology to measure operational loads/strains on a Bell 412 helicopter pitch link as shown in Figure 5 [9].

3.5 Theory of optical strain gauge

The theoretical basis for the optical strain gauge is shown in Figure 6. Straight line segments are recorded in the undeformed and deformed configurations with the end points *MNOP* mapped to the endpoints $M^*N^*O^*P^*$ [10]. The strains along these orthogonal directions are calculated from the following formulas:

$$E_{MN} = \frac{M^*N^* - MN}{MN} \tag{9}$$

$$E_{NO} = \frac{N^*O^* - NO}{NO} \tag{10}$$

$$E_{OP} = \frac{O^*P^* - OP}{OP} \tag{11}$$

$$E_{PM} = \frac{P^*M^* - PM}{PM} \tag{12}$$

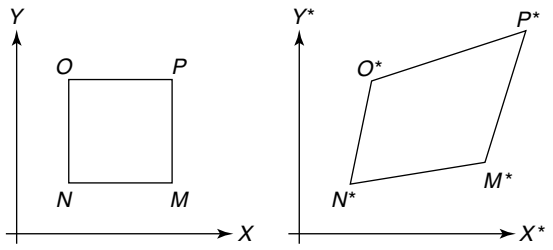


Figure 6. Optical strain gauge in undeformed and deformed configurations.

This type of strain gauge [11] has further properties of serialization as shown in Figure 7. The encoded serialization is shown in the boundaries of the gauge similar to a bar code. The target gauge can be in the form of a thin polymer with the gauge design laser machined into the multilayer polymer to achieve the light and dark patterns or the gauge can be laser bonded onto the surface. The polymer gauge can be affixed to the surface with M-Bond 200, which is the same process used for thin-film electrical resistance gauges. The other alternative application is to laser bond the pattern onto the surface.

This is accomplished by spraying the surface with fine metallic particles suspended in a vehicle of water or alcohol in a normal hand-held aerosol spray can. When the surface has dried, a low-power laser is used to fuse the particles to the surface to create the pattern. The surface is then cleaned with a cloth and water. A third application is to paint the surface with a specialized paint that discolors in response to exposure to a low-power laser. The result is that the gauge is integral to the paint layer. Examples of the

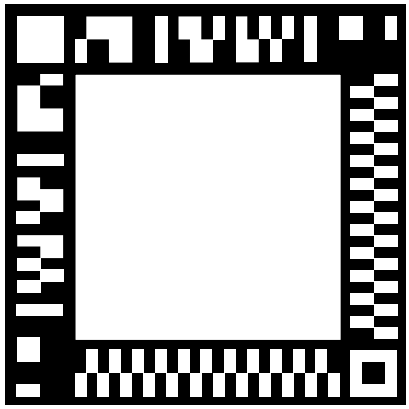


Figure 7. Gauge serialization.

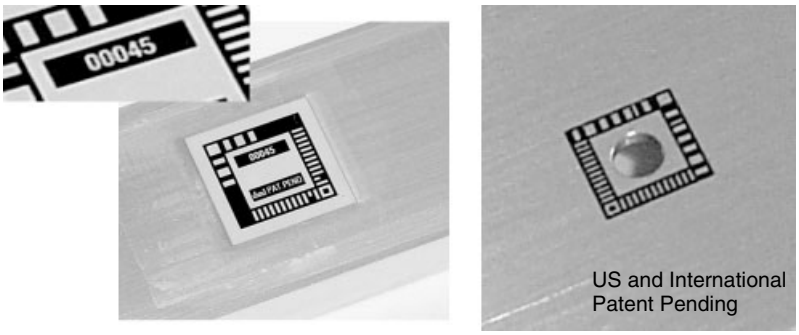


Figure 8. (a) Polymer gauge with serialization and (b) laser-bonded gauge.

polymer gauge and the gauge laser bonded around a hole are shown in Figure 8(a) and (b).

3.6 Features of the technology

(i) Self-contained strain/fatigue monitoring system; (ii) minimal skilled labor required; (iii) nonlinear, multicomponent strain sensor: two orthogonal extensional strains and shear strain from a single gauge; (iv) resting system error $<5\,\mu\varepsilon$; (v) measurement repeatability: $<30\,\mu\varepsilon$; (vi) accuracy increases as strain increases; (vii) temperature compensation enabled; (viii) no fragile electrical components, connection free; (ix) line-of-sight target acquisition; (x) material independent (works on plastics, metals, composites, etc.); (xi) gauge application process does not denigrate material/mechanical properties of host material; and (xii) gauge contains binary encoded information (for gauge location, serial number, or asset management) up to 4 billion unique numbers for each gauge.

3.7 Cold-working validation using optical strain gauge

A 7075-T6 specimen with five fastener holes was used to demonstrate the application of the technology to validate cold working [11]. Wire-free Direct Measurements, Inc. (DMI) gauges were applied around each hole as shown in Figure 9 (front) and baseline readings made and recorded. The specimen

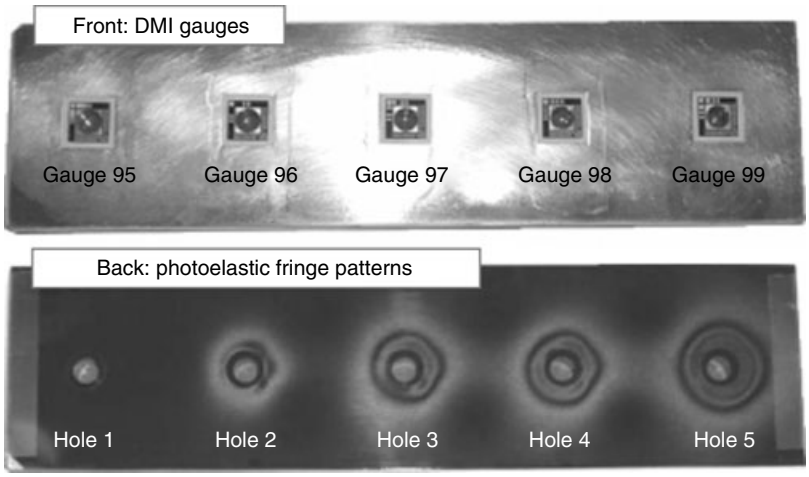


Figure 9. 7075-T6 test specimen front and back.

Table 1. Cold work by hole and gauge

Hole #	CX type	DMI gauge #
1	None	95
2	2%CX (under expanded)	96
3	CB minimum (minimum standard)	97
4	CB maximum (maximum standard)	98
5	CA nominal (over expanded)	99

CX, Cold Work Expansion

was cold worked by Fatigue Technology Inc. (FTI). Prescribed expansions were achieved according to Table 1. After FTI completed cold-work expansion, each gauge was read and the residual strain was measured and recoded. This summary presents the results of the strain measurements and shows the applicability of DMI technology to cold-work expansion.

Residual strain measurements were made on both the inner and outer boundary of each gauge and measured strain components E1, E2, E3, and E4 were recorded according to Figure 9. The holes and corresponding gauge codes are shown in Table 1 and Figure 10.

The DMI gauge outer and inner boundary measurements are shown in Figures 11 and 12. This demonstrates that the residual strain magnitudes follow the trend in photoelastic patterns in Figure 9. Note that the right-hand side fringe order is slightly higher in holes 3, 4, and 5. Accordingly, the strain component E2 measured on the right-hand side of gauges 97, 98, and 99 is also slightly higher. In this figure, gauge 98 fails to indicate higher E2 components due to inner

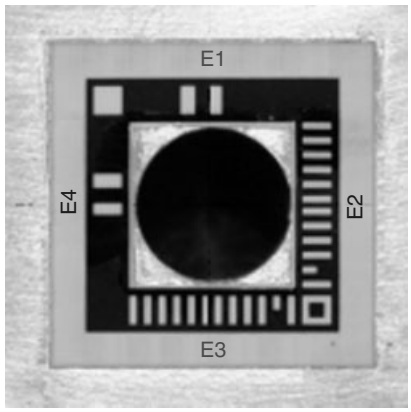


Figure 10. Gauge 96 surrounding hole 2.

boundary gauge damage during cold-work expansion. Nonetheless, the remaining components give sufficient indication of the degree of cold working. In addition, DMI gauges can be produced with dimensions that better match holes sizes, thereby avoiding damage.

4 CONCLUDING REMARKS

Operational load sensors that measure force, torque, pressure, and strain have been successively employed in operating mechanical systems for many years. Specific applications of accelerometers that are used in HUMS systems for rotary wing aircraft measure data associated with exceedance monitoring, rotor track and balance, operational usage, regime recognition, and drivetrain diagnostics. Strain gauges are ideal for dynamic load measurements such as the

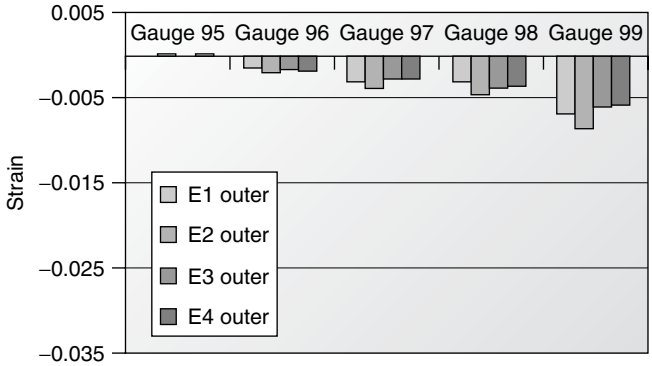


Figure 11. Outer boundary gauge residual strains.

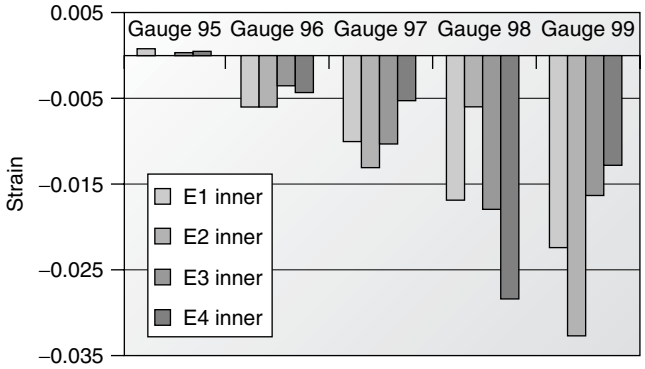


Figure 12. Inner boundary gauge residual strains.

axial load application for the pitch link illustrated in this article. The electrical resistance strain gauge has been an important sensor for measurements in full-scale fatigue tests in both fixed and rotary wing aircraft. The need to develop small scale gauges of this type for operational aircraft has led to the development of energy harvesting wireless strain gauges such as the one by Microstrain. Strain gauge applications of this type measure strains that can be calibrated to axial, bending, pressure, and torque loading. The application of the optical strain gauge while used to measure strains due to cold working are not a measure of loads directly. Cold working of fastener holes increases aircraft fatigue life; however, due to overloads, these compressive residual strains can be relaxed, thus reducing the expected fatigue life. The optical strain gauge measures the compressive strain relaxation and is thus an indirect measure of the effects of overload conditions.

REFERENCES

[1] Curtis D. *Johnson Process Control Instrumentation Technology*. Prentice-Hall: Upper Saddle River, NJ, 1993.

[2] Brandt G, Moon S, Miller SM. Manuever regime recognition development and verification for H-60 structural monitoring. *Presented at the American Helicopter Society 63rd Annual Forum*. Virginia Beach, VA, 1–3 May 2006.

[3] Brandt G. Moon S. Development of a fatigue tracking program for navy rotary wing aircraft. *Presented at the American Helicopter Society 50th Annual Forum*. Washington, DC, 11–13 May 1994.

[4] Dora R, Baker T, Hess R. Application of the IMD HUMS to the UH-60A Blackhawk. *Presented at the American Helicopter Society 58th Annual Forum*. Montreal, QC, 11–13 June 2002.

[5] Hiatt DS, Hayden R. Concepts for certifying a data-driven HUM system. *Presented at the American Helicopter Society 58th Annual Forum*. Montreal, QC, 11–13 June 2002.

[6] Kobayashi AS (ed). *Handbook on Experimental Mechanics*, Society for experimental Mechanics, Prentice-Hall: Englewood Cliffs, NJ, 1987.

[7] Sohn H, *et al. A Review of Structural Health Monitoring Literature from 1996–2001*. Los Alamos National Laboratory, report LA-13976-MS, 2003. http://www.findarticles.com/p/articles/mi_qa5348/is_200403/ai_n21346075/pg_4.

[8] Dally JW, Riley WF. *Experimental Stress Analysis, Fourth Edition*. College House Enterprises, LLC: Knoxville, TN, 2005.

[9] Maley S, Plets J, Phan ND. US Navy roadmap to structural health and usage monitoring—the present and future. *Presented at the American Helicopter Society 63rd Annual Forum*. Virginia Beach, VA, 1–3 May 2007.

[10] Novozhilov V. *Foundations of the Non-Linear Theory of Elasticity*. Graylock Press: New York, 1953.

[11] Ranson WF, Vachon RI. Crack detection and growth monitoring in holes using DMI SR-2 technology. *Paper # 17 2007 SEM Annual Conference*. Springfield, MA, 3–6 June 2007.

Chapter 56

Damage Presence/Growth Monitoring Sensors

Hua Gu[1] and Ming L. Wang[2]

[1] *Advanced Structures Group, Caterpillar Production System Division, Caterpillar Inc., Peoria, IL, USA*
[2] *Department of Civil and Materials Engineering, University of Illinois, Chicago, IL, USA*

1 INTRODUCTION

Most damage detection methods are based on the fact that damage will change the stiffness, mass, or energy dissipation properties of a system. As a result, the measured dynamic response will change as well. Most global damage detection methodologies suffer from lack of sensitivity. Since damage is a local phenomenon, it may not tremendously influence the global response of a structure, which usually deals with lower frequencies measured during vibration tests. Other environmental and operational variations, such as varying temperature, moisture, loading conditions, and the complexity of the structure itself, make damage detection more challenging [1].

The employment of Lamb waves in nondestructive testing has attracted more and more attention in structural health monitoring research community. Lamb waves are able to propagate over a long distance with very little amplitude loss. If a receiving transducer is positioned at a remote location on the structure, the received signal contains information about the integrity of the line between the transmitting and receiving transducers. The test thereby monitors a line rather than a point and so considerable testing time may potentially be saved [2, 3].

Perhaps the most important work in Lamb wave inspection is conducted by three groups of researchers—Peter Cawley's group from the Imperial College, London, Joseph Rose's group from the Pennsylvania State University, and Victor Giurgiutiu's group from the University of South Carolina. Work from these researchers can be found in numerous literatures [2–34].

The excitation and reception of Lamb waves in structures can be carried out by using interdigitated transducers (IDTs). These transducers have a comb structure composed of sets of periodically distributed fingers with a spatial period equal to the excited

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

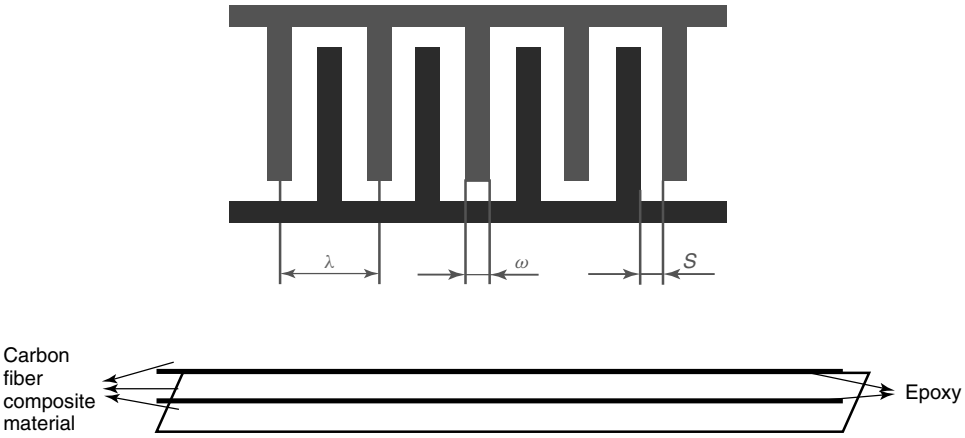


Figure 1. Schematic of a PVDF IDT.

Lamb wavelength [12, 35]. It is a common application to have an interdigital electrode on the surface of a piezoelectric material form an IDT. Owing to its broad bandwidth and low unit price, polyvinyldine fluoride (PVDF) film has envisioned itself to be a suitable substrate for surface and guided-wave transducers that are required to couple electrical energy with the piezoelectric material to detect the wave. The advantages of using PVDF IDTs are as follows. Firstly, PVDF IDTs are low profile and unobtrusive compared with other types of Lamb wave devices. Secondly, PVDF-based IDTs are flexible, so that they can be used on convex or concave surfaces such as pipes and pressure vessels [7].

Lamb wave signals are usually complex and nonstationary. At least two Lamb wave modes can propagate at a given frequency. Conventional fast Fourier transform (FFT) is unable to reveal the nature of Lamb wave signals containing frequency information that varies with time. To solve this problem, time–frequency representations (TFRs) (*see* **Time–frequency Analysis**) are commonly employed for signal analysis in Lamb wave applications. The continuous wavelet transform (CWT) is used as a TFR in this work.

2 DESIGN AND FABRICATION OF A PVDF IDT

A successfully designed PVDF IDT will have a finger pattern sitting on top of a PVDF film as indicated in Figure 1. λ, w, and s represent wavelength, finger width, and finger spacing, respectively. These parameters along with the length of the transducer are the criteria that need to be considered during a design process.

It has been proved that as long as $50\% \leq \frac{w}{w+s} \leq 90\%$ holds true, there is little difference among different finger widths [19]. As for the length of the transducer, a single pair of electrodes/fingers cannot excite surface acoustic waves very efficiently. Therefore, several pairs of electrodes/fingers are commonly seen in an interdigital transducer. A long transducer with many pairs of electrodes/fingers tends to be efficient for exciting and receiving signals only over a narrow frequency range. Consequently, it is also easier to generate a certain wave mode. However, it is practically not possible to build a sensor with infinite length. Besides, the more finger pairs the sensor has, the narrower the bandwidth of the sensor becomes. It is undesirable to design a sensor with very narrow bandwidth that limits its applications. Therefore, it is common to have IDTs with five or six pairs of electrodes/fingers.

The wave length λ is determined depending on the dispersion curves of the materials to be tested. Figure 2 displays the phase velocity dispersion curves of the tested structure used in this study. It is a laminated structure composed of 3 layers of carbon fiber reinforced polymer (CFRP) material. As the dispersion curves have explained themselves, the wave propagation in this structure, composed of anisotropic material, is much less dispersive than that

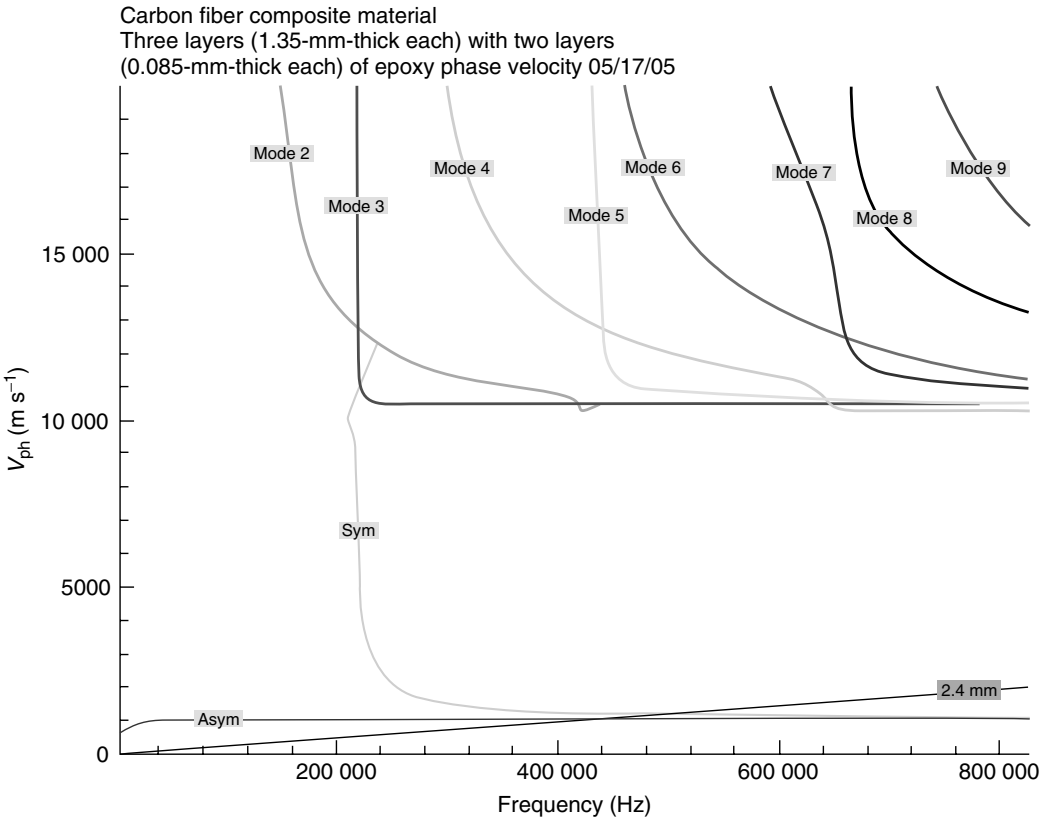


Figure 2. Phase velocity of laminated CFRP plate.

in an isotropic material, such as steel. Since the wave length of IDT is determined by $\lambda = \frac{c_{ph}}{f}$, a straight design line starting from the origin of phase velocity diagram is drawn to help the design process. The slope of the line represents a certain wavelength that if intersects with a dispersion mode will excite that mode at the corresponding frequency. To simplify the wave signal that is to be received, a sensor with a wavelength of 2.4 mm is designed, which ideally only generates symmetric and antisymmetric wave modes in a large frequency band.

The fabrication of PVDF IDTs is realized by using the etch-back photolithography technology adopted from semiconductor industry, which is carried out in a photolithography bay inside a clean room. PVDF films with gold coating on one of the surfaces are acquired from manufacturer. The reason for choosing gold is twofold. Firstly, backing with a high-density material such as gold enhances the performance of PVDF film at low frequencies. Secondly, gold has superb conductivity. After the surface is cleaned, a thin layer of photoresist is evenly spun on top of the gold coating. The finger pattern design is carefully printed out on transparencies and transferred to the photoresist layer through exposing. After being developed, the pattern then stays on the film. The finger pattern is revealed by etching out gold not covered by the resist and the transducer is finalized by removing the photoresist.

3 EXPERIMENTAL SETUP

The test sample structure is 915 mm long, 51 mm wide, and 4.22 mm thick. Through voids of different sizes have been introduced to the middle layer of the laminated structures. They are 0, 1, and 5 mm wide. Figure 3 shows the schematic of the experimental setup. A tone burst sinusoidal signal with a Hanning

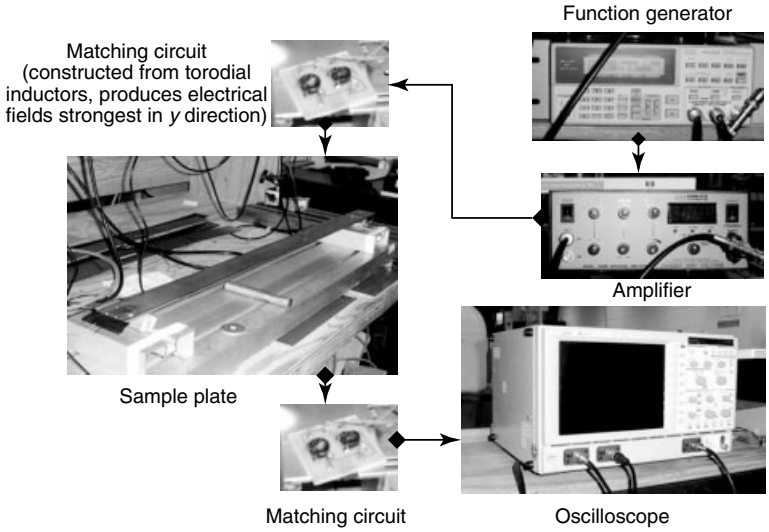


Figure 3. Schematic of the experimental setup.

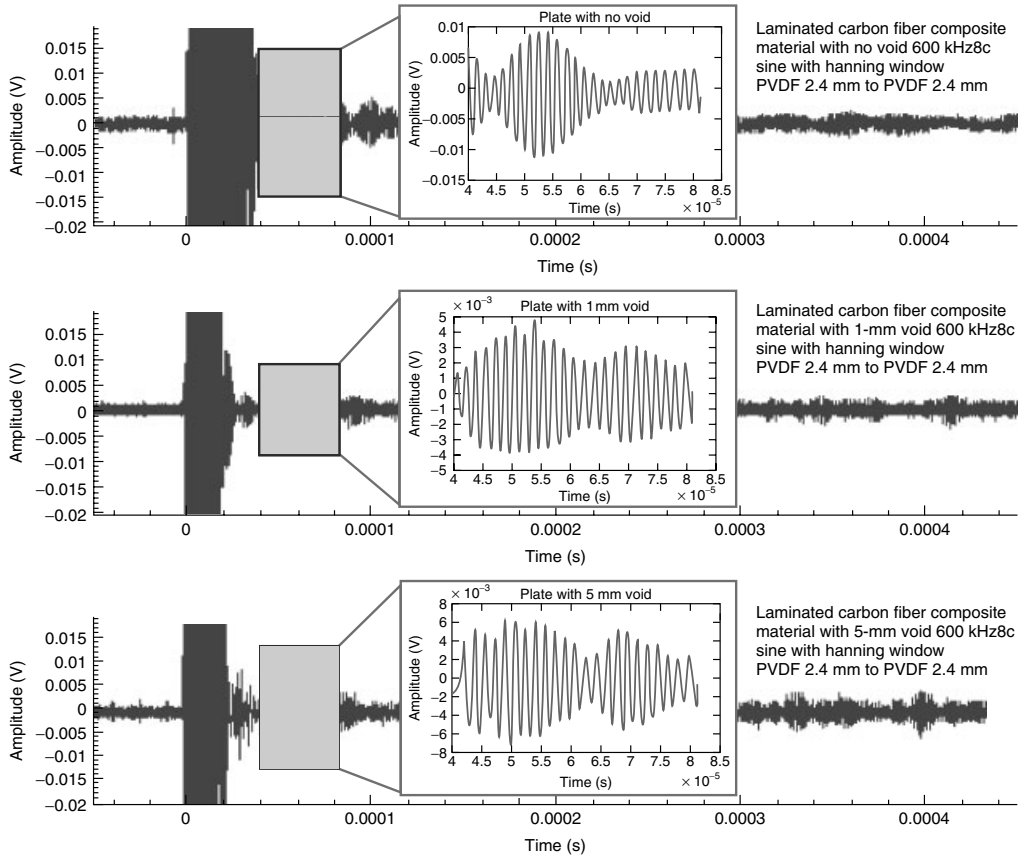


Figure 4. Response from PVDF IDTs on CFRP plates with different void sizes.

window is sent to the function generator as the excitation signal. The signal is then amplified by a high-voltage amplifier before passing through a custom-designed circuit. This circuit applies antiphase voltage on these two groups of fingers of PVDF IDT. As a result, these fingers move in opposite directions, and Lamb waves are generated. After propagating on the plate, wave signals are acquired by another piece of PVDF IDT functioning as a receiver. Another customized matching circuit similar as before is used and converts antiphase voltage into a single-phase voltage before the wave signals are eventually sent to the oscilloscope.

4 RESULTS AND DISCUSSION

Signals received from PVDF IDTs on CFRP plates with no void, 1-mm void, and 5-mm void are plotted in Figure 4. Part of each signal has been extracted and downsampled to reduce the amount of data the computer has to deal with as well as de-noise the signals. Conventional FFT is unable to reveal the nature of signals containing frequency information that varies with time, such as Lamb wave signals. TFRs are usually used for Lamb wave signal processing, among which the CWT is one of the most widely used TFRs for Lamb wave signals. CWT uses a window function called *mother wavelet* to chop the original signal into small sections (*see* **Wavelet Analysis**). Wavelet transform is then applied on each section. The CWT has been performed on the downsampled signals as illustrated in Figure 5.

The results in Figure 5 reveal the fact that if no voids exist in a plate, all the wave modes are integrated in a single pack; once voids appear, wave modes tend to separate from each other; and they get more separated as the size of the void increases.

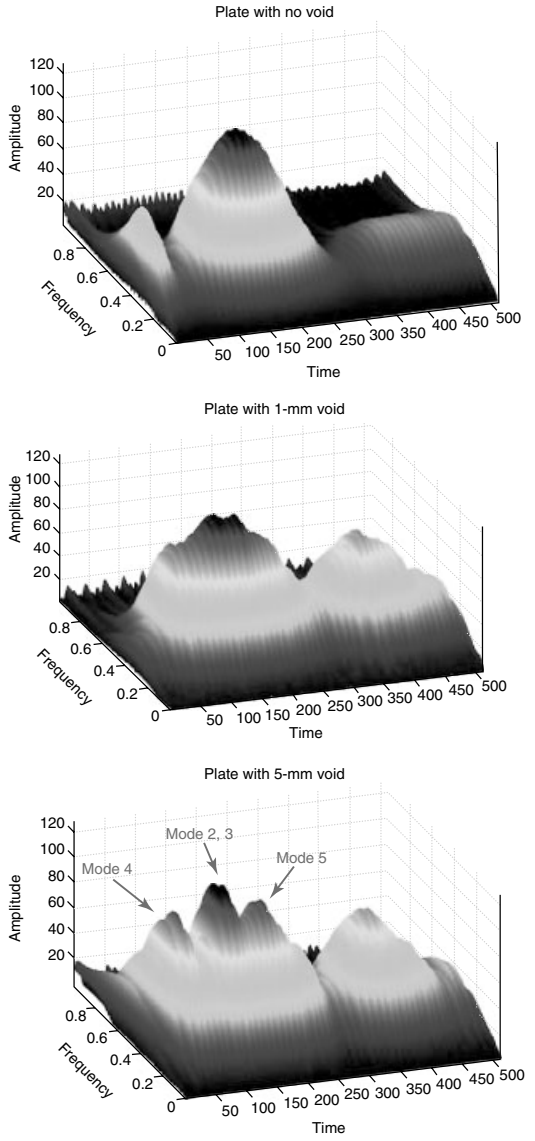


Figure 5. CWT of signals from plates with different void sizes.

5 CONCLUSIONS

A PVDF IDT has been built and used for damage detection in structural health monitoring by generating and receiving Lamb waves. The fabrication of the sensor is achieved on the basis of the etch-back photolithography technology, which gives the sensor a monolithic structure. Results have shown the ability of this PVDF IDT to detect the existence and the severity of damage with the assistance from advanced digital signal-processing methods such as CWT.

ACKNOWLEDGMENTS

This research is supported by the National Science Foundation under grants CMS-0220027. This support is gratefully acknowledged.

REFERENCES

[1] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*. Los Alamos National Laboratory, 2003.

[2] Alleyne DN, Cawley P. The interaction of Lamb waves with defects. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 1992 **39**(3):381–397.

[3] Alleyne DN, Cawley P. Optimization of Lamb wave inspection techniques. *NDT and E International* 1992 **25**(1):11–22.

[4] Dalton RP, Cawley P, Lowe MJS. The potential of guided waves for monitoring large areas of metallic aircraft fuselage structure. *Journal of Nondestructive Evaluation* 2001 **20**(1):29–46.

[5] Monkhouse RSC, Wilcox PD, Cawley P. Flexible interdigital PVDF transducers for the generation of Lamb waves in structures. *Ultrasonics* 1997 **35**(7):489–498.

[6] Monkhouse RSC, Wilcox PD, Lowe MJS, Dalton RP, Cawley P. The rapid monitoring of structures using interdigital Lamb wave transducers. *Smart Materials and Structures* 2000 **9**(3):304–309.

[7] Wilcox PD, Cawley P, Lowe MJS. Acoustic fields from PVDF interdigital transducers. *IEE Proceedings—Science, Measurement and Technology* 1998 **145**(5):250–259.

[8] Wilcox PD, Monkhouse RSC, Cawley P, Lowe MJS, Auld BA. Development of a computer model for an ultrasonic polymer film transducer system. *NDT and E International* 1998 **31**(1):51–64.

[9] Cawley P, Simonetti F. Structural health monitoring using guided waves-potential and challenges. *The 5th International Workshop on Structural Health Monitoring*. DEStech Publications, Stanford University: Stanford, CA, 2005.

[10] Pilarski A, Rose JL. A transverse-wave ultrasonic oblique-incidence technique for interfacial weakness detection in adhesive bonds. *Journal of Applied Physics* 1988 **63**(2):300–307.

[11] Rose JL, Nestleroth JB, Balasubramaniam K. Utility of feature mapping in ultrasonic non-destructive evaluation. *Ultrasonics* 1988 **26**(3):124–131.

[12] Giurgiutiu V, Zagrai A. Damage detection in thin plate and aerospace structures with the electro-mechanical impedance method. *Structural Health Monitoring—An International Journal* 2005 **4**(2): 99–118.

[13] Ditri JJ, Rose JL. Excitation of guided elastic wave modes in hollow cylinders by applied surface tractions. *Journal of Applied Physics* 1992 **72**(7):2589–2597.

[14] Younho C, Hongerholt DD, Rose JL. Lamb wave scattering analysis for reflector characterization. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 1997 **44**(1):44–52.

[15] Shin HJ, Rose JL. Guided wave tuning principles for defect detection in tubing. *Journal of Nondestructive Evaluation* 1998 **17**(1):27–36.

[16] Rose JL, Avioli MJ, Mudge P, Sanderson E. Guided wave inspection potential of defects in rail. *NDT and E International* 2004 **37**(2):153–161.

[17] Hay TR, Rose JL. Fouling detection in the food industry using ultrasonic guided waves. *Food Control* 2003 **14**(7):481–488.

[18] Rose JL. Guided wave nuances for ultrasonic nondestructive evaluation. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2000 **47**(3):575–583.

[19] Rose JL, Pelts SP, Quarry MJ. A comb transducer model for guided wave NDE. *Ultrasonics* 1998 **36**(1–5):163–169.

[20] Hay TR, Rose JL. Flexible PVDF comb transducers for excitation of axisymmetric guided waves in pipe. *Sensors and Actuators, A* 2002 **100**:18–23.

[21] Li J, Rose JL. Implementing guided wave mode control by use of a phased transducer array. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2001 **48**(3):761–768.

[22] Rose JL. A baseline and vision of ultrasonic guided wave inspection potential. *Journal of Pressure Vessel Technology* 2002 **124**(3):273–282.

[23] Giurgiutiu V, Zagrai AN. Characterization of piezoelectric wafer active sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**:959–975.

[24] Giurgiutiu V, Zagrai AN. Embedded self-sensing piezoelectric active sensors for online structural identification. *ASME Journal of Vibration and Acoustics* 2002 **124**(1):116–125.

[25] Giurgiutiu V. Review of smart-materials actuation solutions for aeroelastic and vibration control. *Journal of Intelligent Material Systems and Structures* 2000 **11**:525–544.

[26] Giurgiutiu V, Reynolds A, Rogers CA. Experimental investigation of e/m impedance health monitoring for spot-welded structural joints. *Journal of Intelligent Material Systems and Structures* 1999 **10**:802–812.

[27] Zagrai AN, Giurgiutiu V. Electro-mechanical impedance method for crack detection in thin plates. *Journal of Intelligent Material Systems and Structures* 2001 **12**:709–718.

[28] Giurgiutiu V, Zagrai A, Bao JJ. Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring. *An International Journal of Structural Health Monitoring* 2002 **1**(1):41–61.

[29] Giurgiutiu V, Zagrai A, Bao JJ. Embedded active sensors for in-situ structural health monitoring of thin-wall structures. *ASME Journal of Pressure Vessel Technology* 2002 **124**(3):293–302.

[30] Giurgiutiu V. Embedded NDE with piezoelectric wafer active sensors in aerospace applications. *Journal of Materials* 2003; Special issue on NDE, http://www.tms.org/pubs/journals/JOM/0301/Giurgiutiu/Giurgiutiu-0301.html.

[31] Giurgiutiu V. Embedded ultrasonics NDE with piezoelectric wafer active sensors. *Journal Instrumentation* 2003 **3**(3–4):149–180.

[32] Giurgiutiu V. Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. *Journal of Intelligent Material Systems and Structures* 2005 **16**(4): 291–306.

[33] Yu L, Giurgiutiu V. Advanced signal processing for enhanced damage detection with embedded ultrasonics structural radar using piezoelectric wafer active sensors. *Smart Structures and Systems—An International Journal of Mechatronics, Sensors, Monitoring, Control, Diagnosis, and Maintenance* 2005 **1**(2):185–215.

[34] Giurgiutiu V, Cuc A. Embedded NDE for structural health monitoring, damage detection, and failure prevention. *Shock and Vibration Reviews* 2005 **37**(2):83–105.

[35] Viktorov IA. *Rayleigh and Lamb Waves—Physical Theory and Applications*. Plenum Press: New York, 1967.

Chapter 54

Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors

Mark Blodgett, Eric Lindgren, Shamachary Sathish and Kumar V. Jata

Air Force Research Laboratory, Wright Patterson Air Force Base, Dayton, OH, USA

1 INTRODUCTION

Nondestructive evaluation (NDE) exploits the measurement of changes in physical properties of materials using methods that leave the material in undisturbed state after measurements. The basic NDE methods most often exploit electromagnetic, elastic, and thermal properties of materials. Elastic properties are often measured using acoustic/ultrasonic wave propagation methods, while the electrical and magnetic properties can be measured using eddy current and microwave propagation. The thermal property measurements are often performed using infrared detection.

The eddy-current method utilizes a coil energized by a sinusoidal electromagnetic signal. When the coil is brought near an electrically conductive material, the changing magnetic field induces eddy currents. The eddy currents oppose the changing magnetic field causing changes in the impedance of the coil. The magnitude of the impedance change is directly related to the electrical conductivity of the material. Presence of defects in a material will significantly affect eddy-current generation that is detected as impedance changes in the coil. The frequency of the electromagnetic signal and the electrical conductivity of the material determine the penetration depth of the eddy currents. The resolution and size of the defect that can be detected depends on the diameter of the eddy-current probe. Since eddy current can be generated only in electrically conductive materials, the technique is excellent for crack and defect

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

detection in metallic materials and structures. This article describes in detail the basic principles, coil design, application to defect detection, and advanced developments in material characterization using eddy current.

In general, ultrasonic waves propagated into a material or structure measure the elastic properties (velocity/modulus, attenuation/damping). Whenever the propagating ultrasonic waves are obstructed by defects, the energy is reflected, transmitted, and scattered. Examination of the reflected/transmitted/scattered signals is used to locate and identify the defects in the material. The frequency and wavelength of the ultrasonic wave determine the size of the defect that can be detected and the attenuation in the material determines the penetration depth into the material. In systems that use scanning mode of operation, the physical dimensions of the transducer limits the resolution. One of the major advantages of ultrasonic NDE is the ability of ultrasonic waves to propagate through any type of material. The section on ultrasonic sensors describes in detail the generation, detection, and the interaction of ultrasonic waves with defects in varieties of materials for NDE application.

Some NDE sensors are also used to detect signals emitted by a material when a component is under external load. Strain energy is released when excessive deformation, crack initiation, and certain types of crack propagation occur, and this energy is then converted to acoustic energy. The sudden bursts of acoustic energy can be listened to by an acoustic emission sensor. The acoustic emission monitoring is a passive listening method and the sensors can be attached to the structure permanently and can be used for periodic or continuous interrogation of the material. This method is gaining popularity in structural health monitoring applications.

2 EDDY-CURRENT SENSING

2.1 Background

Eddy currents are closed loops of induced current circulating in planes perpendicular to the magnetic flux, making them useful for nondestructive inspection (NDI), thickness gauging, and position sensing. Typically, eddy currents travel parallel to the exciting coil's windings and their spread is limited by the size of the inducing magnetic field. Eddy currents concentrate in the near-surface layers adjacent to the excitation coil and their strength decreases with distance from the coil according to Maxwell's equations, which show that eddy-current density decreases exponentially with depth, known as the *skin effect*. The depth that eddy currents penetrate into a material is affected by the frequency of the excitation current, the electrical conductivity, and the magnetic permeability of the specimen (Figure 1). The depth of penetration decreases with increasing frequency and increasing conductivity and magnetic permeability. The depth at which eddy-current density has decreased to 1/e, or approximately 37% of the surface density, is called the *standard depth of penetration,* δ. "Standard" penetration denotes the use of a plane-wave approximation to the electromagnetic field interaction with the test sample and is unaffected by the size of the coil. The eddy-current inspection methodology is widely regarded as the most common electromagnetic NDE tool and is used in many applications for detection and quantification of defects in strength critical structures like pipelines and aircraft, thickness measurements, coating measurements, and various applications involving measurement of conductivity.

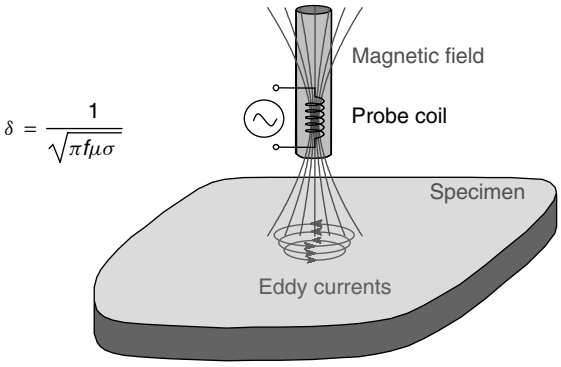


Figure 1. An illustration describing the interaction of the alternating current excited eddy-current probe coil with a conducting specimen to generate eddy currents. The standard penetration is shown as, δ (mm), where $\pi = 3.14$, $f =$ frequency (Hz), $\mu =$ magnetic permeability (H mm^{-1}), and $\sigma =$ electrical conductivity (%IACS) International Annealed Copper Standard.

2.2 Early discoveries

Contemporary eddy-current sensors and instrumentation are the result of decades of development and have an even longer history in terms of the fundamental principles on which the technology is founded. Electromagnetism was discovered in 1820 by Hans Christian Oersted when he demonstrated the effect of electricity in a wire on a compass needle. He discovered that the wire conducting electric current generated invisible magnetic lines of force, which when placed above the compass needle caused it to swing perpendicular to the direction of the current flow and to swing in the opposite direction upon reversal of the current flow. Of course, magnets also have invisible lines of force and several years later, Michael Faraday and Joseph Henry made nearly simultaneous independent discoveries of electromagnetic induction, in 1831. Both men used iron bars bent into a circle and wrapped in copper wire to create an electromagnet, and then in different arrangements demonstrated the principle of electromagnetic induction, namely that the induced current is proportional to the rate of change of magnetic flux through a coil. Some time later, Lèon Foucault discovered in 1851 that a changing magnetic field produces a circulating current upon intersection with a conductor. These swirls or eddies of current produce magnetic fields that oppose the applied magnetic field, which created them in the first place and are only present when the applied magnetic field is either physically moving or changing in strength. Capping off the early discovery era is the influential work of D. E. Hughes [1], who demonstrated in 1879 that eddy currents can be used as a basis to compare and sort materials by virtue of differences in electrical parameters, such as conductivity and permeability. Hughes showed that the principle of electromagnetic induction can be used to measure properties of coinage and the relative conductivities of metals and also demonstrated techniques for enhancing sensitivity. By combining these seminal early discoveries, it is possible to begin to see the foundation of modern eddy-current testing take shape, given (i) an alternating current (ac) used to excite a probe coil that gives rise to a changing magnetic field; (ii) when placed near a conductor the energized probe produces eddy currents in the solid, which have their own magnetic fields; and (iii) the magnetic fields of the induced eddy currents oppose the applied magnetic field thereby affecting the impedance of the probe coil, which is the basis of most modern eddy-current testing.

2.3 Measurement and testing applications

Modern eddy-current testing has evolved over the course of more than a century of research and development and finds its origins in metal sorting and comparing. During this long development period tremendous advances have been made in sensors, electronics, integrated circuitry, microprocessors, and computers leading us to today's handheld, portable eddy-current NDE instrumentation geared for field applications. In the latter part of the 1930s, Friedrich Förster, the person widely regarded as the forefather of modern eddy-current testing, began to make significant advances to the field by developing theory for complex impedance analysis, experimental equipment, calibration procedures, and nondestructive techniques to quantitatively test the properties of metals [2]. These early developments led to the production of commercially available electromagnetic induction testing equipment in the United States in the 1950s. Today's eddy-current test devices and instrumentation are available for a wide range of industrial applications such as aerospace (*see* **Military Aircraft**; **History of SHM for Commercial Transport Aircraft**; **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Monitoring of Aircraft Engines**), energy (*see* **Fiber-optic Sensor Principles**; **Wind Turbines**; **Large Rotating Machines**), marine (*see* **Monitoring Marine Structures**; **Ship and Offshore Structures**), and automobile (*see* **Sensor Technologies for Direct Health Monitoring of Tires**) industries. An extensive accounting of contributions by early researchers in this field is documented in the NDT Handbook on Electromagnetic Testing [3].

Eddy-current testing techniques are primarily used to measure material properties such as conductivity and permeability, discriminating components based on the presence of flaws or discontinuities, and precise measurement of displacement, thickness, positioning, and proximity. Many critical applications require automated computer-controlled handling of eddy-current sensors and test objects due to the large scale of the inspection area, which allows the

inspection area to be mapped to the component requiring inspection. For example, turbine engine components are designed to optimize aircraft performance, while accommodating the adverse effects of high temperatures and dynamic loading conditions. In critical components such as turbine disks, the hardware is closely monitored throughout its service life using eddy current and other NDIs. Some United States Air Force ((USAF)) turbine engines are entirely disassembled and components are subjected to intensive NDIs to ensure that dimensional tolerances are met and surfaces are free from life-limiting flaws. Many individual components are cleaned and prepared for eddy-current inspections in an effort to detect surface-breaking fatigue cracks, fretting damage, and foreign object damage along with other detrimental features such as dents and gouges. Moreover, with appropriate calibration standards and tracking methods based on the initial fabrication conditions, microstructure, and alloy composition, quantitative eddy-current conductivity measurements could be potentially made to complement eddy-current flaw inspections, which are based on probability-of-detection standards, for the same critical engine components.

Eddy-current NDE techniques have been practiced for several decades, primarily as a means of flaw detection in the near-surface of metallic parts. For turbine engine applications, eddy-current testing is the primary method used to inspect the surface for life-limiting surface flaws, and a great array of probes have been manufactured to allow comprehensive inspections of these complex components. The optimal frequency range of a given probe, and therefore its sensitivity and effective penetration depth, may be selected by manipulating the probe's diameter, the number of wire turns in the coil, and the impedance matching circuitry. Eddy-current inspection is based on the electromagnetic induction principle, i.e., that a changing magnetic field will induce electrical currents in a nearby conductor, which in turn will load the coil and affect the phase and magnitude of its impedance. Accurate measurement of these eddy-current loading effects on the probe's impedance is the basis of most eddy-current measurements and provides the means to evaluate materials for near-surface cracking and inclusions, heat treatment verification, and thickness gauging. It is well known that eddy-current conductivity is affected by a number of things including residual stress, chemical composition, microstructure, hardness, surface roughness, and temperature. Therefore, it is essential to use stable quantitative test equipment and reliable procedures to assure integrity of the acquired experimental data and repeatability of the measurements. Calibration and reference standards are also essential for accurate eddy-current measurements and typically consist of materials with either known conductivities (or permeabilities for magnetic materials) for materials property characterization or with known flaw sizes for defect detection and characterization.

Eddy-current inspection is often used to detect corrosion, erosion, cracking, and other changes in tubing. Heat exchangers and steam generators, which are used in power plants, have thousands of tubes that must be prevented from leaking. This is especially important in nuclear power plants where reused, contaminated water must be prevented from mixing with freshwater that will be returned to the environment. The contaminated water flows on one side of the tube and the freshwater flows on the other side. The heat is transferred from the contaminated water to the freshwater and the freshwater is then returned back to its source, which is usually a lake or river. It is very important to keep the two water sources from mixing, so power plants are periodically shut down so that the tubes and other equipment can be inspected and repaired. The eddy-current test method provides high-speed inspections for these types of applications.

2.4 Basic measurement instrumentation

Most modern eddy-current testing equipment consist of a probe coil, an ac source, a voltmeter (or ammeter), and a display for analysis of the data. Most instruments used to measure eddy-current NDI data are based on the electromagnetic principle of mutual induction, which can be illustrated as an electrical circuit between the energized eddy-current probe and the conductive object undergoing testing. Since eddy currents generate their own magnetic fields that affect the primary field of the coil, it is essential that the instrument is capable of measuring the changes in resistance and inductive reactance of the probe coil at a given frequency. The mutual inductance between the probe and the test object is affected by electrical conductivity, magnetic permeability, and liftoff,

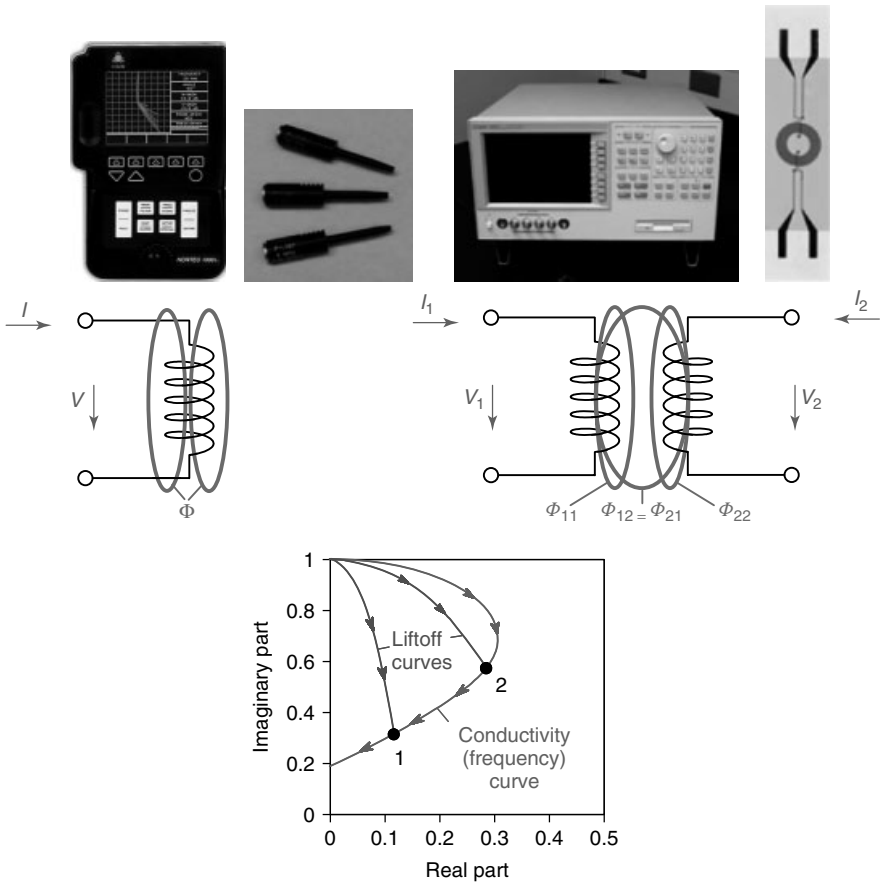


Figure 2. Some commonly available eddy-current probes, flexible coil and associated instrumentation; resonance eddy-current probe circuits and an impedance plane diagram showing the result of an eddy-current scan of a conducting material including the effects of lifting the probe from the surface of the material.

which is the distance between the inspection object and the probe. Most commercially available eddy-current instruments have a broad range of excitation frequencies from, for example, 100 Hz to 6 MHz. The basis for the eddy-current measurement is the (Maxwell) ac bridge, which is used to measure unknown inductances in terms of calibrated resistance and capacitance. Since the phase shift between inductors and capacitors is 90° out of phase, capacitive impedance can be used to balance out an inductive impedance if they are in opposite legs of the ac bridge. The eddy-current instrument display is essentially an impedance plane with resistance along the abscissa and inductive reactance along the ordinate and the measured impedance is the vector addition of the two components.

Eddy-current probes typically consist of a primary coil, which is capable of being brought into close proximity with the test object, and an identical reference coil housed in the body of the probe to accommodate temperature changes and minimize thermal drift (Figure 2). Probes are essentially resonant circuits, but operated well below the resonance to avoid stray capacitance and other parasitic affects. A wide variety of eddy-current probes exist today for different types of inspection applications including absolute probes for conductivity measurement or thickness gauging, differential probes for crack detection complex geometry, and reflection probes for materials characterization. Hybrid probes exist for applications such as detection of tiny surface cracks in aerospace components, where a split-D

type of transducer might be used, which consists of a drive coil encompassing two D-shaped sensing coils. Other types of eddy-current probes also exist, which might use giant magnetoresistive elements, Hall-effect sensors, or flexible thin-film sensors for specialized applications (*see* **Eddy-current *in situ* Sensors for SHM**; **Electric and Electromagnetic Properties Sensing**). Similarly, a wide variety of probe configurations exist for surfaces, bolt-holes, and inner- and outer-diameter inspections of pipes and tubing. Often, probes will have a ferrite core to intensify the magnetic lines of force near the core or take advantage of shielding to limit the spread of the probes magnetic field and therefore that of the eddy currents. In most applications, sensing the eddy currents involves assessing the impedance changes of a detector coil using an ac bridge where the impedance of the probe is compared to known impedances, forming a balancing arm in the bridge circuit. High sensitivity is achieved by matching the impedance of the probe to the impedance of the measuring instrument, otherwise known as *nulling the instrument*. The fundamental science of electromagnetic nondestructive test methods and applications for engineers and students can be found in [4].

3 ULTRASONIC SENSORS

Ultrasound is the commonly used technique in NDE to detect the presence of defects or determine the elastic properties of materials. Ultrasound is a linear elastic mechanical wave that propagates at frequencies greater than 20 kHz. Typically, frequencies between 1 and 10 MHz are used in production NDE, but frequencies in the range of hundreds of megahertz have been used in specialized inspections. The most common method to generate and detect ultrasonic waves is to place a sensor in contact with the material being evaluated. These sensors are called *transducers* and use a variety of material properties to generate and detect the mechanical wave. The most common class of transducers is based on piezoelectricity. However, other types of transducers can be based on magnetostrictive, electromagnetic, or capacitive behavior of the element in the transducer, where the element refers to the component of the transducer that physically moves to generate and detect the ultrasonic mechanical waves and converts this energy into electrical signals that can be viewed and/or recorded on oscilloscopes or similar displays. This article will provide an overview of piezoelectricity and a review of the key components that are typically required in a contact piezoelectric transducer. It will also provide a brief description of other methods to generate and detect ultrasound, plus a summary of noncontact methods, including the use of lasers. It is important to note that this article is a short summary of these items. There is a substantial volume of literature, including books, monographs, and conferences, dedicated to the design, development, and evaluation of ultrasonic transducers [5–8] and the reader is encouraged to explore these for additional information.

Piezoelectricity is described as a material property that converts electrical signals to mechanical deformation and *vice versa*. This behavior was first observed by Jacques and Pierre Curie in 1880 by detecting the presence of an electric charge when certain materials were compressed. Later work indicated that inverse behavior can occur. This behavior is caused by the crystallographic structure of the piezoelectric crystal. The most commonly cited natural piezoelectric material is quartz, which has a hexagonal crystal structure [9]. To obtain the piezoelectric behavior, single crystals of this material are needed and they have to be cut along specific axes to obtain the desired deformation. Figure 3 illustrates the application of an electric field to a single hexagonal crystal to generate a longitudinal ultrasonic wave and Figure 4 shows how a transverse, or shear, wave is generated from a single crystal cut along a different plane.

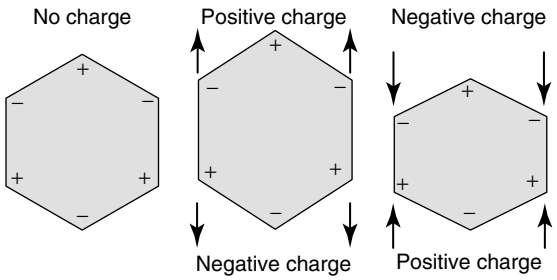


Figure 3. A schematic demonstrating the generation of longitudinal ultrasonic waves by applying an electrical pulse to a hexagonal quartz crystal.

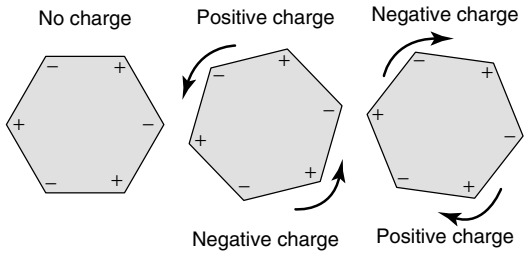


Figure 4. A schematic demonstrating the generation of shear, or transverse, ultrasonic waves by applying an electrical pulse to a hexagonal quartz crystal.

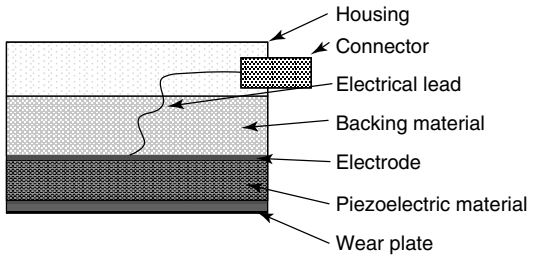


Figure 5. A cross section of a typical commercial transducer showing the typical components of such a transducer.

When an electrical spike is applied to the crystal, it will deform to either attract or repel the electrical charge in response to the asymmetric balance of charges in the signal crystal. The linear deflection of the single crystal generates a longitudinal wave, whereas the rotational deformation will generate a shear, or transverse, wave. Note that a much thicker couplant capable of supporting shear motion must be used when generating or detecting shear waves using a piezoelectric element. Alternative methods, such as those using angle-beam techniques, generate shear waves using refraction, which can occur with less viscous couplants, such as water or inert gels.

Subsequent developments in piezoelectric materials resulted in the discovery and development of ceramic materials, such as lead zirconate titanate (PZT) and lead metaniobate. An advantage of these materials is that they do not need to be single crystals to demonstrate piezoelectric behavior. These materials are poled at elevated temperatures to align their piezoelectric properties, which yield a deformation in these materials when they are exposed to an electrical field. Using this approach, the physical motion of these materials enables the preparation of transducers that can generate and detect both longitudinal and shear waves. This eliminates the need for careful sectioning of a ceramic crystal, as required for quartz, but has the disadvantage that these materials can only be used at temperatures below the depolarization temperature, which can range between 130 and 575 $^{\circ}$C. These materials are very sensitive to surface displacements, which make them ideal transmitters and receiver of ultrasonic energy that remains in the linear elastic regime [10].

Commercially available transducers typically use these materials as their active element. Common features of commercial transducers are shown in a representative cross section of a transducer in Figure 5. The transducer is typically a metallic case. The active element is protected from the inspection surfaces by a wear plate, which is made of a material that minimizes wear and matches the acoustic impedance of the active element as much as possible. Electrical leads are coupled to either surface of the piezoelectric material to apply the voltage spike or tone burst to excite the element. A significant amount of the volume of a transducer is dedicated to the dampening material that minimizes the resonance of the piezoelectric element. The dampening material decreases the number of cycles the piezoelectric material will resonate, which assists in resolving multiple reflections from the material being inspected. For example, this minimizes the interference of these signals when trying to detect damage close to the location of the transducer. Note that electrical wire failure and disbanding of the dampening material are the most common causes of failure in a transducer, which would lead to a diminished response or a perturbation in the received signal from damage.

A more recent development in transducer design is the emergence of composite elements. These transducers, which have been commercially available for approximately 10 years, combine a polymer material together with the piezoelectric material, usually to assist in the dampening process. The polymer-based material is commonly introduced into the transducer by cutting thin channels in the ceramic element and filling the channels with the polymer material. The polymer can dampen the vibration of the ceramic while minimizing the effect on the sensitivity of the ceramic when compared to a traditional dampening backing layer. Thus, composite transducers are frequently used when high sensitivity and resolution are needed.

A multitude of other materials and methods have been explored to generate and detect ultrasonic waves. Several polymer materials, most notably poly(vinylidene fluoride) (PVDF), have been shown to have piezoelectric properties. This material can be configured into relatively thin films and retains its piezoelectric behavior when flexed, making it a desired sensor when trying to conform to a complex geometry. These materials are typically very effective detectors of ultrasonic signals, but are not very efficient transmitters. In addition, their response can be very dependent on temperature, requiring these sensors to be used in a stable environment.

The transducers described above are used in most bulk material ultrasonic NDE applications. However, for acoustic emission (AE), the performance requirements of transducers change as these sensors are tailored to detect ultrasonic signals emitted from crack propagation, delaminations, and similar types of damage. To maximize the detection of these signals, AE transducers typically have a response that is very flat as a function of frequency. With conventional piezoelectric transducers, this performance characteristic is typically obtained by increasing the dampening of the piezoelectric element. Therefore, these sensors are commonly larger than conventional transducers and frequently have broadband preamplifiers to amplify the dampened signal. An alternative approach is to make the contact area of the sensing element very small, which can be exemplified by conical transducers and pinducers. With these small contact areas, extra dampening material is not required as the sensor functions as a point detector. Owing to the requirement that these sensors be very broadband, they are also very inefficient and are not commonly used as transmitters. Note that other broadband sensors, such as PVDF, can also be very effective AE sensors.

Alternative materials have been used for generation and detection of ultrasonic signals. Magnetostrictive materials convert magnetic fields into mechanical deformation by the alignment of their magnetic moments [11]. These materials are commonly used for sound navigation and ranging (SONAR) applications as they are very effective and efficient at low frequencies. Capacitive transducers use changes in capacitance between the sensor and the material in which the ultrasonic wave is propagating to sense ultrasonic signals [12]. As this behavior can occur through air, these transducers do not require couplant, but typically must be in close proximity of the surface of the specimen in which the ultrasonic wave is propagating. Electromagnetic acoustic transducers (EMATs) are used with electrically conductive materials. These sensors use an radio frequency (RF) coil to generate eddy currents in the specimen. When the coil is placed inside a permanent magnetic field, the eddy currents interact with the magnetic field to generate mechanical waves in the material. Note that these three effects occur in reverse order when the ultrasonic signal is detected. Several efforts have used piezoelectric materials for air-coupled transducers. They are most effective in a through transmission mode for composite materials, which have a lower acoustic impedance when compared to metallic structures.

Another noncontact method that has been researched extensively is the use of lasers to generate and detect ultrasound. For generation, it is common to use a short pulse laser that causes very rapid and localized heating to thermoelastically generate a mechanical wave. If too much energy is deposited on the surface, the generation method becomes ablative, which will cause material to be removed from the surface. Detection is typically performed using an optical interferometer, such as Michelson or Fabry–Perot. Extensive literature exists that explores multiple laser-based variations to generate and detect ultrasonic signals [9].

A final type of transducer to consider is array transducers. These transducers include multiple piezoelectric elements sufficiently small to be considered as point sources of ultrasonic signals. These elements are excited in a controlled manner to use the principle of superposition to direct the effective wavefront in different directions. There are limits to the amount an ultrasonic wave can be directed and controlled, which is determined by the material, the number and size of the individual elements, and the physics of ultrasound. However, the use of arrays is growing in popularity, especially in structures that do not contain an excessive number of scattering features that can interfere with the superposition of the ultrasonic waves generated from the point sources of ultrasound.

4 SUMMARY

This article described some introductory ideas of eddy-current coils used for material characterization

and detecting damage in materials and structures. Similarly, ultrasonic transducers were described for the generation and detection of ultrasonic waves. There are extensive volumes of literature that describe the material properties and physics that enable these NDE transducers to work as designed and the reader is strongly encouraged to explore this literature before embarking on research and development efforts in this field.

REFERENCES

[1] Hughes DE. Induction balance and experimental researches therewith. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*. *Fifth Series*, 1879, Vol. 8, No. 46, pp. 50–57.

[2] Förster F. The first picture: A Review of the Initial Steps in the Development of Eight Branches of Nondestructive Material Testing. *Materials Evaluation* 1938 **41**(3):1477–1488.

[3] McMaster RC, McIntire P, Mester ML (eds). *Nondestructive Testing Handbook*, 2nd Edition. The American Society for Nondestructive Testing, 1986, Vol. 4.

[4] Libby HL. *Introduction to Electromagnetic Nondestructive Test Methods*. Wiley-Interscience, 1971.

[5] Silk MG. *Ultrasonic Transducers for Nondestructive Testing*. Adam Hilger, 1984.

[6] Krautkramer J, Krautkramer H. *Ultrasonic Testing of Materials*. Springer-Verlag, 1990.

[7] Bray DE, Stanley RK. *Nondestructive Evaluation*. McGraw-Hill, 1989.

[8] Van Vlack LH. *Elements of Materials Science and Engineering*. Addison-Wesley, 1980.

[9] Kossoff G. The effects of backing and matching on the performance of piezoelectric ceramic transducers. *IEEE Transactions on Sonics and Ultrasonics* **SU-13**(1):20–30.

[10] Savage HT, Clark AE, McMaster OD. *Rare Earth-Iron Magnetostrictive Materials and Devices Using These Materials*, US Patent Number 4,308,474, December 1981.

[11] Schindel DW, Hutchins DA, Zou L, Sayer M. Air-coupled capacitance transducers. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control* **42**(1):42–50.

[12] Scruby CB, Drain LE. *Laser Ultrasonics Techniques and Applications*. Taylor & Francis, 1990.

Chapter 66

Global Navigation Satellite Systems (GNSSs) for Monitoring Long Suspension Bridges

Xiaolin Meng[1] and Wei Huang[2]

[1] *Institute of Engineering Surveying and Space Geodesy, University of Nottingham, Nottingham, UK*
[2] *Intelligent Transportation Systems Research Centre, Southeast University, Nanjing, China*

1 A BRIEF INTRODUCTION TO THE GLOBAL POSITIONING SYSTEM

1.1 GPS constellation

The full term of the well-known acronym GPS is NAVSTAR global positioning system, where NAVSTAR stands for NAVigation System with Time And Ranging [1]. GPS is a satellite-based navigation and positioning system that was designed in the early 1970s by the US military to allow soldiers to autonomously and continuously determine their position within 10–20 m of accuracy, at any point on the earth's surface and under any weather conditions. Providing precise timing information is another important function of GPS. GPS was originally utilized even for the military operations of US military forces; the last two decades have seen a rapid expansion of civilian GPS user groups. The current GPS configuration consists of three segments: the space segment—a constellation of 24 (nominal) satellites distributed in six orbital planes of 55° inclination to the equator with an altitude of 20 200 km above the earth's surface as shown in Figure 1; the control segment—comprises a master control station, worldwide distributed monitor stations, and ground control stations; and the user segment—anyone who receives and uses GPS signal with any type of GPS-enabled devices. More detailed information about GPS segments can be found in [1, 2].

1.2 GPS measurements

Each GPS satellite transmits at least two carrier signals in an L-band, such as L1 and L2 carriers, and more signals are available for the modernized

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

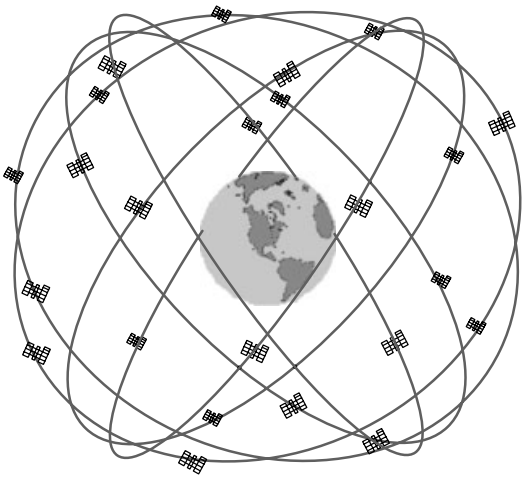


Figure 1. Current GPS constellation (http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/).

GPS satellites. The L1 and L2 carriers are modulated by ranging codes and navigation information such as orbital parameters. The carriers together with ranging codes are used mainly to determine the distance from the antenna of a user's receiver to the GPS satellites in view [3]. These uncorrected distances are called *code* or *carrier pseudoranges* since they include the true geometric distance between the receiver antenna and the tracked satellites plus small range corrections to account for the orbital and clock errors, signal delays caused by the atmosphere, and multipath due to the signal reflection by the surroundings in the vicinity of the GPS antenna [1]. For determining a three-dimensional (3D) location of a GPS antenna, simultaneous tracking to three GPS satellites is adequate. However, since a GPS receiver is normally equipped with a much cheaper clock, compared with the atomic ones used by the GPS satellites, tracking to at least one more satellite is required to solve the clock bias of the GPS receiver for achieving a higher positioning accuracy.

1.3 GPS relative positioning

A prerequisite for achieving GPS positioning accuracy of a few meters (code pseudorange) to centimeters or even subcentimeters (carrier phase) is the effective cancelation of various ranging errors. The number of satellites tracked by a GPS receiver and their spatial distribution are also major affecting factors to the achievable positioning accuracy. Generally, the GPS relative positioning technique involves a pair of GPS receivers simultaneously tracking at least four well-distributed identical satellites for solving the position of an unknown stationary GPS antenna and at least five satellites for a moving antenna. For canceling out spatially correlated GPS ranging errors, one GPS receiver must be installed on a precisely measured benchmark called a *reference station* to help solve the unknown coordinates of another receiver called a *rover*, either in a stationary mode or a continuous motion mode. When relative positioning is utilized in surveying, differencing the GPS ranging measurements made to two satellites simultaneously by a reference receiver and a rover receiver could eliminate both the satellite clock and the receiver clock biases. Carrier and/or code pseudorange measurements can be used in GPS relative positioning and the positioning solutions in the form of 3D coordinate time series can be obtained either in real time or through postprocessing of collected GPS measurements.

1.4 Achievable positioning accuracy and major inhibitors to GPS positioning accuracy

Both code and carrier phase pseudorange measurements are "polluted" by errors that originate from GPS receivers, GPS satellites, signal propagation medium, and observation environment [1, 3]. In the data processing stage, incorrect processing models adopted and/or bugs in the software packages also cause positioning errors. There are many studies in this area on how to effectively reduce, mitigate, and eventually eliminate GPS error sources through mathematical modeling, software development, and digital signal-processing approaches, and hardware design and configuration. Some error sources, which are observation environment oriented and change from place to place, prove very difficult to be modeled or mitigated. For instance, pervasive multipath effect, which is a major error source for both the carrier

phase and the code-pseudorange-based GPS positioning, makes the effort for achieving highly accurate GPS positioning elusive. Most GPS antenna will not be able to detect which signals are directly transmitted by the satellites and which are reflected from the objects surrounding a GPS antenna. As a result of the complexity of objects surrounding an antenna, there is virtually no versatile model or physical antenna design to effectively reduce multipath, which introduces a maximum error of up to about 5 cm for carrier-phase-based positioning, a quarter of its L1 wavelength, and several meters for code-pseudorange-based GPS positioning.

1.5 Integer ambiguity resolution for high accuracy positioning

Centimeter or even subcentimeter positioning accuracy can be achieved through using carrier-phase-based GPS relative positioning under the condition of successful resolution of the initial integer ambiguities. When carrier phase measurements are employed in positioning, only a fraction of the wavelength of the carrier phase is recorded by a GPS receiver, but the whole number of carrier wavelengths or cycles between a GPS antenna and the tracked satellites remains unknown. Reliable and quick resolution of the unknown number of cycles from a tracked satellite to a GPS antenna is vital in using carrier phase measurements for subcentimeter positioning. Until the early 1990s, GPS was only used to monitor static deformation of structures with low dynamics, for example, dams and low buildings. The least squares adjustment, Kalman filtering, statistical process, and other search techniques can resolve this ambiguity and determine the most probable solution [4, 5]. Furthermore, the invention of integer ambiguity on-the-fly (OTF) algorithms by different researchers around the world in the beginning of 1990s made the real-time kinematic global positioning system (RTK GPS) a viable tool for monitoring more flexible structures such as long suspension bridges, high-rise buildings, and slender TV or communications towers [6, 7], or for tracking high-speed moving objects such as low earth orbit (LEO) satellites, because of its high accuracy, productivity, flexibility, and simplicity in data processing (conducted by the embedded receiver firmware). The positioning solutions produced by the internal processor of a GPS receiver can be streamed to a control center via cable connection or through wireless communications for further analysis and visualization of resulting solutions. However, the collected raw carrier phase measurements, which are recorded with memory cards inside the GPS receivers, can be decoded as American Standard Code for Information Interchange (ASCII) data files in a receiver independent exchange (RINEX) format. These data files can be postprocessed in a kinematic mode by using the OTF approach with any commercial or household GPS postprocessing software to output each epoch positioning solution. More accurate positioning results can be obtained because of the possibility to edit and clean GPS measurements.

1.6 Communication links

Radio modems provide wireless communications between GPS reference receivers and rover receivers for carrying out real-time relative positioning. When a reference GPS receiver broadcasts the corrections via its radio modem transmitter, an unlimited number of rover GPS receivers can pick up these data via their own radio modems. The transmission range depends on the power of the radio modem transmitter and also on the terrain and the radio antenna setup. There are many different radio modems in the market and ultrahigh frequency (UHF), very high frequency (VHF), and spread spectrum are the most commonly used in RTK GPS positioning. Government authorization may be required for using certain types of radio modems. In some countries, there are bands that are allocated for public use without the need for any special authorization. For instance, the 900-MHz band in the United States and 2.4 GHz in most European countries are allowed for spread spectrum communications without any special authorization (but there are limitations on the amount of power that one can use to transmit signals, for instance, in the United Kingdom, only 0.5-W carrier power is allowed). More recently, some GPS manufacturers adopted cellular communication technology or the third-generation (3G) wideband digital networks, such as global system for mobile communications (GSM) and general packet radio service (GPRS), as an alternative GPS communication link. Since subscribers only pay for the data amount that they

have actually transmitted or received, the use of GPRS technology is more flexible and much cheaper compared with a GSM communication link. Dedicated local area network (LAN) can also be utilized to transmit GPS corrections from reference stations to the rovers over the Internet. Satellite-based wireless communication approach is becoming popular and has already been employed to transmit corrections to the rovers in the area where there is no GSM/GPRS coverage and the use of radio modems is constrained by the local terrain and transmission power limitation.

2 GPS FOR STRUCTURAL HEALTH MONITORING (SHM)

2.1 Brief history of GPS for SHM

According to [8], the process of implementing a damage detection strategy for aerospace, civil, and mechanical engineering infrastructures is referred to as *structural health monitoring* (SHM) and a damage is defined as changes to the material and/or geometric properties of these systems, including changes to the boundary conditions and system connectivity, which adversely affect the performance of the systems. The SHM process involves the observation of a system over time, using periodically sampled dynamic response measurements from an array of sensors, the extraction of damage-sensitive features from these measurements, and the statistical analysis of these features to determine the current state of system health. The output of this process is periodically updated information regarding the ability of the structure to perform its intended function in light of the inevitable aging and degradation resulting from operational environments. An SHM system should have the capacity to provide, in near real time, reliable information regarding the integrity of the structure after earthquakes or blast loading.

Recent advances in GPS positioning, computer science, telecommunications technologies, and advanced digital signal processing have made GPS a much more robust, reliable, convenient, accurate, and cost-effective tool for the deformation monitoring of natural and artificial structures. To date, GPS is widely used to monitor volcano eruptions [9, 10], crustal movements [11, 12], vertical land movements [13], landslides [14], earth structures [15], dams [11], buildings [16–21], and bridges [22–29]. In general, these applications can be roughly categorized into three levels, i.e. large, medium, and local scales, according to the separations of the GPS stations. A variety of regional continuous global positioning system (CGPS) networks, such as EUREF Permanent Network (as of March 04, 2007, it has 199 permanent GNSS tracking stations, which cover the European continent); the OS Net of the Ordnance Survey in the United Kingdom, which consists of more than 100 permanent stations; and the Geographical Survey Institute's GPS Network, which comprises 1224 GPS real-time stations in Japan (www.euref.eu; www.ordnancesurvey.co.uk; www.gsi.go.jp), are typical examples of the first category applications. These networks play a vital role in monitoring crustal movement, which is crucial for the evaluation process of seismic and volcanic activity as well as the extraction of geodynamic information. On a medium scale, GPS-based volcano eruption monitoring presents a good example. The last level of GPS-based monitoring applications takes place on a variety of natural or artificial structural deformation and deflection monitoring in the local scale, which is the main scope discussed in this article (*see* **Ambient Vibration Monitoring**).

2.2 Case studies of GPS for SHM of long suspension bridges

Under different loading conditions, a suspension bridge generally experiences two distinct types of deformations, i.e., the long-term movement caused by the foundation settlement, bridge-deck creep, and stress relaxation; and the short-term motion of the bridge, or bridge deflection, such as those activated by wind, tidal current, earthquake, or traffic [27, 30]. The latter deformation is recoverable in most cases and the bridge will resume to its original status from the deformation with the release of external forces.

When GPS is utilized to monitor bridges, it has the capacity to detect two different kinds of deformations simultaneously [25, 31]. Analyzing the response of a long bridge to the short-term irregular loading is much more important in terms of risk level of major damage and its significance for the improvement of design code. However, it is more difficult to be measured compared with the identification of

long-term foundation settlement, concrete creep, loss of prestress, and thermal expansion or contraction. In monitoring foundation displacement, averaging GPS time series of a few hours or more can be used, producing positioning accuracy of a few millimeters, since error sources such as multipath (an unwanted indirect signal reflected by the surroundings before reaching a GPS antenna), which characterizes as a low period of movement, could be effectively removed. To measure the deformations occurring over a time interval of a few seconds, there is a much shorter or even no averaging time available for mitigating errors; the time interval may not even be long enough for resolving the critical initial integer ambiguities. This significantly restricts the usage of GPS positioning in monitoring short-term movements of more flexible structures. Also, to measure short-term effects, sampling rates of GPS sensors must be significantly increased, which will incur heavier onboard data processing and communication overloads if RTK GPS positioning is required.

In the last decade or so, with rapid advances in GPS technology, digital signal processing (DSP), telecommunications, and computing science, as well as with the continuous efforts conducted by the researchers around the world, GPS positioning is gradually being accepted by the civil engineering community as a viable monitoring tool and has started to play an important role in SHM. This is mainly due to the advent of OTF integer ambiguity resolution technology. As discussed in the previous section of this article, solving the unknown number of cycles from a tracked satellite to the GPS antenna is vital in using carrier phase measurements for subcentimeter positioning, which is a prerequisite for GNSS-based SHM. The significant increase in the GPS sampling rate from 1 Hz to rates higher than 50 Hz currently is another reason for the acceptance of GPS-based SHM of long suspension bridges.

Structural aging and degradation resulting from changes in the materials and operational environments such as significantly increased traffic volume and weight, as well as the boundary conditions and system connectivity, and flaws in design are the main reasons for many recent bridge closures and failures around the world. Transport agencies and researchers have strong interests in finding appropriate sensor systems and computational models to implement on-line monitoring and diagnosing systems to detect and locate the changes to the materials and/or geometric properties of bridges. Of various monitoring tasks, measuring short-term dynamic behavior of such structures is of particular interest.

The Tsing Ma Bridge in Hong Kong, China, is the sixth longest suspension bridge in the world and it is the longest single-span suspension bridge that carries both road and rail traffic [32]. The Tsing Ma Bridge and other two cable-stayed linking bridges are perhaps the only bridges around the world that are equipped with the most comprehensive and sophisticated sensor systems, forming a part of six whole SHM components. The six integrated modules include the sensory system, the data acquisition and transmission system, the data processing and control system, the structural health evaluation system, the structural health data management system, and the inspection and maintenance system.

The main objective of the Tsing Ma SHM system is to monitor the loading and structural parameters so that the performance of the bridge under current and future loading conditions can be evaluated and predicted. Such evaluated results will facilitate the planning of bridge inspection activities, and help the determination of not only the causes of damages, but also the extent of remedial works, once the damage is identified [33, 34].

There are nine different types of sensors to form the sensory systems, which include anemometers, thermometers, servo-type accelerometers, dynamic weigh-in-motion sensors, GPS receivers, leveling sensing stations, displacement transducers, weldable strain gauges and monitoring CCTV cameras, making a total of 848 sensors on the Tsing Ma and two other adjacent cable-stayed bridges [33]. GPS technology was introduced because of its improved measurement efficiency and accuracy [29]. Of a total of 29 high-grade geodetic-type dual-frequency GPS receivers (using both L-band carries), two GPS receivers were set up as the reference stations. On the Tsing Ma Bridge, 14 GPS receivers were permanently installed as the monitoring stations at the critical locations of the bridge deck where maximum displacements are expected: four pairs on the bridge deck itself, one pair on the each side of the supporting towers, and one pair on the cable at midspan. The remaining 13 GPS receivers were installed onto two cable-stayed bridges near the Tsing Ma Bridge to form a large monitoring system in the region. Both chokering

and lightweight antennas were used for the reference stations and monitoring stations, respectively. Most GPS stations were installed close to the leveling and sensing stations or to the accelerometers for validating and comparing the data sets from different monitoring sensor systems. GPS data were collected at a sampling rate of 10 Hz continuously, and optical fibers were used to transmit the positioning results from each monitoring station to a processing center for further analysis and visualization purposes.

The Akashi Kaikyo Bridge in Japan is the world's longest road suspension bridge. It was constructed using a newly developed wind- and seismic-resistant design and it is necessary to verify the design assumptions and constants during strong winds and severe earthquake loadings [25]. Navigation is extremely difficult and there have been many shipping accidents in the Akashi Strait because of the hostile environmental conditions. To keep the world's longest suspension bridge in a safe operating condition and also to study the dynamic response to various loadings, an experimental trial was carried out by using dual-frequency GPS receivers in 1999 to monitor bridge deformation in a real-time mode. The first results from the monitoring of the Akashi Kaikyo Bridge were presented in [25]. GPS receivers were set up at three locations, constituting a reference station at the anchorage on the Kobe side of the bridge and two monitoring stations on the top of a supporting tower and at midspan, to monitor three-dimensional displacements. Wind loading and temperature data were also collected simultaneously. The displacement time series of six months were analyzed together with recorded time series of wind and temperature loadings. The instantaneous gradients of displacement and temperature were calculated. The regression functions established could be used to identify future abnormal deflections after attacks of strong wind loading or earthquakes. Data from other sensors are also available for this bridge, but for a bridge of this size, three GPS receivers seem inadequate for even picking up global displacements.

China has many of the world's longest suspension bridges, which were constructed in recent years as a result of the high demand for land transport infrastructure. Structural safety has already caused great concern and many SHM systems have been installed in many new bridges. The most commonly used sensors are anemometers, accelerometers, temperature sensors, and strain gauges. Only a very few long suspension bridges in the mainland China such as the Jiangying Bridge (main span of 1385 m) and the Runyang Bridge (main span of 1490 m), both over the Yangtze River were installed with permanent GPS monitoring systems. Like the Akashi Kaikyo Bridge, only a limited number of monitoring GPS stations have been installed at the critical locations on the bridge deck and supporting towers. For instance, there are nine permanently installed dual-frequency GPS receivers forming an important part of the monitoring system for the Jiangyin Bridge. One GPS receiver is used as the reference station and other eight receivers are placed atop the two bridge towers, and at the 1/4, 1/2, and 3/4 points of the bridge span [35]. The sampling rate of the GPS receivers is set to 20 Hz.

In other countries, GPS has been used in a number of bridge monitoring studies, mostly carried out by GPS receivers temporarily installed on the bridge decks or supporting towers to collect sample data sets. For example, GPS receivers were employed to monitor the displacement of France's Normandy Suspension Bridge under controlled traffic loading to verify whether the performance of the bridge was consistent with the design specifications [36]; the Danish Road Directorate used GPS to determine the as-built geometry and assess temporal deformations of Denmark's Storebaelt Bridge and other bridges [37]; and the Applied Research Laboratory (ARL) of the University of Texas at Austin in the United States has conducted a series of research on GPS-based structural deformation monitoring since the beginning of 1990s and published many valuable articles on relevant topics [26, 38–43]. The research conducted by the ARL is perhaps the earliest practice in GPS-based bridge deformation monitoring.

The results of a three-day experiment, in which GPS technology was used to measure the motion of a cable-stayed suspension bridge over the Mississippi River near New Orleans, were presented in [26]. The goals of this experiment were to verify the feasibility of using GPS to achieve centimeter-level accuracy in bridge deflection monitoring, and to evaluate the practicability and limitations of the whole monitoring system. Twelve Trimble 4000 SST GPS receivers were used in the test. Two receivers were located on the shore at known reference sites about 10 m apart, approximately 1.6 km away from the bridge. This approach made it possible

to distinguish real bridge motion from GPS-induced biases or errors through a thorough data processing. The remaining 10 GPS receivers were situated at the critical bridge sites where minimum and maximum bridge displacements were expected. A 10-s sampling rate was chosen because the expected bridge motion during a 10-s period was considered to be negligible, and this allowed all the data to be stored on receivers for the whole monitoring period. The result showed a maximum relative vertical displacement of about 7 cm caused by a change in ambient temperature during the experiment. The results revealed a strong relationship between the vertical position of the bridge roadbed and the change in ambient temperature. Multipath also made significant error contribution to the positioning solution and caused problems in the displacement analysis. It was clear that the cables and metal surroundings, the bridge sites, coupled by the passing vehicles, had created a severe multipath environment and additional multipath analysis and mitigation techniques are apparently essential for achieving high accuracy positioning. Other GPS-based bridge deformation trials were also carried out by the ARL team with the attempt to monitor more dynamic bridge performance induced by traffic and wind loadings with 10-Hz GPS receivers, identify the multipath signature and quantify the measurement quality, compare GPS-measured displacements with those of servo accelerometers, and assess the reliability and performance of system components [42].

Preliminary results from ARL and other research teams around the world proved that GPS is a viable tool for detecting both transient deformations of flexible bridge structures and long-period movements. It was also recognized that the big challenge with GPS-based bridge deformation monitoring was how to effectively eliminate or reduce the impact of multipath. Other factors restraining large-scale GPS applications to structural deformation monitoring are the high price of dual-frequency GPS receivers, the large amount of data generated during data collection when a high sampling rate is used, and the existing gaps between the demanded high precision with what can be offered by current GPS positioning technology of centimeter accuracy.

Researchers at the IESSG in the University of Nottingham have conducted a series of trials using state-of-the-art dual-frequency GPS receivers to monitor the movements of a number of long suspension bridges and other structures in the United Kingdom [22, 28, 44]. In particular, several trials were conducted on the Humber Bridge under both a controlled loading environment and using ambient vibration to monitor the bridge deflection [22, 45]. The viability of GPS technology to monitor the displacements of the Humber Bridge subjected to known loading conditions was verified. The initial research of the IESSG was focused on the viability of GPS technology to monitor both long-term bridge settlements and dynamic response to external loading, and the latest work is on the development of a systematic approach for the sensor integration, field test arrangement, data collection, algorithm, and software development, quality control, advanced signal processing, and multipath mitigation techniques, data analysis and dynamic response identification, and result visualization [46–53] (*see* **Modular Architecture of SHM System for Cable-supported Bridges**; **Monitoring of Bridges in Korea**; **Bridge Monitoring in Japan**; **Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge** and **SHM of a Tall Building**).

2.3 Advantages and limitations of GPS for bridge deformation monitoring

Very few conventional bridge monitoring sensor systems have versatile capacities like GPS. In summary, the advantages in using GPS for SHM of long bridges are as follows [27, 38–41, 54, 55]:

- monitoring both static and dynamic movements;
- fully automatic, real-time, all-weather, and long-life data acquisition;
- 3D absolute displacements linked to a global datum WGS84 (World Geodetic System of 1984 —very important for foundation monitoring);
- accurate velocity and acceleration extractions from 3D displacements, which can be validated by the measurements of other sensors such as an accelerometer;
- precisely synchronized positioning solutions for all GPS monitoring sites;

- a precise time source for time stamping/synchronizing other measurements/sensors (accelerometers, weather station, strain gauge, etc.);
- continuously increasing sampling rate (50 Hz for dual-frequency GPS receivers and up to 100 Hz or higher for single frequency GPS receivers);
- no long-term positioning accuracy degradation like accelerometers;
- a data fusion platform for integrating GPS with other sensors for a more robust monitoring system;
- continuous improvement in space segments, regional augmentation systems, ground tracking facilities/algorithms, and end-user hardware (futuristic GGG—GPS, GLONASS, GALILEO—receivers will be able to track more than 80 satellites);
- in the near future, a single GGG receiver based real-time solution of millimeter positioning accuracy at a sampling rate for at least 100 Hz will be possible.

GPS also has some inherent disadvantages, which are the potential inhibitors to a wide acceptance of GPS for SHM. The main disadvantages are as follows [27, 38–41, 54, 55]:

- The high price of GPS hardware and software is the major inhibitor for a wide application of GPS for SHM; a much cheaper single frequency GPS receiver cannot fix its initial integer ambiguity, which prohibits centimeter positioning accuracy.
- Line of sight to at least five well-distributed satellites poses a current challenge for GPS owing to signal obstruction by surroundings.
- The positioning accuracy of the vertical component is worse than the horizontal one and the ratio of the positioning accuracies of two horizontal components changes with geographical locations.
- Unmodeled residual tropospheric delays, such as those induced by the height differences, might cause a few centimeter errors.
- At the moment, high level of multipath caused by surrounding reflection cannot be effectively mitigated.
- The relatively slow sampling rate is not adequate for monitoring short span bridges or the higher frequency band of a long span bridge.
- Data processing and interrogation can be complicated.
- There has not been a thorough research into the integration of GPS measurements with a computational model, nor has there been a successful demonstrator to justify the high investment in the establishment of the sensor system.

3 IMPLEMENTATION OF GNSS CENTERED SENSOR SYSTEMS FOR SHM

3.1 Reference and monitoring GPS stations

For the establishment of a permanent GPS-based bridge monitoring system, monuments need to be installed for mounting both the reference and monitoring antennas as shown in Figure 2(b, c). Different GPS antennas have different weights, physical dimensions and multipath mitigation capacity, and can

(a) (b) (c)

Figure 2. (a) A chokering antenna; (b) a reference station set up on the Jingyin Bridge in China; (c) a monitoring station atop the Akaishi Kaikyo Bridge in Japan [photos courtesy of Leica Geosystems].

be used for different GPS sites. For instance, a chokering antenna covered with a dome could be used for the reference station since its location is usually lower than the bridge deck. High multipath signature caused by the supporting towers, dense cables, passing vehicles, and buildings is evident and can be mitigated by a chokering antenna. A chokering antenna consists of a series of cylinders that stop reflected signals from the ground and from nearby reflective surfaces as shown in Figure 2(a). To significantly reduce multipath induced by the reference stations and for designing a more secure and robust monitoring system, a reference configuration comprising multiple GPS receivers is highly recommended [27]. An inexpensive and lightweight patch antenna as shown in Figure 2(c) can be used for the monitoring sites on the bridge sites. This antenna has a metallic ground plate, which can effectively stop the reflected signals coming underneath the antenna but will not be able to reject multipath from surrounding objects such as cables, towers, and passing vehicles. This is why the high monuments for monitoring sites are preferred as shown in Figure 2(c).

For temporary bridge deformation monitoring practices, antennas of the reference stations can be mounted on tripods, which are centered to the precisely measured ground marks as shown Figure 3(a) (the Forth Road Bridge). In this monitoring, two adjacent reference stations about 6 m apart were used to reduce multipath caused by the reference stations through a data processing procedure. The other reason for this reference station setup is to reduce the risk of potential hardware failure caused by either of the reference stations. The antennas of the bridge monitoring sites can be locked to the bridge handrail using specially designed clamps. A direct access to power was made available by the bridge authority. Figure 3(b) shows a typical example of a monitoring station setup on the Forth Road Bridge [51].

At the moment, the sampling rate of the operational or experimental bridge monitoring systems is set to 10 Hz or a maximum of 20 Hz. At such a sampling rate, it is estimated that the average data volume in a binary format is $7\,\mathrm{Mb\,h^{-1}}$. For a permanent monitoring system, the GPS receivers at the monitoring sites will conduct real-time coordinate computation and then stream these continuous time series to the control center for a further analysis. These receivers might also be able to log the raw measurements into onboard flash memory cards and send the data sets to the center on a regular basis in case of some important event analysis, for instance, an earthquake. For the experimental tests, raw data is recorded into the flash memory cards and up to 4-GB cards are currently available, which means that raw data sets of more than 574 h (equivalent to an observation of 23 days) can be continuously recorded. However, it might cause computation problem with a postprocessing GPS processing software package due to the huge file size after uncompressing the binary data. It is recommended that a data set of 12 h should be used since this data size can be handled by any commercial GPS software and a standard desktop PC. The use of two flash cards at one site could assure nearly continuous measurements at each site.

(a) (b)

Figure 3. Reference stations setup and a monitoring station on the Forth Road Bridge [51].

3.2 GPS sensor locations on the bridge

As previously discussed in this article, a widely adopted sensor configuration for most bridge monitoring systems is to install the GPS sensors onto both sides of the supporting towers, at the 1/4, 1/2, and 3/4 points of the bridge span, with the attempt to extract structural dynamics [29, 51, 56]. With a sensor configuration as described in Figure 4, special events or structural response can be easily identified. Figure 5 shows the vertical deflection of a 100-t lorry that passed each of the monitoring sites from south to north on the Forth Road Bridge. Whether this configuration is optimal and adequate for detecting global bridge deformation and deflection is still an unanswered question. Some preliminary research was conducted using a short span suspension bridge as a case study for the optimal determination of the number and locations of GPS sensor [56].

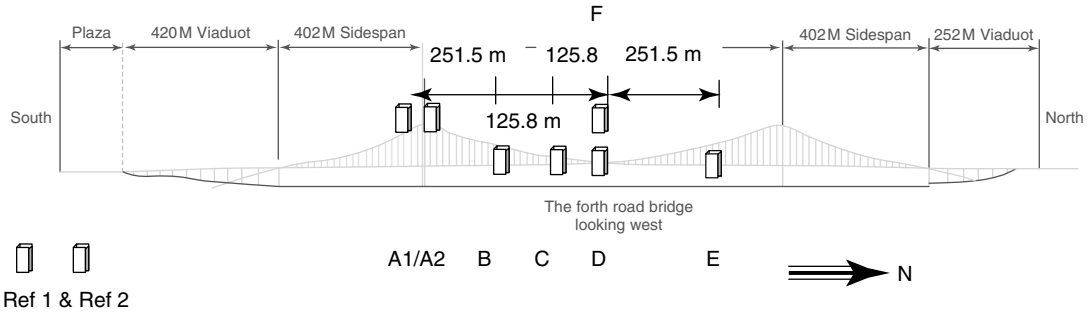


Figure 4. GPS antenna layout for monitoring the Forth Road Bridge [51].

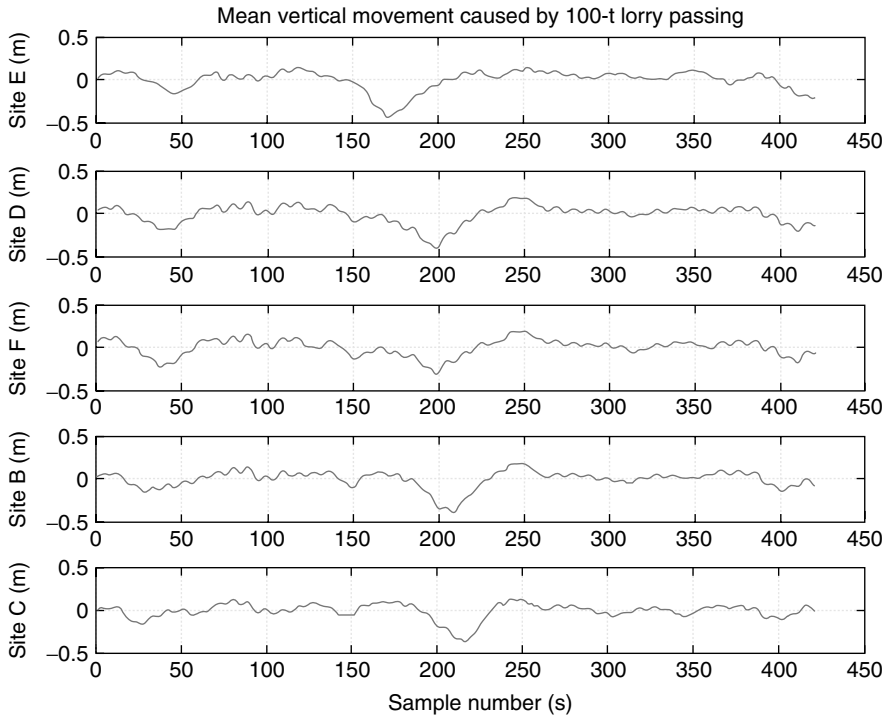


Figure 5. Vertical deflection of five GPS sites caused by a 100-t lorry passing.

3.3 Coordinates and coordinate transformation

The reference datum adopted by GPS positioning is called *the World Geodetic System of 1984*, or *WGS84*, which is a 3D, earth-centered earth-fixed (ECEF) system. Under this reference system, the coordinates determined by a GPS receiver are expressed as either geodetic coordinates (latitude, longitude, and ellipsoidal height) or Cartesian coordinates (x, y, and z). These coordinates do not make sense for many engineering applications such as analyzing 3D bridge deformation, since they are not aligned with the bridge axes. So, these geodetic coordinates within WGS84 must be transformed into rectangular grid coordinates (easting and northing). With the information about the local mean sea level, which defines a local height datum, geoid, the difference between WGS84 (ellipsoid) and the geoid can be estimated, which is called as *geoidal height*. By using this geoidal height the measured ellipsoid height of a point on the earth surface can transformed as the orthometric height, a height referred to as the local geoid. For GPS-based SHM application, using either orthometric height or ellipsoidal height does not make any difference if only one height datum is chosen for whole data analysis. However, the engineering coordinate system of a bridge, or a bridge coordinate system (BCS), is normally defined as a right-handed Cartesian coordinate system using the longitudinal axis of the bridge as x axis, the lateral direction as the y axis, and the vertical direction as z axis [27]. Hence, a second coordinate transformation for transforming the coordinates in a local coordinate system to those in a BCS involving translation and rotation is further conducted through the determination of a rotation angle of the bridge main axis from the north direction and the coordinates of BCS origin in the local grid coordinate system.

3.4 Identification of wind and thermal induced bridge deformation through GPS positioning

GPS is a very useful tool in detecting deformations induced by wind and thermal loadings. The bridge response to wind loading characterizes a short-term vibration but the deformation induced by the diurnal change of temperature is a very slow process. Figure 6(a) shows wind speed measured in a period of 48 h and Figure 6(b) shows the lateral responding deformation at a midspan point of the Forth Road Bridge. The cross-correlation coefficient between wind speed and lateral deformation is 77%, demonstrating a very high level force and response effect [49]. Figure 7 describes the relationship between the thermal loading and the vertical bridge deformation. Delay in the deformation is evident. The calculated correction coefficient is -39%. In considering the deformations caused by other affecting factors, such as traffic and wind loading, this correlation is still very high. From these two figures, the capacity of GPS for monitoring both dynamic and semistatic deformations is further confirmed.

3.5 Deformation analysis and interface with computational models

A simple bridge deformation analysis is to draw 3D deformation time series and compare these deformation curves with safety thresholds. Fast Fourier transform (FFT) algorithm can be used to identify dominant vibration frequencies and maybe the changes of these frequencies could provide some information about the change of the bridge's properties. A more complicated time-frequency analysis approach has attracted the interest of many researchers around the world and can also be employed to analyze the change of these frequencies against time [18, 27, 57]. When all the deformation time series collected from each monitoring site are gathered, it is expected that GPS can be used for the global monitoring of a bridge dynamics with the possibility to connect to localized sensors for detecting potential damage, as proposed in [34]. Hence, correct interpretation of GPS data for structural dynamics forms an important and unique part of an SHM system. From field data collection point of view, any procedures that could assure the measurement quality in a cost-effective manner should be taken. However, more study should be initiated in the near future for the interpretation and analysis of field measurements (*see* **Time–frequency Analysis**).

Some preliminary research has been carried out to extract dynamics from GPS-measured coordinate time series [58]. How the GPS measurements could

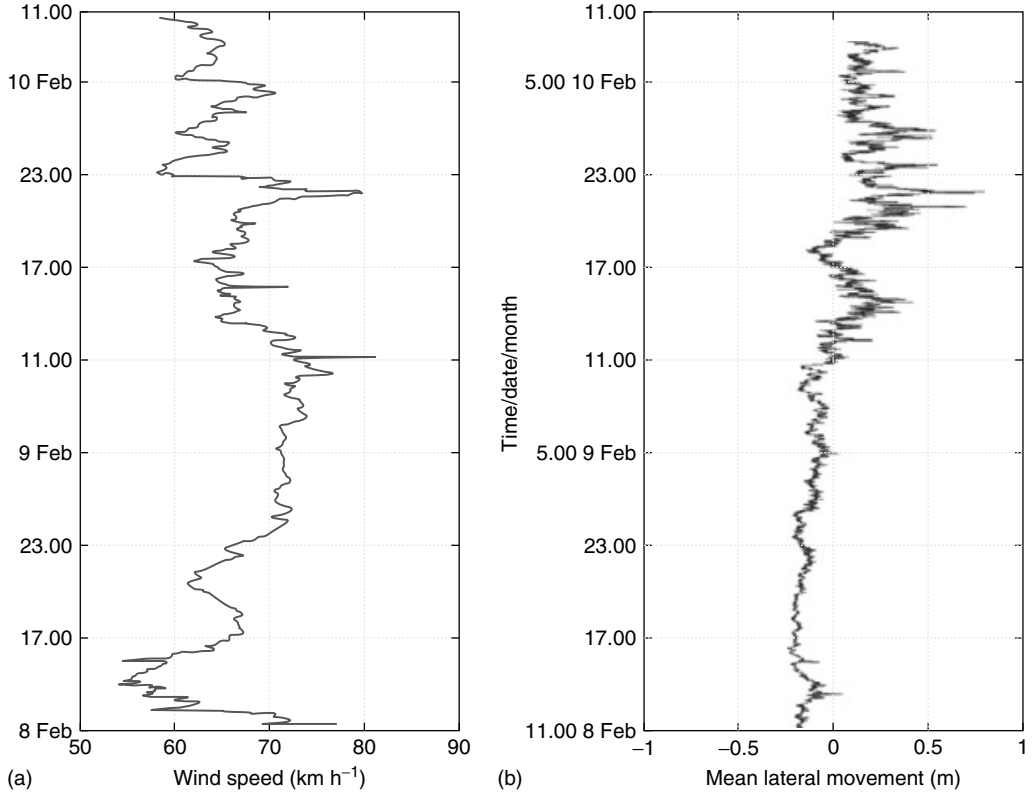


Figure 6. Wind loading and the lateral deflection of the Forth Road Bridge (a 48-h period).

be synergized with a computational model such as a finite element (FE) model is recommended in [58] and the whole loop of a practical SHM comprises the following steps:

Step 1—initial FE model creation;
Step 2—optimal sensor placement;
Step 3—bridge monitoring data collection;
Step 4—model updating (correlation of the FE model to the test data);
Step 5—further bridge monitoring data collection;
Step 6—FE model/real data comparison (comparison between the predictions of an FE model with real bridge deformations during its operational life).

3.6 Integrated sensor systems

Owing to the severe environmental conditions encountered in the bridge deflection monitoring, the instruments used must be lightweight, portable, reliable, and easy to install and the results must be easy to interpret. These are of great importance under extreme loading scenarios such as strong wind, volcanic eruption, and earthquake. At the same time, for correctly interpreting the dynamics of monitored bridges, the measurements should meet accuracy specification requirements. This means the deflections of the bridge should be measurable with available surveying instruments. For instance, to measure centimeter-level deflections the internal accuracy of a GPS receiver should be better than a few millimeters level, and multipath and other error sources should be appropriately mitigated or modeled.

Conventional surveying methods such as leveling have been used in the past to monitor static displacements of engineered structures with millimeter level or higher accuracy, and will certainly continue to be used in the future.

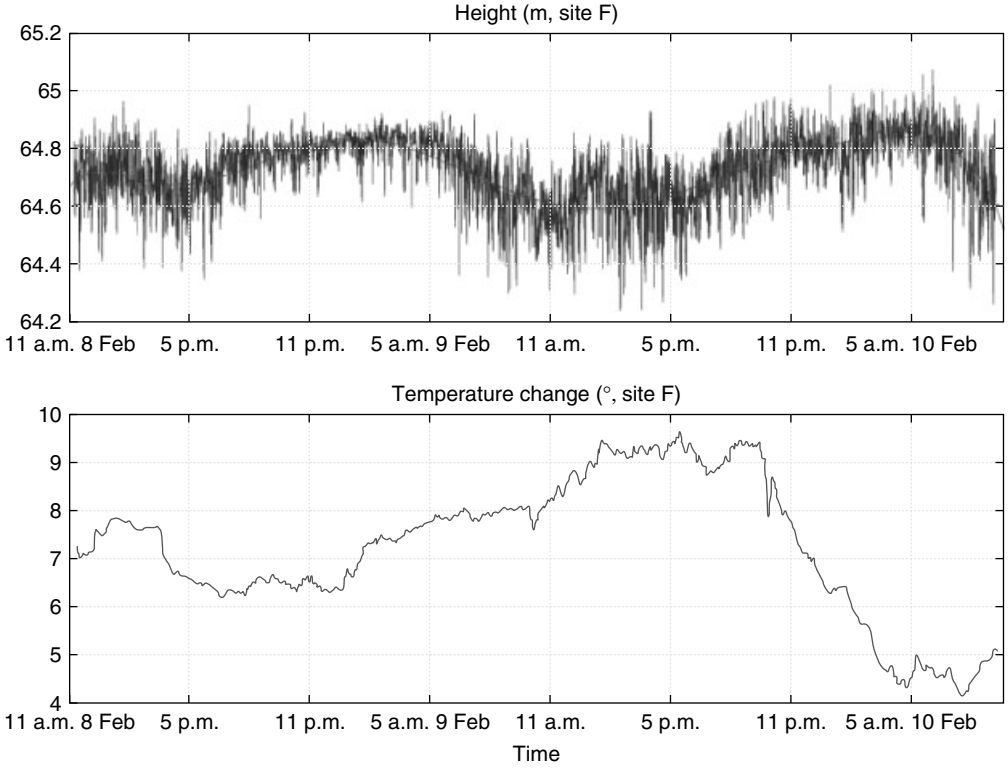


Figure 7. Diurnal temperature change versus the height variation at the midspan of the Forth Road Bridge (a 48-h period).

Modern leveling sensing stations could provide one-dimensional accuracy of about 2 mm at a sampling rate of 2.56 Hz [29]. Surveying robots, electronic distance meter (EDM), theodolites, surveying total stations, photogrammetry, and other surveying instruments could be employed to monitor structural deformation. However, the inherent disadvantages of these terrestrial surveying systems have greatly limited their applications. Previous research reveals that the main disadvantages of these surveying approaches can be summarized as follows:

- long-term intervals between measurements (days or even months);
- averaging of data over a relatively long time span (often some hours are smoothed, which leads to smoothing effects that could hide real movements of the stations);
- relatively low data sampling rate and a poor level of automation;
- batch mode analysis (data is collected, transmitted to a computer, and evaluated a few hours later).

Because of the above limitations of these terrestrial surveying methods, they cannot be employed to monitor structures with dynamic structural deflection and semistatic movements at the same time.

Accelerometer is used as an indispensable sensor in SHM of bridges for the identification of its dynamic characteristics. A triaxial accelerometer could measure three orthogonal accelerations simultaneously. Compared with other surveying systems, a triaxial accelerometer has some special advantages when it is used for bridge monitoring. For instance, the sampling rate can reach several hundreds of hertz depending upon application requirements. Triaxial accelerometers are superior to other sensors since they are not dependent on propagation of electromagnetic waves, and therefore avoid the problems of signal reflection or refraction and line-of-sight connections to the terrestrial or space objects. An

accelerometer could form a completely self-contained monitoring system, utilizing only measurements of accelerations to infer the positions of the system, through integration based on the laws of motion. However, the positional drift of an accelerometer grows extremely rapid with time and can reach hundreds of meters after an interval of several hours [59]. The main error sources come from the instrumental biases and scale factor offsets and the unknown gravity of the earth. Continuous updating such as zero velocity update (ZUPT) or coordinates update (CUPT) is used to avoid error accumulation. It is the need to update that has severely restricted the wide applications of accelerometer technology as a standalone positioning method in surveying. In bridge deflection monitoring, it is impossible to conduct ZUPT; CUPT aided with GPS fixes could be a realistic option to overcome drift problem of accelerometer.

An integrated monitoring system consisting of GPS receivers with triaxial accelerometers could overcome the shortcomings of each individual sensor system and provide a much improved overall monitoring system in terms of productivity and reliability. To eliminate any potential sensor misalignment errors, the researchers in the University of Nottingham designed a cage that hosts both GPS antenna and a triaxial accelerometer as shown in Figure 8 [27]. Detailed GPS and accelerometer data fusion technique was also introduced by the same author and other researchers [30, 60].

As discussed by [47, 50, 61], current GPS constellation causes an uneven distribution of satellites in view: in the equatorial area a more scattered satellite distribution is possible, which means a nearly identical easting and northing positioning accuracies; but in the polar areas, no satellites above 45° elevation angle are visible, which causes degraded 3D positioning accuracies. The worst factor prohibiting wide GPS application for SHM of bridge is that the vertical positioning accuracy is always the worst component in the three coordinates and this might lead to wrong interpretation of actual bridge deformation [50]. In addition to the introduction of high-level multipath, dense cables, supporting towers, and other surroundings also cause satellite visibility problem, making the visible satellites less than adequate for starting a positioning fix. If several ground-based pseudolites could be installed with very low elevation angles,

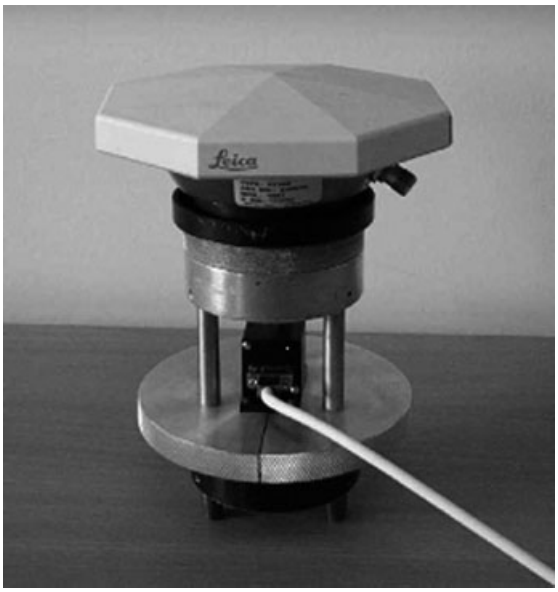


Figure 8. Device for integrated GPS antenna and a triaxial accelerometer.

it will significantly enhance the satellite geometry and positioning accuracy, especially in the vertical direction [47]. Pseudolites were used for the proof of GPS concept in the 1970s, before the launch of real GPS satellites. In the 1990s, due to the high demand in highly precise positioning solutions and the deficiency of current GPS constellation, the pseudolite concept was reinvented for augmenting GPS satellites.

A joint research has been conducted by the University of New South Wales and the University of Nottingham to verify the effect of introducing extra ground-based pseudolite transmitters for bridge deformation monitoring [48, 62–64]. Data processing of both GPS and pseudolite measurements reveals that if they are used correctly pseudolites can augment GNSS satellite geometries and improve 3D positioning accuracies to several millimeters. For the Wilford Bridge monitoring trials in Nottingham, three pseudolites were set up on the northern side of the bridge to fill the hole as shown in Figure 9 for obtaining more complete transmitter geometry. Compared with GPS-only solutions, the actual accuracy improvements in the north, east, and vertical directions are 14, 36, and 46%, respectively. The simulated accuracy improvement in the north, east,

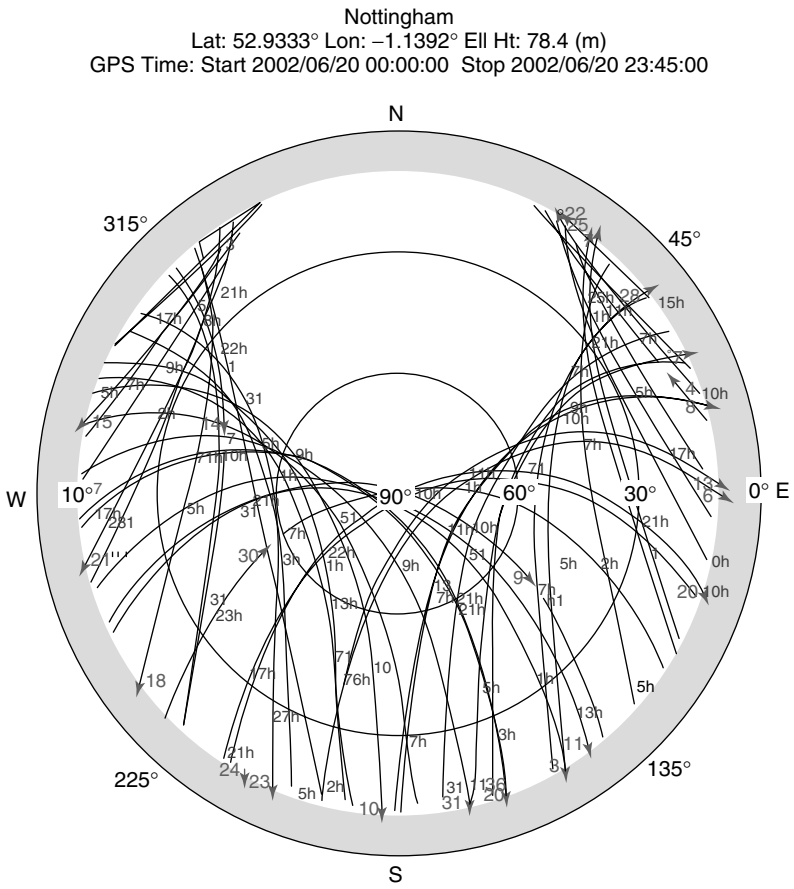


Figure 9. Skyplot of GPS satellites in Nottingham, showing the deficiency of GPS constellation.

and vertical directions are 19, 34, and 59%, which confirms the feasibility of an integrated GPS and pseudolite monitoring system for achieving more robust, accurate, and reliable results, especially in the vertical coordinate. However, since the early pseudolites transmitted on L1 frequency band, which could potentially interfere with GPS signals, the use of pseudolites was banned in many countries and its use was restrained in the United Kingdom. Collaborative research has lead to further cooperation in using a new generation pseudolite, Locatalite, which was pioneered by the Locata; it transmits at the 2.4-GHz frequency band to avoid the interference with the GPS signals [62].

Other integration approaches can also be employed in the SHM of long bridges, such as an integrated monitoring system consisting of an inertial navigation system (INS) and GPS. But the cost of this system could dramatically increase because of the high cost of an INS device.

4 FUTURE VISION

For improving positioning accuracy and availability, system integrity, and reliability, GPS is currently under its modernization through updating its space segment, improving signal quality and computation algorithms. Of course, the revival of Russia's GLONASS, and the development of European Galileo system will significantly increase signal availability and improve space geometry. China is expanding its regional navigation system, Beidou or COMPASS system, to a global coverage. Within a five-year period, the total number of

navigation satellites could reach more than 110 and this will make high and pervasive real-time positioning less of a challenge. For solving the problem of signal obstruction by surroundings and making positioning in severely obstructed areas possible, R&D of regional space-, ground-, and aircraft-based SatNav augmentation systems such as Australia's Ground-Based Regional Augmentation System (GRAS), India's GPS and GEO Augmented Navigation (GAGAN), Japan's Quasi-Zenith Satellite System (QZSS) system and MTSAT Satellite-based Augmentation System (MSAS), the European Geostationary Navigation Service (EGNOS), the United State's Wide Area Augmentation Systems (WAAS), etc., is the main activity of the national space programs of these countries. Establishment of network-based RTK GNSS reference station infrastructure at a national or regional level to support improved positioning accuracy and mobility is also a recent focus in GNSS development. For instance, the Ordnance Survey of the United Kingdom has established a network GNSS facility, which consists of more than 100 permanent stations, covering the whole country. High-quality real-time RTK corrections can be received wirelessly with a GPRS-embedded GPS receiver for achieving 3D positioning accuracy of a few centimeters OTF. For monitoring the deformation of the Humber Bridge, the bridge authority only need to subscribe to the service providers for the precise (The Radio Technical Commission for Maritime Services) RTCM corrections for carrying out continuous monitoring in a more cost-effective and reliable manner. The progress in GNSS receiver and antenna technology will also make future GNSS receivers and antenna more sensible to weak signals and more robust to reject external inference and multipath. However, the following questions need to be further addressed. For instance, is 3D positioning to an accuracy in millimeters high enough for extracting the parameters of bridge deformation at sampling rate of up to 100 Hz or higher? Will network RTK GNSS with more than 110 satellites and other regional augmentation systems be able to provide more reliable, continuous, and robust positioning for monitoring bridges from several meters to several thousand meters of bridge spans? Will future wireless communications fully replace current optic-fiber or cable-based communications for both streaming RTCM corrections from the reference stations to the rover receivers and also sending the positioning solutions to the control center for further analysis? Will GNSS positioning also be able to help predict the existing life span of a bridge effectively? More effort is required to address these questions through individual or collaborative research around world and this will eventually make GNSS positioning a feasible technology for SHM of not only bridges but any civil engineering structures as well.

REFERENCES

[1] Hofmann-Wellenhof B, Lichtenegger H, Collins J. *Global Positioning System: Theory and Practice*. Springer-Verlag: New York, Wien, 1997.

[2] The Institute of Navigation (ION), *Global Navigation System*, 1980.

[3] El-Rabbany A. *Introduction to GPS: The Global Positioning System*. Artech House: London, 2002.

[4] Kim D, Langley RB. GPS ambiguity resolution and validation: methodologies, trends and issues. *Proceedings ION GNSS 2000*. Salt Lake City, UT, September 2000.

[5] Yang Y, Sharpe RT, Hatch RR. A fast ambiguity resolution technique for RTK embedded within a GPS receiver. *Proceedings ION GNSS 2006*. Fort Worth, TX, September 2006.

[6] Frei E, Beutler G. Rapid static positioning based on the fast ambiguity resolution approach FARA: theory and first results. *Manuscripts Geodaetia* 1990 **15**:325–356.

[7] Landau H, Euler HJ. On-the-fly ambiguity resolution for precise differential positioning. *Proceedings ION GPS-92*. Albuquerque, NM, September 1992.

[8] Farrar CR, Worden K. Preface. *Philosophical Transactions of the Royal Society of London, Series A* 2007 **365**:299–301.

[9] Rizos C, Han S, Ge L, Chen H-Y, Hatanaka Y, Abe K. Low-cost densification of permanent GPS networks for natural hazard mitigation: first tests on GSI's GEONET network. *Earth Planets Space* 2000 **52**:867–871.

[10] Rizos C, Han S, Roberts C, Han X. Continuously operating GPS-based volcano deformation monitoring in Indonesia: challenges and preliminary results. Geodesy beyond 2000: the challenges of the first decade. *Proceedings IAG General Assembly*,

ISBN 3-540-67002-5. Springer-Verlag: Birmingham, AL, July 1999, pp. 361–366.

[11] Hudnut KW, Behr JA. Continuous GPS monitoring of structural deformation at pacoima dam, California. *Seismological Research Letter* 1998 **69**(4):299–308.

[12] Hudnut KW, Bock Y, Galetza JE, Webb FH, Young WH. The Southern California integrated GPS network (SCIGN). *Proceedings Deformation Measurements and Analysis, 10th International Symposium on Deformation Measurements*. Orange, CA, March 2001.

[13] Teferle FN, Bingley RM, Dodson AH, Penna NT, Baker TF. Using GPS to separate crustal movements and sea level changes at tide gauges in the UK. In *Vertical Reference Systems*, Drewes H (ed). International Association of Geodesy Symposium, Springer-Verlag, 2001.

[14] Brunner FK, Hartinger H, Richter B. Continuous monitoring of landslides using GPS: a progress report. *Proceedings of Geophysical Aspects of Mass Movements*. Austrian Academy of Sciences: Vienna, 2000, pp. 75–88.

[15] Forward T, Stewart M, Penna N, Tsakiri M. Steep wall monitoring using switched antenna arrays and permanent GPS network. *Proceedings Deformation Measurements and Analysis, 10th International Symposium on Deformation Measurements*. Orange, CA, March 2001.

[16] Breuera P, Chmielewskib T, Orskic PG, Konopkad E. Application of GPS technology to measurements of displacements of high-rise structures due to weak winds. *Journal of Wind Engineering and Industrial Aerodynamics* 2002 **90**:223–230.

[17] Guo J, Ge S. Research of displacement and frequency of tall building under wind load using GPS. *Proceedings ION GPS-97*. Kansas City, MO, September 1997.

[18] Kijewski-Correa T, Kareem A, Kochly M. Experimental verification and full-scale deployment of global positioning systems to monitor the dynamic response of tall buildings. *Journal of Structural Engineering* 2006 **132**(8):1242–1253.

[19] Lovse JL, Teskey WF, Lachepelle G, Cannon ME. Dynamic deformation monitoring of tall structure using GPS technology. *Journal of Surveying Engineering* 1995 **121**(1):35–40.

[20] Nakamura SI. GPS measurement of wind-induced suspension bridge girder displacements. *Journal of Structural Engineering* 2000 **126**(12):1413–1419.

[21] Tamura Y, Matsui M, Pagnini L-C, Ishibashi R, Yoshida A. Measurement of wind-induced response of buildings using RTK-GPS. *Journal of Wind Engineering and Industrial Aerodynamics* 2002 **90**:1783–1793.

[22] Ashkenazi V, Dodson AH, Moore T, Roberts GW. Monitoring the movements of bridges by GPS. *Proceedings ION GPS-97*. Kansas City, MO, September 1997.

[23] Ashkenazi V, Dodson AH, Moore T, Roberts GW. Real time OTF GPS monitoring of the Humber Bridge. *Surveying World* 1996 **4**(4):26–28.

[24] Barnes JB, Rizos C, Wang J, Meng X, Dodson AH, Roberts GW. The monitoring of bridge movements using GPS and pseudolites. *Proceedings 11th International Symposium on Deformation Measurements*. Santorini, May 2003.

[25] Fujino Y, Murata M, Okano S, Takeguchi M. Monitoring system of the Akashi Kaikyo Bridge and displacement measurement using GPS. *Proceedings of SPIE; Proceedings Nondestructive Evaluation of Highways, Utilities, and Pipelines IV*. 2000.

[26] Leach M. Results from a bridge motion monitoring experiment. *Proceedings Sixth International Geodetic Symposium on Satellite Positioning*. The Ohio State University, 1992; pp. 801–810.

[27] Meng X. *Real-time Deformation Monitoring of Bridges Using GPS/Accelerometers*, Ph. D., The University of Nottingham: Nottingham, 2002, http://etheses.nottingham.ac.uk/archive/00000279/.

[28] Roberts GW, Dodson AH, Ashkenazi V. Twist and deflection: monitoring motion of Humber Bridge. *GPS World* 1999 **10**(10):24–34.

[29] Wong KY, Man KL, Chan WY. Monitoring Hong Kong's Bridges: real-time kinematic spans the gap. *GPS World* 2001 **12**(7):10–18.

[30] Li X, Ge L, Ambikairajah E, Rizos C, Tamura Y, Yoshida A. Full-scale structural monitoring using an integrated GPS and accelerometer system. *GPS Solutions* 2006 **10**(4):233–247.

[31] Kashima S, Yanaka Y, Mori K. Monitoring the Akashi Kaikyo Bridge: first experiences. *Structural Engineering International* 2001 **11**(2):120–123.

[32] Wikipedia, *List of World Longest Suspension Bridges*, http://en.wikipedia.org/wiki/List_of_longest_suspension_bridges (accessed Dec 2007).

[33] Wong K-Y. Instrumentation and health monitoring of cable-supported bridges. *Structural Control and Health Monitoring* 2004 **11**:91–124.

[34] Wong K-Y. Design of a structural health monitoring system for long-span bridges. *Structure and Infrastructure Engineering* 2007 **3**(2):169–185.

[35] Leica Geosystems, *Monitoring with GPS RTK Technology*, Jiangyin Bridge, http://www.leica-geosystems.com/corporate/en/ndef/lgs_61984.htm (accessed Dec 2007).

[36] Fairweather V. Measuring bridge movement. *Civil Engineering, ASCE* 1996 **66**(6):48.

[37] Norgard P. Deformation survey of the storebaelt bridge: GPS shows its merits. *Geomatics Info* 1996 **10**(4):37–39.

[38] Duff K. Deformation monitoring with GPS, Part 1: system design and performance. *Proceedings Symposium on Surveying of Large Bridge and Tunnel Projects (FIG)*. Copenhagen, 1997.

[39] Duff K, Hyzak M. Structural monitoring with GPS. *Public Roads* 1997 **60**(4):39.

[40] Hyzak M, Leach M, Duff K. Practical application of GPS to bridge deformation monitoring. *Proceedings of the 64th Permanent Committee Meeting and Symposium of International Federation of Surveyors (FIG)*. Washington, DC, May 1997.

[41] Duff K, Nelson S. Deformation monitoring with GPS, Part 2: performance, affordability, and technology development. *Proceedings Symposium on Surveying of Large Bridge and Tunnel Projects (FIG)*. Copenhagen, 1997.

[42] Hyzak M, Leach M. Bridge monitoring by GPS. *Surveying World* 1995 **3**(3):8–11.

[43] Tolman BW, Craig BK. An integrated GPS/accelerometer system for low dynamics. *Proceedings International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation*. Banff, June 1997.

[44] Meng X, Dodson AH, Roberts GW, Cosser E. Hybrid sensor system for bridge deformation monitoring: interfacing with structural engineers. A window on the future of geodesy. *Proceedings of the International Association of Geodesy. IAG General Assembly*. Springer-Verlag: Sapporo, June 30–July 11, 2003.

[45] Cosser E, Roberts GW, Meng X, Dodson AH. Single frequency GPS for bridge deflection monitoring: progress and results. *Proceedings 1st FIG International Symposium on Engineering Surveys for Construction Works and Structural Engineering*. Nottingham, 28 June–1 July 2004.

[46] Dodson AH, Meng X, Roberts GW. Adaptive FIR filtering for multipath mitigation and its application for large structural deflection monitoring. *Proceedings of International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS 2001)*. Banff, 5–8 June 2001.

[47] Meng X, Dodson A, Roberts GW, Cosser E, Barnes J, Rizos C. Impact of GPS satellite geometry on structural deformation monitoring: analytical and empirical studies. *Journal of Geodesy* 2004 **77**:809–822.

[48] Meng X, Roberts GW, Dodson AH, Cosser E. The use of pseudolites to augment GPS data for bridge deflection measurements. *Proceedings ION GPS 2002*. Portland, ME, September 2002.

[49] Meng X, Roberts GW, Dodson AH, Meo M. GNSS for structural deflection monitoring: implementation and data analysis. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford University, Stanford, CA, September 2005.

[50] Meng X, Noakes C, Dodson AH, Roberts GW. Satellite geometry and its implications for structural deformation monitoring. *Proceedings ION GPS/GNSS 2003*. Portland, ME, September 2003.

[51] Roberts GW, Brown C, Meng X. Bridge deflection monitoring—tracking millimeters across the firth of forth. *GPS World* 2006 **11**(2):26–31.

[52] Roberts GW, Meng X, Dodson AH. Using adaptive filtering to detect multipath and cycle slips in GPS/accelerometer bridge deflection monitoring data. *Proceedings FIG XXII International Congress*. Washington, DC, April 2002.

[53] Roberts GW, Meng X, Dodson AH. Integrating a global positioning system and accelerometers to monitor the deflection of bridges. *Journal of Surveying Engineering ASCE* 2004 **130**(2):65–72.

[54] Genrich JF, Bock Y. Instantaneous geodetic positioning with 10–50 Hz GPS measurements: noise characteristics and implications for monitoring networks. *Journal of Geophysical Research* 2006 **111**(B03403):1–12.

[55] Nickitopoulo A, Protopsalti K, Stiros S. Monitoring dynamic and quasi-static deformations of large flexible engineering structures with GPS: accuracy, limitations and promises. *Engineering Structures* 2006 **28**:1471–1482.

[56] Meo M, Zumpano G. On the optimal sensor placement techniques for a bridge structure. *Engineering Structures* 2005 **27**(10):1488–1497.

[57] Xu L, Guo JJ, Jiang JJ. Time-frequency analysis of a suspension bridge based on GPS. *Journal of Sound and Vibration* 2002 **254**(1):105–116.

[58] Meo M, Zumpano G, Meng X, Roberts GW, Cosser E, Dodson AH. Identification of Nottingham Wilford Bridge modal parameters using wavelet transforms. In *Proceedings of SPIE: Smart Structures*

and Materials 2004. Modelling, Signal Processing, and Control, Smith RC (ed). SPIE, 2004.

[59] Chen W. *Integration of GPS and INS for Precise Surveying Applications*, Ph. D., The University of Newcastle Upon Tyne, 1992.

[60] Chan WS, Xu YL, Ding XL, Dai WJ. An integrated GPS-accelerometer data processing technique for structural deformation monitoring. *Journal of Geodesy* 2006 **80**:705–719.

[61] Santerre R. Impact of GPS satellite sky distribution. *Manuscripta Geodaetica* 1991 **16**:28–53.

[62] Barnes J, Rizos C, Kanli M, Small D, Voight G, Gambale N, Lamance J. Structural deformation monitoring using locata. *Proceedings 1st FIG International Symposium on Engineering Surveys for Construction Works and Structural Engineering*. Nottingham, 28 June–1 July 2004.

[63] Barnes J, Rizos C, Lee HK, Roberts GW, Meng X, Cosser E, Dodson AH. The integration of GPS and pseudolites for bridge monitoring. In *A Window on the Future of Geodesy: Proceedings of the International Association of Geodesy. IAG General Assembly*, Sanso F (ed). Springer-Verlag: Sapporo, June 30–July 11, 2003.

[64] Dodson AH, Meng X, Roberts GW, Cosser E, Barnes J, Rizos C. Integrated approach of GPS and pseudolites for bridge deformation monitoring. *Proceedings of ENC GNSS 2003*. Graz, April 2003.

Chapter 58
Eddy-current *in situ* Sensors for SHM

Neil Goldfine, Vladimir Zilberstein, Darrell Schlicker and Dave Grundy
JENTEK Sensors, Inc., Waltham, MA, USA

1 INTRODUCTION

This article complements **Eddy-current Methods**. The focus is on surface-mounted and embedded eddy-current sensors, arrays, and sensor networks for structural health monitoring (SHM). The reader is expected to have some knowledge of eddy-current methods.

Generally, eddy-current sensors include both scanning sensor arrays and sensors mounted at selected critical locations. This article describes primarily capabilities of surface-mounted and embedded eddy-current sensors for detection and monitoring of damage states as well as for monitoring of usage state variables such as stresses and temperatures.

Eddy-current SHM sensors described here are either permanently or temporarily placed in close proximity of, directly on, or within a structure at selected locations to provide enhanced observability of damage and usage state variables—*specifically to support life management decisions*. Arrays, as defined here, are local sensor constructs with a common drive (primary winding) and multiple one-dimensional or two-dimensional sensing elements that form an array in a plane positioned parallel to the surface of a component. One such sensor is the JENTEK MWM®-Array (MWM stands for meandering winding magnetometer) [1–10]. Networks of sensors and sensor arrays can be distributed throughout a structure to provide coverage of critical locations. Either (i) portable data-acquisition and analysis units, (ii) central onboard electronics, or (iii) distributed electronics modules can be used

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

to acquire data from eddy-current SHM sensor networks.

When used for damage detection and monitoring, embedded or surface-mounted eddy-current sensors perform a function similar to the function of eddy-current sensors for nondestructive testing (NDT) applications described in **Eddy-current Methods**. The advantage of permanently mounting these eddy-current sensors and arrays is that they can be used to inspect difficult-to-access locations without requiring disassembly or surface preparation. The principal disadvantages of such permanently mounted sensors for damage detection and monitoring are that (i) once in place, they cannot be scanned to sample areas that have no damage for relative comparisons with defect indications; (ii) they are limited in surface imaging resolution by the size of the individual sensing elements in an array (note that imaging resolution should not be confused with the crack detection capability since detectable cracks may be either bigger or smaller than the size of the sensing element depending on the signal-to-noise ratio); (iii) they cannot be removed for calibration in air or on standards at the time of the inspection; and (iv) in general, they can only detect damage directly under the sensor (or sensing elements of an array)[a].

Historically, SHM successes have been limited to usage and diagnostic state monitoring. There has been wide use of temperature, pressure, strain, and vibration sensing onboard aircraft and on other high value assets. This includes SHM implementations in aviation, automotive, energy, and many other market sectors. For example, vibration sensing, using inductive sensors, is common. These more traditional SHM uses are not addressed in this article. However, some of these more traditional uses can gain from the sensor technology described here. One example is weight and balance monitoring for aircraft that is limited by the ability of strain gauges or other load monitoring methods to provide reliable measurement of loads over long service periods. Attempts to implement reliable systems have generally not been successful. This article describes the approaches that can address the need for improved stress/load monitoring of both static and dynamic stresses.

To enable *individual component life management* beyond the push over the last few decades for individual aircraft tracking, two new, or at least recently reinvigorated, thrusts in SHM development have emerged. These include the need (i) to replace, or at least supplement NDT, which is typically performed with off-board sensors/probes, by direct onboard damage and condition monitoring, both continuous and intermittent, in difficult-to-access critical locations and (ii) to provide direct monitoring of local stress/strain, temperature, and environmental conditions at the critical locations that are most likely to accumulate damage or experience changes associated with relevant events, e.g., overloads and foreign object damage (FOD) for metals and composites.

To enable life extension and predictable performance for subsystems and components of high-value assets, such as aircraft, rotorcraft, ships, pipelines, bridges, etc., the often-stated goal of replacing or at least reducing the use of conventional NDT by introducing onboard SHM for *direct damage and material condition monitoring* comes from (i) the need to mitigate life cycle cost escalation while improving readiness for aging aircraft and (ii) the need, on new platforms and in platform upgrades, to improve mission readiness/capability rates, reduce field and depot logistics footprints, to enable more aggressive component/platform designs using less conservative safety margins and using advanced materials (composites, functional coatings, multi-material systems), to reduce overall aircraft weight by enabling application of damage tolerance methods by making difficult-to-access areas inspectable, to avoid collateral damage associated with disassembly for inspection, better manage mission loads, to alert to events such as excessive FOD and consequences of inadequate maintenance actions, and to remove the human from the field maintenance loop.

Thus, *the expanded use of onboard sensors for SHM is inevitable*, particularly when advanced surface-mounted and embedded sensors have proven, for some applications, to be far less invasive, with lower total implementation costs, than conventional off-board NDT methods in both logistics support requirements and practical impact on asset readiness/availability.

This article addresses the need for enhanced monitoring of (i) damage states, (ii) usage states, (iii) diagnostic states, including (iv) detection of upset events.

Damage states are scalar, vector, or multidimensional quantities that (i) can be used in models to

predict damage behavior progression, e.g., degree of relaxation of compressive residual stresses intentionally introduced to mitigate damage such as fatigue or stress corrosion cracking or (ii) provide a measure of one or more damage features (e.g., crack size, material loss from corrosion, or composite disbond/delamination area).

Usage state variables of interest include stresses and temperatures that directly affect cumulative exposure to mechanical and/or thermal conditions at critical locations on individual components, for which damage evolution needs to be monitored. Enhanced stress, temperature, and environmental monitoring are continually identified as necessary to achieve the promise of SHM for aircraft and rotorcraft.

Diagnostic states of interest include geometric misalignment, vibration levels, thermal effects, and overload events. The focus of new developments in this area of SHM development is to enable onboard or temporarily installed monitoring capability for such diagnosis at reduced cost and with limited logistics support requirements.

Upset *event detection* through monitoring of *event-dependent states* of a material (e.g., impact or overheating in metallic or composite components) is often identified as a separate requirement that must be addressed. Event detection is particularly important to components managed on a safe-life basis (as practiced by the US Navy).

Of course, there is an inevitable overlap and interdependence in the above definitions, for example, between some diagnostic states and upset events. Upset event identification is certainly a part of diagnostics. Also, detected upset events should be accounted for in the assessment of usage.

This article specifically discusses eddy-current sensor capabilities demonstrated over the last decade to provide material damage monitoring, including fatigue, corrosion, and overloads, as well as stress and temperature monitoring. The focus of this article is on the MWM and MWM-Array eddy-current sensors [11–15, 18–22], since these sensors address the above categories of interest (damage, diagnostics, and usage).

For detection and monitoring of damage or parameters of interest with the available mountable eddy-current sensors, such as MWM sensors or MWM-Arrays, portable units plug into onboard sensors for inspection or short duration monitoring at difficult-to-access locations. These portable units can support onboard NDT, diagnostic, or usage survey function.

Onboard instrumentation, when available, would provide continuous monitoring capabilities for damage, diagnostics, and usage recording. Such onboard resources could be queried in parallel, through prescribed multiplexing. Alternatively, they can be allocated using a concept named *neural plasticity* by analogy to a phenomenon related to human neurological development. Neural plasticity is a concept that enables the use of limited onboard resources that learn and grow in capability with experience by reallocating limited resources.

Onboard SHM applications of eddy-current sensors are beginning to transition to operational fleet use, having been proven extensively in the laboratory and more recently in full-scale component and full-scale aircraft tests. Extensive flight testing of eddy-current sensors for fatigue monitoring and stress monitoring with portable data-acquisition systems is expected to occur within the next few years, followed by transition to fully integrated onboard systems using eddy-current sensors. One example of a potential application would be fatigue-critical locations on landing gear components (*see* **Landing Gear**).

In **Eddy-current Methods**, the concept of mapping and tracking of fatigue damage using time-sequenced scanning with advanced eddy current methods is introduced. For example, in some critical areas in aircraft components, such as engine disk slots, where early damage detection using onboard SHM sensors is not feasible, the only means for early detection is mapping and tracking with improved scanning NDT. On the other hand, integration of onboard SHM sensor data with such mapping and tracking, when practical, would significantly enhance health monitoring capabilities.

In general, (i) onboard SHM sensors for damage detection and monitoring, (ii) traditional scanning NDT methods, (iii) advanced NDT using new mapping and tracking methods for damage evolution, and (iv) enhanced direct load monitoring capability, using onboard SHM sensors, are all needed to implement next-generation adaptive life management programs.

One advantage of using MWM-Array eddy-current sensors in support of adaptive life management is that eddy-current measurements in scanning mode

and with onboard SHM sensors are complementary (providing self-consistent data formats). For fatigue damage, for example, each of these methods provides a measure of the effective electrical conductivity or magnetic permeability variation, with the damage level of interest, for the material under test.

One proposed adaptive life management approach, described by Goldfine *et al.* [13] and modified here to include mounted "onboard" sensors, is outlined in the flow chart in Figure 1.

This approach relies on the availability, reliability, and reproducibility of the NDT scanning/imaging capability and surface-mounted or embedded capability for both coupons and in-service components. Until recently, there were no depot/field inspection tools that could meet this requirement. Now, not only the MWM-Array sensors can meet this need and enable the adaptive life management approach, but other advanced methods, including Ultrasonic Testing (UT), laser UT, thermography, and digital radiography, offer the potential to deliver the previously unattainable level of reproducibility needed to support mapping and tracking of relatively early damage behavior. Also, onboard MWM-Array, UT, and alternative SHM sensors are becoming available. We anticipate that many such technologies will be proven and reach commercial maturity during the next decade.

In the near term, the approach described above is envisioned as an integration of advanced scanning NDT (mapping and tracking) with periodic or scheduled data from onboard SHM sensors recorded using portable (not onboard) data-acquisition units. In a more broadly integrated approach, onboard SHM sensors might provide continuous monitoring, with onboard electronics, along with periodic NDT scanning of selected locations. The goal is to provide the required observability of damage, diagnostics, usage state variables, and significant events.

For aluminum and titanium structural components in aircraft and rotorcraft, there are practical onboard SHM sensing alternatives to scanning NDT inspection methods. This article focuses on these applications for onboard SHM damage monitoring. Also, this article describes the value of surface-mounted and embedded sensors for laboratory fatigue and corrosion testing. Furthermore, new magnetic (eddy current) stress gauge capabilities are described for direct load monitoring.

2 EXAMPLES OF ADVANCED EDDY-CURRENT SHM IMPLEMENTATIONS

In the following five main sections, examples of detection/monitoring of damage and measurements related to diagnostics as well as to usage state variables are described within the following categories: (i) fatigue, (ii) corrosion, (iii) stress, (iv) temperature, (v) mechanical damage, and (vi) composite disbonds/delaminations. These sections rely on MWM-Array implementations to frame the discussion of *in situ* eddy current sensor applications, since the authors are most informed on this advanced sensor implementation. Some of these sections include a discussion of performance evaluation needs and methods for surface-mounted and embedded SHM sensors.

The focus here is on the use of permanently surface-mounted or embedded sensors. However,

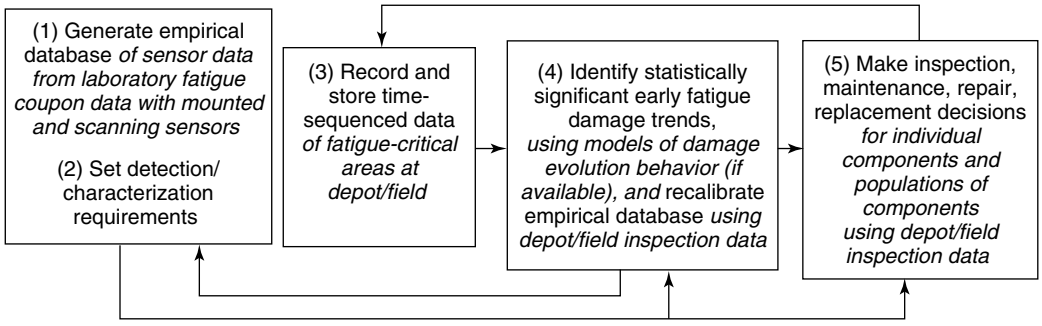


Figure 1. Flow chart of proposed adaptive life management framework.

some discussion is included on integrated methods that combine such SHM data with more conventional NDT data, including mapping and tracking using repeated NDT type imaging.

3 FATIGUE MONITORING

3.1 Fatigue case study 1: linear MWM-Arrays inside bolt holes

In **Eddy-current Methods**, an example of a mapping and tracking application for crack detection and monitoring using scanning MWM-Array bolt-hole inspection, with C-scan imaging, is provided. In Figure 8 of **Eddy-current Methods**, permanently mounted MWM-Array data is also provided for the same coupon test.

As shown in Figures 7 and 8 of the same article (*see* **Eddy-current Methods**), the combination of MWM-Array scanning and permanently mounted sensor data provides valuable insight into the crack growth behavior. This format of coupon testing, using both sensing modes, is useful for generating and calibrating damage behavior databases, as well as for generating real-crack specimens that can be used for sensor performance verification and for building probability of detection curves for both scanning and permanently mounted sensors.

3.2 Fatigue case study 2: detection of small fatigue cracks at a dimple

Also, in **Eddy-current Methods** and in [15], a method of detecting cracks at mechanical damage sites was described (*see* Figure 9, **Eddy-current Methods**). Figure 2 shows the customized titanium coupon designed for that test. In Figure 3(a) and (b), a special scanner is shown for periodically scanning the fatigue-critical area on the coupon to simulate in-service time-sequenced NDT imaging. Figure 3(c) shows a 7-channel MWM-Array designed for this purpose. Figure 2 of **Eddy-current Methods** illustrates the measurement grid approach developed by Goldfine *et al.* [5] for this purpose. This method enables real-time conversion of thousands of digital eddy-current measurements into two images: (i) conductivity image—used to detect cracks and

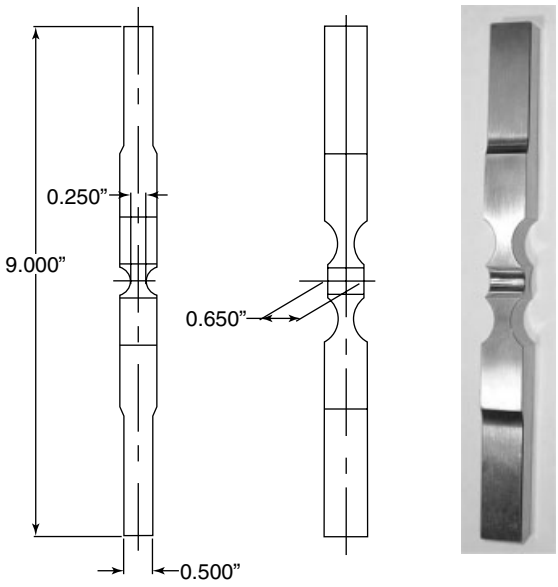


Figure 2. Schematic and photograph of a customized Ti-6Al-4V coupon with exposed fatigue-critical surfaces [8].

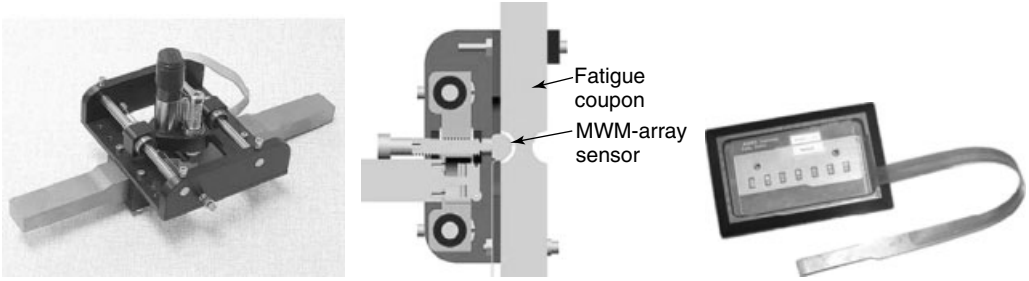


Figure 3. (a) Photograph of the coupon scanner used in support of the fatigue testing, (b) solid model representation of the coupon scanner showing the sensor position, and (c) MWM-Array FA43 sensor.

(ii) liftoff (sensor proximity to the material surface) image—used for self-diagnostics to ensure that liftoff is within an acceptable range at each of the thousands of points within an image.

Furthermore, this approach uses "air calibration," as described in ASTM Standard Practice E2338-04 [16], to improve robustness compared to conventional eddy-current testing (ET) methods that use crack (or simulated crack, e.g., EDM (electrical discharge machining) notch) standards to calibrate. This proven method is now a standard practice at US Navy and US Air Force bases, and is in use by foreign military services (FMS) as well.

This test and the data provided in Figure 9 of **Eddy-current Methods** illustrate the value of a combined approach that integrates NDT and permanently mounted sensor data. This combination could be implemented in fatigue tests and on actual structural members. For example, when localized mechanical damage or a small crack is detected using a scanning array, the damage can be either repaired or left in place, and then a permanently mounted sensor could be applied at this location to monitor initiation and/or growth of cracks at the location. One goal is to estimate the probability of failure before the next inspection for the purpose of scheduling inspection intervals and enabling life extension. The coupon fatigue test data and in-service data obtained from such combined monitoring can also be used to calibrate databases of damage evolution behavior.

Eventually, this should also enable reliable life extension through restoration of parts with localized damage using laser shock processing, low plasticity burnishing, laser additive manufacturing, or other repair methods. It is anticipated that substantial (possibly over additional 100% of design life) life extension for many components is possible if such "health control" actions are implemented.

One key attribute is the portability of experience from one component to the next. If the entire process must be repeated for each new component, then this is not a practical approach. Thus, coupon tests and NDT databases must be sufficiently generic to adapt (i.e., recalibrate) with limited effort for the next application.

Reliable mapping and tracking with MWM-Arrays has been demonstrated for titanium alloy components, in coupon fatigue tests and on engine components. With the use of titanium alloys on both commercial and military aircraft and rotorcraft, the need to detect early fatigue damage and deliver on-condition maintenance is critical. For example, titanium alloys are typically notch sensitive. Thus, small nicks, dings, and scratches introduced during handling, assembly or in service can initiate cracks at fatigue-critical surfaces earlier than predicted by design life estimates. Blending of such surface defects is sometimes allowed followed by, for example, re-shot peening to enhance remaining life. However, many parts are not re-shot peened because of quality control issues; thus, their fatigue resistance is not fully restored. In other cases, components are simply replaced, if surface defect depths exceed a prescribed limit (often just a few thousandths of an inch). This is extremely costly. Thus, methods are needed to provide early detection of damage for critical titanium alloy components, as well as to image and size surface defects to support on-condition maintenance decisions. The combination of surface-mounted sensor monitoring and mapping and tracking has proven to be very effective for such applications. Moreover, surface-mounted sensor monitoring followed by a one-time MWM-Array scanning of the suspect areas can be sufficiently effective, as well.

3.3 Fatigue case study 3: embedded MWM-Arrays for crack initiation and growth monitoring

In a series of fatigue tests, MWM-Arrays embedded between two aluminum alloy layers (Figure 4) monitored crack growth in both 4-hole and 10-hole lap joint tests. These MWM-Arrays retained their integrity as they were removed and reused for a number of successive tests to failure of the lap joints. The four-hole tests were performed under NAVAIR (The Naval Air Systems Command) funding, and the ten-hole tests were performed under US Air Force funding for Lockheed Martin to address a specific need for the F-16 aircraft as described previously [17]. A variety of precrack configurations were tested with larger ("primary") precracks at the center hole and smaller ("secondary") precracks at the other holes. The crack tip progression data are shown in Figure 5. The lines indicate the progression of crack tips from one hole towards another as a function

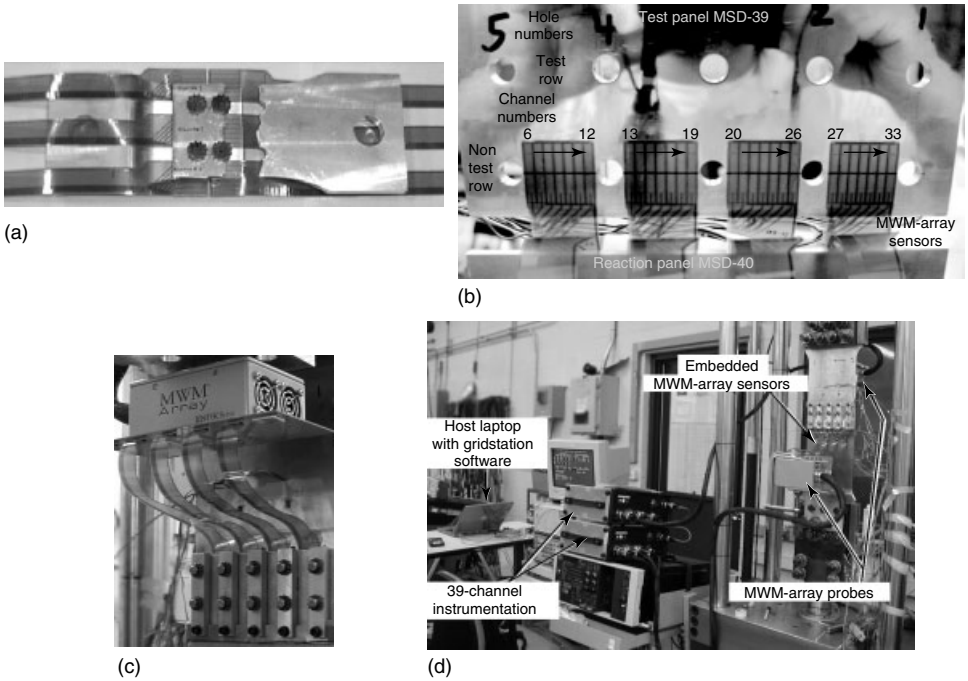


Figure 4. (a) Four-hole lap joint test specimen with embedded MWM-Arrays, (b) the 10-hole specimen with embedded MWM-Arrays shown prior to bolting up, (c) the 10-hole specimen mounted in the load frame, and (d) laboratory fatigue test setup for the 10-hole test.

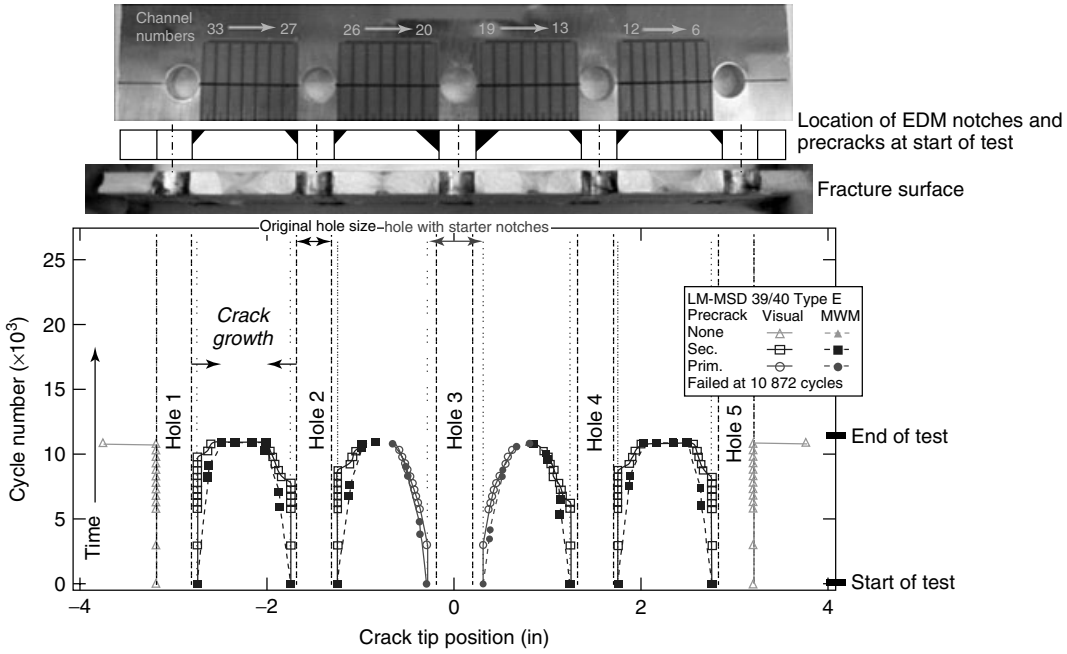


Figure 5. Representative multiple crack growth monitoring data. The specimen had primary (prim.) precracks at hole 3 and secondary (sec.) precracks at the other holes.

of number of cycles. The MWM-Arrays successfully monitored crack growth throughout the test with the sensors embedded at the buried interface between the metal plates.

3.4 Fatigue case study 4: buried cracks at bolt holes with MWM-Rosettes

Also, under a separate program funded by NAVAIR, subsurface cracks were detected and monitored by MWM-Rosettes, which were configured as smart washers and operated in a continuously monitoring mode at low frequencies (see Figure 6). This test is part of a NAVAIR funded program that is to follow by installation on two high-time aircraft for flight testing of this method.

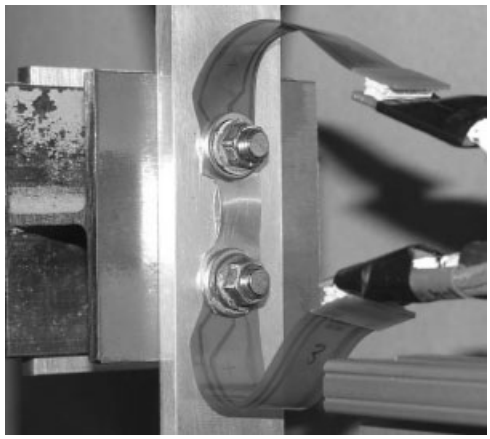


Figure 6. Photograph of the MWM-Rosette fatigue test setup.

For the test described in Figure 7, two MWM-Rosettes were fastened to an approximately 0.2-in.-thick aluminum coupon using bolts and nuts supplied by the Navy. The coupon was manually notched on the far side to promote the initiation of a fatigue crack on the side of the coupon opposite to that of

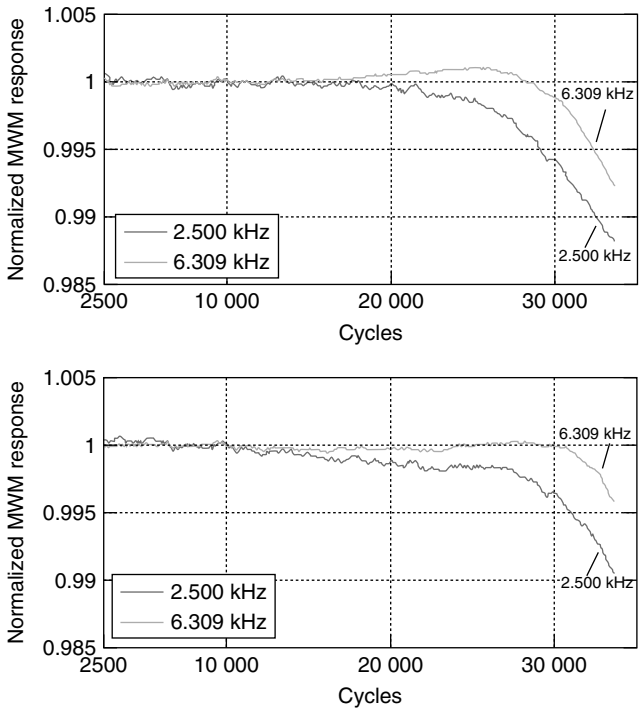


Figure 7. Normalized response of MWM-Rosette in the top hole (top plot) and bottom hole (bottom plot) acquired during a coupon fatigue test. The test data displayed is from approximately cycle 2500 to cycle 33 600. Both sensors were still functional after two fatigue tests and the accumulation of 90 000 total fatigue cycles. The "radial by axial" size of the cracks was 0.10 in. × 0.15 in. (top) and 0.11 in. × 0.14 in. (bottom).

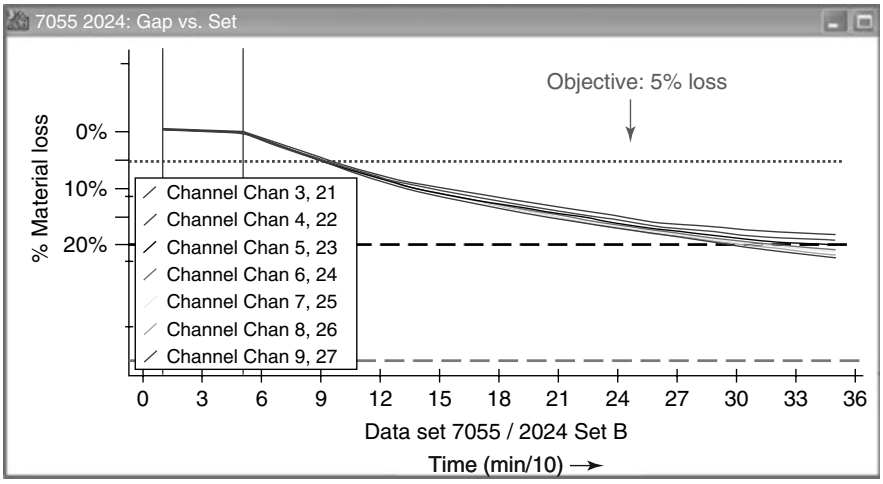


Figure 8. MWM-Array detection and monitoring of hidden corrosion during an accelerated corrosion test.

the MWM-Rosettes. Results of the fatigue test are shown in Figure 7. Cracks initiated at the notches at both holes. The crack extension was then monitored by the MWM-Rosettes. The test was stopped at 33 600 cycles, and acetate replicas of the cracked surfaces were taken. Measurements on the acetate replicas taken after the test showed that at the top hole, the crack extended radially 0.10 in. along the surface and axially 0.15 in. into the hole. At the bottom hole, the crack extended radially 0.11 in. along the surface and axially 0.14 in. into the hole. The cracks produced a reduction of 0.95 to 1.1% in the MWM signal. The MWM-Rosette was able to both detect the buried cracks and monitor their growth.

To illustrate the durability of the MWM-Rosettes operating in the smart washer configuration, a second fatigue test was performed using the same sensors. The sensors accumulated over 90 000 fatigue cycles and were still operable at the end of the test. The results of these tests provide an excellent example of the fatigue monitoring capability of MWM-Array networks.

4 CORROSION MONITORING

Hidden corrosion detection and monitoring has also been demonstrated in various tests using both mounted and scanning MWM-Array sensors [18–20]. The goal is to provide onboard detection and monitoring of hidden corrosion in difficult-to-access locations with mounted, large footprint MWM-Array sensors as well as hidden corrosion detection and monitoring in large accessible areas using wide-area scanning sensor arrays. Figure 8 shows the results of an accelerated corrosion testing program in which hidden corrosion was detected rather early and monitored with a one-dimensional MWM-Array mounted on the opposite side of the wall [19].

Figure 9 shows a surface-mountable two-dimensional MWM-Array for hidden corrosion detection and monitoring. This array is designed to monitor remaining wall thickness under its 6 in. × 6 in. footprint as metal is lost on the far side due to corrosion. Figure 10(a) shows data on a measurement grid for

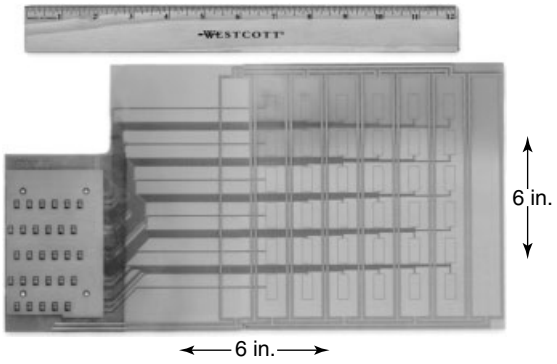


Figure 9. A two-dimensional MWM-Array for onboard monitoring of hidden corrosion.

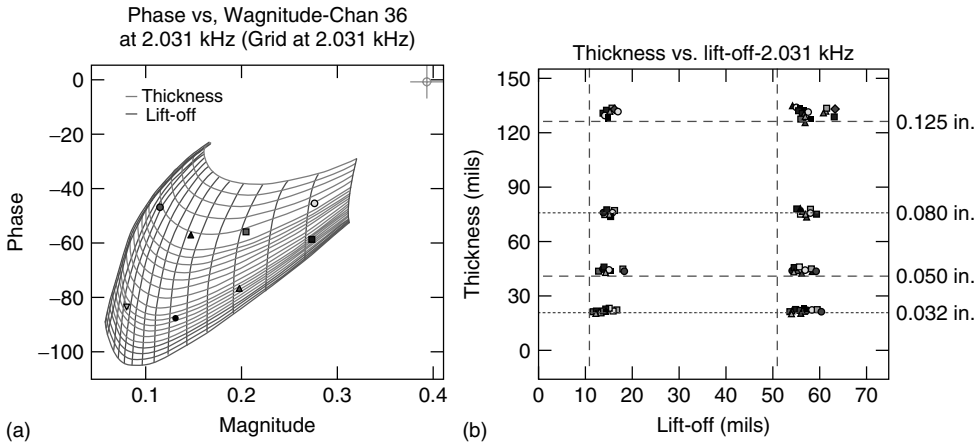


Figure 10. (a) Hidden corrosion monitoring results displayed on a metal (aluminum alloy) layer thickness—lift-off measurement grid and (b) comparison of metal layer thickness measured by MWM-Array with nominal values.

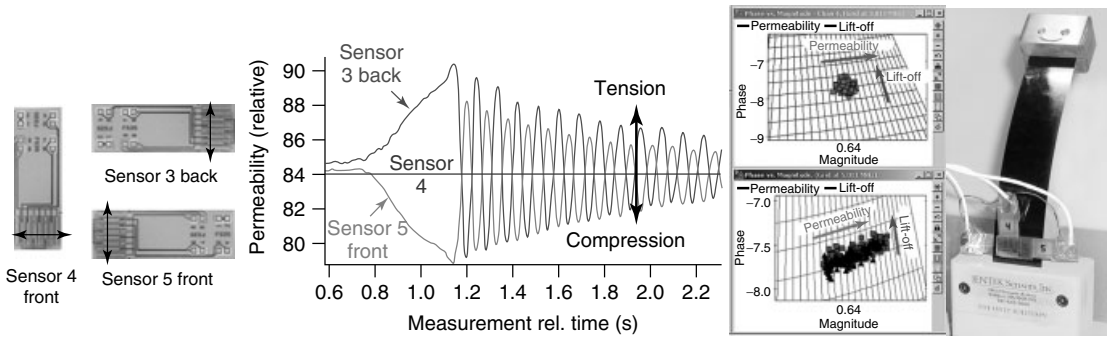


Figure 11. Demonstration of dynamic stress monitoring in a steel strip using networked MWM sensors.

remaining metal layer thickness and liftoff, demonstrating independent measurement of these two variables. The comparison of estimated layer thicknesses with "nominal" values is shown in Figure 10(b). Only a calibration in air was performed prior to making this absolute measurement, using the air calibration methods described in an ASTM standard [16]. This capability to provide onboard monitoring of hidden corrosion loss in fuel tanks, joints (metal–metal and metal–composite), and other difficult-to-access areas can offer substantial maintenance cost reductions by avoiding costly visual corrosion inspections.

5 STRESS MONITORING

Demonstrated capability of MWM sensors and MWM-Arrays to monitor stresses in steel components [8, 21, 22] permits monitoring of dynamic stresses in rotorcraft components using a network of these sensors called *magnetic stress gauges*. This capability is illustrated here for a vibrating steel strip, as shown in Figure 11. The magnetic stress gauges mounted on opposite sides of the strip detect the alternating tensile and compressive stresses and correctly indicate their phase relationship as well as gradual reduction of stresses due to damping. The third magnetic stress gauge was mounted orthogonal to the other two gauges and, consequently, was insensitive to the motion, since it responds to transverse stress that is zero in the case of plane stress, i.e., sufficiently close to the edges of the strip.

In separate overload tests on an actual steel component loaded well outside the elastic range, we have found, however, that a magnetic stress gauge mounted

orthogonal to the principal stress *can* provide an indication of an overload event, which affects permeability in both directions as a result of the changes in residual stresses caused by stress redistribution due to plastic deformation that resulted from such an overload event. Stresses can also be monitored in components fabricated from nonferrous alloys as long as selected regions are coated with a ferromagnetic coating. Dynamic stress monitoring can be performed in a noncontact mode, as has been demonstrated in laboratory tests.

6 TEMPERATURE MONITORING

As shown in Figure 12, a deep penetration MWM-Array has been used to demonstrate the capability to monitor temperature of an aluminum plate separated from the sensor by air gap and an intermediate aluminum plate [14]. This capability of either single- or multiple-frequency MWM-Arrays eddy-current sensors to monitor temperature noncontact and/or through metal layers offers a unique value for the SHM of propulsion systems, energy systems, and other systems with high internal temperatures. Because these sensors require only a conducting path with remote instrumentation, they can also be used to operate in harsh environments or within processing facilities. Figure 13 illustrates that, if the temperature of the intermediate plate is not properly accounted for, the measurement of the buried plate temperature is incorrect. Thus, as illustrated in Figure 13, the use of a three-dimensional lattice (database) for the estimation of three unknowns (the conductivity of the two aluminum plates, and the gap between them) is essential. The relationship between temperature and conductivity is stable up to several hundred degrees. Thus, this is a practical tool for many applications.

7 COMPOSITE DISBOND/ DELAMINATION MONITORING

MWM-Arrays have also demonstrated the capability to provide monitoring of buried disbonds/ delaminations in graphite fiber composites. As shown in Figure 14, the MWM-Array was used to monitor the initiation and growth of damage in a composite–composite joint test performed at Lockheed Martin. This test performed under NAVAIR funding demonstrated the capability of the MWM-Array to monitor disbond growth through over 0.2-in.-thick composite material, during the test, using a surface-mounted MWM-Array.

Unfortunately, eddy-current sensors only detected changes to the fibers (positions, directions, fiber cracking, and bulk conductivity changes associated with density and contacts). Thus, matrix damage is not detectable with this method. Other methods such as ultrasonics and thermography offer the

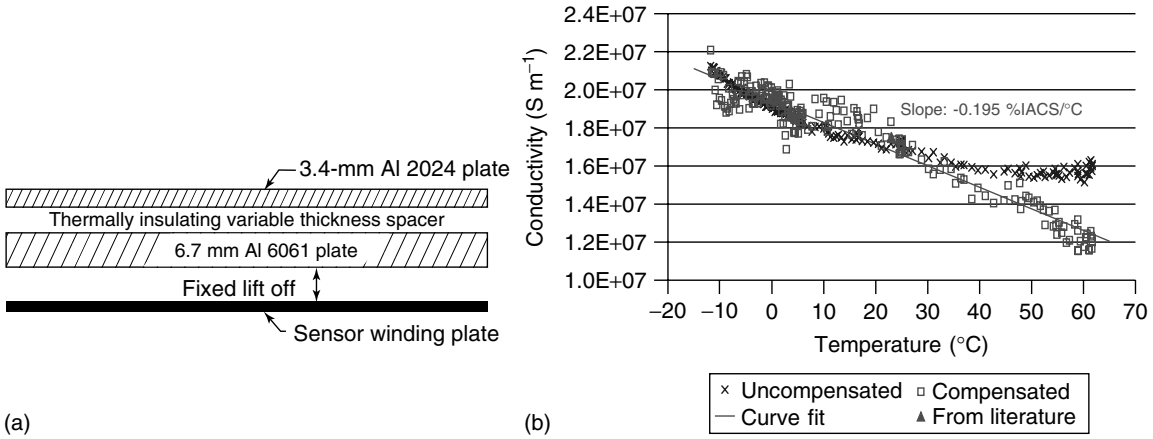


Figure 12. (a) Schematic of the demonstration setup and (b) response of MWM-Array with and without correction for the temperature variation of the intermediate aluminum plate. Note that the temperature of the two plates is measured independently using the MWM-Array at one frequency with two different-shaped magnetic field excitations [14].

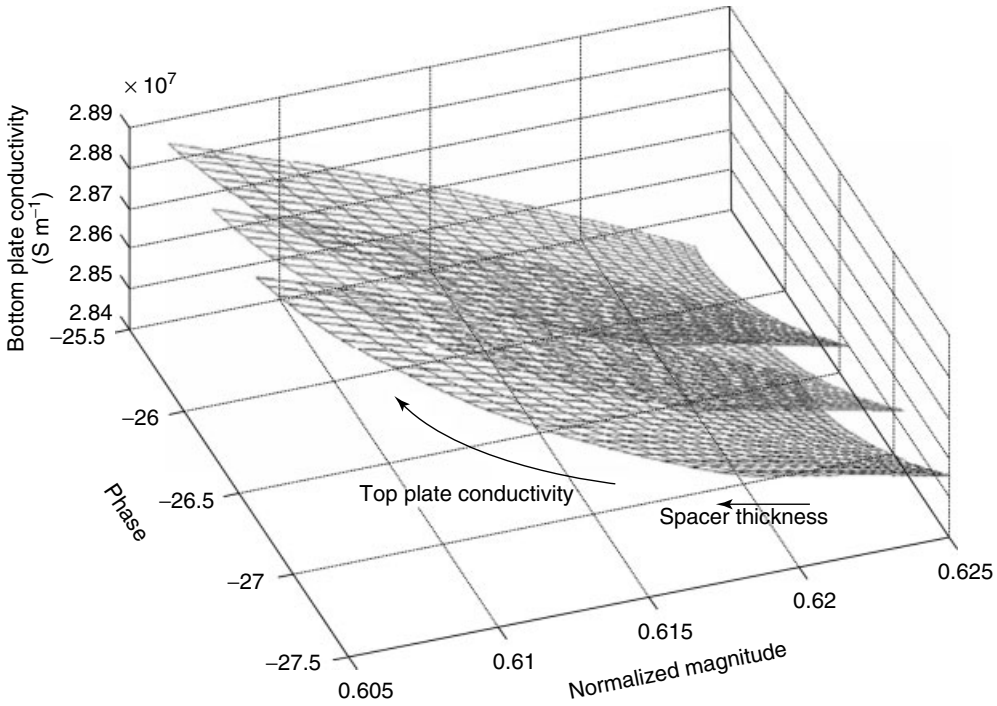


Figure 13. Lattice (precomputed database) for independent estimation of two plate temperatures (via conductivities) and the gap between them.

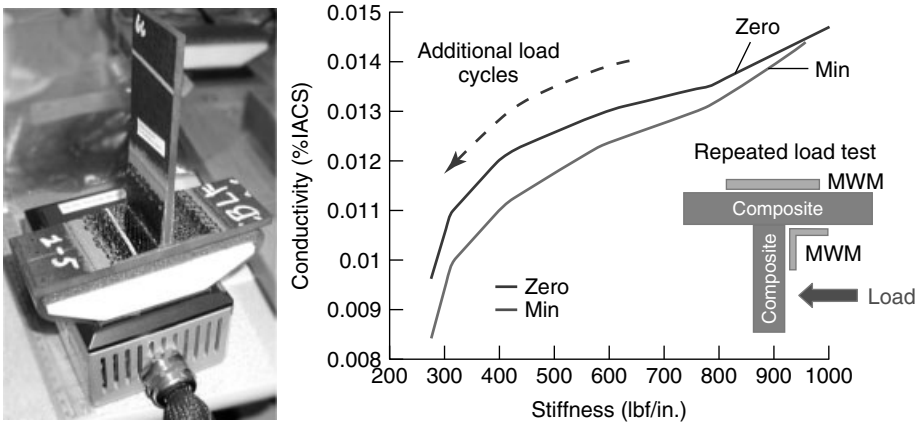


Figure 14. Composite–composite joint test performed at Lockheed Martin. MWM-Array measurements of conductivity are strongly correlated with the decreasing stiffness of the specimen as it undergoes load cycling.

capability to monitor matrix and fiber damage. One interesting method being demonstrated under ongoing NASA funding is Magneto-Thermography™. This promising method, developed by JENTEK Sensors, Inc., takes advantage of the demonstrated capability of MWM-Arrays to monitor buried fiber temperatures, to produce thermography images. This magneto-thermographic method, still in the research stage, can be implemented in a scanning or permanently mounted mode.

8 CONCLUSIONS

Eddy-current *in situ* health monitoring methods are now advancing and transitioning into use. These methods are now used in laboratory fatigue and corrosion test monitoring. In the next couple of years, the broader use of these methods are expected to accelerate owing to planned on-aircraft testing of fatigue sensor networks, using smart washer configurations of the MWM-Rosettes for buried cracks and linear MWM-Arrays for surface-breaking cracks.

Magnetic stress gauges also offer promise for near-term implementation. Full-scale testing and laboratory testing have proven the capability of this method to deliver reliable stress monitoring. This method is also expected to transition into use within the next few years.

Finally, adaptive life management, combining both installed fatigue monitoring and scanning NDT data, is expected to deliver the promise of cost savings and reduced logistics footprints for a variety of future and legacy platforms.

This is an exciting time for both offboard NDT and onboard SHM.

END NOTES

[a.] Note that for materials that exhibit significant variation of effective electrical properties, e.g., electrical conductivity or magnetic permeability with applied or residual stress measured by MWM, the potential to monitor remote damage is possible—by relating such damage occurrence to measured stress changes.

REFERENCES

[1] Goldfine N. Magnetometers for improved materials characterization in aerospace applications. *Materials Evaluation* 1993 **51**(3):396–405.

[2] Sheiretov Y. *Deep Penetration Magnetoquasistatic Sensors*, Ph.D. thesis. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology: Cambridge, MA, June 2001.

[3] Goldfine N. *Hidden Feature Characterization Using a Database of Sensor Responses*, US Patent Number 7,161,351 B2, January 2007.

[4] Goldfine N, Zilberstein V, Grundy D, Weiss V, Washabaugh A. *Fabrication of Samples Having Predetermined Material Conditions*, US Patent Number 7,106,055 B2, September 2006.

[5] Goldfine N, Schlicker D, Washabaugh A, Grundy D, Zilberstein V. *Process Control and Damage Monitoring*, US Patent Number 7,095,224, B2, August 2006.

[6] Schlicker D, Goldfine N, Washabaugh A, Walrath K, Shay I, Grundy DC, Windoloski M. *Test Circuit Having Parallel Drive Segments and a Plurality of Sense Elements*, US Patent Number 7,049,811 B2, May 2006.

[7] Shay I, Goldfine N, Washabaugh A, Schlicker D. *Magnetic Field Sensor Having a Switchable Drive Current Spatial Distribution*, US Patent Number 6,992,482 B2, January 2006.

[8] Goldfine N, Schlicker D, Washabaugh A, Zilberstein V, Tsukernik V. *Surface Mounted and Scanning Spatially Periodic Eddy-Current Sensor Arrays*, US Patent Number 6 952 095 B1, October 2005.

[9] Schlicker D, Goldfine N, Washabaugh A, Walrath K, Shay I, Grundy D, Windoloski M. *Eddy Current Sensor Arrays Having Drive Windings with Extended Portions*, US Patent Number 6,784,662 B2, August 2004.

[10] Goldfine N, Melcher J. *Apparatus and Methods for Obtaining Increased Sensitivity, Selectivity, and Dynamic Range in Property Measurements Using Magnetometers*, US Patent Number 5,629,621, May 1997.

[11] Goldfine N, Melcher J. *Magnetometer Having Periodic Winding Structure and Material Property Estimator*, US Patent Number 5,453,689, September 1995.

[12] Goldfine N. Meandering winding magnetometers: the basics. *1992 ASNT Fall Conference*. Chicago, IL, November 1992.

[13] Goldfine N, Windoloski M, Zilberstein V, Contag G, Phan N, Davis R. Mapping and tracking of damage in titanium components for adaptive life management. *10th Joint NASA/DoD/FAA Conference on Aging Aircraft*. Atlanta, GA, 16–20 April 2007.

[14] Shay I, Zilberstein V, Washabaugh A, Goldfine N. Remote temperature and stress monitoring using low frequency inductive sensing. *SPIE Conference, NDE/Health Monitoring of Aerospace Materials and Composites*. San Diego, CA, 2003.

[15] Goldfine N, *et al.* Damage and usage monitoring for vertical flight vehicles. *AHS 63rd Annual Forum and*

Technology Display. Virginia Beach, VA, 1–3 May 2007.

[16] ASTM Standard Practice E2338-04, *Characterization of Coatings Using Conformable Eddy-Current Sensors without Coating Reference Standards*. ASTM International, Book of Standards, 2004; Vol. 03.

[17] Ball D, Sigl K, McKeighan P, Veit A, Grundy D, Washabaugh A, Goldfine N. An experimental and analytical investigation of multi-site damage in mechanically fastened joints. *USAF Aircraft Structural Integrity Program (ASIP) Conference*. Memphis, TN, December 2004.

[18] Goldfine N, Grundy D, Washabaugh A, Schlicker D, Sheiretov Y, Hugeunin C, Lovett T, Roach D. Corrosion and fatigue monitoring sensor networks. *Structural Health Monitoring Workshop*. Palo Alto, CA, September 2005.

[19] Weiland H, Moran J, Bovard F, Grundy D, Zilberstein V, Lorilla I, Schlicker D, Goldfine N. Corrosion monitoring of lap joints using MWM-array sensors. *ASM AeroMat*. Seattle, WA, May 2006.

[20] Goldfine N, *et al.* Corrosion detection and prioritization using scanning and permanently mounted MWM eddy-current arrays. *Tri-Service Corrosion Conference*. San Antonio, TX, January 2002.

[21] Goldfine N, Grundy D, Washabaugh A, Craven C, Weiss V, Zilberstein V. Fatigue and stress monitoring with magnetic sensor arrays. *2006 Annual Society for Experimental Mechanics (SEM) Conference*. St. Louis, MO, June 2006.

[22] Zilberstein V, Fisher M, Grundy D, Schlicker D, Tsukernik V, Vengrinovich V, Goldfine N, Yentzer T. Residual and applied stress estimation from directional magnetic permeability measurements with MWM sensors. *ASME Journal of Pressure Vessel Technology* 2002 **124**:375–381.

Chapter 68

Miniaturized Sensors Employing Micro- and Nanotechnologies

Kenneth J. Loh and Jerome P. Lynch

Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

1 INTRODUCTION

As structural health monitoring (SHM) systems continue to be deployed to monitor and assess engineered structures, there has been growing interest in the miniaturization of sensing transducers used to record structural behavior. The advantages of miniaturization are multiple; for example, reduction in sensor size and weight is critically important when considering the use of SHM systems in lightweight structures such as aerospace structural systems (e.g., rockets and satellites). Smaller sensors are also easier to install, particularly if the sensor is embedded within structural elements; an example might be thin-film sensors installed within a layered composite material (e.g., carbon fiber-reinforced polymer (CFRP) composites). In addition, miniaturization can lead to potential improvements in sensing accuracy, simultaneous to significant reductions in fabrication costs. Finally, microelectromechanical system (MEMS) sensors are more power efficient than their macroscale counterparts.

Macroscale sensor design concepts can be miniaturized by adopting emerging technologies associated with the MEMSs and nanotechnology fields. Already, MEMS has had a major impact with SHM systems widely using MEMS sensors [1]. MEMS is defined by the use of fabrication methods associated with integrated circuits (ICs) to construct mechanical structures within semiconductor substrates such as silicon (Si), germanium (Ge), and gallium arsenide (GaAs) [2]. By miniaturizing macroscale sensor transduction concepts to the micron-dimensional scale, the approach is commonly termed a *top-down* methodology. Since the initial introduction of MEMS pressure sensors in the 1960s [3], the field of MEMS has rapidly evolved over the past four decades. Today, a plethora of other miniaturized sensing transducers have been proposed by the MEMS community including accelerometers, gyroscopes, gas sensors, and ultrasonic transducers among many others [4]. These examples of MEMS sensors have been shown to offer measurement accuracies on par

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

with macroscale counterparts. MEMS adoption of IC technologies for device fabrication allows computing and wireless communication circuits to be collocated with the MEMS sensor, thereby offering complete system-on-a-chip (SoC) solutions [2]. Furthermore, IC-based manufacturing offers fabrication of MEMS sensors by a batch process, with hundreds of devices fabricated on a single semiconducting wafer [4]. While MEMS sensors have shown tremendous promise, their market adoption has been greatest in automotive (e.g., pressure and acceleration sensing) and inertial sensing markets where high sales volume is able to amortize high fabrication costs [2, 5].

In light of the limitations of MEMS, the interdisciplinary nanotechnology field has emerged to offer chemical tools and processes that permit further miniaturization of sensors designed for SHM applications [6]. Specifically, it is now possible to design materials with specific macroscopic mechanical, electrical, and chemical properties by controlling structure and assembly at the atomistic length-scale, which is nanometers in dimension [7]. This "bottom-up" approach is in stark contrast to the "top-down" design methodology currently adopted by MEMS. Currently, molecular structures such as nanotubes [8], nanoparticles [9], and self-assembled materials [10] are finding their way into the design of sensors [11]. Clearly, future advances in miniaturized sensors will be derived from the technological developments within the nanotechnology domain.

The balance of this review is delineated into two major sections. In the first section, an overview of MEMS is provided. The general methods of MEMS fabrication are described, including bulk and surface micromachining methods. In addition, a plethora of MEMS sensors that have found use in SHM applications are presented. To illustrate the impact MEMS has had on the SHM field, applications of MEMS sensors to SHM problems investigated in the laboratory and field settings are also described (*see* **Microelectromechanical Systems (MEMS)**). In the second section, the tools and processes of the nanotechnology field relevant to the miniaturization of sensors are presented (*see* **Nanoengineering of Sensory Materials**). In essence, nanotechnology allows engineers to address many of the technological limitations currently encountered in MEMS design. A novel suite of sensors assembled at the nanoscale is highlighted to reinforce the future promise of the "bottom-up" sensor design approach for SHM.

2 MICROELECTROMECHANICAL SYSTEM SENSORS

2.1 Fabrication of MEMS structures

There are two major fabrication methods employed in MEMS design to construct three-dimensional structures, namely, bulk and surface micromachining. In the first method, bulk micromachining, the structure is built into the substrate through the removal of substrate material. In contrast, surface micromachining builds the MEMS structure on top of a substrate by depositing materials to the substrate surface. How a MEMS sensor is fabricated is very important to consider at the outset of the design process since each fabrication method uniquely constrains the final geometric shape of the MEMS device. In this section, the general steps of both fabrication approaches are delineated; methods presented are primarily for silicon, which is the preferred substrate material of most MEMS foundries. The interested reader is referred to Kovacs [2] for a more complete presentation of each fabrication approach and for information on how nonsilicon substrates such as Ge and GaAs can be integrated with the MEMS fabrication process.

2.1.1 Bulk micromachining

In bulk micromachining, three-dimensional MEMS structures are fabricated within the substrate by etching. Etching is the process by which substrate material is selectively removed to create structures within the substrate. All etching methods can be classified as part of two major etching categories: wet and dry etching.

In wet etching, liquid chemical compounds react with the silicon substrate so as to remove substrate material. In general, wet etchants are either isotropic or anisotropic depending upon their relationship with the silicon crystalline structure. For example, isotropic etchants remove silicon in all crystal directions at the same rate; the result is structures with rounded features. A common isotropic etchant is HNA, which is a solution of hydrofluoric, nitric, and

acetic acids (HF, HNO_3, and CH_3COOH, respectively). In contrast, anisotropic etchants possess different reaction times depending upon the silicon crystal orientation. Using Miller index notation to denote the various planes of crystalline silicon, it is the (111) plane of silicon that is the slowest to react with anisotropic etchants. Owing to the slow etch time of the (111) plane, it is preferentially exposed during etching. In other words, the (111) plane that is exactly 54.7° relative to the (100) plane is exposed in the final structure after anisotropic etching. Alkali hydroxides (e.g., potassium hydroxide (KOH)) are typically used for anisotropic wet etching of silicon.

In dry etching, noble gas fluorides (e.g., xenon fluoride (XeF_2)) are introduced in a chamber containing a silicon substrate to induce isotropic substrate etching. Dry etching using noble gas fluorides can be tightly controlled through the temperature and pressure of the chamber. An additional benefit of dry etching is that traditional masking can be used to selectively etch the substrate. However, one drawback is the final etched structure is defined by high surface roughness. Another approach to dry etching is reactive ion etching (RIE). An RIE chamber consists of parallel-plate electrodes to which an alternating current (AC) is applied, resulting in an oscillatory electromagnetic (EM) field. As fluorinated gases enter the sealed chamber, the EM field strips electrons and produces highly energized ions. When a silicon wafer is placed upon one of the electrodes, the accelerating ions bombard the substrate surface with high energy, resulting in the removal of substrate material. The ion flux is generally perpendicular to the substrate, resulting in bulk micromachined structures defined by vertical side walls. For silicon, the etch gas used in RIE is generally sulfur hexafluoride (SF_6). An extension of RIE in which polymer is deposited along the sidewalls of the etched structure is termed *deep reactive ion etching* (DRIE). DRIE has gained considerable popularity in recent years owing to the high-aspect ratio structures attainable; smooth, polymer-coated vertical wall structures with aspect ratios as high as 30 : 1 have been achieved [2].

Both wet and dry etchings are assisted using masking and sacrificial layers. For example, silicon dioxide (SiO_2) can be thermally grown upon the surface of a silicon wafer to serve as a masking layer during wet etching. Also, photosensitive resist layers (photoresists) can be deposited upon the surface of a silicon wafer as a masking layer. Using optical lithography, portions of the resist can be exposed to ultraviolet (UV) light; after exposure, the exposed resist is dissolvable in a solvent such as acetone.

2.1.2 Surface micromachining

When employing surface micromachining for fabrication of a MEMS structure, the structure is built upon the substrate using structural thin films, including polycrystalline silicon, aluminum, and silicon nitride (Si_3N_4). Integral to surface micromachining is the incorporation of sacrificial layers of silicon dioxide and photoresists. In the fabrication of MEMS structures, both structural and sacrificial layers can be etched to selectively remove portions of each layer. Typically, wet etching of silicon dioxide is done using HF, while polycrystalline silicon is wet-etched using traditional silicon etchants. Dry etching can also be used in a surface micromachining process. In surface micromachining, MEMS structures are generally defined by large in-plane areas and small out-of-plane thicknesses; in essence, two-and-a-half dimensional (2.5D) structures are formed.

2.1.3 Fabrication of a cantilever

To illustrate the fabrication methods presented above, a simple MEMS structure widely used in the design of MEMS accelerometers is adopted. The steps included in the fabrication of a cantilever beam are detailed for both bulk and surface micromachining processes [12]. The bulk micromachining fabrication process begins with a silicon wafer upon which a layer of silicon dioxide is thermally grown. In addition, a thin layer of photoresist is deposited over the silicon dioxide. Using a photomask containing a pattern of the cantilever, lithography is used to selectively expose the resist layer (Figure 1a). Placement of the wafer in a developer solvent removes the exposed resist (Figure 1b) (in this case, a positive photoresist is employed, where areas exposed to UV light are removed by the solvent). The exposed layer of silicon dioxide can be selectively removed where there is no resist using hydrofluoric acid (Figure 1c). In the last step, the wafer is placed in a wet etchant bath (e.g., KOH) to perform anisotropic etching (Figure 1d). If a [100] wafer is used, a deep trench beneath the cantilever will result with the (111) plane revealed.

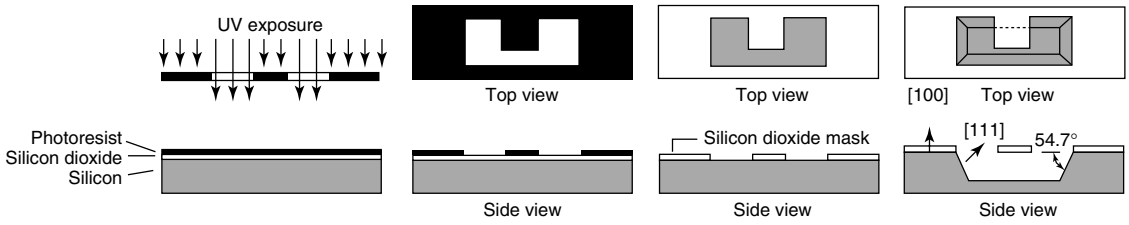


Figure 1. Fabrication of a cantilever by bulk micromachining: (a) silicon wafer with a silicon dioxide layer buried under photoresist is exposed to UV light; (b) resist is dissolved to reveal the underlying silicon dioxide layer; (c) silicon dioxide is etched to reveal the underlying silicon; (d) anisotropic wet etching of silicon results in undercutting of a cantilever element.

The surface micromachining process begins with a silicon wafer upon which silicon dioxide is thermally grown; in this process, silicon dioxide will serve as a sacrificial layer. Positive photoresist is again deposited upon the silicon dioxide and exposed to UV light so as to expose the underlying layer of silicon dioxide (Figure 2a). The silicon dioxide is etched using a wet etchant (such as HF) to reveal the underlying silicon substrate. The remaining silicon dioxide will serve as a sacrificial layer that will be removed during subsequent steps (Figure 2b). A structural layer of polycrystalline silicon is patterned and deposited upon both the silicon and silicon dioxide layers (Figure 2c). To release the cantilever, the remaining silicon dioxide beneath the cantilever is etched (Figure 2d).

2.2 Accelerometers

Accelerometers represent one of the most successful examples of a MEMS sensor in the commercial market. MEMS accelerometers have been used to measure acceleration in automotive (e.g., airbag deployment) [13], aerospace (e.g., inertial sensing) [14], and medical (e.g., pacemaker regulation) [15] applications. The design of all MEMS accelerometers consists of four major functional components within a single sensor package: proof mass, spring, damper, and readout mechanism (Figure 3a). The proof mass is a passive element that undergoes displacement due to inertial forces. The spring restrains the mass motion through a restoring force. Damping is necessary to ensure that the ratio of mass displacement to acceleration amplitude is constant up to the resonant frequency (f_R) of the accelerometer's mass–spring system. Since the displacement of the proof mass will be linearly proportional to the acceleration of the sensor package, an electromechanical transducer is needed to convert the proof mass displacement to an electrical signal. MEMS accelerometers use either capacitive, piezoresistive, or piezoelectric electromechanical transducers as readout mechanisms.

The very first MEMS accelerometer fabricated is the one proposed by Roylance and Angell [16]. The accelerometer is bulk micromachined in silicon using anisotropic wet etching. The device consists of a square proof mass connected to its silicon housing through a thin cantilever element acting as a spring. A scanning electron microscope image of the accelerometer is shown in Figure 3(b); note that the

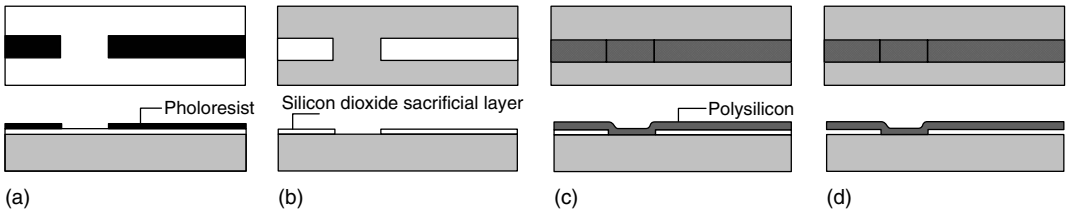


Figure 2. Fabrication of a cantilever by surface micromachining: (a) exposed resist (after lithographic exposure) is dissolved to reveal the underlying silicon dioxide layer; (b) silicon dioxide is etched to reveal the underlying silicon; (c) polycrystalline silicon is deposited over the silicon dioxide sacrificial layer; (d) selective etching of the silicon dioxide is done to release the polycrystalline silicon cantilever.

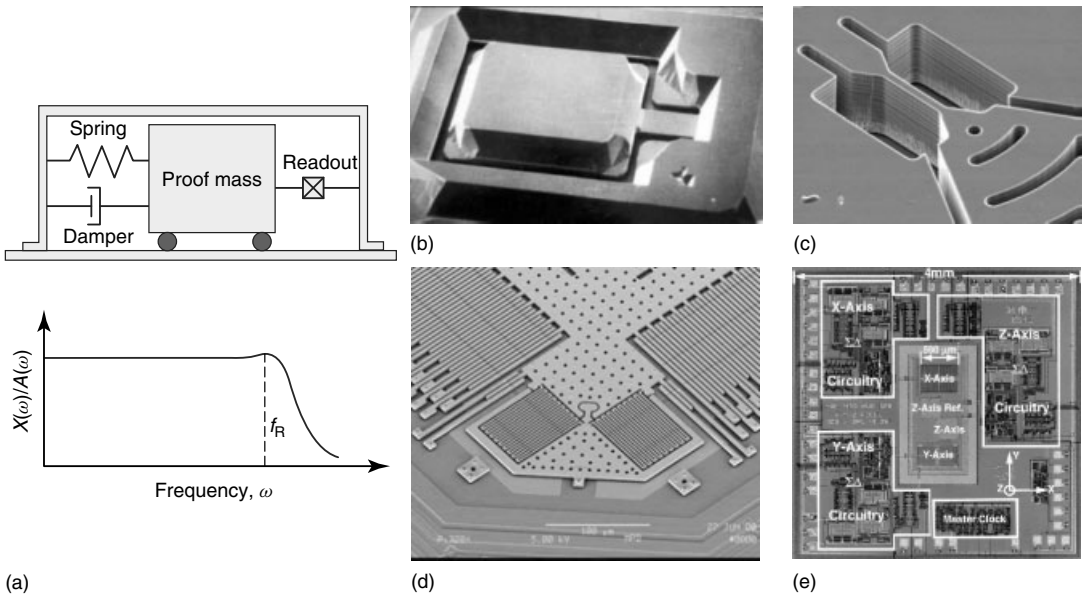


Figure 3. (a) Conceptual components of a MEMS accelerometer including a typical mass displacement–acceleration transfer function $(X(\omega)/A(\omega))$; (b) piezoresistive accelerometer by Roylance and Angell [Reproduced with permission from Ref. 16. © IEEE, 1979]; (c) high-performance planar piezoresistive accelerometer by Partridge *et al.* [Reproduced with permission from Ref. 17. © IEEE, 2000]; (d) close-up microscopic image of the Analog Devices *i*MEMS capacitive accelerometer with differential capacitors and folded springs clearly shown [Reproduced with permission from Ref. 18. © Anolog Devices, Inc., 1999]; (e) three-axis force-balanced accelerometer die with accompanying circuitry by Lemkin *et al.*. [Reproduced with permission from Ref. 19. © IEEE, 1997.]

side walls are at a 54.7° angle with the top surface of the substrate. Air within the sealed device (the silicon structure is sandwiched between two glass plates) provides viscous damping. A piezoresistor is implanted upon the top face of the cantilever to serve as the primary readout mechanism of the sensor; as the proof mass deflects, the cantilever beam experiences strain, which in turn results in a proportional change in resistance of the implanted piezoresistor. The accelerometer is reported to have a low noise floor ($0.001g$), large range ($\pm 200g$), and high bandwidth (2 kHz).

Since the seminal work by Roylance and Angell [16], a number of researchers have proposed piezoresistive MEMS accelerometers. Most recently, Partridge *et al.* [17] have prototyped a high-performance planar accelerometer for shock applications. Their accelerometer uses DRIE to define a pie-slice-shaped proof mass connected to the silicon substrate through a thin cantilever element (Figure 3c). Boron implantation in the cantilever results in a piezoresistive element for the readout of the sensor acceleration measurements. An attractive feature of the accelerometer is that the proof mass displacement is in the plane of the substrate. This restrains undesirable displacements of the proof mass during high-acceleration (shock) motions, thereby protecting the accelerometer from failure. The full dynamic range of the accelerometer is well above $10g$ with a resolution of $20\,\mu g$ at an acceleration bandwidth of 650 Hz.

While piezoresistive accelerometers are easy to fabricate and offer simple electrical interfaces, they exhibit extreme sensitivity to temperature that hampers their use in applications where environmental temperatures vary [20]. As a result, piezoresistive accelerometers have given way to MEMS accelerometers using capacitive readout mechanisms. Capacitive accelerometers have been successfully implemented in the commercial sector by Analog Devices (Norwood, MA). The Analog Devices *i*MEMS accelerometer family adopts a surface micromachining process to fabricate one- and two-axis accelerometers in polycrystalline silicon [18]. The accelerometer's square proof mass is

fabricated in a polycrystalline silicon layer that is attached to the silicon substrate through folded springs (Figure 3d). The proof mass is released after a sacrificial layer of silicon dioxide is removed; "through holes" are drilled in the proof mass to ensure that capillary forces during wet release do not destroy the device. The electromechanical transduction mechanism consists of differential capacitors integrated along the sides of the square proof mass. Various version of the *i*MEMS accelerometer exist with different dynamic ranges (e.g., from ± 1.2 to $250g$). Depending upon the user's application, the bandwidth and noise floor of the sensor can be varied using discrete circuit elements installed along with the sensor. An attractive feature of the *i*MEMS accelerometer is its low cost (approximately \$10 per sensor). Other commercial capacitive accelerometers include those from Silicon Designs (SD Series), Endevco (7290 Series), and Crossbow (CXL-LF Series). A force-balanced MEMS accelerometer based on a capacitive design has also been proposed by Lemkin *et al.* [19] (Figure 3e).

MEMS accelerometers have been applied to SHM applications over the past decade. In the civil engineering sector, MEMS accelerometers are low-cost alternatives to traditional seismic accelerometers (e.g., piezoelectric and force-balanced accelerometers). For example, a dense array of Crossbow CXL-LF accelerometers has been instrumented on the Geumdang Bridge (Icheon, Korea) to measure the bridge response to traffic [21]. The Analog Devices ADXL202 and Silicon Design SD2210 MEMS accelerometers have also been deployed on a pedestrian bridge on the University of California, Irvine campus [22]. The high-performance planar accelerometer proposed by Partridge *et al.* [17] has been validated to measure the acceleration of the Alamosa Canyon Bridge (southern New Mexico) [23]. It is interesting to note that in all of these cases, the MEMS accelerometers have been interfaced to battery-powered wireless sensors. This is due in part to the low-power requirements of MEMS. These applications serve as powerful examples of the success MEMS accelerometers have had in SHM.

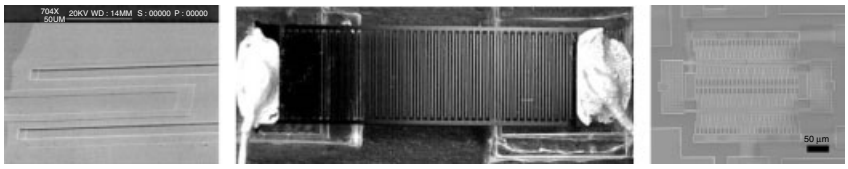
2.3 Strain sensors

Damage (e.g., cracking) is typically localized to a specific location in a structure; as a result, measurement of global structural responses (e.g., acceleration) is generally inadequate for accurate damage detection. Damage detection can be improved if local responses like strain are utilized. Toward this end, a number of researchers have proposed MEMS strain sensors defined by high-accuracy and small-form factors.

MEMS strain sensors have found early use in biological and medical applications. For example, one of the earliest MEMS strain sensors includes a sensor proposed in 1980 for measuring strain in animal tissue [2]. The strain sensor is fabricated from silicon using bulk micromachining techniques to define the sensor geometry and to implant a piezoresistive element. Strain sensors have also been included in the design of MEMS-based neural probes to measure the forces used to penetrate tissue [24]. Recently, MEMS strain sensors proposed for measuring strain in thin films [25–27] have matured with devices now suitable for measuring strain in structural systems.

A novel MEMS strain sensor is proposed by Hautamaki *et al.* [28] for embedment in CFRP composite plates. Their sensor employs a doped polycrystalline silicon element acting as a piezoresistor placed upon a silicon nitride cantilever element (Figure 4a). Upon embedment in the epoxy matrix during construction, strain in the composite panel would be transferred to the cantilever, thereby allowing strain to be accurately measured. Their sensor exhibits a 1–2% change in resistance at $1000\,\mu\varepsilon$.

While piezoresistive MEMS strain sensors are easy to install, they exhibit sensitivity to temperature, thus requiring compensatory circuitry. In contrast, capacitive strain sensors are more stable in varying thermal environments. Aebersold *et al.* [29] propose a bulk micromachined capacitive strain sensor for measuring strain in bending structural elements. On the basis of an interdigitated finger design, the capacitive sensor is fabricated in a 150-μm-thick silicon wafer using DRIE (Figure 4b). The sensor is tested on a structural element loaded in four-point bending, resulting in a repeatable nonlinear change in capacitance as a function of strain. A high-performance MEMS strain sensor based on capacitive sensing is also proposed by Ko *et al.* [31]. Designed for measuring strain in rotating machinery, the device features a resolution of $0.09\,\mu\varepsilon$ and is able to accurately measure static and dynamic strains up to 10 kHz. While silicon is



(a) (b) (c)

Figure 4. MEMS strain sensors: (a) piezoresistive strain sensor fabricated on a silicon nitride cantilever for impregnation in fiber-reinforced composite panels [Reproduced with permission from Ref. 28. © IEEE, 1999]; (b) capacitive strain sensor for measurement of bending strain [Reproduced with permission from Ref. 29. © IOP Publishing Ltd., 2006]; (c) SiC MEMS resonant strain sensor for harsh environments. [Reproduced with permission from Ref. 30. © IEEE, 2007.]

a robust structural material, it can experience wear and corrosion in harsh environments. As a result, a MEMS strain sensor fabricated from polycrystalline silicon carbide (SiC) has been proposed for SHM of high-temperature components (e.g., turbine blades) [30]. This MEMS strain sensor is designed as a comb-driven tuning fork resonator whose resonant frequency varies in linear proportion to strain (Figure 4c); a resolution of 0.11-$\mu\varepsilon$ and 10-kHz bandwidth is achieved in the laboratory.

In the commercial sector, a low-cost MEMS strain sensor has been marketed by Sarcos [32]. Their uniaxial strain transducer (UAST) is a MEMS device in which strain is correlated to signals generated between electrostatic field emitters and measured by on-chip field detectors. This device has been successfully applied to monitor the health of railroad tracks as train cars traverse the track; peak strain is measured and stored for further fatigue analysis [32].

2.4 Acoustic transducers

Acoustic emission (AE) sensing is an important technology in the SHM field. AE sensing strategies seek to capture transient stress waves generated by damage initiation in a structure in order to identify and quantify the damage. Traditionally, AE requires piezoelectric ceramic transducers that are mounted to the surface of the structure to capture stress waves; however, transducers can be both costly and bulky [33]. MEMS represents an enabling technology for the reduction of both the size and cost of AE sensors. Toward that end, a variety of ultrasonic transducers fabricated as MEMS devices have been proposed in the SHM literature. The majority of MEMS sensors proposed for AE has been surface micromachined based on capacitive mechanisms. Jones *et al.* [34] propose the design of a silicon nitride membrane deposited on a silicon wafer to produce a resonant cavity. Deposition of electrodes on the bottom of the cavity and upon the top surface of the membrane creates a capacitor whose capacitance varies with membrane deflection. Square cavity geometries are used ($1\,mm^2$), while the membrane thickness is adjustable from 1 to $2\,\mu m$ with thinner membranes exhibiting greater sensor sensitivity. The authors have conducted various laboratory tests of the AE MEMS sensor using CFRP composite panels. The tests revealed excellent AE detection to ball drops and pencil lead-breaks on the CFRP panel.

Ozevin *et al.* [33] extend on the AE MEMS sensor design proposed by Jones *et al.* [34] by fabricating seven AE detectors on a single die; each detector is tuned to a different resonant frequency spanning from 100 kHz to 1 MHz. The readout mechanism utilized in the AE sensor is a two-plate capacitor with one plate moving in response to the AE stress wave; changes in plate location result in a measurable change in device capacitance. The AE sensor is fabricated using surface micromachining of polycrystalline silicon in the commercial multiuser MEMS processes (MUMPs) foundry. The performance of the fabricated AE MEMS sensor is compared to a commercial piezoelectric transducer during laboratory testing. Both sensors are surface-mounted to a steel beam in which a weld is included in the beam center. Cyclic three-point bending of the beam results in fatigue cracks in the weld; the performance of the MEMS sensor is shown to be comparable to the commercial piezoelectric AE sensor.

2.5 Corrosion sensors

Corrosion is a serious structural deterioration confronting many engineered structures including aircrafts, bridges, and machineries among others. Over the past few years, a number of novel corrosion sensors based on MEMS technology have been proposed in the literature. Niblock *et al.* [35] describe a MEMS fabricated corrosion sensor based on the measurement of linear polarization resistance (LPR); since LPR is correlated to the rate of corrosion, the sensor provides insight into the speed of the corrosion process. The sensor proposed is fabricated from the host metal, which is to be monitored by the MEMS corrosion sensor. The metal is machined into small 25 mm squares and ground to thicknesses as small as 50 μm. The square stock is then chemically etched to form interdigitated fingers separated by a small gap (150 μm). After formation of the fingers, a thermosetting polymer is used to fill the gaps between the electrodes. When a small potential (ΔE) is externally applied across the electrodes, a current flow at the metal surface can be measured (ΔI). The slope of the $\Delta E - \Delta I$ represents the LPR; the LPR can be used to determine the corrosion current directly. Laboratory tests of the LPR sensor reveal its ability to accurately measure the corrosion current in steel specimens.

The Southwest Research Institute (SwRI) has proposed a MEMS sensor for monitoring stress corrosion cracking [36]. Their device consists of a cantilevered beam that is fabricated from the same material for monitoring corrosion. The cantilever is notched along its center so as to allow it to crack under stress and corrosion. Measurement of the beam resistance is directly related to the splitting of the beam (which is assumed to be correlated to stress corrosion cracks developing in the host structure). For additional discussion regarding application of corrosion monitoring for aircraft applications, (*see* **Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control**).

2.6 Wireless interdigital transducers

Another class of MEMS sensors more recently adopted by the SHM community is known as *surface acoustic wave* (*SAW*) *sensors*. SAW sensors are physically small and can be easily designed for wireless interrogation, thereby making them suitable for embedment within structural components. In order to generate SAWs within a piezoelectric substrate material (*see* **Piezoelectricity Principles and Materials** for a discussion on piezoelectricity principles and materials), interdigitated transducers (IDTs) are patterned on the substrate surface using traditional lithography and metallization. Basically, IDTs consist of thin-film comb-shaped electrodes (typically patterned using aluminum) mounted on top of a piezoelectric element (Figure 5). When a voltage is applied to the IDT electrodes, dynamic strain is induced in the piezoelectric substrate, which in turn generates a SAW [5]. The electrode finger spacing (d) is controlled to be half of the SAW wavelength, λ. This spacing ensures that stress waves traveling in both directions from an IDT constructively interfere. The simplest IDT–SAW sensor, known as the *nondispersive delay line*, consists of a pair of IDTs. One IDT connected to an AC source acts as the SAW generator, while the other serves as the sensing element. Depending on the crystallographic cut of the piezoelectric element and the orientation of the IDT electrodes, various types of acoustic waves (e.g., Rayleigh, Love, and shear horizontal) can be generated [37].

In general, SAW sensors correlate a physical phenomenon (e.g., strain) with a change in the

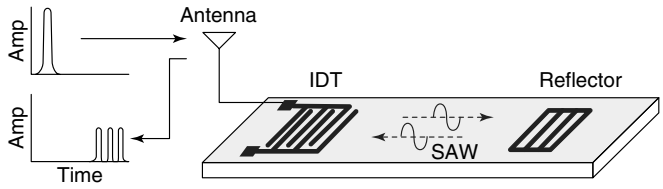


Figure 5. Surface acoustic wave (SAW) sensor with an interdigitated transducer (IDT) patterned upon the piezoelectric substrate along with a reflector element. A wireless interface receives a pulse signal. Reception of the pulse results in multiple reflections that are emitted wirelessly after a short delay.

properties of the SAW elastic wave such as velocity, v, and attenuation. Such changes can be measured by observing changes in the SAW resonant frequency, $f = v/\lambda$, as well as the phase and amplitude differences between the sensor input and output signals [37]. For example, in SAW strain sensors, strain induces a change in the separation distance between reflectors. This results in a measurable phase lag between the sensor's input–output signals, such that the measured phase lag is linearly proportional to strain [5, 38]. Many variations of SAW sensors have been proposed for gas [39], chemical [40], and acceleration [5] measurements, among others [41]. For many of the sensors proposed, antennas are integrated with the SAW sensor to accommodate input excitation and output measurements by a wireless interface [5].

3 NANOSCALE ASSEMBLY OF SENSING MATERIALS

While MEMS processing has offered a plethora of high-performance sensors for SHM, the fundamental design principle is derived from miniaturization of macroscale devices. Despite advances in lithographic patterning, etching methods, and deposition techniques, there exist technological limitations to MEMS and IC manufacturing. For example, lithographic patterning is difficult at nanometer length-scales, thereby retarding continued miniaturization of MEMS structures. Alternatively, nanotechnology offers tools and techniques that allow for the intentional assembly of novel materials at molecular length-scales (*see* **Nanoengineering of Sensory Materials**). Today, the IC and MEMS industries are investing heavily in nanotechnology so that transistors and MEMS structures can continue to be miniaturized [42].

The SHM community will also be a beneficiary of the nanotechnology field. In particular, nanotechnology offers a "bottom-up" assembly approach to the design of multifunctional materials. A multifunctional material is a material system that achieves multiple functional objectives, such as the ability to take load and the capability to sense its response to loading. Integral to the development of multifunctional materials is the use of nanometer-scaled structures (e.g., carbon nanotubes (CNTs) [43], carbon black [44], and nanoparticles [45], among others) included in composite material assemblies. While many of these molecular building blocks have been employed in the design of sensors for SHM, the use of CNTs (Figure 6a) has been very popular due to their impressive physical, mechanical, and electrical properties [43, 46–53].

3.1 Multifunctional carbon nanotube materials

Since the discovery of CNTs by Iijima in 1991 [8], researchers have sought to take advantage of their impressive electrical and mechanical properties for developing novel sensors for SHM [48]. Early realization of CNTs as ideal candidates for strain sensing stems from experimentally measuring the electromechanical response of individually suspended CNTs subjected to localized atomic force microscope probe deformations [49]. Both *in situ* resistance measurements and molecular dynamics simulation results indicate that single-walled carbon nanotube (SWNT) conductance can decrease more than two orders of magnitude while remaining completely reversible over multiple deformation cycles. Although individual CNTs have demonstrated piezoresistive response to applied strain, SHM applications require macroscale sensors that can be installed and queried in engineered structures. In an effort to preserve the inherent piezoresistive properties of individual CNTs while scaling up from a nano- to a macroscale sensing material, researchers have proposed inclusion of CNTs within polymeric matrices. The result is multifunctional materials that can take considerable load [45], yet are capable of self-sensing their strain response to loading [47].

A number of researchers have embedded CNTs within polymeric matrices for strain-sensing applications. Wood *et al.* [50] embedded CNTs within a UV-curable urethane acrylate polymeric thin film. After correcting for temperature-induced strains, they showed that Raman spectra peaks decrease with applied tensile strain. Since Raman spectroscopy requires bulky equipment and is difficult to use in the field, many other researchers have proposed CNT-based "buckypaper" specimens for strain sensing [51–53]. Upon vacuum filtration of a dispersed CNT solution, buckypaper specimens can be peeled off the filter to form the final thin-film strain sensor.

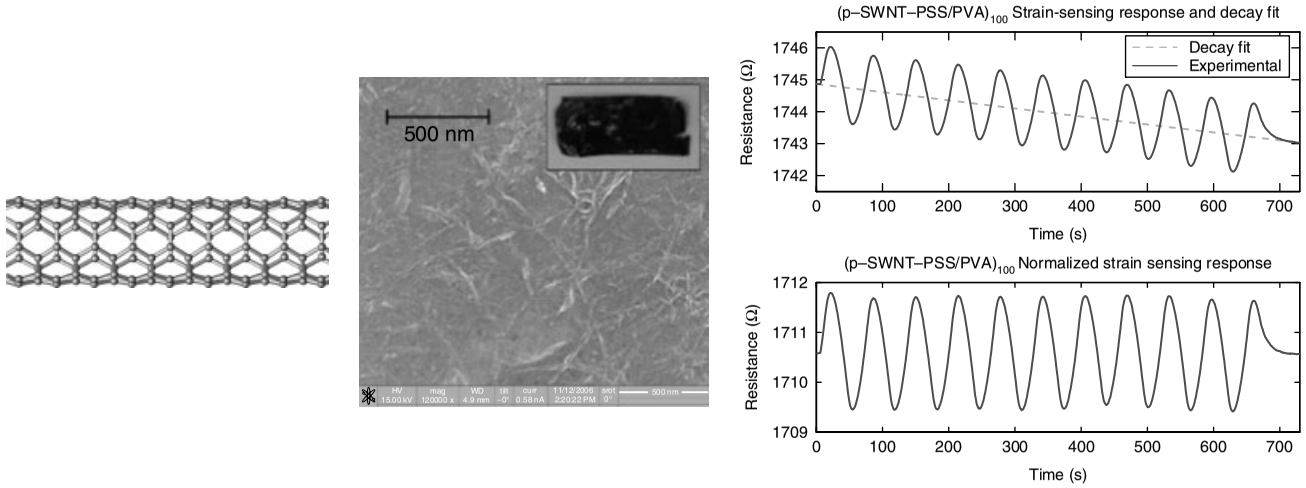


Figure 6. (a) Single-walled carbon nanotube (SWNT) [Courtesy Prof. Vincent Crespi, Penn State]; (b) microscopic view of SWNT-polymeric multifunctional material (insert: final free-standing thin film) [Reproduced with permission from Ref. 46. © IOP Publishing Ltd., 2007]; (c) variation of SWNT-based multifunctional material resistance as a function of mechanical strain. [Reproduced with permission from Ref. 47. © IOP Publishing Ltd., 2007.]

Initial studies conducted by Dharap *et al.* [51] have shown that, using a movable four-point probe technique, SWNT buckypapers exhibit multidirectional linear piezoresistive response up to $\pm 350\,\mu\varepsilon$ strains. Continued work by Li *et al.* [52] measures the Raman wave number shift of the G-band of CNT films and confirms that nanotubes within the buckypaper matrix are strained during applied tensile strains (up to $1000\,\mu\varepsilon$). Using buckypapers fabricated with SWNTs dispersed in dimethyl formamide and then mixed with poly(methyl methacrylate) (PMMA), Kang *et al.* [53] have derived an equivalent electrical circuit model using electrical impedance spectroscopy (EIS) to characterize the SWNT–PMMA thin films. Although SWNT–PMMA thin films exhibit a decrease in gauge factor (strain sensitivity) with increasing CNT weight concentration, pure SWNT buckypapers exhibit a gauge factor of 7 while behaving linearly within $\pm 500\,\mu\varepsilon$ strains. However, strain sensors for structural health monitoring must be robust and reliable over long periods of time. In particular, strain sensors require high sensitivity, linearity, and a wide measurable range (e.g., more than $5000\,\mu\varepsilon$).

Although a variety of carbon nanotube-based thin-film strain sensors have been presented for structural monitoring applications, each aforementioned sensor suffers from some drawbacks. Typically, buckypaper strain sensors exhibit high gauge factors but are very brittle materials; this limits their measurable strain range to within $1500\,\mu\varepsilon$. In response to these limitations, a more homogenous thin-film composite with CNTs within a polymeric matrix can be achieved by layer-by-layer (LbL) self-assembly [45–47]. LbL thin films are assembled by sequential dipping of a charged substrate in polyanionic and polycationic solutions to achieve excellent phase integration and homogenous morphology. Loh *et al.* [47] propose the design of multilayered thin films assembled from solutions containing SWNTs stabilized in poly(sodium 4-styrene sulfonate) (PSS) and poly(vinyl alcohol) (PVA). Free-standing thin films are deposited on a glass substrate with 50–200 layers typically deposited. Subsequent scanning electron microscopy (Figure 6b) reveals good dispersion of SWNTs within the PSS/PVA matrix. Upon application of uniaxial tensile loading, the SWNT-PSS/PVA thin film is shown to be piezoresistive with excellent linearity and high sensitivity (gauge factors span from 1 to 2) (Figure 6c).

4 CONCLUSION

MEMS is a revolutionary technology that positively impacts the field of SHM. MEMSs offer the community microscale sensors that are as accurate as macroscale counterparts at a fraction of the cost and size. Other benefits include the collocation of signal-processing circuitry on the sensor and low-power consumption. As battery-powered technologies such as wireless sensors grow in popularity, low-power MEMS sensors will serve to preserve scarce battery resources. A variety of MEMS sensors have been proposed for use in SHM applications including accelerometers, strain gauges, corrosion sensors, and acoustic transducers. Many novel design and fabrication techniques are constantly being proposed and implemented in the MEMS community to design high-performance sensing devices. On the other hand, the field of nanotechnology also promises to further improve SHM sensing technologies. Unlike the MEMS' "top-down" design methodology, nanotechnology enables molecular manipulation to achieve high-sensitivity and -selectivity, multifunctional sensors. To date, CNT-based multifunctional materials have been developed, which self-sense their response to loading and self-heal in response to damage. In future years, nanotechnology will play an increasingly greater role in the development of next-generation miniaturized sensing technologies. The combination of MEMS low-cost high-throughput fabrication techniques with the versatility of sensor designs originating from the nanotechnology domain can yield novel sensors that outperform those of the current generation.

REFERENCES

[1] Glaser SD, Li H, Wang M, Ou J, Lynch JP. Sensor technology innovation for the advancement of structural health monitoring: a strategic program of US-China research for the next decade. *Smart Structures and Systems* 2007 **3**:221–244.

[2] Kovacs GTA. *Micromachined Transducers Sourcebook*. McGraw-Hill: New York, 1998.

[3] Petersen KE. Silicon as a mechanical material. *Proceedings of the IEEE* 1982 **70**:420–457.

[4] Gardner JW. *Microsensors: Principles and Applications*. John Wiley & Sons: West Sussex, 1994.

[5] Gardner JW, Varadan VK, Awadelkarim OO. *Micro sensors, MEMS and Smart Devices*. John Wiley & Sons: West Sussex, 2001.

[6] Bhushan B (ed). *Springer Handbook of Nanotechnolgy*. Springer: Berlin, 2003.

[7] Nalwa HS. *Nanostructured Materials and Nanotechnology*. Academic Press: San Diego, CA, 2002.

[8] Iijima S. Helical microtubules of graphitic carbon. *Nature* 1991 **354**:56–58.

[9] Klein DL, Roth R, Lim AKL, Alivisatos AP, McEuen PL. A single-electron transistor made from a cadmium selenide nanocrystal. *Nature* 1997 **389**:699–701.

[10] Brinker CJ, Lu Y, Sellinger A, Fan H. Evaporation-induced self-assembly: nanostructures made easy. *Advanced Materials* 1999 **11**:579–585.

[11] Mahar B, Laslau C, Yip R, Sun Y. Development of carbon nanotube-based sensors—a review. *IEEE Sensors Journal* 2007 **7**:266–284.

[12] Senturia SD. *Microsystem Design*. Kluwer Academic Press: Boston, MA, 2001.

[13] Valldorf J, Gessner W (eds). *Advanced Microsystems for Automotive Applications 2005*. Springer: Berlin, 2005.

[14] Cass S. MEMS in space. *IEEE Spectrum* 2001 **38**:56–61.

[15] Panescu D. MEMS in medicine and biology. *IEEE Engineering in Medicine and Biology Magazine* 2006 **25**:19–28.

[16] Roylance LM, Angell JB. A batch-fabricated silicon accelerometer. *IEEE Transactions on Electron Devices* 1979 **ED-26**:1911–1917.

[17] Partridge A, Reynolds JK, Chui BW, Chow EM, Fitzgerald AM, Zhang L, Maluf NI, Kenny TW. A high-performance planar piezoresistive accelerometer. *Journal of Microelectromechanical Systems* 2000 **9**:58–66.

[18] Weinberg H. Dual axis, low g, fully integrated accelerometers. *Analog Dialogue* 1999 **33**:23–24.

[19] Lemkin MA, Boser BE, Auslander D, Smith JH. A 3-axis force balanced accelerometer using a single proof-mass. *Proceedings of the International Conference on Solid-State Sensors and Actuators*, Chicago, IL. IEEE: New York, 1997; Vol. 2, pp. 1185–1188.

[20] Acar C, Shkel AM. Experimental evaluation and comparative analysis of commercial variable-capacitance MEMS accelerometers. *Journal of Micromechanics and Microengineering* 2003 **13**: 634–645.

[21] Lynch JP, Wang Y, Loh KJ, Yi JH, Yun CB. Performance monitoring of the Geumdang Bridge using a dense network of high-resolution wireless sensors. *Smart Materials and Structures* 2006 **15**:1561–1575.

[22] Chung HC, Enotomo T, Loh K, Shinozuka M. Real-time visualization of bridge structural response through wireless MEMS sensors. *Proceedings of SPIE* 2004 **5392**:239–246.

[23] Lynch JP, Partridge A, Law KH, Kenny TW, Kiremidjian AS, Carryer E. Design of a piezoresistive MEMS-based accelerometer for integration with a wireless sensing unit for structural monitoring. *Journal of Aerospace Engineering* 2003 **16**:108–114.

[24] Najafi K, Hetke JF. Strength characterization of silicon microprobes in neurophysiological tissues. *IEEE Transactions on Biomedical Engineering* 1990 **37**:474–481.

[25] Guckel H, Randazzo T, Burns DW. A simple technique for the determination of mechanical strain in thin films with applications to polysilicon. *Journal of Applied Physics* 1985 **57**:1671–1675.

[26] Gianchandani YB, Najafi K. Bent-beam strain sensors. *Journal of Microelectromechanical Systems* 1996 **5**:52–58.

[27] Lin L, Pisano AP, Howe RT. A micro strain gauge with mechanical amplifier. *Journal of Microelectromechanical Systems* 1997 **6**:313–321.

[28] Hautamaki C, Zurn S, Mantell SC, Polla DL. Experimental evaluation of MEMS strain sensors embedded in composites. *Journal of Microelectromechanical Systems* 1999 **8**:272–279.

[29] Aebersold J, Walsh K, Crain M, Martin M, Voor M, Lin JT, Jackson D, Hnat W, Naber J. Design and development of a MEMS capacitive bending strain sensor. *Journal of Micromechanics and Microengineering* 2006 **16**:935–942.

[30] Azevedo RG, *et al.* A SiC MEMS resonant strain sensor for harsh environment applications. *IEEE Sensors Journal* 2007 **7**:568–576.

[31] Ko WH, Young DJ, Guo J, Suster M, Kuo HI, Chaimanonart N. A high-performance MEMS capacitive strain sensing system. *Sensors and Actuators, A* 2007 **133**:272–277.

[32] Lee H, Yun HB, Maclean B. Development and field testing of a prototype hybrid uniaxial strain transducer. *NDT&E International* 2002 **35**:125–134.

[33] Ozevin D, Greve DW, Oppenheim IJ, Pessiki SP. Resonant capacitive MEMS acoustic emission transducers. *Smart Materials and Structures* 2006 **15**:1863–1871.

[34] Jones ARD, Noble RA, Bozeat RJ, Hutchins DA. Micromachined ultrasonic transducers for damage detection in CFRP composites. *Proceedings of SPIE* 1999 **3673**:369–378.

[35] Niblock TGE, Surangalikar HS, Morse J, Laskowski BC, Castro-Cedeno MH, Wilson AR. Development of a commercial micro corrosion monitoring system. *Proceedings of SPIE* 2002 **4934**:179–189.

[36] Brossia CS, Hanson HS. *MEMS Sensor for Detecting Stress Corrosion Cracking*. Patent 6,925,888. U.S. Patent and Trademark Office, 2005.

[37] Polh A. A review of wireless SAW sensors. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2000 **47**:317–332.

[38] Seifert F, Bulst WE, Ruppel C. Mechanical sensors based on surface acoustic waves. *Sensors and Actuators, A* 1994 **44**:231–239.

[39] Avramov ID, Voigt A, Rapp M. Rayleigh SAW resonators using gold electrode structure for gas sensor applications in chemically reactive environments. *Electronic Letters* 2005 **41**:450–452.

[40] Penza M, Antolini F, Antisari MV. Carbon nanotubes as SAW chemical sensor materials. *Sensors and Actuators, B* 2004 **100**:47–59.

[41] Reindl L, Ruppel CCW, Kirmayr A, Stockhausen N, Hilhorst MA, Balendonck J. Radio-requestable passive SAW water-content sensor. *IEEE Transactions on Microwave Theory and Techniques* 2001 **49**:803–808.

[42] Yu B, Meyyappan M. Nanotechnology: role in emerging nanoelectronics. *Solid-State Electronics* 2006 **50**:536–544.

[43] Saito R, Dresselhaus G, Dresselhaus MS. *Physical Properties of Carbon Nanotubes*. Imperial College Press: London, 1998.

[44] Donnet JB, Bansal RC, Wang MJ (eds). *Carbon Black: Science and Technology*. Marcel Dekker: New York, 1993.

[45] Kotov NA (ed). *Nanoparticle Assemblies and Superstructures*. CRC Press: Boca Raton, FL, 2006.

[46] Hou TC, Loh KJ, Lynch JP. Spatial conductivity mapping of carbon nanotube composite thin films by electrical impedance tomography for sensing applications. *Nanotechnology* 2007 **18**: 315501.

[47] Loh KJ, Kim J, Lynch JP, Kam NWS, Kotov NA. Multifunctional layer-by-layer carbon nanotube-polyelectrolyte thin films for strain and corrosion sensing. *Smart Materials and Structures* 2007 **16**:429–438.

[48] Baughman RH, Zakhidov AA, DeHeer WA. Carbon nanotubes—the route toward applications. *Science* 2002 **297**:787–792.

[49] Tombler TW, Zhou C, Alexseyev L, Kong J, Dai H, Liu L, Jayanthi CS, Tang M, Wu SY. Reversible electromechanical characteristics of carbon nanotubes under local-probe manipulation. *Nature* 2000 **405**:769–772.

[50] Wood JR, Zhao Q, Frogley MD, Meurs ER, Prins AD, Peijs T, Dunstan DJ, Wagner HD. Carbon nanotubes: from molecular to macroscopic sensors. *Physical Review B* 2000 **62**:7571–7575.

[51] Dharap P, Li Z, Nagarajaiah S, Barrera EV. Nanotube film based on single-wall carbon nanotubes for strain sensing. *Nanotechnology* 2004 **15**: 379–382.

[52] Li Z, Dharap P, Nagarajaiah S, Barrera EV, Kim JD. Carbon nanotube film sensors. *Advanced Materials* 2004 **16**:640–643.

[53] Kang I, Schulz MJ, Kim JH, Shanov V, Shi D. A carbon nanotube strain sensor for structural health monitoring. *Smart Materials and Structures* 2006 **15**:737–748.

Chapter 67
Nanoengineering of Sensory Materials

Inpil Kang[1], Gunjan Maheshwari[2], YeoHeung Yun[2], Vesselin Shanov[3], Sachit Chopra[3], Jandro Abot[4], Gyeongrak Choi[5] and Mark Schulz[2]

[1] *Division of Mechanical Engineering, Pukyong National University, Busan, South Korea*
[2] *Department of Mechanical Engineering, University of Cincinnati, Cincinnati, OH, USA*
[3] *Department of Chemical and Materials Engineering, University of Cincinnati, Cincinnati, OH, USA*
[4] *Aerospace Engineering, University of Cincinnati, Cincinnati, OH, USA*
[5] *Korean Institute of Industrial Technology, Chonan-Si, South Korea*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

Nanoengineering is important because advanced sensors with high sensitivity, small size, and low cost may be developed on the basis of nanoscale materials. This article gives an overview of the development of nanoscale material sensors for use in structural health monitoring (SHM). The field of nanoengineering of sensors is in the beginning stages and research in this area is exciting and progressing quickly. Nanotechnology has introduced new materials such as carbon nanotubes (CNTs), zinc oxide nanobelts, nanowires (NWs), and metal or semiconductor nanoparticles that are expected to revolutionize the fields of biochemical, electrical, and mechanical sensing. Having 1–100 nm nanoscale size and a high surface to volume ratio, nanomaterials can be combined in host materials to produce changes in the electronic, photonic, catalytic, or other properties of the bulk material for use in virtually all types of sensing applications [1, 2]. Among the nanomaterials available, CNTs are promising for developing unique and revolutionary sensors owing to their structural and electrical characteristics. CNTs have high

strength and high thermal and electrical conductivities and therefore can provide structural and functional capabilities simultaneously, including actuation [3] and sensing [4–6]. CNTs can be synthesized from a few microns to 1.5-cm long with nanometer diameters [7]. For applications at the macroscale, CNT smart materials are usually based on composite materials [8]. The small size of CNTs allows them to have high sensitivity in mechanical environments. Thanks to their smart material properties and many fabrication possibilities, CNTs are producing various kinds of sensors from the nano to the macroscale in size, which are described next.

In nanoscale experimentation, Tombler *et al.* [9] investigated the change of electrical conductivity of a single-wall carbon nanotube (SWCNT) under mechanical deformation using the tip of an atomic force microscope (AFM). Watkins *et al.* [10] used lithography and aligned SWCNT to fabricate an SHM sensor based on strain measurement. This microelectromechanical systems (MEMS) technique can measure small strain and it can also detect very small cracks. For macroscale sensing, Kang *et al.* [11] reported a CNT polymer-based piezoresistive strain sensor for SHM applications. To develop a sensor and actuator embedded in a structure, Jalili *et al.* [12] have been investigating next-generation functional fabrics utilizing SWCNT composites. Functional fabrics and yarns are being developed with distributed actuation/sensing capabilities using CNT-based mono- and multifilament yarns. Having high energy density storage, and the ability to be activated using low power, a nanomaterial-based sensor can be light and flexible and easily embedded into a structure. These characteristics are suitable for developing new sensors required for health monitoring of lightweight structures. *In situ* sensors can monitor the behavior of structures assessing damage or deterioration. Flexible and easily embedded CNT sensors are envisioned to be useful for *in situ* SHM for ageless vehicles that are capable of remaining in "as new" operating condition indefinitely, regardless of use. This concept has been proposed by NASA. The ageless vehicle requires structural self-assessment and repair capabilities to be carried out by distributed sensors [13].

Extraordinary recent nanoscale materials work by Baughman *et al.* [14–18] has been to develop artificial muscles that are powered electrically or by fuel cells, nanotube yarn and sheets that are a reinforcing material that is also used for strain sensing, and biological sensors. Using longer nanotubes to improve the properties of the spun fibers is one of the objectives of the continuing research by Baughman, Zhang, *et al.* Lynch *et al.* [19] have developed an electrical impedance tomography method for damage detection in CNT composite materials. NASA has developed nanotube sensors for mechanical, chemical, and biological applications. A good reference is the book by Meyyappan [20]. Varadan *et al.* have developed a strain sensor using nanotubes with a semiconductor matrix [21]. Yuan and Jin have modeled the elastic properties of CNTs [22], which may be used for modeling nanotubes in composites for strain sensing applications. A strain sensor using nanotube materials is described by Ramaratnam *et al.* in [23]. Getty has formed a magnetic compass using a ferromagnetic needle mechanically coupled to SWCNTs [24]. The needle is deflected in a magnetic field and bends the SWCNTs, which can cause orders of magnitude change in their resistance thus forming a tiny highly sensitive magnetic field sensor.

In the area of civil engineering, significant investigation is underway to determine if nanotubes added to cement can act as bridges across cracks and voids to transfer load and create a material more impervious to bending tensile loading [25, 26]. Nanotubes have a high strain to failure, which helps in the load transfer across cracks. Multiwall carbon nanotubes (MWCNTs), after being functionalized (chemically modified) using H_2SO_4 and HNO_3 solutions, were added to cement matrix composites. The treated nanotubes improved the flexural strength, compressive strength, and failure strain of the cement matrix composite. The porosity and pore size distribution of the composites were also reduced. The phase composition was characterized with Fourier transform infrared spectroscopy. Interfacial interactions were found between the CNTs and the hydrations in the cement (such as C−S−H and calcium hydroxide), which should produce a high bonding strength between the reinforcement and cement matrix.

There is the exciting potential to develop nanoscale hybrid materials that have designed-in properties to meet specific applications. As an example of a novel hybrid material, researchers at Argonne National Laboratory have combined the world's hardest

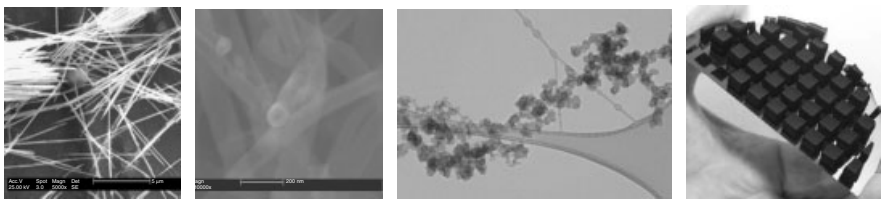
material—diamond—with the world's strongest material—CNTs [27]. The process for "growing" diamond and CNTs together is the first successful synthesis of a diamond–nanotube hybrid material and opens the way for its use in energy-related applications and SHM owing to anticipated piezoresistive behavior.

In the general context of developing new sensory materials, there are many types of starting nanomaterials available. These 1D nanostructures include nanobelts, NWs, nanorods, nanotubes, nanonails, nanoflowers, etc. There are four commonly available nanoparticle materials that are particularly important for use in SHM and the related areas of engineering asset evaluation and condition monitoring. These nanoparticles can also be used to develop sensor systems that perform data mining at the sensor level [28, 29] and that can harvest power from ionic flow [30]. These four materials [31–36] are NWs, carbon nanofibers (CNF), carbon nanosphere chains (CNSCs), and CNT arrays, as shown in Figure 1. NWs (Figure 1a) have semiconducting, conducting, magnetic, and other properties, and have high density compared to other nanoparticles. CNF (Figure 1b) are nested MWCNTs that have a 20° conical shape, larger diameter than nanotubes, good properties, and very low cost. CNSC (Figure 1c) are carbon onions linked together, catalyst-free, highly electrically conductive when post treated, and low cost. With a specific post treatment, CNSC also have weak magnetic properties. Thus, carbon is now a member of the magnetic club. CNT arrays (Figure 1d) are parallel forests of aligned MWCNT up to 1.5-cm long with good electrical and mechanical properties.

Other upcoming nanomaterials not as commonly available are shown in Figure 2. These include telescoping MWCNTs, coiled MWCNT, alloy NWs, piezoelectric NWs, and zinc oxide NWs and nanobelts that are piezoelectric.

Properties of SWCNT are discussed briefly. Other types of nanotubes have properties that are similar but not as good as SWCNT. SWCNT is the strongest and most flexible molecular material known because of the C−C covalent bonding and seamless hexagonal network architecture. SWCNT have a Young's modulus of 1 TPa, which is above the modulus of 70 GPa for aluminum, and 700 GPA for carbon fiber. The strength-to-weight ratio of an SWCNT nanocomposite could be 4 times the same ratio for graphite/epoxy as predicted in Chapter 15 of [37]. The maximum strain of SWCNT can be ∼10%, and the thermal conductivity is $\sim 3000\,\mathrm{W\,mK^{-1}}$ in the axial direction. SWCNT have semiconducting or conducting electrical properties depending on the chirality (angle of twist) of the nanotube. Electrically conductive nanotubes are called *metallic tubes* and armchair nanotubes have electrical conductivity like metal while semiconducting tubes have transistor properties. Nanotubes have a high current carrying capacity and have a high aspect ratio and small tip radius of curvature ideal for field emission. SWCNT have magnetoresistive and piezoresistive properties, but negligible piezoelectric property. They also have electrochemical properties and a supercapacitance property in an electrolyte and can form an electrochemical actuator. Properties of MWCNT are similar to SWCNT but MWCNT are electrical conductors. Electrochemical actuation, a nanobearing, structural reinforcement, a telescoping actuator, and co-ax cable are some of the envisioned applications of MWCNT.

Processing materials with nanoscale features require extra care in most steps of the procedure. In the area of synthesis, researchers are continuously improving control over CNT length, diameter, and array density by careful control over substrate preparation and nanotube synthesis conditions (Chapter 5 of [37]). Dispersion of nanoparticles in polymers is another critical step that is continually being



(a) (b) (c) (d)

Figure 1. Nanomaterials for sensors: (a) Ni NW; (b) CNF; (c) CNSC on TEM grid; (d) MWCNT.

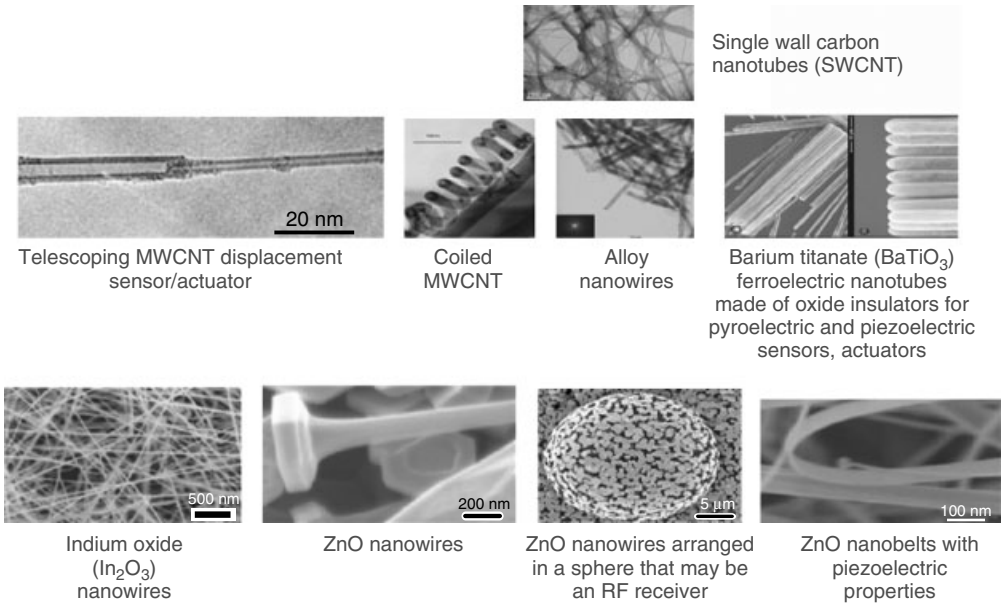


Figure 2. Less common types of nanoscale materials that may form SHM sensors.

improved [38]. High-resolution transmission electron microscopy, X-ray photoelectron spectroscopy, and environmental scanning electron microscopy are the primary tools used to characterize the nanoparticles and their smart materials. Our philosophy in developing nanoparticles and products follows from a Sherlock Holmes quote: "The world is full of obvious things which nobody by any chance ever observes." Careful observation at every step of material processing is needed in the field of nanoengineering.

Some general guidelines are discussed next to aid in the design of nanomaterials for sensors. As the size of the particle goes down, the surface area to volume ratio goes up and the properties of materials often improve. Characteristics of nanoscale materials and devices that may be helpful in sensor design are (i) viscosity increases at the nanoscale; (ii) electrostatic attraction is large at the nanoscale, particles stick together, and surface tension is large; (iii) friction depends on nanoscale surfaces, MWCNT nested nanotubes that are straight are almost friction-free translational and rotational bearings, van der Waals forces cause retraction of tubes or bundling of tubes; (iv) increased heat transfer may change the efficiency of tiny machines; (v) the melting point of nanomaterials may decrease as the thickness decreases; (vi) as size goes down, the frequency of electric circuits built using nanoparticles goes up; (vii) nanotube contact resistance is high and may be reduced by coating the nanotube ends such as with titanium and gold; and (viii) there are other electronic and optical effects described in the literature that may be useful for SHM sensing and power harvesting. In the next sections, several types of nanoparticle-based sensors are described to give an overview of the field. Since this field is rapidly changing, readers are encouraged to survey the literature for new developments.

2 NANOPARTICLE SPRAY-ON SENSORS

Many types of nanoparticles can be dispersed in a solvent or polymer and sprayed onto a structure to form a sensor. Here, we consider fabrication of a long spray-on sensor using CNTs, which can be considered to be a continuous sensor like a dendrite of a sensory or afferent neuron that conveys information from tissue and organs to the central nervous system in the human body. A structural neuron is formed using a CNT continuous sensor (a dendrite) and an analog electronic processor (the

cell body) that collects inputs and generates an electrical output (fire). The output flows through a wire conductor (an axon in the biological system) to a computer (the brain) for analysis. A network of structural neurons can cover large areas and monitor a structure for damage (pain) in real time using parallel processing (as in the biological neural system) and a small number of channels of data acquisition. The CNT-based neuron can be used as an alternative to the piezoelectric ceramic-based neurons used in [39]. Piezoelectric-based neurons are described in **Wind Turbines**. A structural neural system (SNS) can be built with a dispersed MWCNT solution that can be sprayed with an airbrush on a patterned surface of a structure as shown in Figure 3. The CNT neuron is a thin and narrow polymer film sensor that is sprayed, bonded, or taped onto a structure [40]. The electrochemical impedance (resistance and capacitance) of the neuron changes due to deterioration of the structure where the neuron is located.

The neuron has bulk piezoresistivity due to the CNTs in a polymer matrix, which is useful as a strain sensor for engineering applications. The polymer improves interfacial bonding between the nanotubes and enhances the strain transfer, repeatability, and linearity of the sensor. The largest contribution of piezoresistivity of the sensor may come from slippage of overlaying or bundled nanotubes in the matrix, from a macroscopic point of view. Nano interfaces of CNTs in a polymer matrix also contribute to the linear strain response compared to other microsize carbon fillers. Low weight percentages of nanotubes in a polymer not only have a percolation behavior with high sensitivity but also may have a nonlinear response. Buckypaper is formed by dispersing nanotubes and forming a film by solution evaporation. SWCNT are the best nanoparticles to form buckypaper because the van der Waals forces are large for small diameter nanoparticles. However, buckypaper is still relatively weak. Although buckypaper has a large gauge factor for strain sensing, the response is more nonlinear and the strain range is smaller as compared to using a polymer host for the nanotubes. A typical response of a nanotube-polymer sensor on a cantilever beam is shown in [11, 40] and can be modeled as $V = 0.0016 \times S$, where $V =$ voltage, $S =$ microstrain.

2.1 Strain sensing

The CNT-based neuron is modeled as a parallel R−C circuit as shown in Figure 4(a). The neuron can monitor static and dynamic strain as shown in Figure 4(a, b). This testing showed that the neuron has a bandwidth of about 20 Hz and measures the average strain over the length of the sensor. The strain sensitivity is similar to that of a strain gauge. The CNT-based neuron is a practical sensor because the monitoring signals are low voltage and form a simple circuit.

2.2 Crack detection testing

CNT-based neurons can monitor damage by two methods. The first is by the change in impedance of the sensor if a crack propagates through the sensor. In this case, a fine network of micron thin neurons could cover a structure. The second approach is to monitor the change in dynamic response of the structure. The dynamic response will change when damage occurs in the structure. A CNT neuron was tested to measure cracking on a composite beam. Deterioration due to cross-sectional damage or a crack was detected using the dynamic strain response of the neuron. The local stiffness of the structure changes due to damage, which, in turn, influences the dynamic response of the system. The CNT-based neuron can also monitor crack growth because the

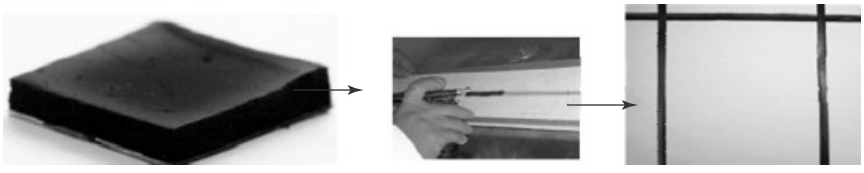


Figure 3. CNTs array dispersed in a solvent or polymer and sprayed onto a structure to form long thin continuous sensors that are analogous to a dendrite of the biological sensory neuron.

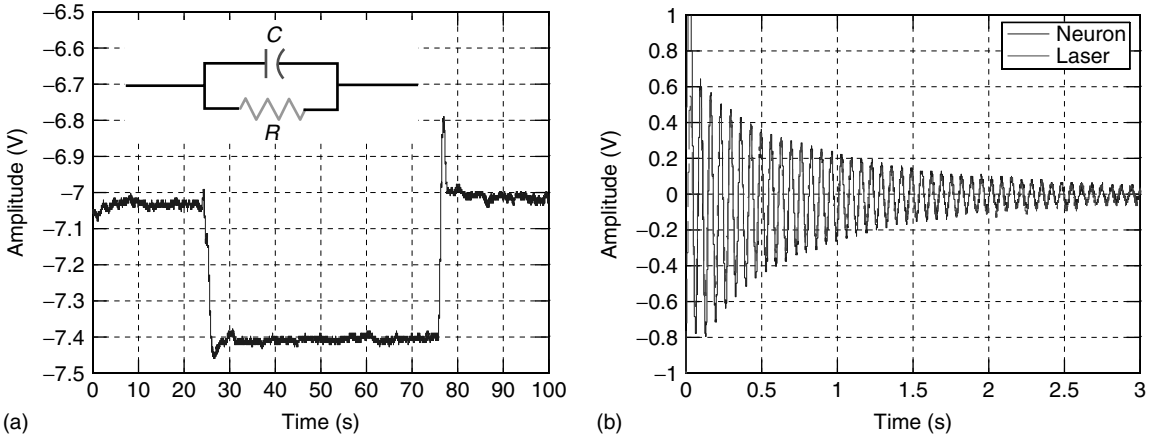


Figure 4. Modeling and strain responses of the CNT-based neuron on a glass fiber cantilever beam: (a) Electrical model and response of the CNT-based neuron under a static load; (b) dynamic strain response of the neuron during free vibration due to an initial displacement, the displacement response of the beam was measured by a laser displacement sensor (Keyence, LC-2400 Series) and almost overlays the strain response of the neuron.

nanotube film will change resistance when a crack propagates through it. Under crack propagation, the resistance of the neuron increases and the capacitance decreases, which changes the voltage response of the neuron, which can be measured using a bridge electrical circuit. The response of the neuron as a crack propagates through it is shown in Figure 5(a). The normalized crack size is defined as the ratio of crack size divided by neuron width when the crack progresses through the neuron normal to its length direction. A crack size of 100% means complete separation of the neuron by the crack. The increased resistance due to damage causes a higher amplitude voltage and the reduced capacitance induces a phase shift of the dynamic response. With this crack sensor approach, damage must change the strain at the neuron. A high density of micron size neurons can be used to detect small damage.

2.3 Corrosion detection testing

The same CNT neuron used for strain measurement can effectively measure corrosion because of the high electrochemical sensitivity of the nanotubes. Corrosion occurring on a metallic structure produces

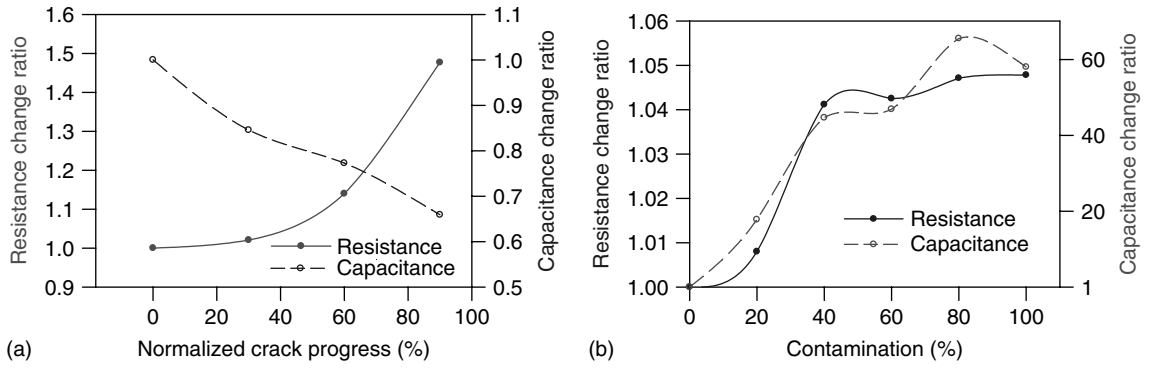


Figure 5. Response of the neuron due to cracking on a vibrating beam: (a) The variation of resistance and capacitance of the neuron as the crack propagates through the CNT neuron for the static beam; (b) The CNT neuron used as a corrosion sensor.

a diffusion layer at the interface between the structure and the CNT neuron. The corrosion ions penetrate into the CNT polymer sensor and form a double layer charge (Chapter 15 [37]) on the nanotube surface. The diffused ions also change the electrical parameters of the CNT sensor much like a doping effect. Having a large surface area, the capacitance of the CNT can be changed by orders of magnitude by the ions produced by corrosion, and the resistance will also change a small amount. A preliminary test of the corrosion sensitivity of the sensor was performed. A chemical buffer solution was used to simulate corrosion. The contamination percent is defined as the percent contaminated area on the neuron. After applying the solution, the variation in resistance and capacitance of the sensor is shown in Figure 5(b). The capacitance of the neuron sensor dramatically increased by a factor of 50 while the resistance decreased by only 5%. For composites, moisture absorption, corrosion at interfaces, and chemical degradation of the polymer may be monitored using the electrochemical response of the neuron.

As shown in the above experiments, the CNT sensor is versatile and can measure multiple physical properties of a structure. The electrical parameters of the CNT consist of a parallel resistor and capacitor and show distinct changes and trends related to physical measurements that are useful for SHM. Strain mostly causes a change of resistance. The capacitance is not affected much by strain. The sensor is moderately sensitive to changes in both resistance and capacitance when monitoring cracking. However, the sensor is dramatically sensitive to the change of capacitance when monitoring corrosion. On the basis of these sensitivities, the multifunctional capability of the neuron is a new approach for SHM. Since corrosion, infiltration of water, cracking, delamination, temperature, and change of pH can all possibly change the impedance of the nanotubes, the neuron is a sensitive barometer of the overall health of the structure. The multifunctionality is expected to effectively detect multiple symptoms of damage occurring at the same time and at the same location in complex structures. CNT neurons can be applied in the field by spraying the material onto the structure. Different types of nanoparticles (SWCNT, MWCNT, CNF, and CNSC), polymers (polymethylmethacrylate (PMMA), epoxy, polyvinyl alcohol (PVA)) and processing methods (dispersion, functionalization, vacuum, pressure, and temperature) were investigated to develop the neuron [8].

3 PIEZORESPONSIVE POLYMERS USING NANOPARTICLES

Piezoresponsive materials change properties due to strain of the material. Three types of piezoresponsive materials are generally being considered to form sensors. The materials are piezoresistive (change electrical resistance with strain), piezomagnetic (change magnetic properties with strain), and piezoelectric (produce a charge with strain). The piezoresponsive materials are usually formed by loading a polymer host material with one or more nanoparticles that have specific sensing properties. This section gives examples of different piezoresistive sensors being developed at the University of Cincinnati and by collaborators based on integrating sensor nanoparticles into polymers. Since the polymer is often used as the matrix of a composite material, the piezoresponsive polymer can be used to form self-sensing nanocomposite materials. CNSC-epoxy nanoskin and CNSC-polyurethane nanoskin are two new materials being developed to provide a means to tailor the surface properties of structures.

A general procedure to fabricate nanocomposite and sensor materials follows four steps: (i) selecting the nanoscale constituent material, synthesis, and properties characterization of nanoparticles including tubes, wires, spheres, and plates, with desired electrical, magnetic, and other properties; (ii) performing intermediate processing to prepare the material for incorporation into a host material, involves purification, thermal treatment, functionalization, spinning thread, and partial self-repair by welding or recrystallization; (iii) forming the bulk material and nanostructured smart materials including nanocomposites, sensors, films, actuators, wires, or casting into polyacrylonitrile/pitch resin (PAN/PITCH) carbon fibers; and (iv) material design for specific applications including mechanical, thermal, electrical, and environmental.

Processing of nanocomposite materials (step ii) is challenging because nanofillers have several, not all good, effects on the bulk material. Nanotubes are so small that they may affect the crystallization of the polymer and the heat transfer of the polymer

affects the curing cycle. Bonding the nanoparticle to the polymer usually requires attaching interface molecules to the surface of the nanoparticle. Modification of the surface of nanoparticles is called *functionalization*, and a good overview is given in [38, 41]. Noncovalent attachment of molecules relies on van der Waals forces or polymer chain wrapping. The noncovalent bonding alters the nanoparticle surface to be compatible with the bulk polymer with the advantage that the mechanical properties of the nanoparticle are not changed. A disadvantage is that the forces between the wrapping molecule and nanoparticle are often weak, which reduces the efficiency of the load transfer between the matrix and nanoparticle.

On the other hand, covalent bonding of functional groups to the nanoparticle produces a strong connection that may improve the efficiency of load transfer between the particle and matrix. Covalent bonding is usually specific to a given system with crosslinking possibilities. On the down side, covalent bonding is done using strong acids or plasma, which will likely introduce defects in the nanoparticles, such as holes in the walls of CNTs. Defects will lower the strength of the nanoparticle and may affect the electrical and other properties of nanotubes. Ultrasonication using a tip sonicator is used to disperse nanotubes, but the sonicator may put defects into the nanotubes. A good practice is to observe the walls of the nanotubes using a high-resolution transmission electron microscope after sonication to check that the damage is small.

Using the processing methods described and different nanoparticles, a number of recipes are being developed and improved for fabricating nanocomposite multifunctional materials (Table 1). We envision that future nanocomposite materials will be custom blended for specific applications. Several nanocomposite materials in development are discussed next.

3.1 Piezoresistive epoxy

Piezoresistive epoxy is formed by adding CNF or CNSC to epoxy. CNF material is commercially available at low cost [33] and probably is the nanomaterial most commonly used to reinforce polymers. The material comes in different grades including prefunctionalized material that is chemically modified using acid treatment to improve bonding to polymers. Results of putting CNF in polymers are described in [37, 41]. The CNSC is a newer material that has

Table 1. Recipes for fabricating nanocomposite self-sensing materials

Nanoscale material	Intermediate processing	Bulk material properties
Nanotube array	Spin into thread	Strong nanocomposite self-sensing and repair
Nanotube array	Cast polymer into array (epoxy, rubber, other)	Polymer nanocomposite (piezoresistive)
Nickel NW	Disperse in polymer	Polymer nanocomposite (piezomagnetic, NDE of composites)
CNSC/CNF	Disperse in polymer	Polymer nanocomposite (piezresistive)
CNSC/CNF	Disperse in cement mix	Nanocement (piezoresistive)
ZnO nanobelt	Align in polymer	Nanocomposite (piezoelectric)
A + B + C	Disperse in D	Nano A,B,C,D (piezo)

an onion-like morphology and is highly electrically conductive and catalyst-free. The CNSC material is commercially available from Clean Technologies International [32]. Modeling nanocomposites using the simple rule of mixtures, dispersion of CNSC in the polymer, and testing CNSC/Epoxy nanocomposites are described in [42]. The basic procedure for manufacturing nanocomposites involves dispersing CNSC in epoxy resin using a shear mixer and ultrasonicator simultaneously, as shown in Figure 6(a). A compression test setup and the compression stress–strain curve are shown in Figure 6(b, c). The mechanical properties of the polymer improve a small amount by addition of 3 wt% of CNSC. The elastic modulus increased from 2.9 to 3.3 GPa, a 13% improvement, and the strength increased from 106 to 113 MPa, a 6% improvement. Dispersion of up to 5 wt% of CNSC in epoxy has been done. Plasma functionalization is also being done to improve dispersion and bonding to the polymer. Studies of the mechanical, electrical, and thermal properties of this material are under way.

3.1.1 Electrical testing of carbon nanosphere chain/epoxy nanocomposites

Nanocomposite button samples were fabricated for electrical testing. Argon plasma treated CNSC were dispersed in epoxy resin and cast in Teflon button shape molds. Epoxy Epon 862 resin with W curing agent was used. Figure 7(a) shows the resistivity of a 2-wt% CNSC-epoxy nanocomposite sample versus stress. It can be seen that the resistivity decreases initially with increasing stress and then approaches a constant value. The beginning part of the graph indicates that one can use the composite for a pressure or strain sensor. Figure 7(b) shows the resistivity of a 10-wt% CNSC-epoxy nanocomposite sample versus stress. The electrical conduction of CNSC-epoxy for

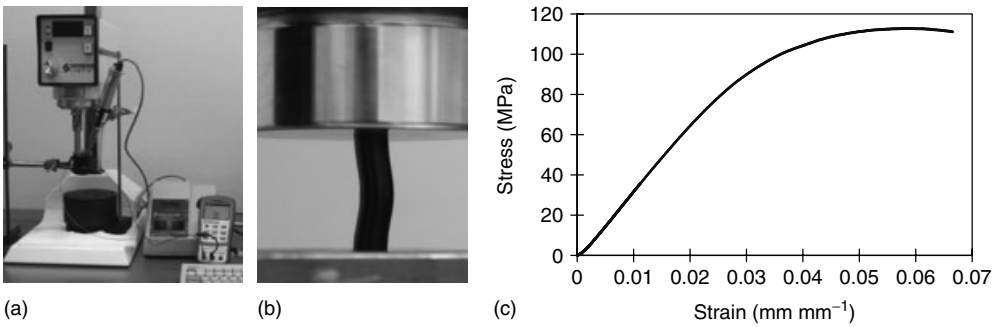


Figure 6. CNSC material processing: (a) mixer, ultrasonicator, and impedance measurement for dispersing nanoparticles; (b) nanocomposite testing; (c) stress–stain CNSC/epoxy 3 wt%.

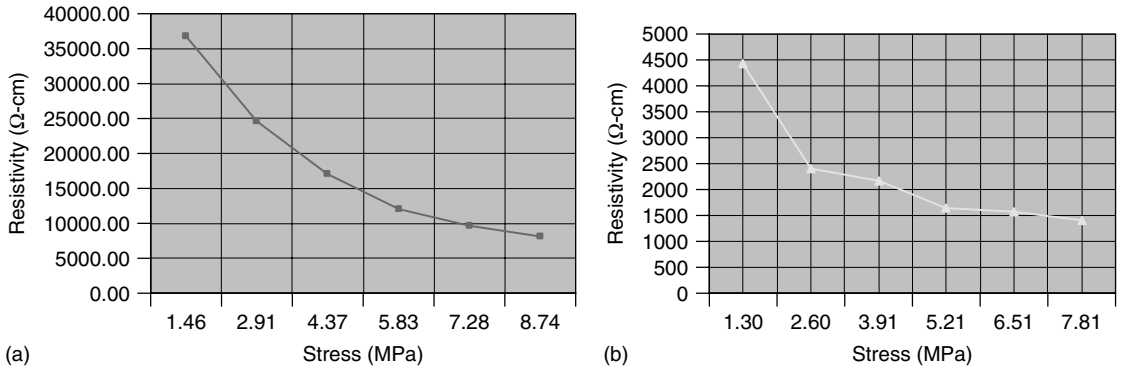


Figure 7. Resistivity versus compressive stress for CNSC/epoxy nanocomposites: (a) 2 wt% CNSC; (b) 10 wt% CNSC.

10 wt% loading is greater than for 2 wt% loading. The mechanical properties of this material have not been evaluated yet. A goal for some applications is to increase the electrical conductivity as much as possible without degrading the mechanical properties. In this case, the resistance between electrodes on the structure can be monitored to detect damage. At higher weight percentages, the viscosity of the polymer becomes very high and may prevent wetting of microfibers in a composite. To sense strain, a low weight percentage of CNSC could be used to provide a high gauge constant. High sensitivity is obtained when the loading is near the percolation level for the material, about 1 wt% or less of CNSC. Functionalization of the CNSC is under investigation to improve dispersion and to prevent reagglomeration when curing.

In general, nanoparticle-enhanced polymer design can provide electrical conductivity to the polymer and can reinforce epoxy and other matrix materials. The CNSC when compacted are highly electrically conductive, and presently about 5% by weight of CNSC can be added to the epoxy to provide electrical conductivity. A limiting factor is to not make the viscosity too high for wetting microfibers when making a laminated composite. The next step in the nanocomposite development is to investigate in detail functionalization (chemically modifying their surface) of CSNC by plasma etching to provide better dispersion and adhesion to the matrix. The plasma treatment will be controlled so that the electrical properties of the CNSC are not significantly affected. In general, a goal is to provide multifunctionality to the material, which means, for example, (i) having the greatest electrical conductivity possible, (ii) self-sensing to detect cracks, corrosion, and delamination in composites, and (iii) to do this without reducing the other properties of the matrix material. It would also be interesting to use a combination of different scales of nanoparticles (smallest to largest, SWCNT, MWCNT, CNSC, and CNF), which might stay dispersed better and might transfer load more gradually to the matrix.

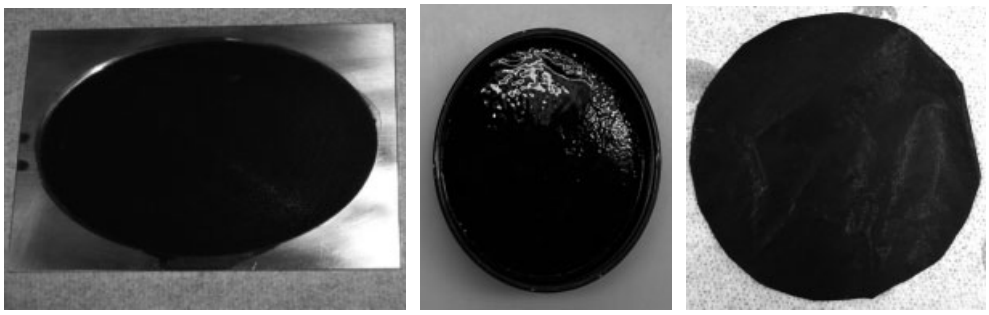
3.2 Carbon nanosphere chain-polymer nanoskin materials

Carbon nanosphere chain epoxy, polyurethane (PU), and PVA nanoskin materials are discussed in this section. It has been shown that the electrical conductivity of nanoparticles depends on mechanical contact pressure. Higher pressure increases conductivity. When forming nanocomposites by dispersing nanoparticles in a polymer, it is difficult to apply contact pressure because of the shape of the mold and because the phase diagram changes (e.g., the material may not cure). Therefore, a skin material is considered because it can be held together by pressure during curing, which improves the electrical properties. CNSC are used to form the nanoskin in these examples, but other materials can be used. CNSC have high electrical conductivity after post treatment. Oxidization in air starts at about 500 °C. The CNSC are thus useful for developing moderate temperature range sensors. Advantages of CNSC are low cost, large quantities of the material can be produced, and it is catalyst-free, which is useful in obtaining a large signal and to avoid toxicity. CNSC magnetic properties are under investigation.

3.2.1 Carbon nanosphere chain epoxy nanoskin

Nanoskin is formed by (i) a solution casting a CNSC film or other nanoparticles in a mold; (ii) coating the film with epoxy or other polymers; and (iii) pressure casting in a mold. The nanoskin material formed has good electrical and thermal conduction, piezoresistive and possibly piezomagnetic properties, and possibly power harvesting properties when a plasma gas or electrolyte is passed over the skin. An interesting property of nanoskin formed in this way is that it is electrically conductive on one side of the skin and electrically insulating on the other side of the skin. The film is shown in Figure 8(a). It is expected the skin could be made electrically conductive through the thickness by increasing the pressure and percentage of CNSC in the mold. Areas for improvement of CNSC-epoxy are to increase the wt% in dispersion with polymers and using higher compression casting to push the electrical conductivity closer to that of the compressed powder material. The electrical conductivity of nanoskin is quite below the conductivity of aluminum, but is being improved by postprocessing and by improved casting procedures.

The electrical resistance of a section of the epoxy nanoskin was measured by polishing the surfaces



(a) (b) (c)

Figure 8. CNSC polymer multifunctional nanoskin: (a) bottom of mold to form CNSC-Epoxy nanoskin; (b) CNSC-Polyurethane nanoskin; (c) CNSC-Polyvinyl Alcohol nanoskin.

of the nanoskin to expose the nanoparticles and applying copper foil electrodes to the two sides of the skin. The skin was compressed and the force and resistance values were recorded and range from 64 N and 2000 Ω to 383 N and 877 Ω. This data shows that the resistance decreases with increasing load because the contact resistance of the copper and nanotube interface is reduced with pressure, and possibly because the nanoparticles are coming into closer contact in the epoxy. There is potential to greatly reduce the resistivity of the nanoskin by using a larger volume of CNSC and higher pressure during casting. The powder form of the CNSC under compaction has a resistivity of 0.5 Ω cm for the as-grown material. Post-treated material has orders of magnitude lower resistivity. The lowest resistivity possible for the nanoskin is the resistivity of the power material.

3.2.2 Carbon nanosphere chain-polyurethane nanoskin

A PU elastomer film was prepared using CNSC. The procedure was to disperse CNSC in N,N-Dimethylformamide (DMF) solvent, and evaporate the solvent in a mold leaving a thin film of CNSC on the mold. Then PU was poured over the CNSC and cured. The resulting film shown in Figure 8(b) was elastic, electrically conductive on the CSNC side, and electrically insulating on the topside. CNSC-PU material may be used for damage detection by sputtering thin film electrodes on the surface and monitoring the electrochemical impedance between the electrodes to detect cracks or corrosion. Applications of the electrically conductive PU elastomer might be an anti-icing heater film, erosion resistant coatings, and a low impedance piezoresistive sensor, and there may be medical applications such as to repair the body and build artificial organs. Electrical and mechanical characterizations are being carried out.

3.2.3 Carbon nanosphere chain-polyvinyl alcohol nanoskin

PVA nanoskin was formed by dispersing CNSC in water using a magnetic stirrer and a long mix cycle. The film was cured at room temperature in air for 48 h. The resulting film is shown in Figure 8(c). The electrical resistance of the PVA nanoskin was 150 Ω for 10% concentration of CNSC by weight. The electrical measurement was made using a two-point probe and multimeter. Resistivity will be determined using a four-point probe.

3.3 Piezoresistive and piezomagnetic nickel nanowire polymers

Nickel NWs and the general physics of NWs are discussed briefly to provide ideas for future sensor applications. NWs differ from their corresponding bulk (3D) materials because of an increased surface area to volume of the material and because of quantum confinement effects. NWs also exhibit different optical and electrical properties from the bulk material, e.g., transistor and magnetic properties, but here only the magnetic attraction and electrical conduction properties are considered for developing sensors. First, there are two basic approaches to

synthesize NWs: nontemplate-assisted growth and template-assisted growth. A simple way to make NWs is to use a mold or template [43]. In this experiment, nickel NWs are grown inside the pores of an alumina filter and then the filter is removed by etching to yield magnetic NWs. The nanoporous membranes used were designed for healthcare applications including virus filtration, sample preparation, and liposome manufacture (http://www.whatman.com). These alumina membranes are manufactured by applying a large electrical potential to a piece of aluminum metal submerged in an acid. Aluminum is oxidized to alumina (Al_2O_3) and pores are created. The size of the pores depends on the applied potential. In this experiment, membranes with a pore size of $0.02\,\mu m$ are used as templates.

Template-assisted growth used here requires the following steps: template preparation; filling the template with NW material; and etching away the template. Ni NWs are formed by electrochemical deposition of Ni on a cathode electrode in the electrolyte solution. One possible application for magnetic NWs filled in a polymer is for SHM of composites. The electrical or magnetic properties of the composite can be monitored and changes can be related to damage. It may be possible to adapt the eddy-current method from aircraft nondestructive evaluation (NDE) to Ni NW composites. The integration of Ni NWs in composites is in the beginning stages. Some characteristics of NWs are that they are much denser than nanotubes and initially seem more difficult to disperse. Also, the NWs affect the thermal conduction when processing the composite. Functionalization of the NWs to bond to epoxy must be explored. NWs should be much easier to align in a polymer because of their large magnetic attraction. Electrical and magnetic characterization will be performed for Ni NW nanoskin and nanocomposites.

4 ELECTRICALLY CONDUCTIVE CEMENT USING CARBON NANOFIBERS

Developing smart materials for civil infrastructure applications is an important area of research because of the huge volume of material that might be used. CNF are considered for this application because CNF are produced in large volumes in a continuous chemical vapour deposition (CVD) process using an iron catalyst. The nested carbon cones (Figure 9a) or spiral cones are electrically conductive, 70–150-nm diameter, micrometer long, have good strength, are low cost \$45/kg, and are available already functionalized (PR–24–XT oxidized and debulked). Applied Sciences Inc. (Cedarville Ohio) observed very good machinability at 1–3 wt% of nanofiber in cement as evidenced by the image in Figure 9(b) of a thread on a machined cement bolt. The University of Cincinnati has pressure cast CNF into mortar mix to provide electrical conductivity and piezoresistive sensing, but brittleness of the material is a challenge.

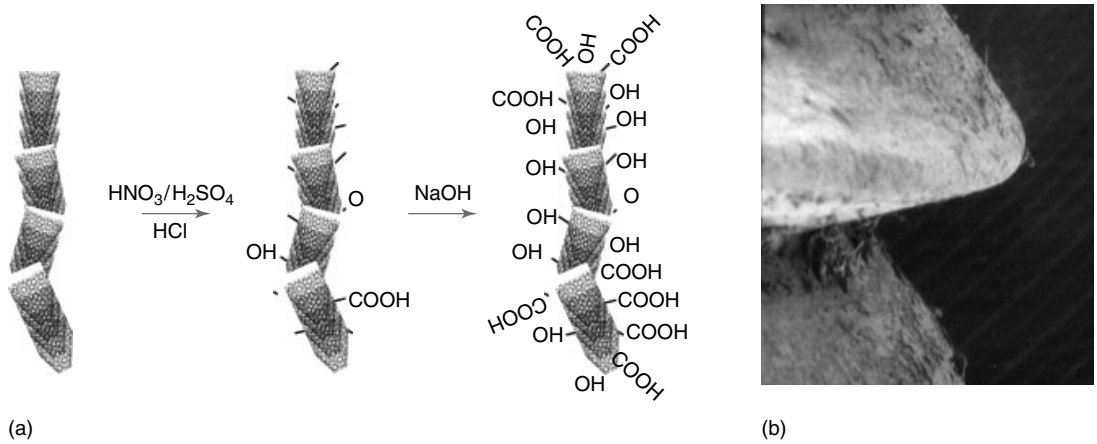


Figure 9. CNF-nanocement: (a) CNF functionalized; (b) screw threads in a nanocement bolt (image from Applied Sciences, Inc.).

The recipe to fabricate nanocement is (i) mortar mix, (ii) CNF (5% wt), (iii) ball mill 3 h, (iv) add water and mix, (v) pressure cast, (vi) heat treat, (vii) apply electrodes, and test. The CNF must be mixed with the cement as a powder because the small amount of water used would not allow a high enough percent of CNF to disperse. The cement/CNF powder is mixed by ball milling. Then water is added in slight excess and mixed to form a paste. The paste is put into a cylindrical mold and pistons are used to compress the nanocement mixture to remove air voids and excess water. Final curing is done in an oven. Electrodes are placed on the cement to compute the electrical resistance. An increase in resistance between any two electrodes indicates damage, corrosion or cracking. A specimen of the nanocement is shown in Figure 10. The resistivity decreases with increasing load until a resistivity of about 10 Ω cm is reached. This is an initial result and the dispersion and mechanical properties must be evaluated. Improvement in the properties of this initial sample is possible.

CNSC are the other material that seems ideal for integration into cement due to the high volume of material that can be produced at low cost and CNSC have high electrical conductivity and no catalyst. Nanotubes may be too expensive to use for reinforcing cement. Presently, high-grade cement uses chopped fiberglass with microscale fibers. It may be possible to develop hybrid cement using different types of fibers. The cement may be used to heat and melt ice, as an antenna material, and as a supercapacitor material. When wet, the cement may tend to increase corrosion of steel reinforcement bars (rebar) by providing an electrical path for corrosion to occur. This aspect has to be considered further.

5 ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY FOR SHM

The analysis method to detect cracking and corrosion in nanocomposites including polymer, elastomer, and cement is electrochemical impedance spectroscopy (EIS) as shown in Figure 11.

A sine wave with a zero or nonzero dc voltage is applied to the sensor. The current is measured and the complex impedance is calculated. A redox chemical is sometimes used to reduce the impedance at a particular dc potential at which the redox reaction occurs. As an example of the use of EIS, a gold electrode was functionalized and used in an electrolyte with a redox couple. The EIS of the electrode was computed over a frequency range of 0.1 Hz–300 KHz. EIS was performed using a three-electrode cell, consisting of a gold electrode as the working electrode, an Ag|AgCl reference electrode, and a platinum wire counter electrode. EIS measurements were performed using a Gamry Potentiostat (model: PCI4/750) coupled with a Gamry EIS300

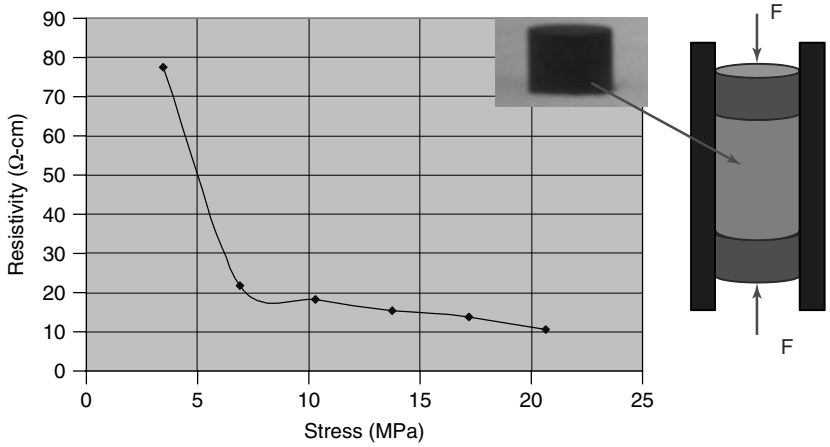


Figure 10. CNF and nanocement sample showing resistivity versus stress, the sample (inset) and schematic of pressure casting CNF in cement to improve the strength/electrical conductivity.

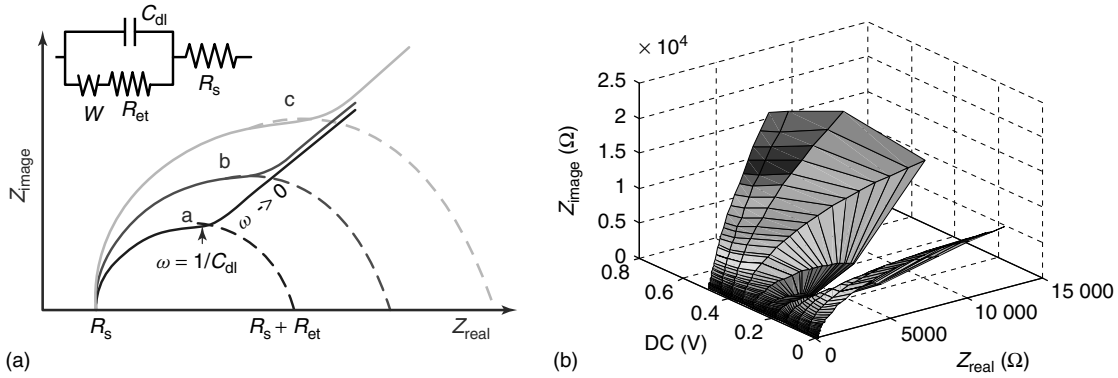


Figure 11. Electrochemical impedance spectroscopy to detect corrosion and cracking: (a) EIS model of an electrode and the response; (b) 3D EIS of a gold electrode in phosphate buffer saline with 5.0 mM K3Fe(CN)6 and 5 mM of K4Fe(CN)6. The effect of dc potential is shown.

software. Figure 11 shows the EIS response. Note that at the dc potential of 0.2 V the redox reaction greatly lowers the impedance. The EIS response can be modeled using the Randal Warburg equation $Z(\omega) = R_s + [R_{et}/(1 + \omega^2 R_{et}^2 C_{dl}^2)] - [j\omega R_{et}^2 C_{dl}/(1 + \omega^2 R_{et}^2 C_{dl}^2)]$, where $R_{et}, C_{dl}, R_s, \omega$ are the electron transfer resistance, double layer capacitance, solution resistance, and frequency, respectively. The EIS method can be adjusted to detect chemical degradation such as from corrosion, UV deterioration, hygrothermal degradation, and resistive degradation because of cracking and delamination of electrically conductive composites. A solid polymer electrolyte can also be used to further tailor the EIS signature of a composite. EIS monitoring is possibly a simpler approach than propagating waves to detect damage in large complex structures. Electrically conductive polymer, elastomer, or cement nanocomposites are multifunctional materials whose electrochemical properties change with damage or corrosion. The use of EIS is a new approach that takes advantage of the electrical conductivity properties of composites to simplify SHM.

6 NANOTUBE THREAD WITH BUILT-IN MULTIFUNCTIONALITY

Another approach for developing a multifunctional sensor material is to develop an intermediate product that can be used as a sensor material and as a fiber reinforcing material in composites. The proposed material is CNT thread. The thread is made from spinning long CNT. Long CNT are produced on wafers as shown in Figure 12. Properties of various fibers for comparison to CNT are given in [37] and show that SWCNT have higher strength, and stiffness, and are lighter than any other fiber. If a thread could be made that has similar properties to SWCNT, many new structural and electronic applications would open up. The mechanical properties of fiber nanocomposites [37] show that a polymer nanocomposite may have significantly better properties than conventional composite materials. Spinning long nanotubes into threads for reinforcement, self-sensing, and self-repair may produce a variety of versatile new smart materials.

SHM, data mining, and structural materials development have traditionally been separate enterprises, but this is changing. These disciplines are coming together to meet the needs of developing advanced integrated vehicle health monitoring systems for aerospace vehicles and other applications from nanomedicine to mars exploration. The logical way to attack this need is through the bottom-up design of a structural material that has multifunctional properties. Properties that are important for this material include high strength, high stiffness, light weight, electrical conductivity, self-sensing for damage, limited self-repair of damage, and a sensor architecture that enables data mining and other special properties depending on the application. These properties can be achieved by building a nanocomposite material based

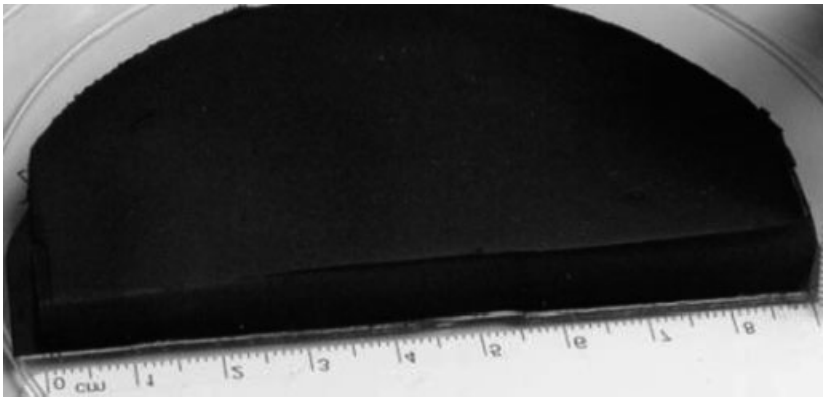


Figure 12. Carbon nanotubes grown up to 1-cm long on a silicon wafer. The CNT are used for spinning thread. (CNT grown by First Nano Inc. using UC substrate.)

on nanotube thread. Long nanotubes and different types of nanoparticles can be blended to spin a thread. In this article, we have discussed blending short nanoparticles in a polymer. The polymer and thread may also be combined to form a nanocomposite with unique properties, such as high electrical conductivity and high strength. The blended materials approach makes all the capabilities of nanotechnology virtually available for developing enabling new materials for advanced applications. In particular, nanotube thread can bring multifunctionality to composite materials.

Designing smart materials from the bottom up can provide multifunctional material properties and practical application for SHM. However, the sensing capability should be designed into the material. In other words, all the properties of the material must be designed together. Threads are necessary to provide bulk reinforcement because CNT cannot be grown beyond approximate centimeter length, currently. Thus, centimeter-long nanotubes must be spun into thread to provide a strong bulk material. The thread can be made completely of nanotubes or it can be blended using two or more materials. For example, to provide intermediate properties and reasonable cost, long nanotubes could be blended with polyester to form the thread. In another approach, the CNT might be blended with small percentages of CNSC, or a variety of other nanoparticles, to give the desired properties of the thread. It is also possible that the CNT could be functionalized to modify their adhesion or other properties before spinning thread. The goal is to provide a strong thread with smart material properties that cannot be achieved by any other material system on earth. Multiple threads will be woven together to form a fiber. The fibers can be used to form unidirectional prepreg plies or woven into a fabric to provide two-directional properties. EIS can be used to monitor the electrical properties of the thread for damage. A smart fabric design is also possible using the thread. Smart fabric can be made by weaving the nanotube thread into cloth. Smart fabric should have several interesting applications including reinforcing fabric in composites, as tough materials for garments for soldiers, firefighters, and first-responders, as electromagnetic radiation shielding materials, and other applications. Mixing the different types of nanotube or nanoparticles and in some cases conventional thread provides a large design space to build in properties so that the material will do what we want it to.

Spinning nanoscale thread is a new area of research that is expected to become very important in the future. In general, fibers with small diameter generally have better mechanical properties. Most commercial fibers are in the micron diameter range due to economic considerations. This suggests that small diameter nanotube thread could have good mechanical properties and might replace microfibers. This idea has produced considerable interest in spinning nanotubes into thread and multistrand fibers using different approaches. Electrospinning is a method to produce continuous nanofibers from polymer solutions in high electric fields [44]. A thin polymer jet is ejected when the electric force on induced charges on the polymer liquid overcomes the surface

tension. The charged jet is elongated and accelerated by the electric field, dries, and is deposited on a substrate as a random nanofiber mat. Over a hundred synthetic and natural polymers were electrospun into fibers with diameters ranging from a few nanometers to micrometers. The resulting nanofiber samples are often uniform and do not require expensive purification, and the electrospun nanofibers are continuous. Electrospinning has the potential for low-cost electromechanical control of fiber placement and integrated manufacturing of two- and three-dimensional nanofiber assemblies.

The assembly of CNT into continuous fibers has been achieved mostly through postprocessing methods. Fibers of nanotubes or nanotube-polymer blends have been drawn or spun from solutions or gels. A thread of nanotubes can be dry-drawn from an aligned assembly on a silicon substrate as a result of van der Waals interactions [17]. However, direct spinning of CNT is also being done. Nanotube thread has been formed after the pyrolysis of hexane, ferrocene, and thiophene directly in a furnace [45]. By mechanically drawing CNT directly from the gaseous reaction zone, continuous fibers were wound without an apparent limit to the length. Continuous spinning requires rapid production of high-purity nanotubes to form an aerogel in the furnace hot zone and forcible removal of the product from the reaction by continuous wind-up. Ferrocene and thiophene are used in the process. The overall question of spinning is discussed next.

6.1 To spin or not to spin

Rarely do we find noncompromising solutions to materials problems. The impasse with nanotechnology is bringing the properties of nanoscale materials to the macroscale. A key question is, how can we use CNT in the real world of airplanes and other structures for sensing and reinforcement? Spinning nanotubes into thread is one way to bring the properties of short fibers or nanotubes to the macroscale, but spinning is a compromise solution because the properties of thread are generally below the properties of the fiber or nanotube. However, thread also has advantages of energy absorption, tailorable stiffness, multicomponent material, and others as compared to nanotubes alone. When spinning, the longer the nanotube is, the smaller the thread angle becomes and possibly the greater the properties of the thread. Another compromise solution is linking nanotubes by using a binder material or by welding nanotubes to each other. There is no binder material that has mechanical properties in the order of the properties of nanotubes and unless the load path is continuous through nanotubes, the thread properties will be low. Welding nanotubes to tungsten has been done but welding nanotubes to nanotubes is difficult, based on initial investigations. There appears to be some potential to weld nanotubes to each other using a welding filler material and this technique is under investigation.

Not spinning would be the ideal solution because the full properties of nanotubes would be available at the macroscale. The problem with not spinning becomes one of synthesizing nanotubes to meter lengths. In theory, we believe that synthesis of continuous nanotubes is possible, but in practice the catalyst turns off after the nanotubes are in the centimeter-length range. Also, the quality of the nanotubes decreases with length. Considerable effort is being expended to grow the longest nanotubes possible [46, 47]. Lengths beyond the centimeter range and high quality are expected. Recently, forming nanotubes into threads or films is being done by several industries including Industrial Nano, CNT Technologies, Nanocomp, General Nano, and universities including Cambridge University, UT Dallas, and the University of Cincinnati for composites applications and possibly for use in a space elevator [48]. Spinning for simultaneous reinforcement, self-sensing, and limited self-repair may produce a versatile new smart material. Initial blend thread produced at the University of Cincinnati is shown in Figure 13. Properties of the initial thread are being evaluated. Without post processing, electrical resistivity is about $0.02\,\Omega\,\text{cm}$. This thread has promise for applications in hundreds of commercial products.

7 A STRUCTURAL NEURAL SYSTEM USING CARBON NANOTUBES

In the case of SHM of aerospace systems, many sensors are typically needed to provide sensitivity to small damage on large structures, and different

Figure 13. Long CNT called *Black Cotton* spun into thread may provide reinforcement, sensing, and partial self-repair simultaneously. Figures show spinning of thread from long CNT arrays. Two or three threads can be rolled into one yarn ~0.2-mm diameter. A spinning machine is used to spin the thread based on the approach developed by Zhang *et al.* [17].

types of sensors are needed to monitor different types of components. Also, aerospace and other systems are becoming more and more complex. This means that onboard sensors and signal processing instrumentation are needed to simultaneously monitor the integrity of their multiple subsystems that contain structural, mechanical, and electrical components. Besides integrity monitoring, large sensor networks are also being used in qualification testing, monitoring unmanned vehicles, and for feedback control and damage mitigation of systems. In all these applications, the integration and mining of increasing amounts of data from many sensors and systems is becoming impractical based on size, weight, reliability, and cost of the monitoring and support systems. A new approach to sensing is presented next wherein only anomalous responses are obtained from sensors as opposed to longtime histories of sensor raw data.

7.1 The structural neural system

An SNS was developed for anomalous event detection and to significantly reduce the complexity and cost of SHM and data mining by selectively processing signals at the sensor level. The SNS is formed using continuous sensors and a biomimetic signal processing architecture. A continuous sensor is a one-circuit connection of multiple sensor nodes that has one output signal that is a combination of signals from the individual sensor nodes. An anomalous response from any sensor node is detected within the combined output signal from all the sensors by filtering and thresholding. Thus, a continuous sensor has only one output signal and can replace many individual sensor nodes and still detect abnormal events. A trade-off of the continuous sensor is that the exact time response of each sensor node is not available. Only hallmark events are captured by the SNS based on the analog electronics that are used to define an anomalous event. The SNS can be customized by combining many continuous sensors into a network that is designed on the basis of the particular configuration (geometry and materials) of the component. The SNS has a biomimetic architecture that processes signals like the biological neural system—in a highly distributed massively parallel fashion. This means that many continuous sensors (which are akin to biological neurons) can operate all the time, but the neural system processes only signals that contain unusual events. This effectively reduces the burden of data mining because only anomalous events are passed to the data logging system by the SNS. In a concept future digital version of the SNS, tens of continuous sensors each with potentially 10 or more nodes (hundreds of sensors in total) can be monitored using a grid pattern and only four channels of analog to digital signal conversion. An SNS using nanotube spray-on film neurons would detect cracking and corrosion by changes in the electrochemical impedance of the neurons, thus locating and quantifying the size of the damage. More details of the SNS and testing using piezoelectric ceramic sensors are given in **Wind Turbines**, and also in [49].

7.2 Multistate continuous sensors

The problem of monitoring different types of components and damage may require that several types of sensors be used in the SNS. Therefore, a recent

advance in the capability of the SNS is the development of multistate continuous sensors with different types of sensor nodes. Continuous sensors can use almost any type of sensor node, but within each continuous sensor the sensor nodes must be the same type, because the analog electronics that control the output are matched to the specific sensor type. This means that the output of the sensor network will provide the same information on the location of damage and characteristics of the aberrant waveform, no matter which sensor types are used. By knowing which neuron is "firing", the waveform can be decoded into the appropriate sensor information in the computer. This allows one generic sensor network to be used for SHM, regardless of the type of sensor and system being monitored. Almost any type of sensor node such as accelerometers and strain gauges can be used to form a continuous sensor, but the interface electronics are usually specific for each sensor type. Several types of nanotube type materials can be integrated within a composite material and used as continuous sensors.

7.3 Applications of CNT thread

In many applications, the sensor material can be integrated into composite materials and provide multifunctional properties such as reinforcement, thermal conduction, and damage monitoring. In the aerospace and defense sectors, there are hundreds of likely components that could benefit from simultaneous structural improvement and monitoring using CNT thread. Use of the SNS with CNT neurons has potential for damage detection on large structures such as aircraft, building, bridges, and health monitoring of a cable for the space elevator. Leaving the Planet by Space Elevator [50] is a grand challenge problem for space exploration systems, SHM, and data mining systems. Organizations supporting the space elevator concept include NASA, Black Line Ascension, and the Spaceward Foundation. CNT thread may form a ribbon for the space elevator and EIS might be used to evaluate the condition of the ribbon. A video camera on the elevator will also monitor the ribbon. Data mining from impedance spectra and video will be needed. Long CNTs called *Black Cotton*™ are being spun into thread, which may provide reinforcement, sensing, and partial self-repair simultaneously. Potential applications of CNT are quite open. In complex composite materials, CNT thread can be put where we want it inside composite materials in sections of changing thickness and curvature, joints, around holes, in bond lines, and anywhere health monitoring is required. The thread is mostly inert and will be nanometer or micron thick and will not affect structural integrity or weight. Sensor thread can be put inside all types of structures where other sensors are impractical, such as in flexbeams and rotor blades of helicopters, in composite pressure vessels, in cement, on the surface of bridges, aircraft, space vehicles, and in elastomers.

8 FUTURE EMBEDDED WIRELESS NANOSENSORS

There are many compelling reasons for wanting to embed small sensors or devices inside composite materials. Thus far, technologically, industry has been able to develop radio frequency identification (RFID) tags that can identify a product or object, or micromotes that are battery powdered and report some physical variable. However, these devices are just the beginning. Nanoscale materials are opening up the possibility to build revolutionary devices and tiny sensors that can go inside materials (and the human body) and do what we want them to. This section presents initial analyses toward building an active microsensor that can go inside structural materials and detect damage early. Active microsensors perform electrochemical measurements and produce an electronic signal that is used to detect cracking, delamination, or electrolytes due to corrosion or water ingestion in structural materials. The active microsensor also has a feedback mechanism to increase its sensitivity or to activate a control function. The active microsensor will be built using larger electronic components first, and finally pushed down in size using nanoscale materials. Building small sensors and developing a way to communicate with them is an extraordinary difficult technological problem. To attack this problem, nanoscale electronic components (nanotubes, NWs, capacitors, inductors, solenoids, antennas, transistors, and actuators) and a four-arm nanomanipulator operating under an environmental scanning electron microscope can be used to assemble the components into prototype devices.

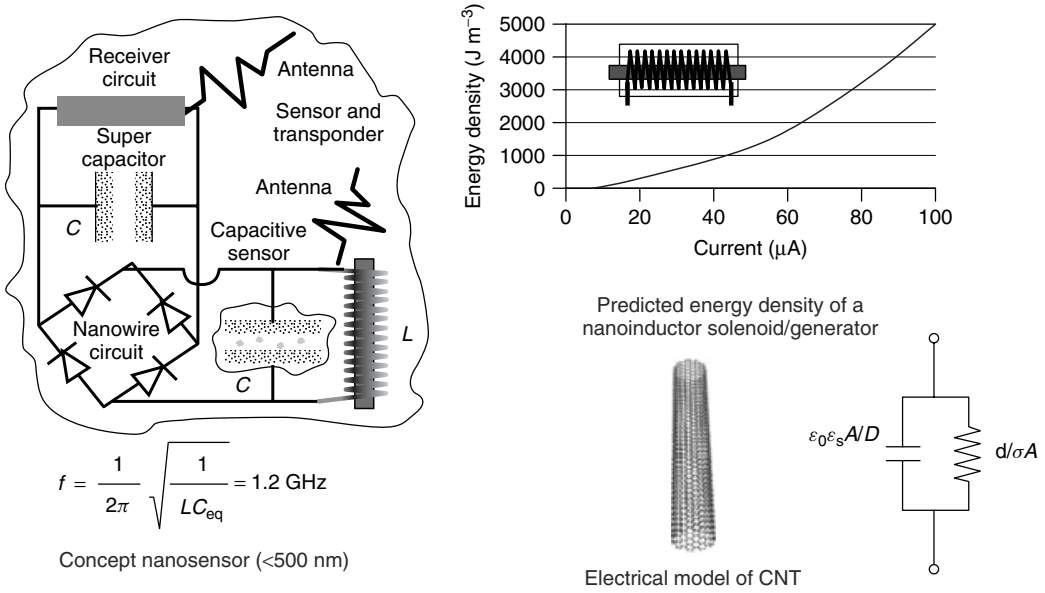


Figure 14. Initial design of a sensor transponder for asset evaluation and condition monitoring of composite structures and later for in-body detection of disease. The sensor works by receiving an RF signal, storing charge, and discharging into a transmitter circuit whose resonant frequency depends on the supercapacitance of nanotube sensors in an electrolyte.

As an example of a novel device, a transponder sensor is proposed (Figure 14) to detect damage in composites.

Developing *in situ* sensors holds great promise for making revolutionary advances in self-sensing nanocomposite materials. The circuit was analyzed using standard electronics circuit and component equations with nanoscale components. Refinement of the model is needed and details of the antenna and transistor circuit are under consideration. Once the sensor design is verified, a method to mass produce the sensors by chemical vapor deposition synthesis will be investigated.

9 SUMMARY AND CONCLUSIONS

This article introduced nanoengineering of sensor materials and discussed their potential for applications in the near future. A major advantage of sensors based on nanomaterials is that several nanomaterials can be blended to fabricate a sensor material that can do what we want it to. The blending can be in a polymer to form a polymer-based sensor material or nanoparticles can be blended in the intermediate form of a thread. The polymer and thread can be used together in composite materials to form self-sensing materials. Prospects for developing new sensors that have high sensitivity are wide open, making nanoengineering of sensors from the bottom up an exciting new field. Engineering asset evaluation and condition monitoring using nanomaterials may have hundreds of applications in components where smart nanocomposites can be used for multifunctional properties improvement, self-sensing of damage, and limited self-repair. In particular, long CNT arrays with further optimization are expected to be an enabling material for structural, electrical, sensing, and thermal applications. "Nanoizing" materials and structures is a new technological science that SHM and NDE engineers should pay attention to.

ACKNOWLEDGMENTS

This work was sponsored by Clean Technologies International Corp., North Carolina A&T SU through the ONR, the Institute for Nanoscale Science and Technology at UC, NSF grant CMS-0510823, and the Korea Institute of Industrial Technology.

Development of the SNS was supported by the National Renewable Energy Laboratory under subcontract number XCX-2-31214-01. Alan Laxson is the technical monitor of the project. Henry Westheider and Douglas Hurd built the fixtures for the smart materials casting and testing. Tom Baca, John Hurtado, and Todd Simmermacher of Sandia National Laboratories sponsored early work on health monitoring. Bradley Edwards of Industrial Nano provided suggestions about SHM of the space elevator ribbon. All this support is gratefully acknowledged.

REFERENCES

[1] Gouma P, Sberveglieri G. *Novel Materials and Applications of Electronic Noses and Tongues*, 2004, http://www.mrs.org/publications/bulletin, MRS BULLETIN/OCTOBER.

[2] Katz E, Willner I. Integrated nanoparticle–biomolecule hybrid systems: synthesis, properties, and applications. *Angewandte Chemie International Edition* 2004 **43**:6042–6108.

[3] Baughman RH, *et al.* Carbon nanotube actuators. *Science* 1999 **284**:1340–1344.

[4] Peng S, O'Keeffe J, Wei C, Cho K, Kong J, Chen R. Carbon nanotube chemical and mechanical sensors. *3rd International Workshop on Structural Health Monitoring*. Stanford, CA, 15–17 September 2003.

[5] Wood JR, Wagner HD. Single-wall carbon nanotubes as molecular pressure sensors. *Applied Physics Letters* 2000 **76**(20):2883–2885.

[6] Kong J, Frankin NR, Zhou C, Chapline MG, Peng S, Cho K, Dai H. Nanotube molecular wires as chemical sensors. *Science* 2000 **287**:622–625.

[7] Shanov VN, Schulz MJ. *Nanoworld and Smart Materials and Devices Laboratories*. University of Cincinnati, 2008, http://altmine.mie.uc.edu/mschulz/public_html/smartlab/smartlab.html.

[8] Kang I, Jung JY, Choi GR, Park H, Lee JW, Yoon KW, Yun Y, Shanov V, Schulz MJ. Developing carbon nanocomposite smart materials. *Solid State Phenomena* 2007 **119**:207–210.

[9] Tombler TW, Zhou C, Alexseyev L, Kong J, Dal H, Liu L, Jayanthl CS, Tang M, Wu SY. Reversible electromechanical characteristics of carbon nanotubes under local-probe manipulation. *Nature* 2000 **405**:769–772.

[10] Watkins AN, Ingram JL, Jordan JD, Wincheski RA, Smits JM, Williams PA. Single wall carbon nanotube-based structural health monitoring sensing materials. *NSTI Conference*. Nanotech, 2004; Vol. 3.

[11] Kang I, Schulz MJ, Kim JH, Shanov V, Shi D. A carbon nanotube strain sensor for structural health monitoring. *Smart Materials and Structures* 2006 **15**(3):737–748.

[12] Jalili N, Goswami BC, Dawson DM. Distributed sensors and actuators via electronic-textiles. *National Textile Center Research Briefs—Materials Competency*, NTC Project: M04-cL05. National Textile Center, June 2005.

[13] Abbott D, *et al. Development and Evaluation of Sensor Concepts for Ageless Aerospace Vehicles: Development of Concepts for an Intelligent Sensing System*, NASA/CR-2002-211773. NASA Langley Research Center, 2002.

[14] Ebron VH, *et al.* Fuel powered artificial muscles. *Science* 2006 **311**(5767):1580–1583.

[15] Zhang M, Shaoli F, Zakhidov AA, Lee SB, Aliev AE, Williams CD, Atkinson KR, Baughman RH. Strong, transparent, multifunctional, carbon nanotube sheets. *Science* 2005 **309**(5738):1215–1219.

[16] Baughman RH. Materials science. Playing nature's game with artificial muscles. *Science* 2005 **308**(5718):63–65.

[17] Zhang M, Atkinson KR, Baughman RH. Multifunctional carbon nanotube yarns by Downsizing an ancient technology. *Science* 2004 **306**(5700): 1358–1361.

[18] Baughman RH. Materials science. Muscles made from metal. *Science* 2003 **300**(5617):268–269.

[19] Tsung-Chin H, Loh KJ, Lynch JP. Electrical impedance tomography of carbon nanotube composite materials. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, *Proceedings of the SPIE*, Masayoshi T, Chung-Bang Y, Giurgiutiu V (eds). SPIE, 2007; Vol. 6529, pp. 652926.

[20] Meyyappan M. *Carbon Nanotubes: Science and Applications*. CRC Press: Boca Raton, FL, 2005.

[21] Soyoun J, Ji T, Xie J, Jining J, Varadan VK. *A Novel Strain Sensor using Carbon Nanotubes-Organic Semiconductor Matrix Composite on Polymeric Substrates; Smart Structures, Devices, and Systems III*, *Proceedings of the SPIE*, Al-Sarawi SF (ed). SPIE, 2007; Vol. 6414.

[22] Jin Y, Yuan FG. Simulation of elastic properties of single-walled carbon nanotubes. *Composites Science and Technology* 2003 **63**:1507–1515.

[23] Arun R, Nader J, Himanshu R. Development of a novel strain sensor using nanotube-based materials with applications to structural vibration control. In *Sixth International Conference on Vibration Measurements by Laser Techniques: Advances and Applications*, *Proceedings of the SPIE*, Tomasini EP (ed). SPIE, 2004; Vol. 5503, pp. 478–485.

[24] Getty S. *SWCNT Nanocompass for Next-Generation Magnetometry. Presented at SAMPE*. Materials Engineering Branch, NASA Goddard Space Flight Center, June 2007.

[25] Lia GY, Wang PM, Zhao X. Mechanical behavior and microstructure of cement composites incorporating surface-treated multi-walled carbon nanotubes. *Carbon* 2005 **43**(6):1239–1245.

[26] Makar J, Margeson J, Luh J. Carbon nanotube/cement composites—early results and potential applications. NRCC-47643. *3rd International Conference on Construction Materials: Performance, Innovations and Structural Implications*. *Vancouver*, 22–24 August 2005; pp. 1–10.

[27] Xiao X, Elam JW, Trasobares S, Auciello O, Carlisle JA. Synthesis of a self-assembled hybrid of ultrananocrystalline diamond and carbon nanotubes. *Advanced Materials* 2005 **17**(12):1451–1565.

[28] Schulz M. A structural neural system for data mining and anomaly detection. *Data Mining in Aeronautics, Sciences, and Exploration Systems Conference (DMASES)*. Computer History Museum. Mountain View, CA, 26–27 June 2007, http://ase.arc.nasa.gov/projects/dmases/2007/.

[29] Schulz M, Kirikera G, Yun Y, Shanov V, Mullapudi S, Maheshwari G, Allemang R. A structural neural system with multi-state sensors for integrated systems health management. *Proceedings of The 2nd World Congress on Engineering Asset Management (EAM) and The 4th International Conference on Condition Monitoring*. The Cairn Hotel, Harrogate, UK, 11–14 June 2007; pp 1739–1750.

[30] Ghosh S, Sood AK, Kumar N. Carbon nanotube flow sensors. *Science* 2003 **299**:1042–1044.

[31] The Easy Tube System, First Nano. *Carbon Nanotube Synthesis*, Ronkonkoma, NY, http://www.firstnano.com. 2007.

[32] Clean Technologies International Corporation, 2008, http://www.cleantechnano.com/CleanTechNano/.

[33] Applied Sciences, Inc. and Pyrograf Products, Inc., Cedarville, OH, 2008, www.apsci.com.

[34] Carbon Nanotechnologies, Inc., 2008, http://cnanotech.com.

[35] Nanolab, Inc., 2008, info@nano-lab.com.

[36] Shanov V, Gorton A, Yun Y, Schulz M. *Catalyst and Method for Manufacturing Carbon Nanostructured Materials*, Invention disclosure: UC 107-044 (patent pending), 17 October 2006.

[37] Schulz MJ, Kelkar A, Sundaresan M. *Nanoengineering of Structural, Functional and Smart Materials*. CRC Press, 2006.

[38] Andrews R, Weisenberger M. *Carbon Nanotube Polymer Composites: A Review of Recent Developments*. University of Kentucky Center for Applied Energy Research, 2004, www.isr.us/Spaceelevator-conference/.

[39] Kirikera GR. *An Artificial Neural System for Structural Health Monitoring*, MS thesis, University of Cincinnati, 2003.

[40] Kang I, *et al.* Introduction to carbon nanotube and nanofiber smart materials. *Composites Part B: Engineering* 2006 **37**(6):382–394.

[41] Koo J. *Polymer Nanocomposites: Processing, Characterization, and Applications*, *First Edition*, McGraw-Hill, April 18 2006.

[42] Shanov VN, Choi G, Maheshwari G, Seth G, Chopra S, Li G, Yun Y, Abot J, Schulz MJ. Structural nanoskin based on carbon nanosphere chains. *SPIE Smart Structures Conference*. San Diego CA, March 2007.

[43] Bentley AK, Farhound M, Ellis AB, Lisensky GC, Nickel A-M, Crone WC. Template synthesis and magnetic manipulation of Nickel nanowires, *Journal of Chemical Education* 2005 **82**:765–768.

[44] Spivak AF, Dzenis YA, Reneker DH. A model of steady state jet in the electrospinning process, *Mechanics Research Communications* 2000 **27**(1): 37–42.

[45] Zeng LX, *et al.* Ultralong single-wall carbon nanotubes. *Nature Materials* 2004 **3**:673–676.

[46] Press release by NSF on long nanotube growth, 2007, http://www.nsf.gov/news/news_summ.jsp?cntn_id=108992&org=NSF&from=news.

[47] *Press Release on Long Nanotube Growth. University of Cincinnati Researchers Grow Their Longest Carbon Nanotube Ever*, 2007, http://www.uc.edu/news/NR.asp?id=4811.

[48] Black Line Ascension, Inc., 2007, www.blacklineascension.com.

[49] Kirikera GR, Shinde V, Schulz MJ, Ghoshal A, Sundaresan MJ, Allemang RJ, Lee JW. A structural neural system for real-time health monitoring of composite materials, *Structural Health Monitoring: An International Journal* 2008 **7**: 65–83.

[50] Ragan P, Edwards B. *Leaving the Planet by Space Elevator*. Lulu.com, 2006.

Chapter 52
Piezoelectricity Principles and Materials

Victor Giurgiutiu
Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA

1 INTRODUCTION

Piezoelectricity (discovered in 1880 by Jacques and Pierre Curie) describes the phenomenon of generating an electric field when the material is subjected to a mechanical stress (direct effect), or, conversely, generating a mechanical strain in response to an applied electric field. The *direct piezoelectric effect* predicts how much electric field is generated by a given mechanical stress. This *sensing effect* is utilized in receiving piezoelectric sensors. The *converse piezoelectric effect* predicts how much mechanical strain is generated by a given electric field. This *actuation effect* is utilized in transmitting piezoelectric sensors. Piezoelectric active sensors act as both transmitters and receivers of elastic waves.

Piezoelectric properties occur naturally in some crystalline materials, e.g., quartz crystals (SiO_2) and Rochelle salt. The latter is a natural ferroelectric material, possessing an orientable domain structure that aligns under an external electric field and thus enhances its piezoelectric response. Piezoelectric response can also be induced by electrical poling in certain polycrystalline materials, such as piezoceramics. In recent years, affordable high-performance piezoceramics have become commercially available at affordable prices.

Piezoelectric materials are some of the major building blocks of the structural health monitoring (SHM) sensors. The intrinsic active behavior of these materials, which change dimensions in response to electric fields, and produce electricity in response to mechanical strain and stress, makes them ideal for actuation and sensing as required in SHM applications.

2 BASIC EQUATIONS

2.1 Piezoelectric equations

Electroactive materials can be distinguished into piezoelectric and electrostrictive. *Linear piezoelectric materials* obey the constitutive relations between the mechanical and electrical variables, which can be

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

written in tensor notations as

$$S_{ij} = s^E_{ijkl} T_{kl} + d_{ijk} E_k + \delta_{ij} \alpha^E_i \theta \quad (1)$$

$$D_j = d_{jkl} T_{kl} + \varepsilon^T_{jk} E_k + \tilde{D}_j \theta \quad (2)$$

where S_{ij} is the mechanical strain, T_{kl}, mechanical stress, E_k, electrical field, and D_j, electrical displacement. The variable s^E_{ijkl} is the mechanical compliance of the material measured at zero electric field ($E = 0$), ε^T_{jk} is the dielectric permittivity measured at zero mechanical stress ($T = 0$), and d_{kij} is the piezoelectric coupling between the electrical and mechanical variables. The variable θ is the temperature, and α^E_i is the coefficient of thermal expansion under constant electric field. The coefficient $\tilde{D}_j$ is the coefficient that connects electric displacement with temperature. The stress and strain variables are second-order tensors, while the electric field and the electric displacement are first-order tensors. Since thermal effects only influence the diagonal terms, the respective coefficients, α_i and $\tilde{D}_j$, have single subscripts. The term δ_{ij} is the Kroneker delta ($\delta_{ij} = 1$ if $i = j$; zero otherwise).

Compressed matrix notations (Voigt notations) are often used in engineering practice in lieu of the tensorial equations (1) and (2). The stress and strain tensors are arranged as six-component vectors, with the first three components representing *direct* stress and strain, and the last three representing *shear* stress and strain. When written in compact form, equations (1) and (2) become

$$S_p = s^E_{pq} T_q + d_{kp} E_k + \delta_{pq} \alpha^E_q \theta, \quad p, q = 1, \ldots, 6 \quad k = 1, 2, 3 \quad (3)$$

$$D_i = d_{iq} T_q + \varepsilon^T_{ik} E_k + \tilde{D}_i \theta, \quad q = 1, \ldots, 6 \quad i, k = 1, 2, 3 \quad (4)$$

Please note that the piezoelectric matrix in equation (3) is the transpose of the piezoelectric matrix in equation (4). Equations (3) and (4) can also be written in matrix format, i.e.,

$$\{\mathbf{S}\} = [\mathbf{s}]\{\mathbf{T}\} + [\mathbf{d}]^t\{\mathbf{E}\} + \{\boldsymbol{\alpha}\}\theta \quad (5)$$

$$\{\mathbf{D}\} = [\mathbf{d}]\{\mathbf{T}\} + [\boldsymbol{\varepsilon}]\{\mathbf{E}\} + \{\tilde{\mathbf{D}}\}\theta \quad (6)$$

Electromechanical coupling coefficient is the square root of the ratio between the mechanical energy stored and the electrical energy applied to a piezoelectric material, i.e.,

$$k^2 = \frac{\text{Mechanical energy stored}}{\text{Electrical energy applied}} \quad (7)$$

For direct actuation, $k^2_{33} = d^2_{33}/s_{33}\varepsilon_{33}$, where Voigt matrix notations are used. For transverse actuation, $k^2_{31} = d^2_{31}/s_{11}\varepsilon_{33}$; and for shear actuation, $k^2_{15} = d^2_{15}/s_{55}\varepsilon_{11}$. For uniform in-plane actuation, one uses the planar coupling coefficient, $\kappa_p = \kappa_{13}\sqrt{2/(1-\nu)}$, where ν is the Poisson ratio. The piezoelectric response in axial and shear directions is depicted in Figure 1.

2.2 Electrostrictive equations

In contrast to linear piezoelectricity, the *electrostrictive response* is quadratic in electric field. Hence, the direction of electrostriction does not change as the polarity of the electric field is reversed. The general constitutive equations incorporate both piezoelectric and electrostrictive terms as follows:

$$S_{ij} = s^E_{klij} T_{kl} + d_{kij} E_k + M_{klij} E_k E_l \quad (8)$$

$$D_m = d_{mkl} T_{kl} + \varepsilon^T_{mn} E_n + 2M_{mnij} E_n T_{ij} \quad (9)$$

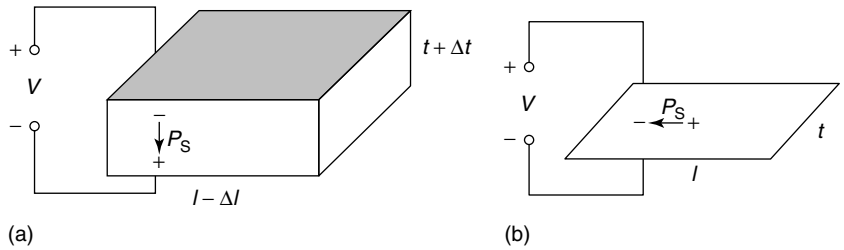


Figure 1. Induced-strain responses of piezoelectric materials: (a) axial strain and (b) shear strain.

Note that the first two terms in each equation are similar to the linear piezoelectric behavior. However, in electrostrictive materials, the linear piezoelectric response is much weaker than in piezoelectric materials. The third term represents the strong nonlinear electrostrictive behavior. The coefficients M_{klij} are the electrostrictive coefficients.

2.3 Magnetostrictive equations

Most magnetoactive materials are based on the magnetostrictive effect. The *magnetostrictive constitutive equations* contain both linear and quadratic terms as follows:

$$S_{ij} = s^{E}_{ijkl} T_{kl} + d_{kij} H_k + M_{klij} H_k H_l \qquad (10)$$

$$B_m = d_{mkl} T_{kl} + \mu^{T}_{mk} H_k + 2 M_{mnij} E_n T_{ij} \qquad (11)$$

where, in addition to the already defined variables, H_k is the magnetic field intensity, B_j is the magnetic flux density, and μ^{T}_{jk} is the magnetic permeability under constant stress. The coefficients d_{kij} and M_{klij} are defined in terms of magnetic units. The magnetic field intensity in a rod surrounded by a coil with n turns per unit length depends on the coil current, I, i.e.,

$$H = nI \qquad (12)$$

Magnetostrictive material response is quadratic in the magnetic field, i.e., the magnetostrictive response does not change sign when the magnetic field is reversed. However, the nonlinear magnetostrictive behavior can be linearized about an operating point through the application of a bias magnetic field. In this case, *pseudo-piezomagnetic* behavior, in which response reversal accompanies field reversal, can be obtained. In Voigt matrix notations, the equations of linear piezomagnetism are as follows:

$$S_i = s^{H}_{ij} T_j + d_{ki} H_k, \quad i, j = 1, \ldots, 6; \quad k = 1, 2, 3 \qquad (13)$$

$$B_m = d_{mj} T_j + \mu^{T}_{mk} H_k, \quad j = 1, \ldots, 6; \quad k, m = 1, 2, 3 \qquad (14)$$

where, S_i is the mechanical strain, T_j is the mechanical stress, H_k is the magnetic field intensity, B_m is the magnetic flux density, and μ^{T}_{mk} is the magnetic permeability under constant stress. The coefficient s^{H}_{ij} is the mechanical compliance measured at zero magnetic field ($M = 0$). The coefficient μ^{T}_{mk} is the magnetic permeability measured at zero mechanical stress ($T = 0$). The coefficient d_{ki} is the *piezomagnetic constant*, which couples the magnetic and mechanical variables and expresses how much strain is obtained per unit applied magnetic field.

3 FERROELECTRIC PEROVSKITES

The popular piezoceramic materials owe their piezoelectric behavior to the ferroelectric phenomenon observed in a class of crystalline materials called *ferroelectric perovskites*.

3.1 Ferroelectricity

Ferroelectricity is the property of having permanent electric polarization, and of being able to alter it by the application of an external electric field. The term *ferroelectricity* was derived by analogy with the term *ferromagnetism*, which describes how permanent magnetization is altered by the application of an external magnetic field. As the electric field applied to a ferroelectric material is increased beyond a critical value called *coercive field*, E_c, the polarization suddenly increases to a high value. This value is more or less maintained when the electric field is decreased, such that at zero electric field the ferroelectric material retains a permanent spontaneous polarization P_S. When a negative electric field is applied beyond the negative value $-E_c$, the polarization suddenly switches to a large negative value, which is roughly maintained as the electric field is decreased. At zero electric field, the permanent spontaneous polarization is $-P_S$. As the electric field is again increased into the positive range, the polarization is again switched to a positive value, as the field increases beyond E_c. Characteristic of this behavior is the high hysteresis loop traveled during a cycle. The ferroelectric behavior can be explained through the existence of aligned internal dipoles that have their direction switched when the electric field is sufficiently strong.

3.2 Perovskite structure

Perovskites are a large family of crystalline oxides with the metal-to-oxygen ratio 2 : 3. Perovskites derive their name from a specific mineral, perovskite, first described in 1839 by the geologist Gustav Rose, who named it after the famous Russian mineralogist Count Lev Aleksevich von Perovski (1792–1856). The simplest perovskite lattice has the expression, $X_m Y_n$, in which the X atoms are rectangular close-packed and the Y atoms occupy the octahedral interstices. The rectangular close-packed X atoms may be a combination of various species, X^1, X^2, X^3, etc. For example, in the barium titanate perovskite, $BaTiO_3$, we have $X^1 = Ba^{2+}$ and $X^2 = Ti^{4+}$, while $Y = O^{2-}$ (Figure 2). In the lattice structure, the Ba^{2+} divalent cations are at the corners, the Ti^{4+} tetravalent cation is in the center, while the O^{2-} anions are on the faces. The Ba^{2+} cations are larger, while the Ti^{4+} cations are smaller. The size of the Ba^{2+} cation affects the overall size of the lattice structure. Perovskite arrangements like in $BaTiO_3$ are generically designated ABO_3. Their commonality is that they have a small, tetravalent metal ion, e.g., titanium or zirconium, in a lattice of larger, divalent metal ions, e.g., lead or barium and oxygen ions (Figure 2). Under conditions that confer tetragonal or rhombohedral symmetry, each crystal has a dipole moment.

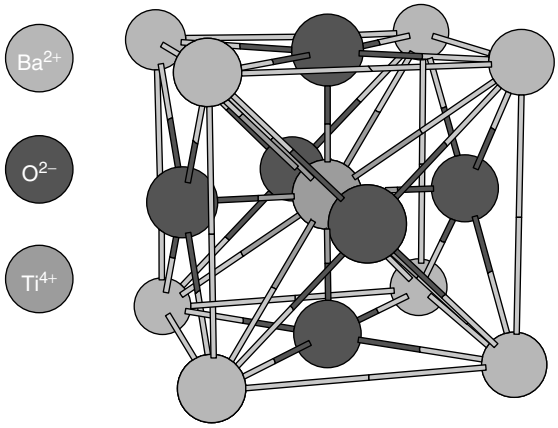


Figure 2. Crystal structure of a typical perovskite, $BaTiO_3$: the Ba^{2+} cations are at the cube corners, the Ti^{4+} cation is in the cube center, and the O^{2-} anions on the cube faces.

3.3 Spontaneous strain and spontaneous polarization of ferroelectric perovskites

At elevated temperatures, the primitive perovskite arrangement is symmetric face-centered cubic (fcc) and does not display electric polarity (Figure 3a). This symmetric lattice arrangement forms the *paraelectric phase* of the perovskite, which exists at

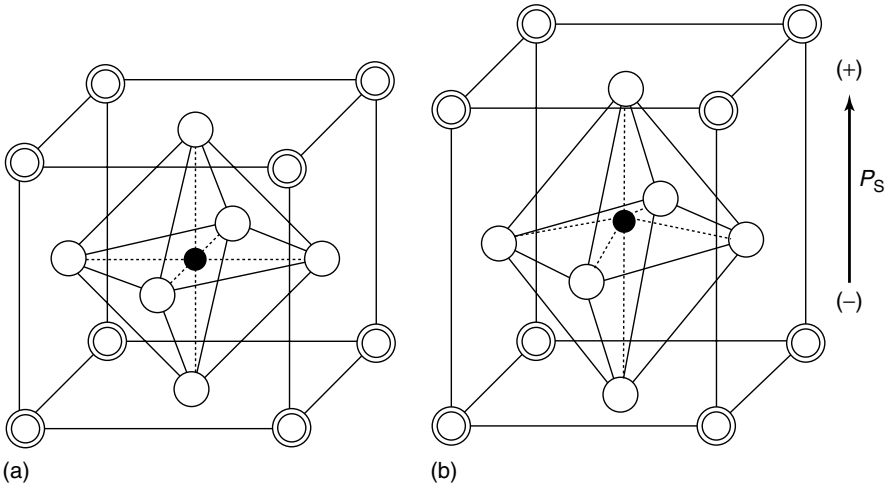


Figure 3. Spontaneous strain and polarization in a perovskite structure: (a) above the Curie point, the crystal has a cubic lattice, displaying a symmetric arrangement of positive and negative charges and no polarization (paraelectric phase) and (b) below the Curie point, the crystal has a tetragonal lattice, with an asymmetrically placed central atom, thus displaying polarization (ferroelectric phase).

elevated temperatures. As the temperature decreases, the lattice shrinks and the symmetric arrangement is no longer stable. For example, in barium titanate, the Ti^{4+} cation snaps from the cube center to other minimum-energy locations situated off center. This is accompanied by the corresponding motion of the O^{2-} anions. Shifting of the Ti^{4+} and O^{2-} ions causes the structure to be altered, creating strain and electric dipoles. The crystal lattice becomes distorted and slightly elongated in one direction, i.e., tetragonal (Figure 3b). In barium titanate, the distortion ratio is $c/a = 1.01$, corresponding to 1% strain in the c direction with respect to the a direction. This change in dimensions along the c axis is called *spontaneous strain*, S_S. The orthorhombic tetragonal structure has polarity because the centers of the positive and negative charges no longer coincide, yielding a net electric dipole, i.e., spontaneous polarization P_S. This polar lattice arrangement forms the *ferroelectric phase* of the perovskite, which exists at lower temperatures. The transition from one phase into the other takes place at the phase transition temperature, commonly called the *Curie temperature*, T_C. In barium titanate, $BaTiO_3$, the phase transition temperature is around 130 °C. As the perovskite is cooled below the transition temperature, T_C, the paraelectric phase changes into the ferroelectric phase and the material displays spontaneous strain, S_S, and spontaneous polarization, P_S; vice versa, when the perovskite is heated above the transition temperature, the ferroelectric phase changes into the paraelectric phase, and the spontaneous strain and spontaneous polarization are no longer present.

4 TYPICAL ELECTROACTIVE CERAMICS

Ceramics are hard, brittle, heat resistant, and corrosion-resistant materials made by shaping and then firing a nonmetallic mineral, such as clay, at a high temperature. Ceramics are polycrystalline materials. Commonly, ceramics are electrical and thermal insulators.

Electroactive ceramics are a class of ceramics that display very strong piezoelectric and/or electrostrictive response. Electroactive ceramics consist of polycrystalline structures of ferroelectric perovskites with strong piezoelectric and/or electrostrictive properties. The electroactive ceramics are synthetic compounds that can be fabricated with engineered properties tailored to meet specific operational requirements. On a macroscale, ferroelectric ceramics are given single-crystal symmetry by poling. Poled ferroelectric ceramics are commonly called *piezoelectric ceramics* (ANSI/IEEE 176). Most piezoelectrics are crystalline solids. Some piezoelectric materials are single crystals, either natural or synthetic. Others are polycrystalline materials that are given macroscale single-crystal-like symmetry through poling.

Electroactive ceramics display significant mechanical response under applied electric field, and electrical response under applied mechanical action. Typical strain response of commercially available electroactive ceramics is around 0.1% in the quasi-linear range. Higher strain response can be obtained by taking the electroactive ceramics in the strongly nonlinear range at high electric fields. Nonlinear strains of up to 0.2% have been reported in certain electroactive ceramics. The operation in the high nonlinear range is usually associated with a marked increase in hysteresis. This results in significant internal heating under high-frequency operation. Operation in the high nonlinear domain also results in a marked decrease in fatigue life.

The ceramic perovskites display both piezoelectric and electrostrictive behavior. One or the other of these two properties is usually enhanced through chemical formulation and processing. Some of these electroactive ceramics can display, according to detailed formulation and processing, either a predominantly piezoelectric or a predominantly electrostrictive behavior.

4.1 Fabrication and poling of electroactive ceramics

Conventional fabrication of ferroelectric ceramics is done in several stages as follows:

1. synthesis of the ferroelectric perovskite powders;
2. sintering and compaction of the perovskite powders into ferroelectric ceramics;
3. electric poling of the ferroelectric ceramics.

More novel methods for fabrication of ferroelectric ceramics include the following:

1. coprecipitation
2. sol–gel (alkoxide hydrolysis)
3. thin-film fabrication
4. single-crystal growth.

During cooling, the lead zirconate titanate (PZT) ceramic undergoes phase transformation from paraelectric state to ferroelectric state. This transformation takes place as the material cools below the Curie temperature, T_C. The resulting ferroelectric ceramic has a polycrystalline structure (grains) with randomly oriented ferroelectric domains (Figure 4a). If the grains are large, ferroelectric domains can exist even inside each grain. Owing to the random orientation of the electric domains, individual polarizations cancel each other, and the net polarization of the virgin ferroelectric ceramic is zero.

This random orientation can be transformed into a preferred orientation through *poling*. Poling aligns the dipole domains and gives a net polarization to the piezoceramic material. A poled ferroelectric ceramics behaves more or less like a single crystal. Poling of piezoceramics is attained at elevated temperatures in the presence of a high electric field. The application of a high electric field at elevated temperatures results in the alignment of the crystalline domains. This alignment is locked in place when the piezoceramic is cooled with the poling electric field still applied, thus resulting in permanent polarization. During poling, the orientation of the piezoelectric domains also produces a mechanical deformation. When the piezoceramic is cooled, this deformation is locked in place (permanent strain). Poling is performed in silicon oil bath at elevated temperature under a dc electric field of $1-3\,\mathrm{kV\,mm^{-1}}$.

4.2 PZT piezoceramics

PZT is a ferroelectric perovskite consisting of a solid solution of $Pb(Zr_{1-x}Ti_x)O_3$. In the PZT perovskite unit cell, lead, Pb^{2+}, occupies the corners, oxygen, O^{2-}, the faces, and titanium/zirconium, Zr^{4+}/Ti^{4+}, the octahedral voids. To date, many PZT formulations exist, the main differentiation being between "soft" (e.g., PZT 5-H) and "hard" (e.g., PZT 8). PZT attains the highest piezoelectric coupling and the maximum electric permittivity near the morphotropic phase boundary. This corresponds to the change in the crystal structure from the tetragonal phase to the rhombohedral phase, which happens when the Zr/Ti ratio is approximately 53/47. The explanation for this phenomenon is as follows: above the Curie temperature, PZT has a cubic lattice and is paraelectric. The Curie temperature varies with the alloying proportion, from $\sim 250\,^{\circ}\mathrm{C}$ for pure $PbZrO_3$ to $\sim 500\,^{\circ}\mathrm{C}$ for pure $PbTiO_3$. Below the Curie temperature, PZT is ferroelectric; but its lattice can be either tetragonal or rhombohedral, according to the alloying proportion.

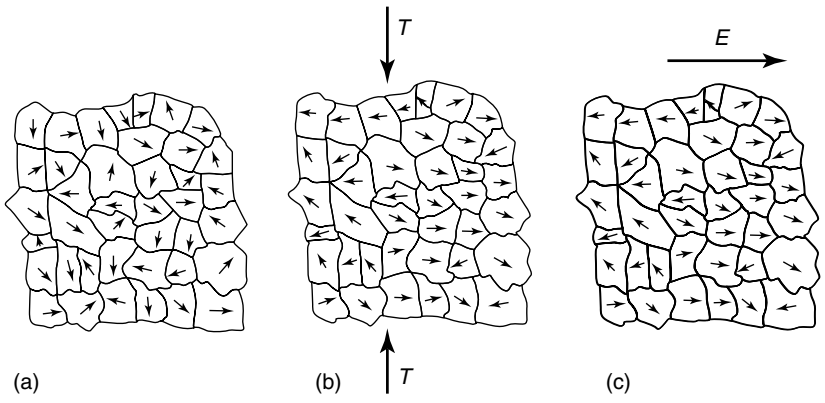


Figure 4. Piezoelectric effect in polycrystalline perovskite ceramics: (a) in the absence of stress and electric field, the electric domains are randomly oriented; (b) application of stress produces orientation of the electric domains perpendicular to the loading direction. The oriented electric domains yield a net polarization; and (c) application of an electric field orients the electric domains along the field lines and produces induced strain.

On the phase diagram, the line separating the two phases is called the *morphotropic phase boundary*. At room temperature, this boundary is placed around the 47/53 alloying ration.

Within the linear range, PZT-like piezoelectric ceramics produce strains that are more or less proportional to the applied electric field or voltage. Induced strains in excess of 1000 $\mu\varepsilon$ (0.1%) are encountered (Figure 5).

4.3 Electrostrictive ceramics—relaxor ferroelectrics

Electrostrictive ceramics are perovskite materials in which the electrostrictive response is dominant. Perovskites that display a large electrostrictive response are the disordered complex perovskites, which have high electrostrictive coefficient, M, with respect to the electric field and a diffuse transition temperature. Such ferroelectric ceramics are also called *relaxor ferroelectrics*, because they display large dielectric relaxation, i.e., frequency dependence of the dielectric permittivity. In a relaxor ferroelectric, the permittivity decreases as the test frequency increases. In addition, the value of temperature at which the permittivity peaks shifts upward. This behavior is in contrast to that of conventional ferroelectrics, for which the temperature at which the permittivity peaks hardly changes with frequency. The dielectric relaxation phenomenon can be attributed to the presence of microdomains in the crystal structure.

In relaxor materials, the transition between piezoelectric behavior and loss of piezoelectric capability

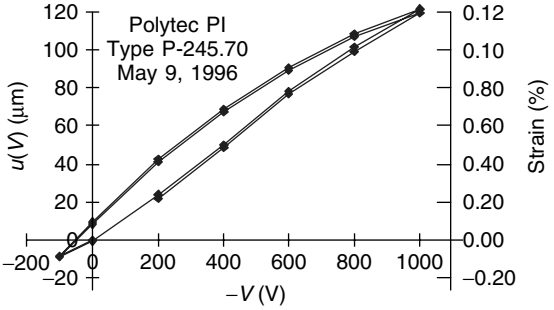


Figure 5. Induced-strain displacement versus applied voltage for Polytec PI model P-245.70 PZT actuator of 100-mm nominal length; induced strain of around 0.12% is observed.

does not occur at a specific temperature (Curie point), but instead occurs over a temperature range (Curie range). Thus, electrostrictive ceramics have a rather diffused phase transition that spans a temperature range around the transition temperature. Hence, the temperature dependence of electrostrictive ceramics around the transition temperature is markedly less than that of normal perovskite solid solutions. Sometimes, their transition temperature range is lower than the room temperature, which is beneficial for stable operation at elevated temperatures.

Lead magnesium niobate/lanthanum formulations, and lead nickel niobate currently are among the most studied relaxor electrostrictive ferroelectrics. They have very high dielectric permittivity and polarization. The coercive field of electrostrictive ceramics is much smaller than that of piezoelectric ceramics.

A common electrostrictive ceramics is lead magnesium niobate, $Pb(Mg_{1/3}Nb_{2/3})O_3$, also known as *PMN*. Another commonly used electrostrictive ceramic is lead titanate, $PbTiO_3$, also known as *PT*. Combination of these two formulations are also common, under the designation PMN–PT. Another electrostrictive ceramic is (Pb, La)(Zr, Ti)O_3, also known as *PLZT*. Other ferroelectric ceramic systems that have been found to display strong electrostrictive behavior include lead barium zirconate titanate, (Pb, Ba)(Zr, Ti)O_3, and barium stannate titanate, Ba(Sn, Ti)O_3. To obtain large electrostriction, it is essential that ferroelectric microdomains in the ceramic structure are generated. Various methods are used in this property, such as the doping with ions of a different valence or ionic radius, or the creation of vacancies, which introduce microscopic spatial inhomogeneity.

The strain versus electric field curves of electrostrictive ceramics display the typical quadratic behavior (Figure 6a). On such curves, a positive mechanical strain is obtained for both positive and negative electric fields. However, the strain field curve is strongly nonlinear, as appropriate to quadratic behavior. What is remarkable about electrostrictive ceramics is their very low hysteresis. Figure 6(a) shows that the increasing and decreasing curves superpose almost everywhere, with only a small exception in a limited region at relatively low fields.

Commercially available PMN formulations are internally biased and optimized to give quasi-linear behavior. In this situation, they display much less

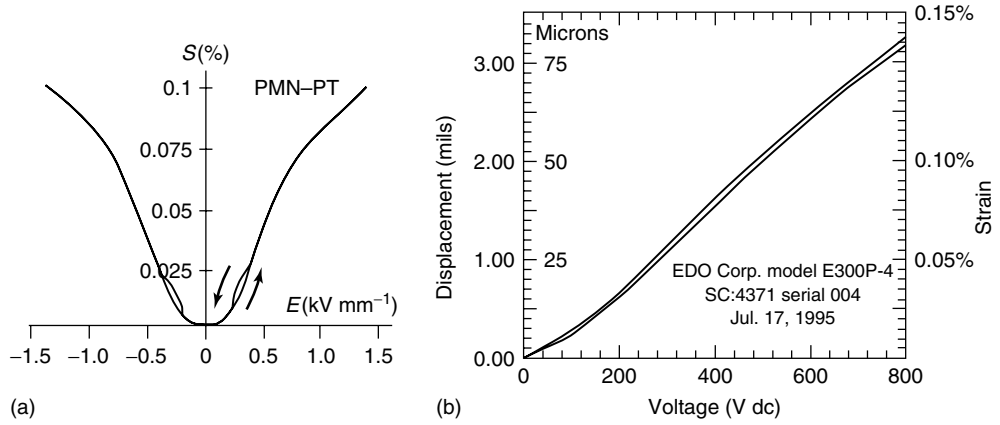


Figure 6. Electrostrictive ceramic: (a) field induced strain in 90–10 PMN–PT and (b) induced strain and displacement versus applied voltage for a 57-mm-long stack of EDO Corporation EC-98 electrostrictive PMN ceramic.

nonlinearity than the conventional quadratic electrostriction, and resemble more the conventional linear piezoelectricity (Figure 6b). Linearized electrostrictive ceramics retain the very low hysteresis of quadratic electrostrictive ceramics. Thus, from this standpoint, they are superior to conventional piezoelectric ceramics. However, linearized electrostrictive ceramics do not accept field reversal. After linearization, the constitutive equations of electrostrictive ceramics resemble those of conventional piezoceramics, i.e.,

$$S_{ij} = s^E_{ijkl} T_{kl} + \tilde{d}_{ijk} E_k \tag{15}$$

$$D_m = \tilde{d}_{mkl} T_{kl} + \varepsilon^T_{mn} E_n \tag{16}$$

The symbol $\sim$ indicates that the piezoelectric constants, $\tilde{d}_{ijk}$, of equations (15) and (16) are different from the corresponding constants d_{ijk} in the original equations (8) and (9). This is due to the linearization process. In equations (8) and (9), the d_{ijk} constants were quite small, since the main effect was due to the quadratic effects represented by the m_{klij} constants. In equations (15) and (16), the $\tilde{d}_{ijk}$ constants are quite significant, since they represent the effect of the linearization of equations (8) and (9).

4.4 Piezopolymers

Piezoelectric polymers are polymers that display piezoelectric properties similar to those of quartz and piezoceramics. Piezoelectric polymers are supplied in the form of thin films. They are flexible and show large compliance. Piezoelectric polymers are cheaper and easier to fabricate than piezoceramics. The flexibility of the polyvinylidene fluoride (PVDF) overcomes some of the drawbacks associated with the brittle nature of piezoelectric ceramics. Its applicability has been proven in keyboards, headphones, speakers, and high-frequency ultrasonic transducers. A typical piezoelectric polymer is the PVDF or PVF_2. This polymer has strong piezoelectric and pyroelectric properties. Its chemical formulation is $(–CH_2–CF_2–)_n$. This polymer displays a crystallinity of 40–50%. The PVDF crystal is dimorphic, the two types being designated I (or β) and II (or α). In the β phase (i.e., type I), PVDF is polar and piezoelectric. In the α phase, PVDF is not polar and is commonly used as electrical insulator. To impart piezoelectric properties, the α phase is converted into β phase and then polarized. Stretching α-phase material produces the β phase. The symmetry of PVDF is $mm2$. Remarkable progress has recently been made in developing piezoelectric polymeric materials through the use of copolymers. The copolymer VDF/TrFE consists mostly of piezoelectric β phase with high crystallinity ($\sim$90%). The film is then cut into various sizes to form piezopolymer sensors.

The piezopolymer surface is metallized to produce the surface electrodes. Silver ink can be screen-printed in patterns onto clear PVDF film. Leads are attached according to customers' specifications. Crimp or eyelet lead attachments are used. After surface metallization, polarization is obtained through

the application of a strong electric field. Improved polarization is obtained through *field cooling*, i.e., cooling with the electric field applied to the sample. To achieve this, the PVDF specimen is to be held at 90–130 °C for 15–120 min under a field of 500–1000 kV cm^{-1}. It is then cooled to room temperature while maintaining the same electric field level.

One important advantage of piezopolymer films is their low acoustic impedance ($Z = \rho c$). The acoustic impedance values of piezopolymers are closer to those of water, biotissue, and other organic materials than the acoustic impedance of piezoceramics. For example, the acoustic impedance of PVDF film is only 2.6 times larger than that of water, whereas the acoustic impedance of piezoceramics is typically 11 times that of water. When the acoustic impedance of two media has similar values, the transmission between the two media is enhanced, and reflection at the interface is reduced.

4.5 Magnetostrictive materials

In simple terms, *magnetostriction* is the material property that causes certain ferromagnetic materials to change shape when an external magnetic field is applied. Magnetostrictive materials expand in the presence of a magnetic field, as their magnetic domains align with the field lines. Magnetostriction was initially observed in nickel, cobalt, iron, and their alloys but the values were small ($<50\,\mu\varepsilon$). Large strains ($\sim 10\,000\,\mu\varepsilon$) were observed in the rare-earth elements terbium (Tb) and dysprosium (Dy) at cryogenic temperatures (i.e., below 180 K). Large room-temperature magnetostriction exists in terbium–iron alloy $TbFe_2$. The binary alloy Terfenol-D ($Tb_{0.3}Dy_{0.7}Fe_{1.9}$), developed at Ames Laboratory and the Naval Ordnance Laboratory (now Naval Surface Weapons Center), displays magnetostriction of up to $2000\,\mu\varepsilon$ at room temperature and up to 80 °C and higher. Current Terfenol-D binary alloy formulations are of the form $Tb_{1-x}Dy_xFe_{1.9-2}$ where x is the relative proportion of dysprosium, while the proportion of iron can vary between 1.9 and 2.

5 SUMMARY AND CONCLUSIONS

This article has reviewed the equations of piezoelectricity and piezomagnetism, given a brief physical explanation of the phenomenon origin, and briefly discussed basic types of electroactive and magnetoactive materials.

Electroactive and magnetoactive materials are materials that modify their shape in response to electric or magnetic stimuli. Such behavior is essential in SHM sensors applications. On the one hand, elastic wave sensing with electroactive and magnetoactive materials creates direct conversion of mechanical energy into electric and magnetic energy. Under dynamic conditions, strong and clear voltage signals are obtained directly from the piezosensor without the need for intermediate gauge bridges, signal conditioners, and signal amplifiers. This direct sensing effect is especially significant at high frequencies when the rapid alternation of polarity prevents significant charge leaking. On the other hand, elastic wave actuation is achieved with active materials that display dimensional changes when energized by electric or magnetic fields.

Piezoelectric, electrostrictive, and magnetostrictive materials have been presented and analyzed. Of these, piezoelectric (PZT), electrostrictive (PMN), and magnetostrictive (Terfenol-D) materials have been shown to have excellent frequency response and good induced-strain capabilities (~0.1%). Figure 7 compares induced-strain response of some commercially available piezoelectric, electrostrictive, and magnetostrictive materials. It can be seen that the electrostrictive materials have less hysteresis, but more nonlinearity. The little hysteresis of electrostrictive ceramics can be an important addition in certain applications, especially at high frequencies. However, one should be aware that this low hysteresis is strongly temperature dependent. As the temperature decreases, the hysteresis of electrostrictive ceramics increases, such that, below a certain temperature, the hysteresis of electrostrictive ceramics may exceed that of piezoelectric ceramics. In general, since the beneficial behavior of the electrostrictive ceramics is related to the diffuse phase transition in the relaxor range, their properties degrade as the operation temperature gets outside the relaxor phase transition range.

In summary, one can conclude that the potential of active materials for sensing and actuation SHM applications has been demonstrated in several successful applications. However, this field is still in its infancy and further research and development is

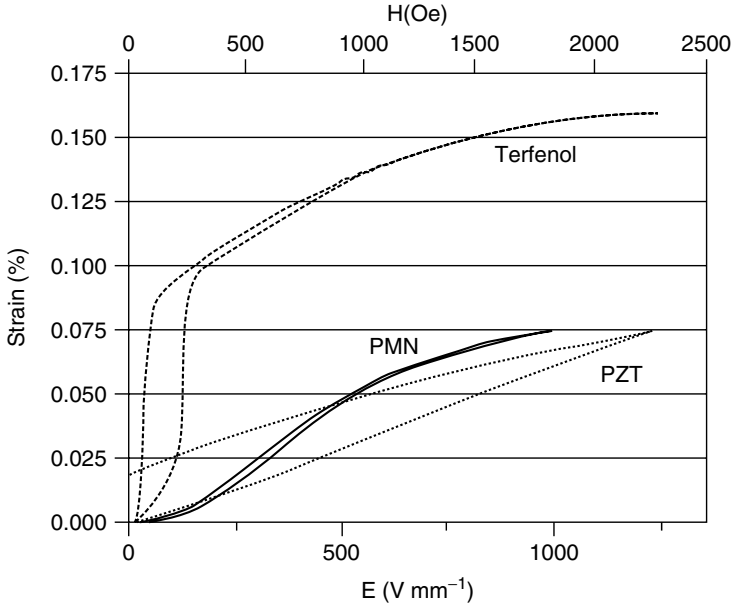


Figure 7. Strain versus electric-field behavior of currently available induced-strain materials.

being undertaken to establish these active materials as reliable, durable, and cost-effective options for large-scale engineering applications.

RELATED ARTICLES

Electromechanical Impedance Modeling

Lamb Wave-based SHM for Laminated Composite Structures

Piezoelectric Impedance Methods for Damage Detection and Sensor Validation

Piezoceramic Materials—Phenomena and Modeling

Constitutive Modeling of Magnetostrictive Materials

Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators

Damage Detection Using Piezoceramic and Magnetostrictive Sensors and Actuators

Probabilistic Approaches to Sensor Layout Design, Data Processing, and Damage Detection

Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors

Piezoelectric Wafer Active Sensors

Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection

Miniaturized Sensors Employing Micro- and Nanotechnologies

Design of Active Sensor Network and Multilevel Data Fusion

Energy Harvesting and Wireless Energy Transmission for SHM Sensor Nodes

Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications

Thermal Protection System Monitoring of Space Structures

Development of an Active Smart Patch for Aircraft Repair

Design, Analysis, and SHM of Bonded Composite Repair and Substructure

Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control

Wind Turbines

Integrated Sensor Durability and Reliability

FURTHER READING

ANSI/IEEE Std 176, *IEEE Standard on Piezoelectricity*. The Institute of Electrical and Electronics Engineers: New York, 1987.

Engdahl G. *Handbook of Giant Magnetostrictive Materials*. Academic Press, 2000.

Giurgiutiu V. *Structural Health Monitoring with Piezoelectric Wafer Active Sensors*. Elsevier Academic Press: New York, 2008.

Ikeda T. *Fundamentals of Piezoelectricity*. Oxford University Press, 1996.

Lines ME, Glass AM. *Principles and Applications of Ferroelectrics and Related Materials*. Clarendon Press: Oxford, 2001.

MIL-STD-1376B(SH), *Military Standard Piezoelectric Ceramic Materials and Measurements Guidelines for Sonar Transducers*. Defense Quality and Standardization Office: Falls Church, VA, February 1995.

du Tremolet de Lacheisserie E. *Magnetostriction: Theory and Applications of Magnetoelasticity*. CRC Press, 1993.

Uchino K. *Piezoelectric Actuators and Ultrasonic Motors*. Kluwer Academic Publishers, 1997.

Chapter 57

Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection

Yunfeng Zhang

Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA

1 INTRODUCTION

Structural defect inspection and corresponding retrofit actions lead to a prolonged life and enhanced reliability of structural systems. However, in practice, manual inspection is still the most commonly used approach; this can be quite costly and time consuming, and therefore unsuitable for rapid assessment of structural conditions. It is thus desirable to replace manual inspection with on-line structural defect detection techniques. For example, existing nondestructive evaluation (NDE) techniques for fatigue crack detection include dye penetrants, eddy current, radiography, acoustic emission (AE), magnetic particle inspection, ultrasonic testing, etc. All these NDE techniques have limited use for on-line fatigue crack monitoring due to either one or a combination of the following problems: accessibility, automation, bulky size, power supply, environmental noise, and long-term durability. In particular, most NDE instruments give results that need to be interpreted by a skilled operator, with a potential for subjective mistakes.

In the past decade or so, sensors made of piezoelectric materials have gained increasing popularity in the field of structural health monitoring (e.g. [1, 2]; *see* **Integrated Sensor Durability and Reliability**). For example, recent research works by Ayres *et al.* [3], Park *et al.* [4], Giurgiutiu [5], and Yu [6] (*see* **Piezoelectric Wafer Active Sensors**) on piezoelectric active wafer sensor have made it quite a promising technique for NDE applications. Piezoelectric materials can be broadly classified into three major categories [7]: ferroelectric ceramics, piezoelectric polymers (e.g., polyvinylidene fluoride, PVDF), and piezoelectric composites. Because of their electromechanical coupling property, piezoelectric materials have been widely used for sensing and actuation applications [8–11]. Piezoelectric ceramics

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

are perhaps the most popular piezoelectric materials for sensing purpose. As of today, most piezoelectric sensors are made of piezoelectric ceramics like lead zirconate titanate (PZT). However, limitations of the mechanical properties of PZT such as brittleness and cracking at large deformations could impede its use in certain applications. Additionally, prefabricated PZT wafers do not fit well on surfaces with complex geometry such as welded joints.

To overcome these problems associated with conventional piezoelectric materials such as PZT, polymer-based piezoelectric paints have been investigated by a few researchers as a potential substitute for piezoelectric ceramics in certain sensing applications [12–22]. Piezoelectric paint typically consists of tiny piezoelectric particles randomly dispersed within a polymer matrix and therefore belongs to the 0–3 piezoelectric composite family. Table 1 summarizes the composition and piezoelectric activity of the piezoelectric paint formulations developed by different researchers. Like piezoelectric ceramics, piezoelectric paint can be used for ultrasonic NDE. Piezoelectric paint that can be directly deposited onto the surface of host structures has several advantages over conventional piezoelectric ceramics. This article reviews the past and current research work in the area of piezoelectric paint as well as its potential use as sensing material for ultrasonic NDE, particularly distributed AE sensor or ultrasonic guided wave–based embedded sensor in metal or fiber reinforced polymer (FRP) structures. Although a comprehensive experimental program still needs to be carried out to fully investigate the potential of piezoelectric paint sensor technology, it can be concluded on the basis of preliminary experimental results that piezoelectric paint sensor appears to provide a promising inexpensive NDE technique for ultrasonic-based monitoring of structural defects such as fatigue cracks in structures.

2 COMPOSITION AND FABRICATION OF PIEZOELECTRIC PAINT

Piezoelectric composite materials consisting of ferroelectric ceramics and polymer [26–31] have received considerable interest as sensing elements because of their favorable material properties that often cannot be obtained in single-phase materials. Through judicious selection of the polymer matrix, the composite properties of the piezoelectric paint can be tailored to meet the specific requirements of application conditions. For example, piezoelectric paint might be more suitable for use in fiber reinforced polymer composite structures because of its improved acoustic impedance matching property compared to piezoelectric ceramics.

The arrangement of the component phases within a composite is critical for the electromechanical properties of composites [22]. Newnham *et al.* [32] have developed the concept of connectivity to describe the manner in which the individual phases are self-connected. In a diphasic system, there are 10 types of connectivities in which each phase is continuous in zero, one, two, or three dimensions. The 10 connectivities are denoted as the following: 0–0, 0–1, 0–2, 0–3, 1–1, 1–2, 2–2, 1–3, 2–3, and 3–3. It is conventional for the first digit to refer to the piezoelectrically-active phase. Piezoelectric paint typically consists of tiny piezoelectric particles mixed within polymer matrix and therefore belongs to the "0–3" piezoelectric composite. The "0–3" means that the piezoelectrically active ceramic particles are randomly dispersed in a three dimensionally connected polymer matrix. Conceivably, the advantages of "0–3" piezoelectric composites over other connectivity types are their ease of fabrication into complex shapes including large flexible thin sheets, extruded bars and fibers, and molded shapes; they may conform to any curved surface. In general, the 0–3 composite family has been found to exhibit high hydrostatic piezoelectric voltage coefficients and "figure of merit" when compared to the properties of conventional single-phase materials [12].

Piezoelectric paint is comprised of three major components: piezoelectric ceramics powder with average particle size in the range of 3–20 μm, a polymer binder to carry the powder in suspension during application and bind it together on curing, and additives including defoamer, dispersants, and surfactants to enhance the paint mixing, deposition, and curing properties. The properties of 0–3 composites are strongly dependent on both the piezoelectric and polymer phases utilized, as well as the fabrication method employed [26]. Among all piezoelectric materials, PZT, $BaTiO_3$, $PbTiO_3$, $LiTaO_3$ ceramics,

Table 1. Summary of piezoelectric paint properties in prior research (number in bracket indicates volume fraction of the respective constituent phases)

Researchers	Ingredients		Fabrication method		Piezoelectric activity d_{33} (pC N^{-1})
	Ceramics	Polymer	Application/cuing	Poling	
Hanner *et al.* [12]	PZT/$PbTiO_3$ (60–70%)	Acrylic copolymer in suspension form or polyurethane	Spreading/24-h drying in air +24-h drying in vacuum oven at 110 °C	Oil bath poling at 100 °C with 10–15 MV m^{-1} or corona poling	26–39
Egusa and Iwasawa [13]	PZT (53%)	Epoxy	Screen printing/3-day drying in air	Conventional poling at 20 MV m^{-1} for 30 min	Unknown
Wenger *et al.* [14]	PTCa (65 or 60%)	P(VDF-TrFE) or epoxy	Hot-rolling/unknown curing method	Conventional poling at 10–25 MV m^{-1} for 30 min at 100 °C	33 or 30
Badcock and Birt [17]	PZT-5H (60%)	Epoxy	Spreading/unknown curing method	Oil bath poling at 19 MV m^{-1} for 5 min at 60 °C	16.5
Sakamoto *et al.* [18]	PZT (59%)	Polyurethane	Pressed at 20 MPa into thick film	Conventional poling at 3 MV m^{-1} for 1 h at 100 °C	12
Li *et al.* [19]	PZT (35–70%)	Cement	Spread into 3-mm-thick film/cured in concrete curing room (23 °C, 100% humidity) for 26 days	Oil bath poling at 20 MV m^{-1} for 1 h at 150 °C	7–33
Hale *et al.* [23]	PZT (40%)	Acrylic paint	Spray painting/unknown curing method	Conventional poling at 60 °C up to 1 h	20
Zhang and Li [24]	PZT (30%)	Epoxy	Spreading/cured at elevated temperature 50 °C for 3 days	Conventional poling (5 MV m^{-1}) at 70 °C for 30 min	5
Kobayashi *et al.* [25]	PZT	Unknown sol–gel solution	Spray/drying and firing by heat gun	Corona poling at 120 °C for 10 min	30
Lahtinen *et al.* [22]	PZT (60%)	Unknown sol–gel solution	Spreading with paint applicator/unknown curing method	Conventional poling (12.5 MV m^{-1}) at room temperature for 15 min	Unknown

and the mentioned ceramics with different dopings have been used for 0–3 composites, while PZT is by far the most popular active ingredient for piezoelectric paints. Marin-Franch *et al.* [33] used calcium modified lead titanate/poly(ether ketone ketone) (PTCa/PEKK) composites to detect AE when mounted on carbon fiber reinforced composite panels. PTCa ceramic is selected because it has the appropriate piezoelectric properties together with a low Curie temperature 260 °C, so it can be readily poled at temperatures appropriate to the polymer.

Han *et al.* [34] developed a colloid processing method to improve the microstructural homogeneity and decrease the chance of void formation in 0–3 composites. With this technique, piezoceramic powder was dispersed in a dilute polymer solution, allowing a polymer coating to be absorbed onto the powder surface. Colloidally processed composites composed of coprecipitated PT-BF powder and Eccogel polymer were measured to have the largest d_{33} ($65\,\mathrm{pC\,N^{-1}}$) and highest figure of merit ($dg = 6000 \times 10^{-15}\,\mathrm{m^2\,N^{-1}}$) of all "true" 0–3 composites [26]. Han *et al.* also studied the effect of the particle size on dielectric and piezoelectric properties of 0–3 composites. Lau *et al.* [35] studied incorporation of nanosized PT powder into a poly(vinylidene fluoride - trifluoroethylene) [P(VDF-TrFE) 70/30 mol%] matrix to form a 0–3 nanocomposite with a 0.2 volume fraction of PT. A thin film of 5-μm thickness was prepared by spin coating the composite on a glass substrate. The transducer was 1.2 mm in diameter with a center frequency of 40 MHz.

The polymer matrix material is critical to the ease of fabrication and performance characteristics of piezoelectric paint sensor. For example, to investigate the effect of the polymer on resistivity and dielectric properties of 0–3 composites, Han *et al.* [34] prepared PT composites with Eccogel polymer, PVDF copolymer, and ethylene-propylenediene monomer (EPDM) polymer. It was found that although higher poling conditions could be applied to the PVDF copolymer and EPDM composites, the highest d_{33} value was obtained from the epoxy composites. The higher electrical conductivity of the polymer matrix may have created more electric flux paths between the ceramic particles. This in turn increased the electric field acting on the ceramic filler and made poling of the ceramic easier [26]. The epoxy gave the highest figure of merit ($d_h g_h = 5600 \times 10^{-15}\,\mathrm{m^2\,N^{-1}}$) and the EPDM gave the lowest figure of merit ($600 \times 10^{-15}\,\mathrm{m^2\,N^{-1}}$).

The poling of ferroelectric materials to induce piezoelectricity is an important stage in the fabrication of piezoelectric paint. Poling can be carried out by two rather different techniques—the conventional dc poling method and the Corona discharge method. Poling of composites having a polymer matrix with 0–3 connectivity is especially difficult because the electric field within the high-dielectric-constant grains is far smaller than in the low-dielectric-constant polymer matrix [36]. Since most polymers have a lower dielectric constant compared with piezoelectric ceramic materials, most of the applied electric field will pass through the lower dielectric constant phase. Moreover, local breakdown at weak spots short-circuits the electrodes and prevents further poling since very large electric fields are required to pole these types of composites. One way to resolve this difficulty with poling is to introduce a third conductive phase between the piezoelectric particles. Sa-Gong *et al.* [37] prepared such composites by adding carbon, germanium, or silicon to PZT. Sakamoto *et al.* [18] doped 0–3 piezoelectric composites of PZT powder and vegetable-based polyurethane (PU) with small amounts of carbon powder (vol% in the range of 0.5–2.0%) to ease the poling by creating a continuous electric flux path between PZT grains.

The corona poling technique has been proposed by Waller and Safari [36] to pole flexible piezoelectric composites such as piezoelectric paints. This technique has been successful in poling PVDF films. In the Corona poling method, electric charge from a corona point is sprayed onto the unelectroded surface of the sample, creating an electric field between the sample faces. Because of the absence of electrodes, there is no short-circuiting of the sample at weak spots. This is particularly beneficial for poling of piezoelectric paint sensors because short-circuited piezoelectric paint sensors have to be manually removed from the host structure after poling. Generally speaking, heating lowers the coercive field and makes poling easier. The piezoelectric properties of ceramics and composites poled by the Corona method are comparable or better than those poled by the conventional poling technique.

3 REVIEW OF PIEZOELECTRIC PAINT SENSOR DEVELOPMENT

In their early studies of thin-film 0–3 polymer/piezoelectric ceramic composites, Hanner *et al.* [12] prepared two formulations of piezoelectric paints—one based on an acrylic copolymer and the other based on a PU and the electrical properties of these piezoelectric paints were characterized. Both formulations were loaded with 60–70% volume fraction of PZT and a coprecipitated $PbTiO_3$. The paint samples, with a film thickness ranging between 200 and 500 μm, were placed in a vacuum oven at 110 °C for 24 h to remove residual water or solvent. Poling of the paints was accomplished by both the conventional oil bath and the corona discharge techniques. The values of the piezoelectric charge constant d_{33} for the PZT-filled composites were found to be 25–28 pC N^{-1}, while those for the coprecipitated $PbTiO_3$/acrylic copolymer composites were 35–38 pC N^{-1}.

Egusa and Iwasawa [13] prepared an epoxy resin-based piezoelectric paint with 53 vol% PZT powder as filler. The paints were spread onto one side of a 30-mm-wide aluminum beam. The 152-μm-thick paint film was then cured in air at room temperature for at least three days or at 150 °C for 45 min. The effects of poling field, film thickness, and cure temperature on the paint's piezoelectric activity were investigated [38]. The piezoelectric sensitivity of the paint film as an AE sensor was also evaluated in their study [39]. A nearly flat frequency response of the paint film to AE waves was observed in the frequency range above 0.3 MHz.

Hale and Tuck [15], White *et al.* [20], and Hale *et al.* [23] fabricated and tested a piezoelectric paint-based strain sensor for use in structural vibration monitoring. The paint formulation comprised of milled PZT ceramic powder mixed in a water-based acrylic paint, based with PZT powder concentrations of up to 80% by weight (approximately 40% by volume). The cured paint was poled at 600 V for 1 h. For trouble-free spraying and high-quality paint films, it was found necessary to perform high-shear mixing for half an hour prior to coating application. It was found from the tests by White *et al.* [20] that the most successful spray system was a miniature DeVilbiss 0.8-mm air-atomizing spray gun operating at approximately 1.7 bar using a gravity feed paint cup. The piezoelectric charge coefficient d_{33} was found to vary between 10 and 30 pC N^{-1} with a predominance around 20 pC N^{-1} [23].

Paints belonging to the class of 0–3 piezoelectric composites, consisting of a ferroelectric ceramic powder of calcium modified PTCa dispersed in a polymer matrix, have been fabricated and their ferroelectric properties have been investigated by Wenger *et al.* [14]. Two formulations of 0–3 composites, one with a copolymer of P(VDF-TrFE) and the other with a thermosetting epoxy resin, were studied. The composite of the ceramic with the epoxy, 60/40 vol% PTCa/epoxy, was prepared by gradually adding the ceramic powder to the resin while stirring continuously to ensure an even mixture [40]. The ceramic/copolymer composite PTCa-P(VDF-TrFE) with a 65/35 volume fraction was prepared by a hot-rolling technique. The composite films were subsequently poled in a dc field of $1–2.5 \times 10^7$ V m^{-1} for 30 min at 100 °C in an insulating silicone oil bath. The d_{33} coefficients for the composites of PTCa/P(VDF-TrFE) and PTCa/epoxy are 33 and 26 pC N^{-1} respectively. Surface-mounted AE sensors were fabricated using the composite films thus obtained and their frequency response was evaluated using a face-to-face technique over the frequency range of 300 kHz to 50 MHz. Embedded transducers, constructed from the piezoelectric composite films were used to detect plate waves within a laminate glass/epoxy plate measuring 304.8 mm × 304.8 mm × 1.9 mm. It was found in their study [14] that, in the case of the embedded transducers, the PTCa/epoxy films seem to be the better choice of transducer material, producing signals comparable in amplitude to those of an embedded PTCa/P(VDF-TrFE) film but more clearly defined.

Sakamoto *et al.* [18] studied a flexible piezoelectric composite with 0–3 connectivity, made from PZT power and castor oil-based PU, which was doped with a small amount of fine-grained carbon powder to facilitate poling at relatively low electric field in a short time span. The carbon particles located between the PZT particles help create a continuous electric flux path. The PZT powder was mixed with carbon by a vibrator for 30 min prior to introducing this mixture in the PU matrix. The composite was placed between two paraffin papers and pressed at room temperature. The applied pressure was about 20 MPa and it was possible to obtain samples in the thickness range

of 250–350 μm. All the samples were poled with a 3 MV m^{-1} electric field at 100 °C for 1 h. After poling, the d_{33} coefficient was measured and it was observed that the highest value of the d_{33} coefficient (around 12 pC N^{-1}) was achieved with PZT/C/PU samples with 59/1/40 vol% composition. On adding 1.0 vol% of carbon, the d_{33} coefficient increased by 25% in comparison with the composite without the semiconductor phase. The disadvantage associated with carbon doping appears to be the relatively low electrical breakdown voltage of about 6 MV m^{-1}. The PZT/C/PU composite has shown the ability to detect both extensional and flexural modes of simulated AE at a distance of up to 8.0 m from the source [41].

Badcock and Birt [17] have examined the use of piezoelectric 0–3 composite as embedded Lamb wave sensors for damage detection in carbon fiber reinforced composite plates (E-glass/913 and T800/924 aerospace composites). The 0–3 composite consists of PZT-5H powder dispersed in an epoxy matrix. Cured PZT/epoxy materials with a PZT volume fraction of 60% was immersed in an oil bath for poling at a predefined temperature in the range of 50–120 °C. It was found that the greatest poling efficiency is achieved at 60 °C (most samples were poled at a field of 19 kV mm^{-1} for 5 min), with sensitivities of around 17 pC N^{-1} for the piezoelectric charge constant d_{33}. Comparative study of three different transducers—a conventional PZT ultrasonic transducer, a 500-μm-thick PVDF element, and a 600-μm-thick piezoelectric 0–3 composite film—were carried out to receive the Lamb wave signal at a fixed distance from the transmitting transducer. It was found the piezoelectric 0–3 composites were twice as sensitive when compared with PVDF films.

Li *et al.* [19] developed cement-based 0–3 piezoelectric composites containing up to 70 vol% of PZT particles by normal mixing and spreading method. Cement and PZT particles were first thoroughly mixed together without adding water, and then water and superplasticizer were added to the mixture. The mixture was then spread on a glass plate to form a disk with a diameter of ∼10 mm and thickness of ∼3 mm. The samples were put in the curing room at a temperature of 23 °C and relative humidity of 100% for 28 days before measurement. Poling was conducted in a silicon oil bath at a temperature of ∼150–160 °C for 1 h. The d values were found to vary from 7.2 pC N^{-1} for the 35 vol% of PZT formulation to 33.4 pC N^{-1} for the 70 vol% of PZT formulations. The cement-based 0–3 piezoelectric composites were shown to have a slightly higher piezoelectric factor and electromechanical coefficient than those of 0–3 PZT/polymer composites with a similar content of PZT particles.

A sol–gel spray technique was used by Kobayashi *et al.* [25, 42] to produce thick-film broadband ultrasonic transducers that are based on piezoelectric 0–3 composites. The piezoelectric particles of PZT and $LiTaO_3$ were dispersed in a sol–gel solution with a viscosity suitable for spray painting [42]. The fabrication process involved spray coating, firing at 420 °C using a heat gun or gas torch, and annealing at 650 °C to optimize the PZT sol–gel. Multiple layers were made to reach the desired transducer thickness. After sputtering the platinum top electrodes, the piezoelectric films of a 50–60-μm thickness were poled at 380 °C with a 7–9 MV m^{-1} electric field. The films were observed to be able to withstand more than 10 thermal cycles between the room temperature and 250 °C without being detached from the steel rod substrate. The d_{33} values of the films were measured to be 0.1 pC N^{-1}. Ultrasonic pulse–echo signals with a signal-to-noise ratio of more than 25 dB have also been received by the PZT/$LiTaO_3$ composite films at elevated temperatures up to 250 °C. More recently, Kobayashi *et al.* [25] used the sol–gel spray technique to deposit PZT/PZT sol–gel composite directly onto selected substrates including graphite/epoxy, aluminum, and steel. The Corona poling technique was used to pole the cured film for 10 min at 120 °C. The piezoelectric constants d_{33} and d_{31} of the films were 30×10^{-12} and -26×10^{-12} m V^{-1}, measured by an optical interferometer and optical coherence tomography, respectively.

Zhang and Li [24] developed an epoxy-based flexible piezoelectric paint that cures at room temperature. The paint film from this recent formulation is very compliant after curing, as can be seen from the bent shape of the paint sample in Figure 1. PZT-5A powder was chosen as the active piezoelectric material in this paint formulation. The epoxy resin used in this study is a liquid resin comprised mainly of diglycidyl ether of bisphenol A (DGEBPA). Additives such as dispersing agents and defoamers were also added to the paint mixture during the paint grinding and mixing process. Curing of the paint was

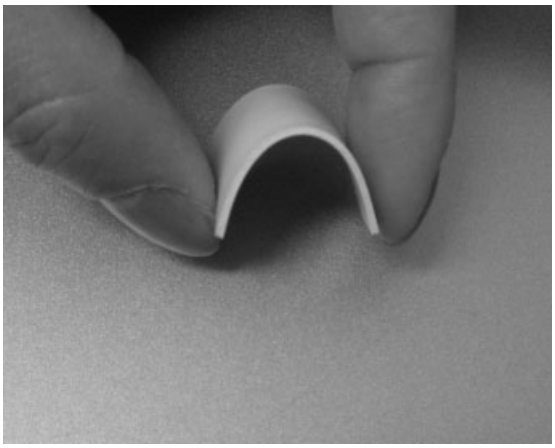


Figure 1. Sample of flexible piezoelectric paint.

done at elevated temperature (50 °C) for three days before applying conductive paint-based electrode to the top side of the paint. The volume fraction of PZT power in cured paint was about 30%. After the curing and electroding process had been completed, poling of the piezoelectric paint was carried out to activate its piezoelectric effect. In this study, poling was performed using a conventional electrode poling device at a temperature of 70 °C for 30 min at an electrical field of 5 MV m^{-1}. The average value of d_{33} was measured to be 5 pC N^{-1}. The ultrasonic sensing characteristic of the piezoelectric paints from this study is described in the next section.

Lahtinen *et al.* [22] developed a piezopaint-based vibration sensor by mixing piezoceramic powder with epoxy resin, which has been shown to be a very sensitive vibration monitoring device. The paint was mixed with 60 vol% PZT pigment concentration. The PZT powder was PZ27 produced by Ferroperm Electroceramics A/S, Denmark. The size range of the powder particles was 0–25 μm, with average particle size of 5 μm. The paint was applied with a laboratory applicator and the dry film thickness was 80 μm. Poling of the piezopaint layer was carried out by applying a 1-kV potential over the piezo layer for a period of 15 min at room temperature. Dynamic loading test of the piezopaint vibration sensors was conducted both in the laboratory and in field conditions. The thermal stability of the piezopaint sensor is excellent compared with piezoelectric polymers because of the fact that the piezoelectricity is a property of a ceramic material in the composite. The thermal stability is limited to the properties of the polymer matrix and the electrode materials. The thermal stability could possibly be further enhanced by applying more stable polymer matrix materials such as Polyetheretherketone (PEEK).

Hale and Lahtinen [43] conducted a program of testing to evaluate the susceptibility of piezoelectric paint to environmental degradation and so to evaluate its suitability for use in shock and vibrations sensors on large outdoor structures. Controlled weathering trials show that the sensors drop in sensitivity over the first few months from manufacture, but thereafter maintain a constant sensitivity irrespective of exposure to sunlight, rain, or frost, etc. Field trials on river crossing bridges in United Kingdom and Finland have shown that the sensors can survive harsh outdoor conditions and remain functional for six years, with no sign of an end to their lives at that time.

4 ULTRASONIC SENSING CHARACTERISTICS OF PIEZOELECTRIC PAINT

Like all other piezoelectric materials, piezoelectric paint can produce electric voltage signals proportional to the applied mechanical deformation in its film plane. Without loss of generality, we assume that piezoelectric paint sensor is only subjected to unidirectional strain in its film plane. The dielectric displacement D_3 is related to the generated electric charge by the following relationship,

$$q = \iint D_3 \, \mathrm{d}A_3 \tag{1}$$

where $\mathrm{d}A_3$ is the differential electrode area in the 1–2 plane of the piezoelectric paint. The electric charge generated by the sensor is

$$q = \mathrm{d}_{31} Y_C b_C \int_{lc} \varepsilon_1 \, \mathrm{d}l \tag{2}$$

where Y_C is the Young's modulus of the piezoelectric paint, and l_c and b_c are the length and width of the piezoelectric paint sensor, respectively.

Sirohi and Chopra [44] studied the behavior of PZT and PVDF strain sensor over a frequency range of 5–500 Hz. In their work, a typical piezoelectric

patch sensor is considered as a parallel plate capacitor, which stores the electric charge generated by the piezoelectric sensor when mechanical deformation is applied. It is shown that if only one-dimensional in-plane deformation is considered, the voltage generated across the electrodes of the capacitor can be related to the average in-plane strain by a sensor sensitivity parameter and the sensor capacitance. If a charge amplifier is connected to the sensor electrode, then the output voltage is

$$V_{\text{out}}(t) = -\frac{q}{C_f} = -\frac{d_{31} Y_C b_C}{C_f} \int_{lc} \varepsilon_1 \, dl \qquad (3)$$

Therefore, the sensor output voltage is proportional to the integral of the in-plane strain in the 1-direction.

AE consists of a propagating elastic wave generated by a sudden release of energy within a material (see, e.g. [45–47]; *see* **Acoustic Emission**; **Applications of Acoustic Emission for SHM: A Review**). In recent years, AE has been shown to provide real-time information on the progression of damage within metallic structures. The elastic wave generated by structural damage propagates through the solid to the surface where it can be detected by surface-mounted sensors. The sensor captured AE waveforms contain information about the damage source, like location, size and type, etc.; it is possible to extract such information by analysis of those waveforms. In a manner different from ultrasonic active sensing, which intentionally excite elastic waves in a solid, AE sensors passively listen for acoustic signals generated by crack initiation and progression. AE has been proved to be a useful NDE method for the investigation of local damage in structural members.

The AE sensing capability of piezoelectric paints has been verified by a number of researchers including Egusa and Iwasawa [39], Wenger *et al.* [14, 40], Sakamoto *et al.* [41], Kobayashi *et al.* [25, 42], Marin-Franch *et al.* [33], and Li and Zhang [48]. Egusa and Iwasawa [39] evaluated the sensitivity of piezoelectric paint film with a PZT volume fraction of 53% to film thickness and poling field as an AE sensor in the frequency range of 0–1.2 MHz. Wenger *et al.* [14] examined the suitability of piezoelectric 0–3 composites embedded into glass-reinforced laminate plates as AE sensors. Piezoelectric bimorph sensors for embedded AE sensing in glass–epoxy laminate platelike structures have been fabricated and investigated by Wenger *et al.* [40]. The bimorphs can differentiate between two types of plate wave propagation—flexural and extensional modes of Lamb waves, although the differentiation is attributed to the wavelength of the acoustic wave relative to the sensor dimensions. Sakamoto *et al.* [41] have shown that the piezoelectric 0–3 composite film with a PZT volume fraction of 49% can successfully detect both extensional and flexural modes of simulated AE at a distance up to 8 m from the source on a fiberglass reinforced plate. Marin-Franch *et al.* [33] developed a PTCa/PEKK composite to detect AE from delamination in carbon fiber reinforced composite panels. It is demonstrated that in a square array they can be used to locate AE sources with good accuracy.

We recently conducted a series of ultrasonic tests including pitch–catch ultrasonic test and pencil break test to examine the sensing capability of piezoelectric paint sensor over ultrasonic frequency range. The piezopaint sensor loaded with 30% by volume of PZT powder was deposited to a 609.6 mm × 609.6 mm × 1.59 mm aluminum plate. A layer of conductive silver paint was applied on top of the piezopaint having a size of 7 mm × 7 mm. The paint film thickness was 1 mm. After curing and electroding, poling of the piezoelectric paint was carried out to activate its piezoelectric effect. A 7 mm × 7 mm × 0.2 mm piezoelectric ceramic patch was also bonded along the centerline of the piezoelectric paint sensors as the actuator to excite Lamb waves in the aluminum plate. A five-cycle windowed toneburst excitation signal was generated by a function generator at various frequencies. Generally, guided waves propagating through metal structure surface are highly dispersive and tend to decay very fast with the increase of the propagation distance. Piezoelectric paint sensors were used to measure the guided-wave propagation at corresponding sensor locations. An AD745 high-speed operational amplifier circuit with a voltage gain of 100 was connected to the piezoelectric paint sensor as a preamplifier and then the amplified signal was recorded with a Tektronix TDS 2024 oscilloscope connected to a computer via a general purpose interface bus (IEEE488) (GPIB) adapter.

In our test, the PZT wafer was excited by 10 V peak-to-peak voltage signal, and ultrasonic lamb waves were received by a piezoelectric paint sensor

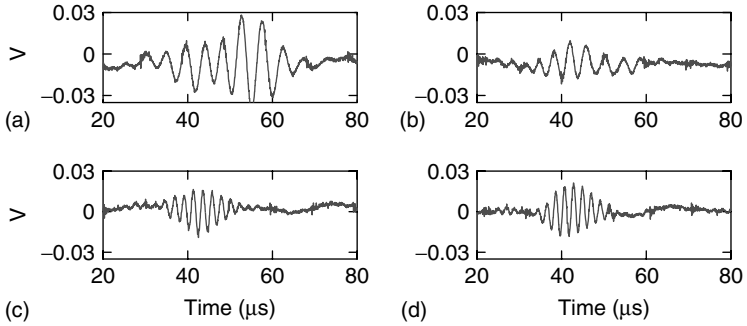


Figure 2. Raw data received by piezoelectric paint sensor: (a) 40-kHz excitation, (b) 60-kHz excitation, (c) 80-kHz excitation, and (d) 100-kHz excitation.

which was 25 mm away from excitation. The received raw time traces at excitation frequencies of 40, 60, 80, and 100 kHz are shown in Figure 2. The amplitude of the received signals at piezoelectric paint sensors was in the range of 20 mV peak-to-peak. Before raw sensor data can be interpreted to determine structural health condition, signal processing is necessary to ease the interpretation process. This is especially true for piezoelectric paint sensor because of its low sensitivity and hence low signal-to-noise ratio compared with a PZT sensor. In the present research, discrete wavelet transform (DWT) technique was used to filter the signal captured by piezoelectric paint sensor. Figure 3 shows the DWT-filtered signals (using a db (Daubechies) 6 level 5 wavelet) at piezoelectric paint sensors using wavelet transform. The filtered signals clearly show the propagation of Lamb waves over time. This experimental study conforms that piezoelectric paint sensor can be used to detect ultrasonic wave motion in platelike structures.

To verify the capability of piezoelectric paint sensor as an AE sensor, we performed pencil break tests on the same aluminum plate. Pencil break source (also called *Hsu-Neilsen source*) is a common approach to simulate AE due to its simplicity and reproducibility. In our study, a pencil lead with a 0.5-mm diameter was broken by pressing against the aluminum plate to simulate AE signal from a microcrack. The same data-acquisition equipment as in the ultrasonic test was used here. For comparison purpose, a 7 mm × 7 mm × 0.2 mm PZT sensor has been placed right next to the piezoelectric paint sensor on the aluminum plate to capture the AE signal. The same amplifier used in the ultrasonic test was applied to the piezoelectric paint sensor. The pencil lead was fractured at the same distance of 127 mm (∼5 in.) away from both the PZT and piezoelectric paint sensor. Figure 4(a) plots the captured AE signals collected at a sampling rate of 2.5 MHz. After filtering with the db 6 level 5 wavelet, the filtered

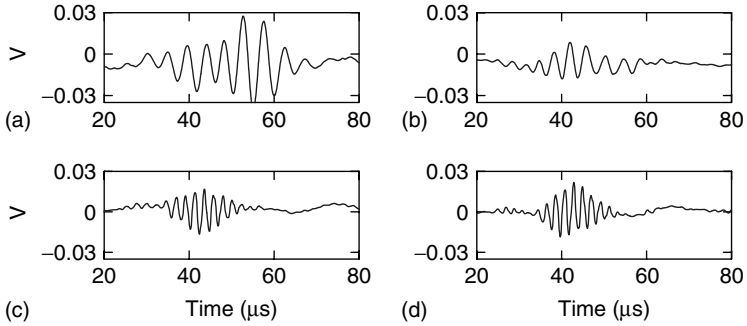


Figure 3. Filtered signal received by piezoelectric paint sensor: (a) 40-kHz excitation, (b) 60-kHz excitation, (c) 80-kHz excitation, and (d) 100-kHz excitation.

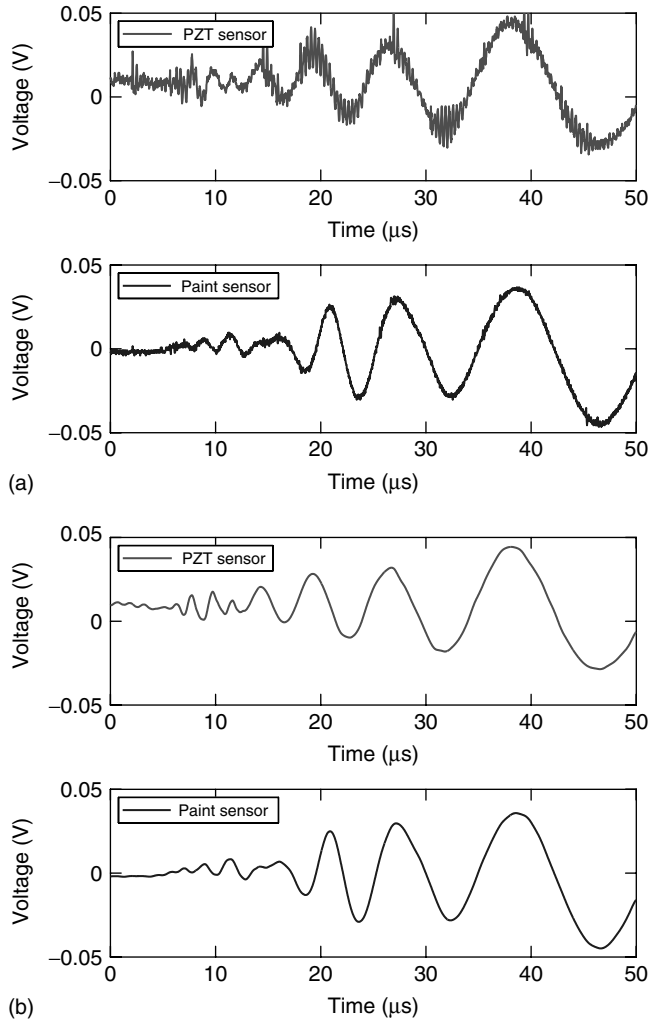


Figure 4. Acoustic emission signal received by piezoelectric sensors in pencil break test: (a) raw data and (b) DWT-filtered data.

AE signals are shown in Figure 4(b). The piezoelectric paint sensor (with amplifier) gave a peak-to-peak response close to that of the PZT sensor (without amplifier) both at the amplitude of 70 mV peak-to-peak. The two signals agree with each other very well.

5 CONCLUSIONS

The state of the art in piezoelectric paint sensor technology and its potential application to ultrasonic-based damage detection are reviewed in this paper. Piezoelectric paint is a composite piezoelectric material with adjustable material properties that are not easily attainable in a single-phase material. Through judicious selection of the polymer matrix, the composite properties of the piezoelectric paint can be tailored to meet the specific requirements of an application condition. Various formulations for piezoelectric paints have been developed over the last 20 years, and its ability to sense ultrasonic signals has been experimentally verified. Compared with conventional piezoelectric ceramics and piezoelectric polymers, piezoelectric paint has several advantages

for ultrasonic sensing application: (i) it enables true distributed sensing since piezoelectric paint can be applied to the surface of the structure in potential hot spots; (ii) painting represents an inexpensive way to apply this novel sensor technology over large surface areas of a host structure, especially for those with curved surfaces or complex geometries; and (iii) piezoelectric paint is more thermally stable than piezoelectric polymer such as PVDF. Additionally, a thickness range varying between 60 μm and 2 mm for piezoelectric paint film can be obtained, while the thickness of PVDF film is limited because of its high poling voltage.

Preliminary results from ultrasonic tests including pitch–catch tests and simulated AE tests that were carried out to verify the ultrasonic sensing capability of piezoelectric paint are reported in this paper and based on these preliminary results, piezoelectric paint sensor appears to offer a promising embedded sensor technology for ultrasonic-based NDE of structural defects or damages in either metallic or composite structure. However, extensive work still needs to be done to fully investigate the performance of piezoelectric paint sensor technology before this technique can be satisfactorily used in practical applications.

REFERENCES

[1] Giurgiutiu V, Cuc A. Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention. *The Shock and Vibration Digest* 2005 **37**(2):83–105.

[2] Jata KV. Piezoelectric transducers. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons: Chichester, 2008.

[3] Ayres JW, Lalande F, Chaudhry Z, Rogers CA. Qualitative impedance based health monitoring of civil infrastructures. *Smart Materials and Structures* 1998 **7**(5):599–605.

[4] Park G, Sohn H, Farrar CR, Inman DJ. Overview of piezoelectric impedance-based health monitoring and path forward. *The Shock and Vibration Digest* 2003 **35**(6):451–463.

[5] Giurgiutiu V. Embedded ultrasonics NDE with piezoelectric wafer active sensors. *Journal of Instrumentation, Mesure, Metrologie* 2003 **3**(3–4): 149–180.

[6] Yu L. Piezoelectric wafer active sensors. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons: Chichester, 2008.

[7] Ikeda T. *Fundamentals of Piezoelectricity*. Oxford University Press: Oxford, 1990.

[8] Polla DL, Francis LF. Processing and characterization of piezoelectric materials and integration into microelectromechanical systems. *Annual Review of Materials Science* 1998 **28**:563–597.

[9] Damjanovic D, Muralt P, Setter N. Ferroelectric sensors. *IEEE Sensors Journal* 2001 **1**(3):191–206.

[10] Niezrecki C, Brei D, Balakrishnan S, Moskalik A. Piezoelectric actuation: state of the art. *The Shock and Vibration Digest* 2001 **33**(4):269–280.

[11] Gautschi G. *Piezoelectric Sensorics*. Springer-Verlag: Berlin, 2002.

[12] Hanner KA, Safari A, Newnham RE, Runt J. Thin film 0–3 polymer/piezoelectric ceramic composites: piezoelectric paints. *Ferroelectrics* 1989 **100**: 255–260.

[13] Egusa S, Iwasawa N. Piezoelectric paints: preparation and application as built-in vibration sensors of structural materials. *Journal of Materials Science* 1993 **28**:1667–1672.

[14] Wenger MP, Blanas P, Shuford RJ, Das-Gupta DK. Acoustic emission signal detection by ceramic/polymer composite piezoelectrets embedded in glass-epoxy laminates. *Polymer Engineering and Science* 1996 **36**(24):2945–2954.

[15] Hale JM, Tuck J. A novel thick-film strain transducer using piezoelectric paint. *Proceedings of the Institution of Mechanical Engineers, Part C* 1999 **213**:613–622.

[16] Papakostas T, White N. Screen printable polymer piezoelectrics. *Sensor Review* 2000 **20**(2):135–138.

[17] Badcock RA, Birt EA. The use of 0–3 piezocomposite embedded Lamb wave sensors for detection of damage in advanced fiber composites. *Smart Materials and Structures* 2000 **9**:291–297.

[18] Sakamoto WK, de Souzab E, Das-Gupta DK. Electroactive properties of flexible piezoelectric composites. *Materials Research* 2001 **4**(3):201–204.

[19] Li Z, Zhang D, Wu K. Cement-based 0–3 piezoelectric composites. *Journal of the American Ceramic Society* 2002 **85**(2):305–313.

[20] White JR, de Poumeyrol B, Hale JM, Stephenson R. Piezoelectric paint: ceramic-polymer composites for vibration sensors. *Journal of Materials Science* 2004 **39**(9):3105–3114.

[21] Zhang Y. In situ fatigue crack detection using piezoelectric paint sensor. *Journal of Intelligent Material Systems and Structures* 2006 **17**(10):843–852.

[22] Lahtinen R, Muukkonen T, Koskinen J, Hannula S-P, Heczko O. A piezopaint-based sensor for monitoring structure dynamics. *Smart Materials and Structures* 2007 **16**:2571–2576.

[23] Hale JM, White JR, Stephenson R, Liu F. Development of piezoelectric paint thick-film vibration sensors. *Proceedings of the Institution of Mechanical Engineers, Part C* 2005 **219**:1–9.

[24] Zhang Y, Li X. Test-bed implementation of piezopaint-based acoustic emission sensor for crack initiation monitoring. *Proceedings of the 4th International Conference on Bridge Maintenance, Safety, and Management (IABMAS'08)*. Seoul, 13–17 July 2008; accepted for publication.

[25] Kobayashi M, Jen C-K, Moisan JF, Mrad N, Nguyen SB. Integrated ultrasonic transducers made by the sol-gel spray technique for structural health monitoring. *Smart Materials and Structures* 2007 **16**:317–322.

[26] Safari A. Development of piezoelectric composites for transducer. *Journal de Physique III* 1994 **4**:1129–1149.

[27] Chilton JA. Electroactive composites. In *Special Polymers for Electronics and Optoelectronics*, Chilton JA, Goosey MT (eds). Chapman & Hall: London, 1995.

[28] Gomez TE, Espinosa FM, Levassort F, Lethiecq M, James A, Ringgard E, Millar CE, Hawkins P. Ceramic powder—polymer piezocomposites for electroacoustic transduction: modeling and design. *Ultrasonics* 1998 **36**:907–923.

[29] Tresseler JF, Alkoy S, Newnham RE. Piezoelectric sensors and sensor materials. *Journal of Electroceramics* 1998 **2**(4):257–272.

[30] Dias CJ, Igreja R, Marat-Mendes R, Inacio P, Marat-Mendes JN, Das-Gupta DK. Recent advances in ceramic-polymer composite electrets. *IEEE Transactions on Dielectrics and Electrical Insulation* 2004 **11**(5):35–40.

[31] Akdogan EK, Allahverdi M, Safari A. Piezoelectric composites for sensor and actuator applications. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2005 **52**(5):746–775.

[32] Newnham RE, Skinner DP, Cross LE. Connectivity and piezoelectric-pyroelectric composites. *Materials Research Bulletin* 1978 **13**:525–536.

[33] Marin-Franch P, Martin T, Fernandez-Perez O, Tunnicliffe DL, Das-Gupta DK. Evaluation of PTCa/PEKK composites for acoustic emission detection. *IEEE Transactions on Dielectrics and Electrical Insulation* 2004 **11**(1):50–55.

[34] Han KH, Safari A, Riman RE. Colloid processing for improved piezoelectric properties of flexible 0–3 ceramic/polymer composites. *Journal of the American Ceramic Society* 1991 **74**(7):1699–1702.

[35] Lau ST, Li K, Chan HLW. PT/P(VDF-TrFE) nanocomposites for ultrasonic transducer applications. *Ferroelectrics* 2004 **304**:19–22.

[36] Waller D, Safari A. Corona poling of PZT ceramics and flexible piezoelectric composites. *Ferroelectrics* 1988 **87**:189–195.

[37] Sa-Gong G, Safari A, Jang SJ, Newnham RE. Poling flexible piezoelectric composite. *Ferroelectrics Letters* 1986 **5**:131–142.

[38] Egusa S, Iwasawa N. Piezoelectric paints as one approach to smart structural materials with health-monitoring capabilities. *Smart Materials and Structures* 1998 **7**:438–445.

[39] Egusa S, Iwasawa N. Application of piezoelectric paints to damage detection in structural materials. *Journal of Reinforced Plastics and Composites* 1996 **15**:806–817.

[40] Wenger MP, Blanas P, Shuford RJ, Das-Gupta DK. Characterization and evaluation of piezoelectric composite bimorphs for *in situ* acoustic emission sensors. *Polymer Engineering and Science* 1999 **39**(3):508–518.

[41] Sakamoto WK, Marin-Franch P, Tunnicliffe D, Das-Gupta DK. Lead zirconate titanate/polyurethane (PZT/PU) composite for acoustic emission sensors. *IEEE Annual Conference on Electrical Insulation and Dielectric Phenomena*. Kitchener, 2001; pp. 20–23.

[42] Kobayashi M, Olding TR, Sayer M, Jen C-K. Piezoelectric thick film ultrasonic transducers fabricated by a sol-gel spray technique. *Ultrasonics* 2002 **39**:675–680.

[43] Hale JM, Lahtinen R. Piezoelectric paint: effects of harsh weathering on aging. *Plastics Rubber and Composites* 2007 **36**(9):419–422.

[44] Sirohi J, Chopra I. Fundamental understanding of piezoelectric strain sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**(4): 246–257.

[45] Ono K. Current understanding of mechanisms of acoustic emission. *Journal of Strain Analysis* 2005 **40**(1):1–14; special issue.

[46] Gorman MR. Acoustic emission in structural health monitoring. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons: Chichester, 2008.

[47] Wevers M. Applications of acoustic emission for structural health monitoring: a review. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons: Chichester, 2008.

[48] Li X, Zhang Y. Piezoelectric paint sensor for ultrasonic NDE. *Proceedings of the Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. SPIE: San Diego, CA, 19–21 March 2007; SPIE Vol. 6529.

Chapter 69
Wireless Sensor Network Platforms

Reinhard Bischoff, Jonas Meyer and Glauco Feltrin
Structural Engineering Research Laboratory, Empa, Swiss Federal Laboratories for Materials Testing and Research, Dübendorf, Switzerland

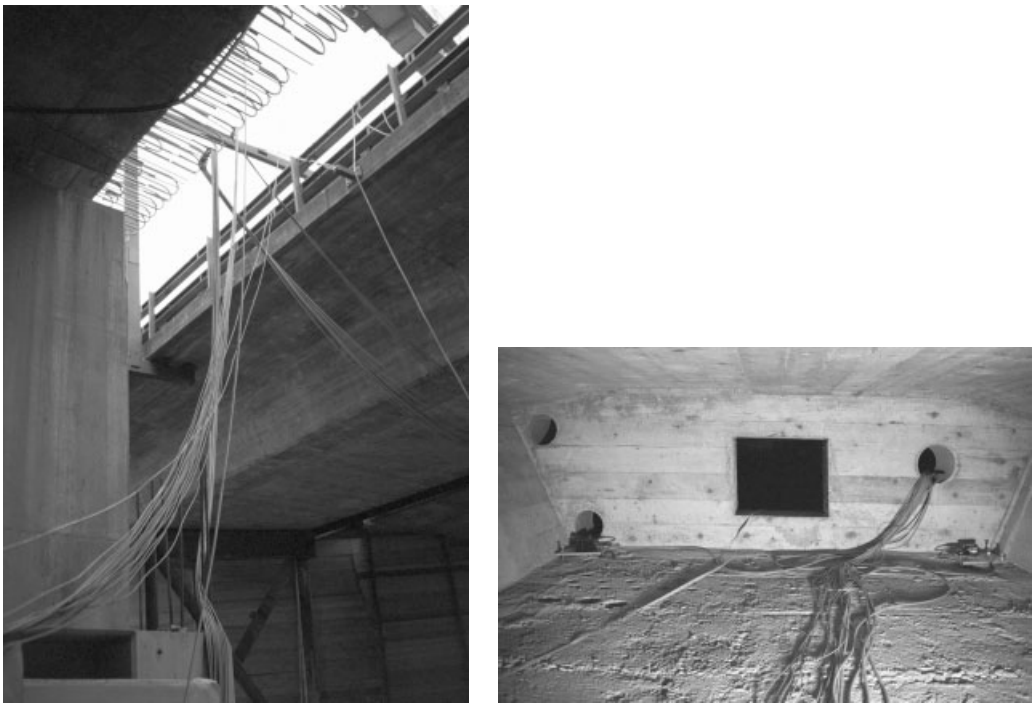
1 INTRODUCTION

Basically, every structural health monitoring (SHM) system is made up of various sensors measuring specific physical parameters, a data acquisition unit, and a storage device to save the acquired data. Traditional SHM systems show a starlike topology where each deployed sensor is connected via long cable runs to a central computer acting as data acquisition and storage device. The installation of such systems tends to be time consuming and therefore expensive (Figure 1). Especially in the field of civil engineering where the structures are typically large, the sensors can be located long way away from the data acquisition unit, resulting in high installation costs. These costs have proved to be a major issue, preventing a broad application of monitoring techniques to large-scale infrastructure. Furthermore, long cable runs are prone to pick up noise, reducing the effective accuracy of the acquired data, and hence require expensive high-quality cables. Moreover, these cables are susceptible to mechanical damage involving considerable maintenance efforts. Cabled systems also tend to offer a limited flexibility in terms of rearrangement of sensors and scalability. The adoption of wireless sensor network (WSN) techniques to SHM applications promises to overcome these drawbacks.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1.1 Wireless sensor network

A WSN is essentially a computer network consisting of many small, intercommunicating computers equipped with one or several sensors. Each small computer represents a node of the network. These nodes are called sensor nodes or motes. The communication within the network is established using radio frequency transmission techniques. The sensor nodes typically form a multihop mesh network by establishing communication links to neighbor nodes. Multihop networks offer different advantages when monitoring data has to be transmitted over long distances. Mainly, the network robustness to sensor node failure and the high power efficiency [1] make multihop networks attractive for monitoring applications. Figure 2 illustrates a schematic multihop



(a) (b)

Figure 1. Traditional, wired SHM installation.

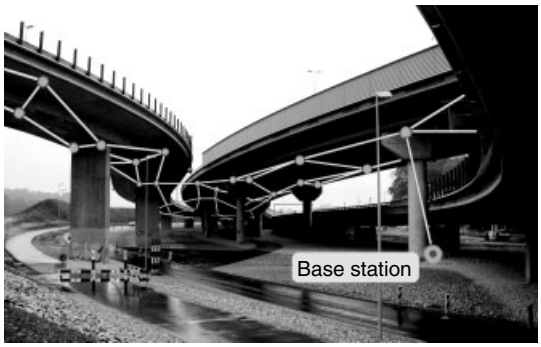


Figure 2. Wireless sensor network deployed on a road bridge. The spots illustrate the sensor nodes, the straight lines the communication links.

network deployed on a road bridge. The network consists of several dozens of sensor nodes. Theoretically, the number of sensor nodes is unlimited. All sensor nodes are equipped with specific sensors tailored to their measurement tasks. On the one hand, these nodes act as data sources, and on the other hand, they act as relaying stations, receiving and forwarding data from adjacent nodes. One or more particular sensor nodes act as base station and represent the data sink in the network. It aggregates all the data generated within the network. In addition, the base station establishes a communication link to a data logging unit or a remote site (e.g., control center), using standard wired or wireless communication technologies like universal mobile telecommunications system (UMTS) or wireless local area network (WLAN).

The initial research into WSNs was mainly driven by military applications like battlefield reconnaissance and surveillance, nuclear, biological, and chemical attack detection, etc. These projects focused on *ad hoc*, multihop WSNs that consisted of thousands of immobile nodes randomly distributed over a large geographical area (e.g., Smart Dust). The nodes were tiny (hardly noticeable), severely resource constrained, and homogeneous (identical hard- and software). Subsequently, the emergence of civilian applications of WSNs in different fields (environmental monitoring, home automation, health applications, production, inventory, delivery control,

etc.) produced a significant diversification of requirements with respect to deployment, mobility, size, cost, network topology, lifetime, etc., and therefore a flourishing of academic and commercial WSN platforms. To cope with these requirements, the platforms increased in size, computational resources, and hardware, as well as in software complexity.

The first commercial platforms appeared in the late 1990s. The most important platform was Crossbow's Rene mote, which emerged from the weC mote developed at the University of California, Berkeley, and which evolved later to the popular Mica platform. These platforms were the precursors of the recent Mica2 and MicaZ platforms (Table 1). A major reason for the popularity of Crossbow's early mote platforms was their open source policy with both hard- and software design open to the public. This policy built the base for the widespread diffusion of TinyOS as operating system for WSNs. Today, various commercial platforms with different characteristics in terms of computing resources, sensor interfaces, software architecture, etc., are available, which allow to cope with a wide spectrum of civilian applications.

2 HARDWARE ARCHITECTURES

The sensor nodes are the fundamental components of a WSN. To enable WSN-based SHM applications, the sensor nodes have to provide the following basic functionality (Figure 3):

- signal conditioning and data acquisition for different sensors;
- temporary storage of the acquired data;
- processing of the data;
- analysis of the processed data for diagnosis and, potentially, alert generation;
- self monitoring (e.g., supply voltage);
- scheduling and execution of the measurement tasks;
- management of the sensor node configuration (e.g., changing the sampling rate and reprogramming of data processing algorithms);
- reception, transmission, and forwarding of data packets;
- coordination and management of communication and networking.

2.1 General architecture

To provide the functionality described above, a sensor node is composed of one or more sensors, a signal conditioning unit, an analog-to-digital conversion module (ADC), a processing unit with memory, a radio transceiver, and a power supply (Figure 4).

If the sensor nodes are actually deployed in the field, especially in harsh environments like construction sites, they have to be protected against chemical and mechanical impacts. Therefore, an adequate packaging of the hardware is required (*see* **Microelectromechanical Systems (MEMS)**).

2.2 Hardware platform categories

Sensor node hardware platforms can be divided into three categories [2]. Each category shows a different hardware setup matched to diverse monitoring applications.

- **Adapted general-purpose computers**

These platforms are low-power personal computers (PCs), embedded PCs, and personal digital assistants (PDAs). These platforms mainly run on Windows CE, Linux, or other operating systems developed for mobile devices. These platforms are predominantly equipped with standard wireless communication devices like Wireless LAN (IEEE 802.11) and/or Bluetooth (IEEE 802.15.1). Because of the high processing ability and the high bandwidth communication, these platforms offer the opportunity to use higher level programming languages, which makes it easier to develop and implement software components. But in turn, they consume a considerable amount of energy and this can be prohibitive in some application scenarios. Additionally, they support networking protocols like Internet Protocol (IP). This simplifies the integration into a monitoring system.

- **Embedded sensor modules**

These platforms are assembled from commercial off-the-shelf (COTS) Chips. Using COTS offers several benefits. These components are widely used, making them cheap because of big production quantities, and are well supported by the manufacturers and communities. The microcontroller unit (MCU) of these platforms is mostly programmed in C. This

Table 1. Selection of wireless sensor network platforms

Name		Tmote	Mica2	MicaZ	Imote2	JN5121	Sun SPOT	Agile (V-Link)	BTnode rev3
MCU	Chip manufacturer	Texas Instrument	Atmel	Atmel	Intel	OpenCores	ARM		Atmel
	Chip model	MSP430F1611	ATMega 128L	ATMega 128L	PXA271 XScale	OpenRISC1000	ARM920T		ATmega 128L
	Frequency (MHz)	8	7.383	7.383	13–416	16	180		7.383
	Type (bit)	16	8	8	32	32	32		8
	ROM, RAM (kB)	48, 10	128, 4	128, 4	32 MB, 32 MB	64, 96	4M, 512		64 + 180, 128
	Interfaces	I^2C, UART, SPI	I^2C, UART, SPI	I^2C, UART, SPI	UART, SPI, I^2C, AC97, I2S, Camera	SPI, UART			ISP, UART, SPI, I^2C
	A/D, D/A	8, 2	8, 0	8, 0					
Data acquisition	A/D channels	8	8	8		4	6	8	
	Maximum sampling rate (kHz)		1	1				2	
	Resolution (bit)	12	10	10		12		12	
	D/A channels	2				2			
	Maximum sampling rate (kHz)								
	Resolution	12				11			

Radio	Chip manufacturer	Chipcon	Chipcon	Chipcon	Chipcon		Chipcon		Zeevo, Chipcon
	Chip model	CC2420	CC1000	CC2420	CC2420		CC2420		ZV4002, CC1000
	Frequencies (kHz)	2400	310, 433 or 868/916	2400	2400	2400	2400		433 or 868/916, 2400
	Raw data rate (kbps)	250	38.4	250	250		250		
	Standard (IEEE)	802.15.4		802.15.4	802.15.4	802.15.4, ZigBee	802.15.4	802.15.4	Bluetooth, 802.15.1
	Range outdoor (m)[a]	125		100	30			70	
External memory	Chip manufacturer	ST	Atmel	Atmel					
	Chip model	M25P80	AT45DB41B	AT45DB41B					
	Size (kB)	1024	512	512				2048	
Power	Supply voltage min, max (V)	2.1, 3.6	2.7, 3.3	2.7, 3.3	3.2, 5	2.7, 3.6	3.7	3.2, 9	0.85, 5
	Current consumption (normal, radio off) (mA)[b]	21.8, 1.8	39, 12	29.4, 12	44–66, 31	50, 5	90, 25	25, 25	32, 12
Dimensions	(cm × cm × cm)	6.6 × 3.2 × 0.7	5.8 × 3.2 × 0.7	5.8 × 3.2 × 0.7	4.8, 3.6, 0.9	3.0, 1.8, 1.0	3.5, 2.5	7.2, 6.5, 2.4	5.8, 3.3
Manufacturer		Moteiv	Crossbow	Crossbow	Crossbow	Jennic	Sun Microsystems	Microstrain	ETH Zürich

[a] Using integrated antenna.
[b] Values declared by manufacturer or typical datasheet values (power consumption computed by individual summation of system core, flash memory, and radio component values).

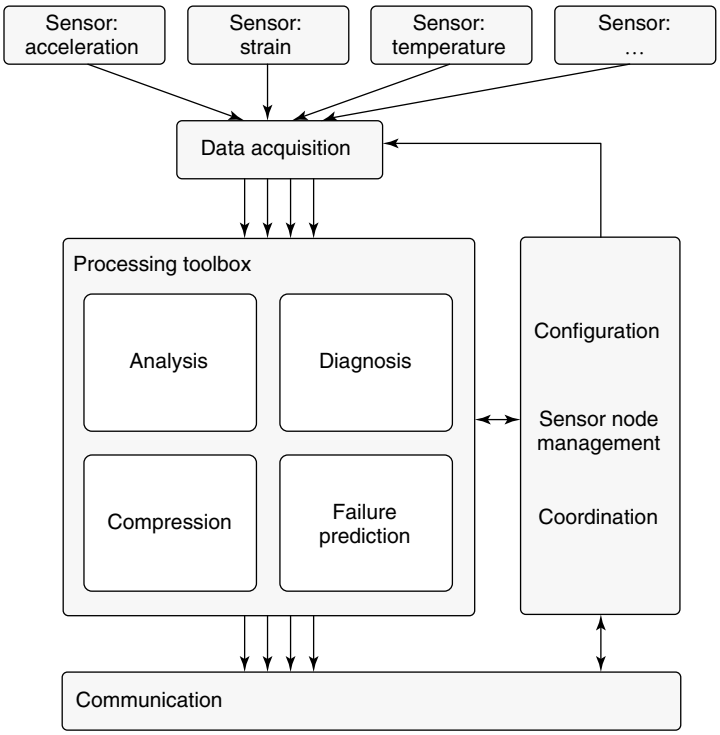


Figure 3. Basic functionality of a sensor node.

enables the development of a tight code that fits the limited memory size. Application developers have full access to hardware, but at the same time need to take care of all the resources. Examples from this category are Tmote from Moteiv, Mica2, MicaZ (Mica family), and Imote2 from Crossbow.

- **System on chip (SoC)**

These platforms integrate micro electromechanical systems (MEMS) sensors, microcontroller, and wireless transceiver technologies on one chip, an application-specific integrated circuit (ASIC). Because of this integration, platforms of this category are extreme low-power devices and have a small footprint/size. The smart dust node [3] is such an example.

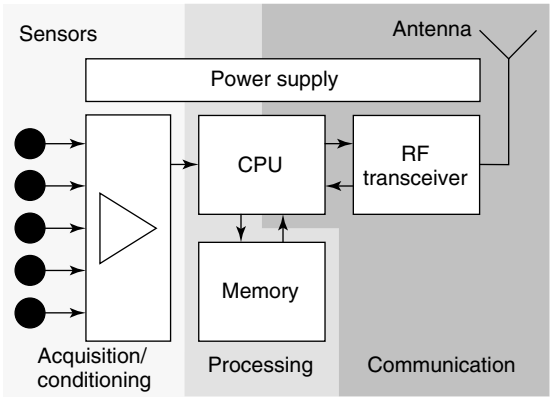


Figure 4. Sensor node (mote): hardware structure of a sensor node.

2.3 Energy-related aspects

The advantages of WSNs over wired sensing systems only have an effect if an unattended operation of the motes for a reasonably long period of time can be achieved. In terms of energy resources, this calls for a self-contained power source and has shown to be the most restrictive requirement in WSN applications. One approach to provide sufficient energy to operate the device over the desired period is to estimate the

total amount of energy that will be consumed and to equip the mote with adequately dimensioned energy storage upon deployment. Although this approach is a viable solution for some lowest-duty cycle applications, the required energy storage tends to become too big for long-term SHM systems. For monitoring applications that target overall system lifetime of several years, the dissipated energy has to be regenerated from mote-extern sources. This process is referred to as energy harvesting or scavenging. These sources absorb energy from the mote's environment and convert it to electrical energy.

Numerous different types of energy storage and harvesting concepts exist, but only the most important ones for SHM systems are presented here. An overview of other, partly inconvenient concepts is given in [4].

2.3.1 Communication versus computation

In terms of power consumption, wireless data transmission is much more expensive than data processing. To extend system lifetime, it is preferable to preprocess the raw sensor readings to reduce the data items needed to be transmitted to the base station. Many recent WSN-based SHM systems transmit the raw data streams to the base station and analyze them in the traditional centralized way. Without introducing huge batteries, this is not a viable solution if a system lifetime of several months to years is targeted. For long-term monitoring applications, distributed analysis algorithms have to be introduced, which allow for decentralized data reduction or even condition assessment.

2.3.2 Energy storage devices

- **Batteries**

The most popular energy storage is batteries. Many battery types with different characteristics have been developed. Every battery has its own advantages and drawbacks and the suitable battery technology has to be selected according to the application requirements. Rechargeable batteries are utilized if energy is harvested from the mote's environment.

- **Ultracapacitors**

The features of supercapacitors lie between those of capacitors and rechargeable batteries. Supercapacitors exhibit virtually unlimited charge–discharge cycles like capacitors, but offer a much higher capacity. These components are adequate as energy storage if it is emptied and replenished in short intervals.

- **Fuel cells**

This is a more recent but promising technology. Fuel cells oxidize hydrogen or hydrocarbon fuels and convert the heat into electrical energy. Currently, the commercially available fuel cells are too big in terms of size and converted energy to be applied to motes. However, much effort is being put into the development of small fuel cells for laptops and mobile phones. These devices will suit WSN applications well.

2.3.3 Energy harvesting and scavenging devices

Because of their nature, environmental energy scavenging devices do not provide a constant energy flow. Therefore, these devices are predominantly operated in conjunction with a storage device like a supercapacitor or a rechargeable battery. It stores excess energy and provides it later, when not enough can be harvested from the environment.

- **Solar cells**

The most popular energy scavenging sources are solar cells. A reasonably small panel delivers enough energy to power a sensor node. Solar cells are predominately operated in conjunction with a supercapacitor or a rechargeable battery. This energy storage is needed to provide energy, when the panel does not. Obviously, solar cells are only an option for outdoor applications.

- **Wind mills**

More unusual energy scavenging devices are small-scale wind mills or turbines. Like solar cells, this concept is only suitable for outdoor applications.

- **Vibration**

An energy harvesting method that is considered for civil engineering applications is to convert vibration energy. Civil engineering structures contain a lot of vibration energy, but it is extremely hard to extract it. The energy levels that current prototypes provide are far to low for monitoring applications. But it could evolve to an interesting source in the future.

2.4 Overview of recent architectures

Various hardware platforms for WSNs are available today and new ones emerge regularly. This diversity offers the possibility to choose a platform that best fits the needs of a specific application. An overview of recently used platforms is given in Table 1. This table only shows a selection. Further platforms are presented in [5] and regularly updated lists in the Internet can be found at [6–9].

2.5 Tmote Sky from Moteiv Corporation

Tmote [10] from Moteiv Corporation is presented as an example of a popular WSN platform (Figure 5). Many comparable platforms with similar hardware setups exist today. All these platforms are based on the Texas Instruments microcontroller family MSP430 and the Chipcon radio CC2420.

The main components of Tmote sensor node platform are the TI MSP430F1611 microcontroller, the FTDI FT232BM USB interface, which allows for programming the microcontroller over USB, and the Chipcon CC2420 low-power radio chip for the wireless communication.

The ultra-low-power microcontroller features 10 kB of RAM and 48 kB of program memory (flash). This 16-bit processor features several power-down modes with extremely low sleep-current consumption that permits the sensor node to run for a long period of time from a limited energy resource. The MSP430 has an internal, digitally controlled oscillator (DCO) that may operate up to 8 MHz. The microcontroller may be turned on from sleep mode in 6 µs, which allows for short reaction time upon the occurrence of an event. When the DCO is off, the MSP430 is clocked from an external 32 768-Hz watch crystal.

The MSP430 has eight external 12-bit ADC ports of which six are accessible on a pin header on the Tmote. The ADC input ranges from 0 to 3.0 V. The maximum total sampling rate for all ports is 200 kHz at 12-bit resolution. The internal ADC ports may be used to monitor the internal processor temperature and the supply voltage. A variety of peripherals are available, including serial peripheral interface (SPI) and universal asynchronous receiver/transmitter (UART), enabling the communication to digital output sensors, digital I/O ports, a watchdog timer, and timers with capture and compare functionality. The I^2C port, which is also integrated into the microcontroller, is mainly used to communicate to additional sensors and signal conditioning boards. The MSP430 also includes a 2-port 12-bit digital-to-analog converter (DAC) module, a supply-voltage supervisor, and a 3-port direct memory access (DMA) controller. Detailed features of the MSP430F1611 are presented in the Texas Instruments MSP430x1xx Family User's Guide [11].

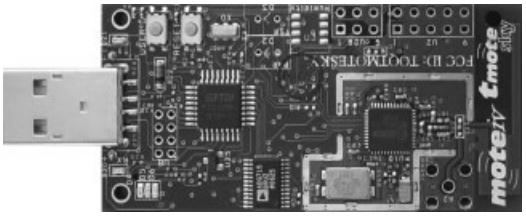


Figure 5. Picture of a Tmote Sky form Moteiv (top view).

The Tmote platform is equipped with the Chipcon CC2420 radio, enabling IEEE802.15.4 standard compliant wireless communication. It offers reliable wireless communication and power management capabilities to ensure low-power consumption. The CC2420 is controlled by the TI MSP430 microcontroller through the SPI port and a series of digital I/O. The radio may be shut off by the microcontroller for reducing the power consumption. The CC2420 provides a digital receive signal strength indicator (RSSI) that may be read at any time. The programmable transmitter output power enables to optimize the power consumption. The theoretically achievable maximum data throughput rate of the system is 250 kbps, without framing and packet headers.

3 SOFTWARE PLATFORMS

Unlike general-purpose operating systems for standard PCs such as Windows or Linux, the WSN software platforms are highly tailored to the limited node hardware. These WSN software frameworks are not full-blown operating systems, since they lack a powerful scheduler, memory management, and file system support. However, these frameworks are widely referred to as *WSN operating systems*. Therefore, this term is retained in the following section.

TinyOS [12], one of the most widespread operating systems, is presented in more detail in the following section. Other operating systems developed for WSNs are Contiki [13], Mantis [14], and SOS [15].

3.1 TinyOS

TinyOS is written in nesC [16], an extension to the C language, which supports event-driven component-based programming. The basic concept of component-based programming is to decompose the program into functionally self-contained components. These components interact by exchanging messages through interfaces. The components are event-driven. Events can originate from the environment (a certain sensor reading exceeds a threshold) or from other components, triggering a specific action. The main advantage of this component-based approach is the reusability of components.

The nesC language extension introduces several additional keywords to describe a TinyOS component and its interfaces. nesC and TinyOS are both Open Source projects supported by a fast growing community.

TinyOS has been ported to over a dozen WSN platforms (Table 1) and is also the native operating system of the presented Tmote platform. It provides a concurrency model and mechanisms for structuring, naming, and linking software components into a robust network embedded system. Today, TinyOS is a sort of *de facto* standard in WSN programming and widely used in the WSN community. As a result, a huge amount of software components for various sensors, network protocols, algorithms, and other WSN related topics is freely available on the Internet.

RELATED ARTICLES

Sensor Network Paradigms

Nondestructive Evaluation of Cooperative Structures (NDECS)

On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System

Energy Harvesting using Thermoelectric Materials

REFERENCES

[1] Karl H, Willig A. *Protocols and Architecture for Wireless Sensor Networks*, ISBN 0-470-09510-5. John Wiley & Sons: Chichester, 2005, pp. 60–62.

[2] Zhao F, Guibas L. *Wireless Sensor Networks: An Information Processing Approach*, ISBN 1-5860-914-8. Morgan Kaufmann: San Francisco, CA, 2004, pp. 240–245.

[3] Kahn JM, Katz RH, Pister K. Mobile networking for smart dust. *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 99)*. Seattle, WA, 17–19 August 1999.

[4] Roundy S, Steingart D, Frechette L, Wright P, Rabaey J. Power sources for wireless sensor networks. In *Proceedings of the 1st European Workshop on Wireless Sensor Networks (EWSN)*, *Lecture Notes in Computer Science*, Karl H, Willing A, Wolisz A (eds). Springer: Berlin/Heidelberg, 2004; Vol. 2920, pp. 1–17.

[5] Lynch JP, Loh K. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest*. Sage Publications, 2005, pp. 91–128.

[6] SNM—The Sensor Network Museum.™ http://www.btnode.ethz.ch/Projects/SensorNetworkMuseum.

[7] Bokareva T. *Mini Hardware Survey*, http://www.cse.unsw.edu.au/~sensar/hardware/hardware_survey.html.

[8] Body Sensor Networks. http://ubimon.doc.ic.ac.uk/bsn/m206.html.

[9] Wireless Sensor Network (WSN) Wiki. http://wsn.oversigma.com/wiki/index.php?title=WSN_Platforms.

[10] Polastre J, Szewczyk R, Culler D. *Telos: Enabling Ultra-Low Power Research*. Information Processing in Sensor Networks/SPOTS: Berkeley, 2005.

[11] MSP430x1xx Family User's Guide. http://focus.ti.com/lit/ug/slau049f/slau049f.pdf, 2006.

[12] Levis P, *et al.* TinyOS: an operating system for wireless sensor networks. *Ambient Intelligence*. Springer-Verlag: New York, 2005.

[13] Dunkels A, Gronvall B, Voigt T. Contiki—a lightweight and flexible operating system for tiny networked sensors. *Proceedings of the 29th Annual IEEE international Conference on Local Computer Networks (Lcn'04)*. LCN IEEE Computer Society: Washington, DC, 2004; pp. 455–462.

[14] Abrach H, Bhatti S, Carlson J, Dai H, Rose J, Sheth A, Shucker B, Deng J, Han R. MANTIS: system support for multimodal networks of in-situ sensors. *Proceedings 2nd ACM International Conference on Wireless Sensor Networks and Applications*. ACM Press: New York, 2003; pp. 50–59.

[15] Han C, Kumar R, Shea R, Kohler E, Srivastava M. SOS: a dynamic operating system for sensor nodes. *Proceedings of the 3rd international Conference on Mobile Systems, Applications, and Services (Seattle, Washington, June 06–08, 2005). MobiSys'05*. ACM Press: New York, 2005; pp. 163–176.

[16] Gay D, Levis P, von Behren R, Welsh M, Brewer E, Culler D. The nesC language: a holistic approach to networked embedded systems. *Proceedings of the ACM SIGPLAN 2003 Conference on Programming Language Design and Implementation*. San Diego, CA, 2003; pp. 1–11.

Chapter 70
Sensor Placement Optimization

Robert J. Barthorpe and Keith Worden
Department of Mechanical Engineering, University of Sheffield, Sheffield, UK

1 INTRODUCTION

The basic problem of fault detection is to deduce the existence of a defect in a structure from measurements taken at sensors distributed upon it. The quality of these measurements and thus the quality of structural health monitoring (SHM) achieved is, to a large extent, dependent upon where sensors are placed on the structure. Cost and practicality issues preclude the instrumentation of every point of interest on the structure and lead to the selection of a smaller set of measurement locations. The purpose of this article is to state the problem of sensor placement optimization (SPO) and describe the approaches that have been investigated for its solution. The following discussion focuses on sensor placement techniques for the category of SHM methods based upon the analysis of structural dynamics, although many of the issues raised are of wider relevance.

Traditionally, successful sensor placement has been heavily reliant upon the knowledge and experience of those performing the testing. Practical methods, for example, choosing locations near the antinodes of low-frequency vibration modes, are combined to create *ad hoc* sensor distributions. Where resources allow, several combinations of possible sensor configurations may be experimentally tested with the one that performs best chosen as the final design. While this is certainly a significant improvement on arbitrary placement, recent research has attempted to formalize the location process by casting it as a problem of optimization.

This research effort has led to the development of a variety of approaches to SPO, which are suited to different applications. An overview of the problem is given in Section 2, a review of the technical literature in Section 3, and an outline theory for a selection of the most influential techniques is given in Section 4.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

A case study to illustrate an application of SPO is presented in Section 5.

2 OVERVIEW OF THE SENSOR PLACEMENT OPTIMIZATION (SPO) PROBLEM

The aim of the sensor placement exercise can be stated as the need to select a subset of measurement locations from a large finite set of candidate locations, in order to represent the system as accurately as possible using the limited number of degrees of freedom (DoFs) available. This can be viewed as a three-step decision process:

1. Sensor quantity—How many sensors need to be installed on the structure to allow successful dynamic testing?
2. Sensor placement optimization—Where should these sensors be located to most accurately capture the required data?
3. Evaluation—How can the performance of different sensor configurations be measured?

In general, the first aspect would have been resolved during pretest planning. The minimum requirement for the system to be observable is that the number of sensors required cannot be less than the number of mode shapes to be identified, with an upper limit usually imposed either by the cost or availability of equipment. In practice, a greater number of sensors are likely to be required to allow the mode shapes to be visualized, and in cases where there are surplus sensors available, a decision must be made as to whether they are best used for improving visualization or as backup against sensor failure.

The second aspect is the area that has attracted the majority of research interest, and is the primary focus of this article. For the limited number of sensors available, the problem is the development of a suitable sensor placement performance measure to be optimized and the selection of an appropriate method with which it can be optimized. Some approaches require a single calculation to be performed, some are iterative, and many others take the form of an objective function to which an optimization technique must be applied. The majority of this article deals with presenting the available alternatives.

The third and final aspect includes several possibilities for assessing the performance of chosen sensor sets. While there is an ever-present temptation to leap into data collection at the earliest opportunity, time spent assessing the effectiveness of the chosen network is invariably well spent.

Throughout this article, the *candidate set* refers to the set of all DoFs that are available as sensor locations. The *measurement set* refers to the DoFs employed as sensor locations. The *full model* includes the measurements at all available points and the *reduced model* includes only the measurements available at the DoFs specified in the measurement set.

3 LITERATURE REVIEW

The problem of determining the optimum locations for sensor placement has been addressed from a variety of perspectives. It appears to have been first addressed by control engineers before finding broad application in the field of structural dynamics. In the recent years, there has been increasing interest in developing placement methods specific to the needs of SHM, typically the optimal identification of structural characteristics sensitive to damage. This development is reflected in the literature.

An influential class of techniques has emerged from the concept of assessing all the locations of the candidate sensor set against some objective function, and then iteratively deleting those sensors that perform least well until the required number of measurement locations remain. For the effective independence (EI) method introduced in [1], based upon earlier work in [2], the sensors are ranked according to their contribution in maintaining the determinant of the Fisher information matrix (FIM, described in Section 4). The location that contributes least at each iteration step is selected for deletion. The kinetic energy (KE) method, see for example [3], assumes that the sensors will have maximum observability of the modes of interest if the sensors are placed at points of maximum KE for that mode. It has often been noted that EI and KE methods can produce similar results, especially in structures with homogeneous mass distributions, and the inherent mathematical connection between the two methods is revealed in [4].

Further examples of the iterative approach include the eigenvalue vector product (EVP) [5], average

driving-point residue (ADPR) [6], effective independence driving-point residue (EI-DPR) [7], strain energy distribution [8], and modal assurance criterion (MAC)-based [9] methods. The comparative performance of several of the iterative techniques is investigated in [10, 11].

It should be noted that the iterative process does not necessarily need to be a reduction from the candidate set. In [12], an alternative EI approach is presented whereby the sensor set is iteratively expanded to include those sensor locations that offer the greatest increase in the determinant of the FIM. The expansion approach reduces the computational expense incurred when the candidate set is large and allows for any desired sensor locations to be specified at the outset, with the remaining sensors placed optimally.

In [13], optimal sensor placement is formulated as a mixed variable programming (MVP) problem. The emphasis is on the development of a general framework using MVP optimization that could be applied to any number of objective functions. The MVP formulation allows variables to be categorical, taking their values from a predefined set or list. For the sensor placement problem, a categorical position variable is defined for each sensor.

A further class of methods takes advantage of developments in the field of combinatorial optimization. Perhaps the first use of genetic algorithms (GAs) for the sensor placement problem is presented in [14]. The authors propose the GA as an alternative to the EI method, with the determinant of the FIM chosen as the objective function (the *fitness* function in GA terminology). In [15], the fitness function is taken as a product of two terms: the first term measures the fitness from the point of view of observability; the second component is geometric and penalizes the clustering of sensors.

In one of the first sensor placement approaches specific to SHM [16], a structural damage localization approach based on eigenvector sensitivity is adopted and an SPO approach using the same method is developed. The EI approach is applied to the sensitivity matrix that is to be used for damage localization, with the DoFs providing the greatest amount of information for localization retained. In [17], the difficulties that can occur in solving the sensitivity matrix are highlighted. The problem is reformulated and solved using an improved GA.

In [18], the selection of optimal sensor locations for impact detection in composite plates is approached using a GA to optimize a fitness function based upon the concept of mutual information. The concept is used to eliminate redundancies in information between selected sensors and rank them based on their remaining information content. In [19], an equally spaced configuration of measurement points is assumed, and the sensor spacing is varied to minimize the average mutual information between measurement locations. In [20], a statistical method for optimally placing sensors for the purposes of updating structural models for subsequent damage detection and localization is presented. The optimization is performed by minimizing information entropy, a unique measure of the uncertainty in the model parameters. The uncertainties are calculated using Bayesian techniques, and the minimization is realized using a GA.

In one of the first studies of sensor placement for an SHM problem, a neural network (NN) for the location and classification of faults and a GA for the determination of an optimal (or near optimal) sensor configuration were used [21]. The NN is trained using mode-shape curvatures provided by an FE model of a cantilever plate. The probabilities of misclassification for the different damage conditions are obtained from the NN, and the inverse of this measure is employed as a fitness function for the GA. The use of simulated annealing (SA) for the optimization has also been investigated [22].

In [23], an NN/GA approach was used to place sensors for the location and quantification of impacts on a composite plate. An artificial NN is trained to locate the point of impact, and a second NN employed to estimate the impact force. A GA was used to select an optimal set of measurement locations from the candidate set, using the estimation of the impact force provided by the NN as the parameter to be maximized. The work is extended to cover fail-safe sensor placements [24]. In [25], a new damage location method is presented that combines classical triangulation procedures with experimental wave velocity analysis and GA optimization.

The NN problem studied in [23] was investigated using a different optimization technique reported in [26]. Here, an ant colony metaphor was used to place sensors based upon the fitness functions generated by the NN, and the results were compared with those

from the GA and exhaustive search. This work is demonstrated in Section 5 (*see* **Artificial Neural Networks**).

4 THEORY

A large variety of performance indices have been developed for the problem of sensor placement, but it is only comparatively recently that the problem has been considered from an SHM perspective. Rather than attempting to cover this multiplicity of approaches in full, the intention of this section is to highlight some of the key considerations made in the selection and development of objective functions appropriate to the SHM practitioner. Several influential general approaches (notably the EI and KE methods) are covered, as are more recent techniques developed for the specific purpose of damage detection. The descriptions have been kept mathematically light intentionally; full descriptions of the algorithms may be found in the references.

4.1 Effective independence (EI)

The EI method makes use of the FIM, which offers a measure of the information that a sampled random variable contains about an unknown parameter; formally, Fisher information is the variance of the score with respect to the unknown parameter. Where there are multiple unknown parameters, it may be stated in matrix form with elements

$$(I(\boldsymbol{\theta}))_{ij} = E\left[\frac{\partial}{\partial\theta_i}\ln f(X;\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\ln f(X;\boldsymbol{\theta})\right] \quad (1)$$

where

- $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_N]$ is the vector of unknown parameters
- $(I(\theta))_{ij}$ is the Fisher information with respect to the unknown parameters θ_i and θ_j
- X is the sampled random variable
- $f(X;\boldsymbol{\theta}) = L(\boldsymbol{\theta})$ is the likelihood function of $\boldsymbol{\theta}$
- E denotes the expectation.

For the SPO problem, the target mode shapes may be regarded as the unknown, sought parameters, with the sampled data being that available from the given sensor distribution. Every DoF in the candidate set is ranked according to its contribution to the determinant of the FIM, and the lowest ranked DoF is eliminated. The new, reduced set is then re-ranked, and the process repeated in an iterative manner until the desired number of sensors remains. This is adopted as the optimal measurement set. Maintaining the determinant of the FIM leads to the selection of a set of sensor locations for which the mode shapes of interest are as linearly independent as possible, while retaining sufficient information about the target modal responses. The approach is based on the EI distribution vector $\mathbf{E}_{\mathrm{D}}$, defined as the *diagonal of the prediction matrix*, $\mathbf{E}$:

$$\mathbf{E} = [\boldsymbol{\Phi}]\{[\boldsymbol{\Phi}]^{\mathrm{T}}[\boldsymbol{\Phi}]\}^{-1}[\boldsymbol{\Phi}]^{\mathrm{T}} \quad (2)$$

where $\boldsymbol{\Phi}$ is the matrix of FE target modes, in this case partitioned according to a given sensor distribution. Each diagonal element is the fractional contribution of each sensor location to the rank of E, which can only be full rank if the target mode partitions are linearly independent. The algorithm is iterative; at each step, terms in $\mathbf{E}_{\mathrm{D}}$ are sorted to give the least important sensor, which is then deleted. The corresponding elements in $\boldsymbol{\Phi}$ are also deleted. The iteration concludes when the required number of sensors is obtained.

4.2 Average driving-point residue (ADPR)

A drawback of the EI approach is that the algorithm can select sensor locations that display low signal strength, making the system vulnerable to noisy conditions. The ADPR offers a measure of the contribution of any point to the overall modal response. If $j = 1 \ldots N$ modes of interest are to be measured and ω_j is the eigenvalue of jth mode, the ADPR at the ith DoF can be calculated from FE data as

$$ADPR_i = \sum_{i=j}^{N} \frac{\Phi_{ij}^2}{\omega_j} \quad (3)$$

4.3 Effective independence driving-point residue (EI-DPR)

The values given by the EI algorithm are weighted by the ADPR values to give the EI-DPR vector. For

the ith DoF

$$E_{D_i}^{\text{EI-DPR}} = E_{D_i}^{\text{EI}} ADPR_i \tag{4}$$

This adaptation leads to a greater likelihood of sensors being placed in areas of high signal strength. In addition to improving signal-to-noise ratios, this tends to result in the selection of relatively uniformly spaced sensor locations.

4.4 Kinetic energy method (KE)

The KE method assumes that the sensors will have maximum observability of the modes of interest if the sensors are placed at points of maximum KE for that mode, and accordingly ranks sensor locations based on their dynamic contribution to the target mode shapes. It follows a similar procedure to that used in the EI method, the key difference being that a KE measure, rather than the determinant of the FIM, is maximized. It is alternatively known as the *modal kinetic energy* (*MKE*) *method* or the *kinetic energy method* (*KEM*) in the literature.

KE indices are calculated for all candidate sensor locations as follows:

$$KE_{ij} = \Phi_{ij} \sum_s M_{is} \Phi_{sj} \omega_j^2 \tag{5}$$

where **M** is the mass matrix. The sensor locations that offer the highest KE indices are selected as the measurement locations. As the method selects those sensor locations with the largest available signal amplitudes, the signal-to-noise ratios tend to be high, making the method attractive for use in noisy conditions. However, in contrast to EI, the KE method does not consider the linear independence of the target modes, an important consideration for both modal identification and test–analysis correlation.

4.5 Eigenvalue vector product (EVP)

The EVP method computes the product of the eigenvector components for candidate sensor location for the range of modes to be measured N: a maximum for this product is a candidate measurement point. Some modification may be required if a point is a node of one of the modes. The EVP of the ith DoF is calculated as

$$EVP_i = \prod_{j=1}^{N} |\Phi_{ij}| \tag{6}$$

4.6 Mutual information

Mutual information gives a measure of how much information one sensor location "learns" from another. If there are two sets of measurement locations, A and B, the amount of information learned by a_i about b_j is represented by the mutual information $I(a_i, b_j)$,

$$I(a_i, b_j) = \log_2 \left[\frac{P_{\text{AB}}(a_i, b_j)}{P_{\text{A}}(a_i) P_{\text{B}}(b_j)} \right] \tag{7}$$

where

- a_i and b_j are the measurements from locations A and B, respectively
- $P_{\text{A}}(a_i)$ and $P_{\text{B}}(b_j)$ are the individual probability densities for A and B
- $P_{\text{AB}}(a_i, b_j)$ is the joint probability density for measurements A and B.

If the measurement of a_i is completely independent of the measurement of b_j, $I(a_i, b_j)$ becomes zero. The average mutual information between A and B is calculated by averaging over all the sensor locations, and the optimal sensor location determined by minimizing the mutual information between sensors.

4.7 Information entropy method

Optimal sensor placement is achieved by minimizing the change in the information entropy $H(D)$, given by

$$\begin{aligned} H(D) &= E_\theta[-\ln p(\theta|D)] \\ &= -\int p(\theta|D) \ln p(\theta|D) \, \mathrm{d}\theta \end{aligned} \tag{8}$$

where

- θ is the uncertain parameter set (e.g., stiffness parameters, modal parameters, etc.)

- D is the dynamic test data
- E_θ is the mathematical expectation with respect to θ.

A rigorous mathematical description is given in [27].

4.8 Sensitivity-based methods

In the SHM-specific method proposed in [16], the prediction matrix **E** used in the EI method is adapted to use the sensitivity matrix developed for damage location. The modified matrix is given by

$$E = \boldsymbol{F}(\boldsymbol{K})[\boldsymbol{F}(\boldsymbol{K})^T \boldsymbol{F}(\boldsymbol{K})]^{-1} \boldsymbol{F}(\boldsymbol{K})^{\mathrm{T}} \qquad (9)$$

where

- $\boldsymbol{F}(\boldsymbol{K})$ is the vector of sensitivity coefficients of the mode shape changes with respect to a damage vector.

As for the EI approach, the diagonal terms of E provide the fractional contribution of the corresponding measurement location to the rank of E. The location that contributes the least is removed, and the process is repeated iteratively until the required quantity of sensors remains.

5 CASE STUDY—SENSOR PLACEMENT OPTIMIZATION USING AN ANT COLONY METAPHOR

A comparatively recent addition to the canon of combinatorial optimization algorithms are those based on "ant colony metaphors". These are founded on the cooperative interaction of simple computational agents termed *ants*. The basic forms of the algorithms—"ant-density", "ant-quantity", and "ant-cycle"—were introduced in [28]. Of these basic forms, ant-cycle proved to be the most effective and this was renamed *ant-system* and discussed in more detail in [29]. The ant algorithms have proved to be a useful addition to the set of methods, combining aspects of greedy search with population-based cooperative search. The work put forward in [26] is presented here to illustrate an interesting approach to the problem of sensor placement.

5.1 The ant algorithm

Real ants are well known to be capable of finding the shortest path to a food source from their nest without using visual cues [30]. This is done by exploiting pheromone information. Ants deposit pheromone as they move and follow pheromone trails deposited by previous ants. If a trail has higher pheromone information, an ant will follow it in preference to other trails with less pheromone. In other words, an ant will follow a trail with higher pheromone, with higher probability. Consider Figure 1. Initially, the ants leave the nest and have no preference as to which path they will follow and they will choose a path with equal probability. Assuming that all ants travel with equal speed, the ants taking the lower path will reach the food before those taking the upper one. When returning to the nest, they will choose the lower path with higher probability because the upper path ants have not arrived yet to lay trail. This reinforces the pheromone path on the lower trail. When the lower path ants arrive back at the nest, new ants will initially see twice as much trail on the lower path and choose it with higher probability. It is not difficult to see that a positive feedback mechanism emerges, which reinforces the desirability of the lower trail. Eventually, all ants will follow the shorter path. This simple example illustrates the idea behind ant algorithms, which is that simple agents can communicate, using distributed memory about the problem, to cooperate and solve an optimization problem. Because the aim here and elsewhere in the engineering literature is to solve an engineering problem and not to model real ant colonies, various simplifications are assumed. Artificial ants differ from real ants in the following ways:

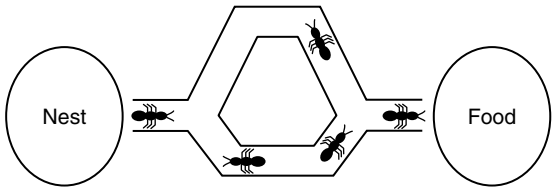


Figure 1. Ant foraging activity.

1. They are completely blind.
2. They have some memory.
3. They live in a discrete-time environment.

In order to explain the original ant-system algorithm, it is convenient to do so in the context of the classic traveling salesperson problem (TSP), which provided the original motivation for the algorithms. The basic TSP is formulated as follows: given N cities distributed randomly within the plane, find the shortest tour that visits each city only once and returns to the original city. The distance $d(i, j)$ between cities i and j with coordinates (x_i, y_i) and (x_j, y_j) is computed using the standard Euclidean norm. One can think of the problem in terms of a graph, where each pair of cities is potentially joined by an edge (i, j). The ants move around in this graph laying a pheromone trail with intensity $\tau(i, j)$ for each edge (i, j) they traverse. This trail can be updated locally as the ants move or can be updated globally at the end of an iteration when all ants have completed a tour. The global updating rule, which forms the basis of the ant-system algorithm, was shown to be superior to local rules in [29]. The ants are forced to complete a tour by maintaining a *tabu list* for each ant, which simply contains a list of the cities that the ant has already visited, and the ant is forbidden to revisit any city in the tabu list.

Given a state of the system at time t in a specific iteration, each ant k is at one of the cities. The probability that an ant will make the journey from city i to city j is assumed to be

$$P_k(i, j) = \frac{[\tau(i, j)]^\alpha [\eta(i, j)]^\beta}{\sum_{j \in J_k(t)} [\tau(i, j)]^\alpha [\eta(i, j)]^\beta} \tag{10}$$

where $J_k(t)$ are the allowed cities for ant k at time t, i.e., the complement of the tabu list. If j is not in $J_k(t)$, the transition is forbidden by setting $P_k(i, j) = 0$. In equation (10), $\tau(i, j)$ is the trail intensity associated with the edge (i, j). The *visibility* $\eta(i, j)$ is the inverse of the distance between cities i and j. This is included in order to allow a degree of greediness in the algorithm; the ants are more likely to move to a nearer city at a given step. This is not guaranteed to lead to a tour with overall minimum length as at step N; the ant may be very far from its starting city. The parameters α and β control the relative importance of trail and visibility. Note that the inclusion of visibility means that the ants are not totally blind in this variant of the algorithm. After N steps and the ants have completed a tour, the pheromone levels for each edge are updated using the rule

$$\tau(i, j) \longrightarrow \rho\tau(i, j) + \sum_{k=1}^{m} \tau_k(i, j) \tag{11}$$

where

$$\tau_k(i, j) = \frac{Q}{L_k} \tag{12}$$

if edge (i, j) is in the tour of the kth ant and $\tau_k(i, j) = 0$ otherwise. Q is a user-defined constant and L_k is the total tour length for ant k in that iteration. The total number of ants is m. The parameter ρ is also user defined. A number between 0 and 1 represents the trail persistence between iterations; $1 - \rho$ is a measure of trail *evaporation*.

This is how the ant-system algorithm is used to solve the TSP. In order to benchmark the code constructed to solve the sensor optimization problem here, the algorithm was applied to a known TSP problem—the Oliver30 problem. This was chosen as it was the problem used in the original ant-system paper [29]. Overall, 25 runs of the ant-system algorithm were carried out. The agreement between the runs was impressive. By 500 iterations, 16 of the runs had arrived at the best previously known solution, with a tour length of 423.74 (Figure 2).

Having established that the algorithm worked properly on the TSP problem, some modification was required before it could be applied to the SPO problem. The first important difference is that one is looking for an optimum subset of the candidate sensor locations and this means that a tour does not visit all locations. The second major difference is that the ants are truly blind. The sensor distributions can only be evaluated when all the members of the tour are known; there is no analog of visibility. The ant algorithm for SPO proceeds as follows. The ants are distributed randomly on the sensors with each edge initialized with the same trail intensity. At each step,

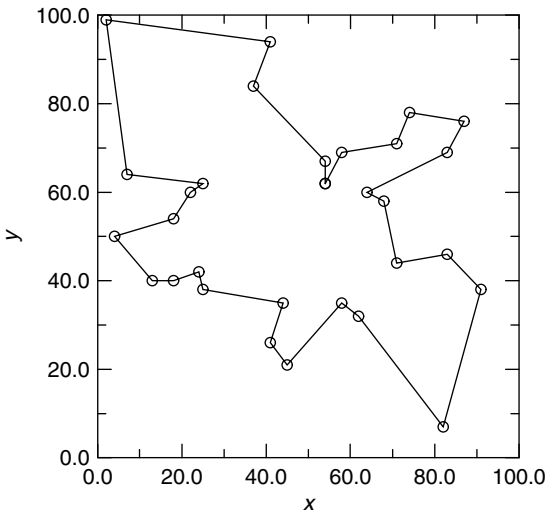


Figure 2. Best solution found to Oliver30 TSP problem.

each ant chooses a sensor to visit by means of a transition probability,

$$P_k(i, j) = \frac{\tau(i, j)}{\sum_{j \in J_k(t)} \tau(i, j)} \tag{13}$$

i.e., α is set arbitrarily to unity (and β is irrelevant). If the sensor has already been visited, transition is forbidden as before. The iteration ends when the required number of candidate locations has been visited. The (global) updating rule is very similar to the TSP variant,

$$\tau(i, j) \longrightarrow \rho\tau(i, j) + (1 - \rho) \sum_{k=1}^{m} \tau_k(i, j) \tag{14}$$

where $\tau_k(i, j) = F_k$, if (i, j) is in the tour for the kth ant and zero otherwise. F_k is the fitness of the kth tour, i.e., the value of the objective function to be maximized for the sensor distribution.

A little trial and error yielded 0.2 as a good initial trail level and the number of ants was taken as 10 simply because that led to optimum solutions in the Oliver30 TSP problem. The rationale behind the value of *persistence* ρ used is given later.

5.2 Impact test of the composite panel

A simple impact experiment was performed to study the effectiveness of NNs for the impact identification problem and also to experimentally investigate the optimal sensor location problem.

The structure under examination consisted of a rectangular 530 mm × 300 mm composite plate made from laminated carbon fibre reinforced plastic (CFRP) and four aluminum channels. The structure is shown in Figure 3. The top flanges of the channels were attached to the plate by a line of rivets; the bottom flanges were fixed rigidly with screws to a pneumatic measuring table. This box structure was intended to simulate the skin panel of an aircraft. The composite plate was instrumented with 17 piezoceramics (*PZT Sonox P5,* 15 mm × 15 mm) (*see* **Integrated Sensor Durability and Reliability**) fixed on the lower surface of the plate, the impacts being performed on the upper surface. The piezoceramics were used as strain sensors. Figure 4 shows the total distribution of the sensors used.

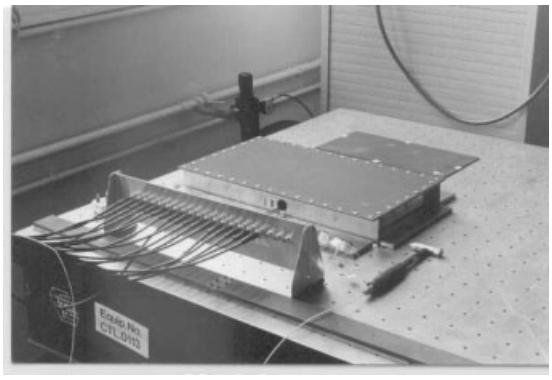


Figure 3. Composite box structure.

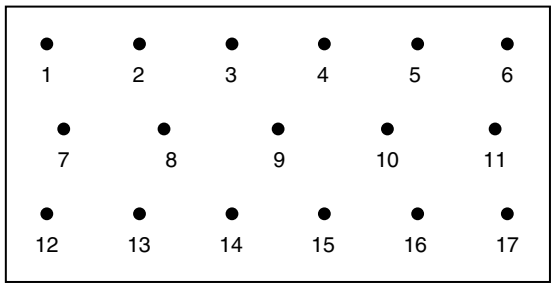


Figure 4. The candidate sensor locations.

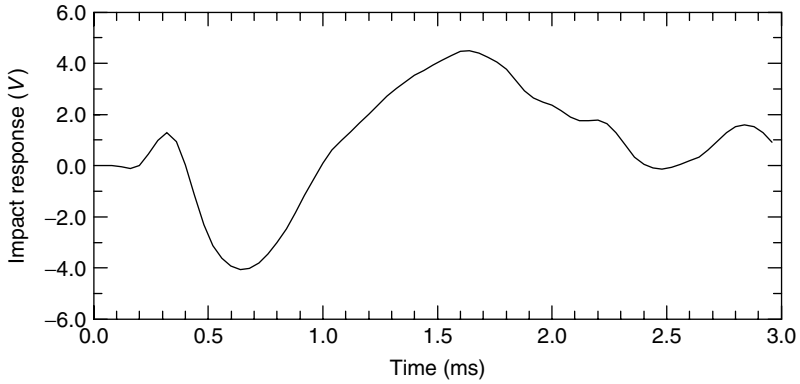


Figure 5. Example of a strain time-history.

The impacts were applied using a modally-tuned impact hammer. The levels of force applied were kept below 0.1 N in order not to damage the plate. As discussed below, two sets of measurements were made. The first comprised 80 impacts at random positions on the plate and was intended for use as network training data. The second set of measurements contained 95 impacts placed on a regular grid, and these were divided between the network validation and test sets: 48 in the former and 47 in the latter. In the validation and testing sets, each location was subjected to impacts of varying magnitudes.

The strain data were recorded using a DIFA SCADAS II 24 channel measuring system running the LMS 3.4.04 data-acquisition software. For each impact, 8192 samples were recorded at a frequency of 25 kHz. Figure 5 shows an example of the strain data recorded.

5.3 Impact identification strategy using neural networks

Before carrying out the subset optimization, a network training exercise was carried out using all 17 candidate sensors. This was in order to establish a lower bound for the error of the impact locator.

The NN paradigm used for this study was the standard multilayer perceptron (MLP) trained with the back-propagation learning rule. The particular implementation is described in more detail in [23].

The first problem in establishing the network analysis was to determine the appropriate features for the diagnosis, i.e., which data should be used to train the network. As described in [23], a preliminary study considered several different time and frequency domain features, namely, (i) time after impact of maximum response, (ii) magnitude of maximum response, (iii) peak-to-trough range of the response, and (iv) real and imaginary parts of the response *spectrum*, integrated over frequency. The features were investigated alone and in combination, and it was found that the best results were obtained using features A and B in tandem. This meant that each sensor contributed two features and the dimension of the pattern vectors for training was thus 34.

The next problem was to design the training strategy. A principled approach demanded the availability of three data sets. The first, the *training* set, is used to determine the network weights. The second, the *validation* set, is used to investigate the optimum structure for the network and the final *testing* set is used to assess the effectiveness of the optimized network. Part of the network training problem is to determine the best structure for the network, i.e., the number of layers and number of neurons per layer. The following approach was adopted as described in [31] and numerous structures were assessed. The number of neurons in the input and output layers of the network was fixed by the number of measurement features and diagnostic outputs, respectively. In this case, the network must have 34 inputs. As the network considered here was required to signal the location of damage, two outputs were required, namely, the x and y coordinates of the impact site. A single hidden layer was assumed, and the number of neurons in the layer was established by optimizing the network performance over the validation set; the

objective function used was the product of the mean x and y errors, i.e., the mean "area error". This was normalized to give a percentage of the plate area. Because the data sets are comparatively small, noise was added during training as a form of regularization. The root-mean-square (RMS) level of this noise was also established by minimizing the validation error.

For the location problem, the optimization procedure produced a minimum error over the validation set when there were eight neurons in the hidden layer and the RMS noise level was 0.1. When the corresponding network was evaluated on the testing set, the mean (modulus) of the x error was 23.1 mm and the mean y error was 25.7 mm. This gives an area corresponding to 1.5% of the plate area.

In order to have a tractable optimization problem, it was assumed that eight hidden units would also be the best choice when only a subset of the candidate sensor data was used for training. A similar assumption was made for the noise level. Following trial and error experiments, a training run of 100 000 iterations was applied.

5.4 Results

The algorithm summarized by equations (13) and (14) was used for the SPO problem here. The fitness F_k was evaluated as follows. If an ant k produced a certain tour at a given iteration, i.e., (1, 3, 7) in the search for an optimum three-sensor distribution, the NN was trained only with data from the sensors 1, 3, and 7. If this yielded an area error δA, the fitness was returned as $F_k = 1/\delta A$. Because the network structure was fixed, δA was taken from the validation set.

The main parameter controlling the efficiency of the algorithm was found to be the persistence ρ. Several values were experimented with, using a three-sensor search and the corresponding results from exhaustive search from [23] for comparison. Values of 0.95, 0.9, 0.8, and 0.5 were considered. As it was found that only $\rho = 0.9$ leads to the optimum under 59 iterations, this was chosen as the value for the runs. As observed above, 10 ants were used in each run and these were distributed randomly among the cities at initiation. A maximum of 50 iterations were allowed; this bounded the number of function evaluations at 500.

The method was demonstrated for finding a three-sensor distribution. The reason for this is that the number of candidate three-sensor distributions is 684 and it was feasible to compare the results of the optimization algorithms with those from exhaustive search. Also, three sensors are the minimum number that would be needed to locate the impact event from time-of-flight data using triangulation.

As the NN returns a value for the fitness that depends on the starting conditions for training, six separate exhaustive searches were carried out with different initial conditions. The best distribution found over the six runs was found to be 3 : 10 : 12 with an error of 1.99% (Table 1). As observed in [23], this solution agreed with engineering intuition in spacing the sensors in such a way as to effectively triangulate over a large area of the plate (Figure 6). When the GA was applied using the six sets of network starting conditions used for the exhaustive searches, it returned the same distribution (3 : 10 : 12) as the exhaustive search in every instance. In most of the cases, the GA found the optimal result faster than the exhaustive search would have.

For comparison, in 10 runs, the ant algorithm found the above solution once out of 10 runs, but

Table 1. Best three-sensor distributions from six exhaustive searches [23]

Search no.	Distribution	Area error (%)
1	3 : 12 : 17	2.16
2	8 : 11 : 12	2.15
3	1 : 3 : 17	2.15
4	3 : 12 : 14	2.19
5	3 : 7 : 11	2.13
6	3 : 10 : 12	1.99

Reproduced from Ref. 23. © Blackwell, 2000.

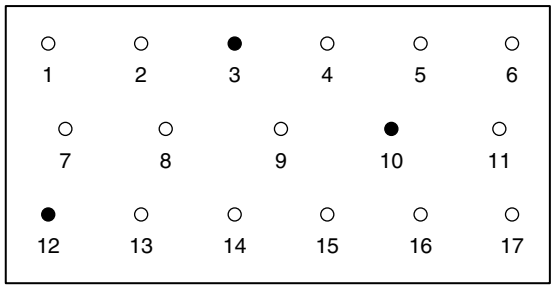


Figure 6. The optimal three-sensor distribution found by exhaustive search and by the genetic algorithm.

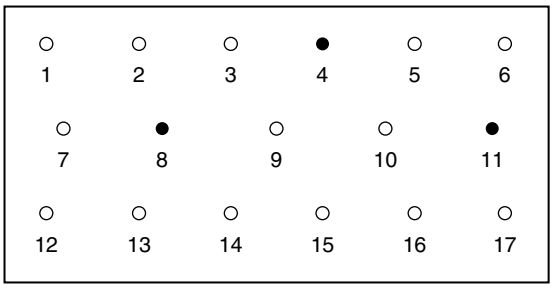


Figure 7. The optimal three-sensor distribution found by the ant colony algorithm.

also found a better solution with an error of 1.89% (distribution $4:8:11$ as shown in Figure 7). The average area error over the 10 runs was 2.11%. The ant algorithm found the optimal distributions faster than the exhaustive search would have.

5.5 Conclusions

The main conclusion of this article is that ant colony metaphors provide an effective solution to the SPO problem. The algorithm is shown to find the optimum in a problem with three sensors, which is amenable to an exhaustive search. Further, the optimal solution is discovered faster than when a GA is used. This result was not entirely expected, as the effect of the parameters α and β was investigated in [29] and it was found that searches based on pheromone alone ($\beta = 0$) were suboptimal.

6 REMARKS

The need for SPO techniques specific to SHM has been recognized, and promising approaches are emerging. Future methods are likely to be specific to the SHM methodology employed (as in the case investigated in [16] and the various NN approaches) and should be developed alongside them. A great number of techniques developed for other structural dynamics applications are also of use to the SHM practitioner. Techniques based upon maximizing the FIM determinant are widely used and represent a significant improvement on the *ad hoc* approaches of the past, with approaches based upon combinatorial optimization appearing to offer further significant advantages. This said, engineering judgment will always play a vital part in successful sensor placement and is pervasive in the application of SPO techniques.

In terms of the development of new algorithms, it is desirable that they should, wherever feasible, be compared with the results of an exhaustive search. This is admittedly limited to cases with low numbers of sensors in the candidate and measurement sets. Where the computational demands of exhaustive search prove prohibitive, "optimal" solutions should be validated both against randomly generated configurations, and against the configurations suggested by the symmetry of the analyzed structure. The robustness of the placement algorithm should be investigated to assess performance when factors such as sensor quantity and analytical model fidelity are altered. There remains a shortage of work concentrating on comparative studies of different approaches and validation of results. The former will give some guidance in engineering applications; the latter will increase confidence in the SPO methodology.

Wherever possible, successful numerical simulation should be followed by experimental study. Where sensor placement is being optimized for the purpose of SHM, extending the experimental study to cover the results of an SHM study would be informative, especially where comparison can be made between placement algorithms and exhaustive search results.

REFERENCES

[1] Kammer DC. Sensor placement for on-orbit modal identification and correlation of large space structures. *Journal of Guidance, Control, and Dynamics* 1991 **14**(2):251–259.

[2] Shah PC, Udwadia FE. A methodology for optimal sensor locations for identification of dynamic systems. *Journal of Applied Mechanics-Transactions of the Asme* 1978 **45**(1):188–196.

[3] Heo G, Wang ML, Satpathi D. Optimal transducer placement for health monitoring of long span bridge. *Soil Dynamics and Earthquake Engineering* 1997 **16**(7–8):495–502.

[4] Li DS, Li HN, Fritzen CP. The connection between effective independence and modal kinetic energy methods for sensor placement. *Journal of Sound and Vibration* 2007 **305**(4–5):945–955.

[5] Jarvis B. Enhancements to modal testing using finite elements. *Sound and Vibration* 1991 **25**(8):28–30.

[6] Penny JET, Friswell MI, Garvey SD. Automatic choice of measurement locations for dynamic testing. *AIAA Journal* 1994 **32**(2):407–414.

[7] Imamovic N. *Validation of Large Structural Dynamics Models using Modal Test Data*, PhD thesis. Imperial College London, 1998.

[8] Hemez FM, Farhat C. An Energy based Optimum Sensor Placement Criterion and Its Application to Structural Damage Detection. *Proceedings of the 12th International Modal Analysis Conference*. Honolulu, HI, 1994; pp. 1568–1575.

[9] Breitfeld T. A method for identification of a set of optimal points for experimental modal analysis. *Proceedings of the 13th International Modal Analysis Conference*. Nashville, TN, 1995.

[10] Larson CB, Zimmerman DC, Marek EL. A comparison of modal test planning techniques: excitation and sensor placement using the NASA 8-bay truss. *Proceedings of the12th International Modal Analysis Conference*. Honolulu, HI, 1994; pp. 205–211.

[11] Meo M, Zumpano G. On the optimal sensor placement techniques for a bridge structure. *Engineering Structures* 2005 **27**(10):1488–1497.

[12] Kammer DC. Sensor set expansion for modal vibration testing. *Mechanical Systems and Signal Processing* 2005 **19**(4):700–713.

[13] Beal JM, Shukla A, Brezhneva OA, Abramson MA. Optimal sensor placement for enhancing sensitivity to change in stiffness for structural health monitoring, *Optimization and Engineering* 2008 **9**(2):119–142.

[14] Yao L, Sethares WA, Kammer DC. Sensor placement for on-orbit modal identification via a genetic algorithm. *AIAA Journal* 1993 **31**(10): 1922–1928.

[15] Frauchi CG, Gallieni D. Pre-test optimisation by genetic algorithm. *Proceedings of the 19th International Seminar on Modal Analysis*. Leuven, 1994.

[16] Shi ZY, Law SS, Zhang LM. Optimum sensor placement for structural damage detection. *Journal of Engineering Mechanics* 2000 **126**(11): 1173–1179.

[17] Guo HY, Zhang L, Zhang LL, Zhou JX. Optimal placement of sensors for structural health monitoring using improved genetic algorithms. *Smart Materials and Structures* 2004 **13**(3):528–534.

[18] Said WM, Staszewski WJ. Optimal sensor location for damage detection using mutual information. *11th International Conference on Adaptive Structures and Technologies*. Nagoya, 2000; pp. 428–435.

[19] Trendafilova I, Heylen W, Van Brussel H. Measurement point selection in damage detection using the mutual information concept. *Smart Materials and Structures* 2001 **10**(3):528–533.

[20] Papadimitriou C, Beck JL, Au SK. Entropy-based optimal sensor location for structural model updating. *Journal of Vibration and Control* 2000 **6**(5):781–800.

[21] Worden K, Burrows AP, Tomlinson GR. A combined neural and genetic approach to sensor placement. *Proceedings of the 13th International Modal Analysis Conference*. Nashville, TN, 1995; pp. 1727–1736.

[22] Worden K, Burrows AP. Optimal sensor placement for fault detection. *Engineering Structures* 2001 **23**(8):885–901.

[23] Worden K, Staszewski WJ. Impact location and quantification on a composite panel using neural networks and a genetic algorithm. *Strain* 2000 **36**(2):61–70.

[24] Staszewski WJ, Worden K, Wardle R, Tomlinson GR. Fail-safe sensor distributions for impact detection in composite materials. *Smart Materials and Structures* 2000 **9**(3):298–303.

[25] Coverley PT, Staszewski WJ. Impact damage location in composite structures using optimized sensor triangulation procedure. *Smart Materials and Structures* 2003 **12**(5):795–803.

[26] Overton G, Worden K. Sensor optimisation using an ant colony metaphor. *Strain* 2004 **40**(2):59–65.

[27] Papadimitriou C. Optimal sensor placement methodology for parametric identification of structural systems. *Journal of Sound and Vibration* 2004 **278**(4–5):923–947.

[28] Dorigo M, Maniezzo V, Colorni A. *Positive Feedback as a Search Strategy*, Report No. 91-016, Politecnico di Milano, 1991.

[29] Dorigo M, Maniezzo V, Colorni A. The ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics. Part B* 1996 **26**(1):1–13.

[30] Hölldobler B, Wilson EO. *The Ants*, Springer-Verlag: Berlin, 1990.

[31] Tarassenko L. *A Guide to Neural Computing Applications*, Arnold: London, 1998.

Chapter 71
Sensor Network Paradigms

Charles R. Farrar, Gyuhae Park and Kevin M. Farinholt
Engineering Institute, Los Alamos National Laboratory, Los Alamos, NM, USA

1 INTRODUCTION

Structural health monitoring (SHM) is the process of detecting damage in structures. The goal of SHM is to improve the safety and reliability of aerospace, civil, and mechanical infrastructure by detecting damage before it reaches a critical state. To achieve this goal, technology is being developed to replace qualitative visual inspection and time-based maintenance procedures with more quantifiable and automated damage assessment processes. These processes are implemented using both hardware and software with the intent of achieving more cost-effective, condition-based maintenance. A more detailed general discussion of SHM can be found in [1, 2].

The authors believe that all approaches to SHM, as well as all traditional nondestructive evaluation procedures (e.g., ultrasonic inspection, acoustic emissions, and active thermography) can be cast in the context of a statistical pattern-recognition problem [2, 3]. Solutions to this problem require four steps: (i) operational evaluation, (ii) data acquisition, (iii) feature extraction, and (iv) statistical modeling for feature classification; these form the statistical pattern-recognition paradigm for SHM. A specific topic that has not been extensively addressed in the SHM literature [4, 5] is the development of mathematically and physically rigorous approaches to designing the SHM sensing system that is used to address the data-acquisition portion of the problem. To date, most SHM system designs are done in a somewhat *ad hoc* manner where the engineer picks a sensing system that is readily available and that he or she is familiar with, and then attempts to demonstrate that a specific type of damage can be detected with that system. If an appropriate level of damage-detection fidelity cannot be obtained, then the system is modified in some empirical manner with the hope that the fidelity improves. Alternatively, as new sensing systems are developed by engineers outside the SHM field, researchers in this

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

field apply these systems to their respective SHM studies in an effort to see if these systems provide an enhanced damage-detection capability. Through these approaches, several sensor network paradigms for SHM have emerged, and this article summarizes and compares these paradigms. When making such a comparison, it should be noted that the authors do not believe there is one sensor network paradigm that is optimal for all SHM problems. All these paradigms have relative advantages and disadvantages. Also, the paradigms described are not at the same level of maturity and, hence, some may require more development to obtain a field-deployable system while others are readily available with commercial off-the-shelf solutions.

This article first addresses the data-acquisition portion of the paradigm, where the various parameters of the system that must be considered in its design and subsequent field deployment are summarized. Several sensor systems that have been developed specifically for SHM are then discussed in terms of these parameters. These sensor systems suggest the definition of three general SHM sensor network paradigms that are then described along with a summary of their relative attributes and deficiencies. A fourth sensor network that is currently under development is proposed that provides an alternative approach to sensing for SHM. This article also summarizes the practical implementation issues for SHM sensor systems in an effort to suggest a more mathematically and physically rigorous approach to future SHM sensing system design. The article concludes by referring the reader to fundamental axioms of SHM that have been proposed [6] and more specifically the subset of these axioms that address sensing issues for SHM.

2 DATA ACQUISITION FOR SHM

The *data-acquisition* portion of the SHM process involves selecting the excitation methods, the sensor types, number and locations, and the data-acquisition/storage/processing/transmittal hardware. The actual implementation of this portion of the SHM process is application specific. A fundamental premise regarding data acquisition and sensing is that these systems do not measure damage. Rather, they measure the response of a system to its operational and environmental loading or the response to inputs from actuators embedded with the sensing system. Depending on the sensing technology deployed and the type of damage to be identified, the sensor readings may be more or less directly correlated to the presence and location of damage. Data interrogation procedures (feature extraction and statistical modeling for feature classification) are the necessary components of an SHM system that convert the sensor data into information about the structural condition. Furthermore, to achieve successful SHM, the data-acquisition system has to be developed in conjunction with these data interrogation procedures.

Inherent in the data acquisition, the feature extraction and statistical modeling portions of the SHM process are data normalization, cleansing, fusion, and compression. As it applies to SHM, data normalization is the process of separating changes in sensor reading caused by damage from those caused by varying operational and environmental conditions [7]. Data cleansing is the process of selectively choosing data to pass on to, or reject from, the feature selection process. Data fusion is the process of combining information from multiple sensors in an effort to enhance the fidelity of the damage-detection process. Data compression is the process of reducing the dimensionality of the data, or the feature extracted from the data, in an effort to facilitate efficient information storage and to enhance the statistical quantification of these parameters. These four activities can be implemented in either hardware or software and usually a combination of these two approaches is used.

3 SHM SENSOR SYSTEM DESIGN CONSIDERATIONS

All sensor systems are deployed for one of the following applications:

1. detection and tracking problems;
2. model development, validation, and uncertainty quantification;
3. control systems.

SHM is a detection and tracking problem. The goal of any SHM sensor system development is to make

the sensor reading as directly correlated with, and as sensitive to, damage as possible (detection) and then have the sensor readings and associated damage-sensitive features extracted from these data change in a monotonic fashion with increasing damage levels (tracking). At the same time, one also strives to make the sensors as independent as possible from all other sources of environmental and operational variability. To best meet these goals for the SHM sensor and data-acquisition system, the following sensing system properties must be defined:

1. types of data to be acquired;
2. sensor types, number, and locations;
3. bandwidth, sensitivity, and dynamic range;
4. data acquisition/telemetry/storage system;
5. power requirements;
6. sampling intervals (continuous monitoring versus monitoring only after extreme events or at periodic intervals);
7. processor/memory requirements;
8. excitation source (active sensing).

Note that some sensor system properties are not independent. For example, increasing sensitivity usually is associated with decreasing dynamic range and increasing bandwidth is typically associated with decreasing frequency resolution. There can be even more issues that must be addressed when developing the sensing portion of the SHM process. Fundamentally, there are four issues related to a specific SHM application that control the selection of hardware to address these sensor system design parameters:

1. the length scales on which damage is to be detected;
2. the time scale on which damage evolves;
3. how varying and/or adverse operational and environmental conditions affect the sensing system;
4. cost.

In addition, the feature extraction, data normalization, and statistical modeling portions of the process can greatly influence the definition of the sensing system properties. Before such decisions can be made, two important questions must be addressed.

First, one must answer the question, "What is the damage to be detected?" The answer to this question must be provided in as quantifiable a manner as possible and address issues such as (i) type of damage (e.g., crack, loose connection, and corrosion), (ii) threshold damage size that must be detected, (iii) probable damage locations, and (iv) anticipated damage growth rates. The more specific and quantifiable this definition, the more likely it is that one will optimize one's sensor budget to produce a system that has the greatest possible fidelity for damage detection. Second, an answer must be provided to the question, "What are the environmental and operational variability that must be accounted for?" To answer this question, one will not only have to have some ideas about the sources of such variability, but one will also have to have thought about how to accomplish data normalization. Typically, data normalization is accomplished through some combination of sensing system hardware and data interrogation software. However, these hardware and software approaches are not optimal if they are not done in a coupled manner.

In summary, from the discussion in this section, it becomes clear that the ability to convert sensor data into structural health information is directly related to the coupling of the sensor system hardware development with the data interrogation procedures.

4 CURRENT SHM SENSING SYSTEMS

Sensing systems for SHM consist of some or all of the following components:

1. transducers that convert changes in the field variable of interest (e.g., acceleration, strain, and temperature) to changes in an electrical signal (e.g., voltage, impedance, and resistance);
2. actuators that can be used to apply a prescribed input to the system (e.g., a piezoelectric transducer bonded to the surface of a structure);
3. analog-to-digital (A/D) converters that transform the analog electrical signal into a digital signal that can subsequently be processed on digital hardware; for the case where actuators are used, a digital-to-analog (D/A) converter is also needed to change the prescribed digital signal to an analog voltage that can be used to control the actuator;
4. signal conditioning;
5. power;

6. telemetry;
7. processing;
8. memory for data storage.

The number of sensing systems available for SHM is enormous and these systems vary quite a bit depending upon the specific SHM activity. Two general types of SHM sensing systems are described below.

4.1 Wired systems

Here, wired SHM systems are defined as ones that telemeter data and transfer power to the sensor over a direct wired connection from the transducer to the central data analysis facility, as shown schematically in Figure 1. In some cases, the central data analysis facility is then connected to the internet such that the processed information can be monitored at a subsequent remote location. There are a wide variety of such systems. At one extreme is peak-strain or peak-acceleration sensing devices that notify the user when a certain threshold in the measured quantity has been exceeded. A more sophisticated system often used for condition monitoring of rotating machinery is a piezoelectric accelerometer with built-in charge amplifier connected directly to a hand-held, single-channel fast Fourier transform (FFT) analyzer. Here, the central data storage and analysis facility is the hand-held FFT analyzer. At the other extreme is custom-designed systems with hundreds of data channels containing numerous types of sensors that cost on the order of multiple millions of dollars such as the sensing system deployed on the Tsing Ma bridge in China [8].

There are a wide range of commercially available wired systems, some of which have been developed for general-purpose data acquisition and the others for SHM applications. Those designed for general-purpose data acquisition can typically interface with a wide variety of transducers and also have the capability to drive actuators. Most of these systems have integrated signal conditioning, data processing, and data storage capabilities and run off of AC power. Those designed to run off batteries typically have a limited number of channels and they are limited in their ability to operate for long periods of time.

One wired system that has been specifically designed for SHM applications consists of an array of piezoelectric lead zirconate titanate (PZT) patches embedded in a Mylar sheet that is bonded to a structure [9]. The PZT patches can be used as either an actuator or a sensor. Damage is detected, located, and, in some cases, quantified by examining the

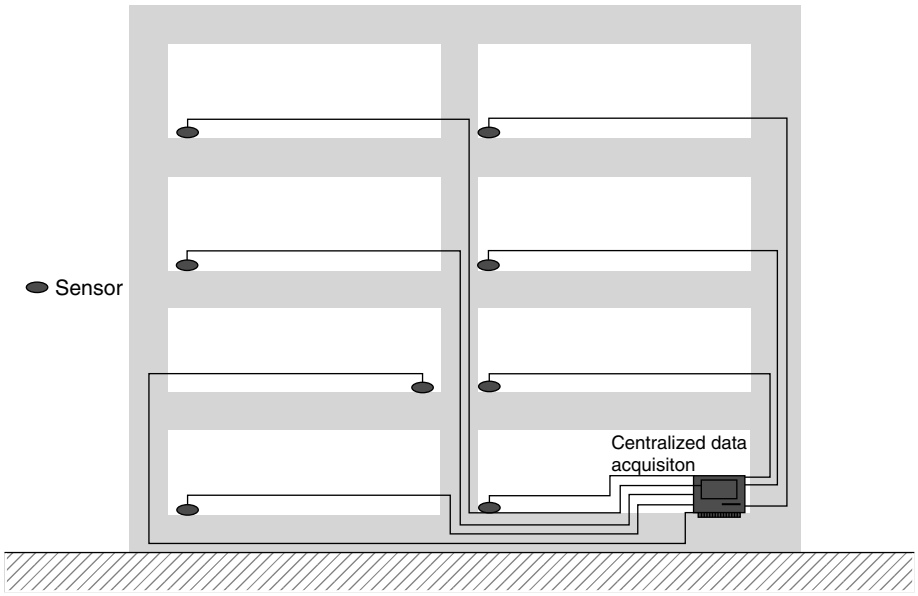


Figure 1. Paradigm I: a wired sensor network connected to a central data-acquisition system running off ac power.

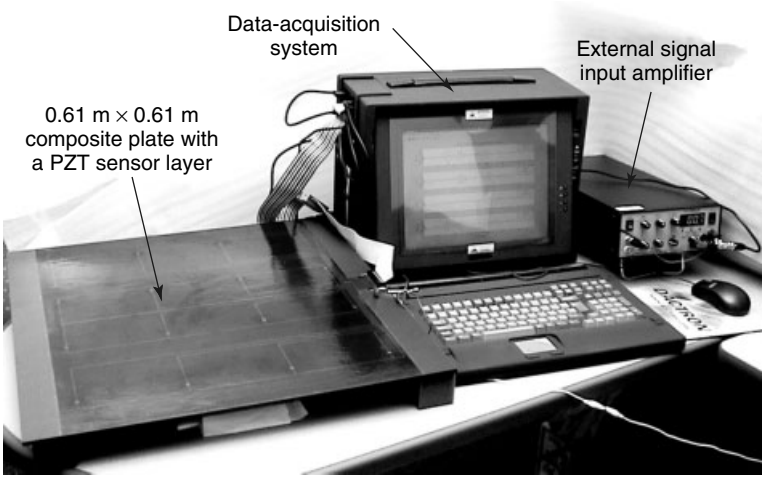


Figure 2. An example of a wired data-acquisition system designed specifically for SHM applications. This system consists of 16 piezoelectric patches in a Mylar sheet. The sensors are connected to a data-acquisition system through the ribbon wire.

attenuation of signals between different sensor–actuator pairs or by examining the characteristics of waves reflected from the damage. An accompanying computer is used for signal conditioning, A/D and D/A conversion, data analysis, and display of final results. The system, which runs on ac power, is shown in Figure 2. This system is described in more detail in **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**.

4.2 Wireless transmission systems

More recently, researchers have been adapting general-purpose wireless embedded sensor nodes for SHM applications. Tanner *et al.* [10] modified an SHM algorithm to the limitations of commercial off-the-shelf wireless sensing and data processing hardware. A wireless sensing system of *motes* running TinyOS operating systems developed at the University of California, Berkeley, was chosen because of their commercial availability and their built-in wireless communication capabilities. A mote consists of modular circuit boards integrating a sensor, microprocessor, A/D converter, and wireless transmitter, all of which run off two AA batteries. A significant reduction in power consumption can be achieved by processing the data locally and only transmitting the results. The system was demonstrated using a small portal structure with damage induced by loss of preload in a bolted joint. The tested mote system is shown in Figure 3. However, the processor proved to be very limited, allowing only the most rudimentary data interrogation algorithms to be implemented. Another application of these sensor nodes to civil-engineering infrastructure can be found in **Wireless Sensor Network Platforms**.

Lynch *et al.* [11] presented hardware for a wireless peer-to-peer SHM system. Using off-the-shelf components, the authors couple sensing circuits and

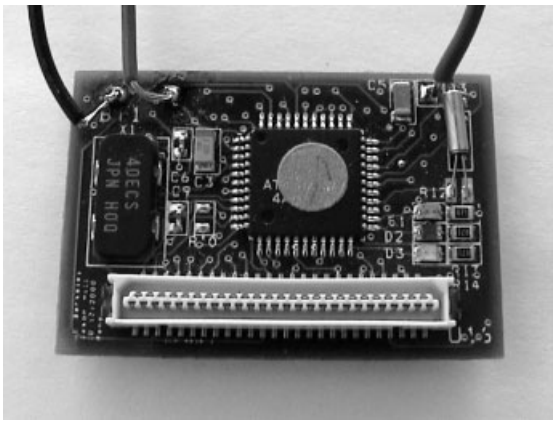


Figure 3. A mote sensor node that includes a microprocessor, sensor, A/D converter, and radio. A penny has been placed on the node for scale.

wireless transmission with a computational core allowing a decentralized collection, analysis, and broadcast of a structure's health. The final hardware platform includes two microcontrollers for data collection and computation connected to a spread spectrum wireless modem. The software is tightly integrated with the hardware and includes the wireless transmission module, the sensing module, and application module. The application module implements the time-series-based SHM algorithm. This integrated data interrogation process requires communication with a centralized server to retrieve model coefficients. The close integration of hardware and software with the dual microcontrollers strives for a power efficient design.

Spencer [12] provides the state-of-the-art review of current "smart sensing" technologies that includes the compiled summaries of wireless work in the SHM field using small, integrated sensor, and processor systems. A smart sensor is defined as a sensing system with an embedded microprocessor and wireless communication. Many smart sensors covered in this article are still in the stage where they simply sense and transmit data. The mote platform is discussed as an impetus for development of the next generation of SHM systems and a new generation of mote is also outlined. The authors also raised the issues that current smart sensing approaches scale poorly to systems with densely instrumented arrays of sensors that will be required for future SHM systems.

To develop a truly integrated SHM system, the data interrogation processes must be transferred to embedded software and hardware that incorporate sensing, processing, and the ability to return a result either locally or remotely. Most off-the-shelf solutions currently available, or in development, have a deficit in processing power that limits the complexity of the software and SHM process that can be implemented. Also, many integrated systems are inflexible because of tight integration between the embedded software, the hardware, and sensing. More recently, researchers have implemented distributed data interrogation algorithms where processing is done across the sensor network to enhance the computational capabilities of these sensor systems [13, 14].

To implement computationally intensive SHM processes, Farrar *et al.* selected a single board computer as a compact form for increased processing power [15]. Also included in the integrated system is a digital signal-processing board with six A/D converters providing the interface to a variety of sensing modalities. Finally, a wireless network board is integrated to provide the ability for the system to relay structural information to a central host, across a network, or through local hardware. Figure 4 shows the prototype of this sensing system. Each of these hardware parts are built in a modular fashion and loosely coupled through the transmission control protocol or Internet protocols. By implementing a common interface, changing or replacing a single component does not require a redesign of the entire system. By allowing processes developed in the Graphical Linking and Assembly of Syntax Structure (GLASS) client to be downloaded and run directly in the GLASS node software, this system became the first SHM hardware solution where new processes can be created and loaded dynamically. This modular nature does not lead to the most power optimized design, but instead achieves a flexible development platform that is used to find the most effective combination of algorithms and hardware for a specific SHM problem. Optimization for power is of secondary concern and will be the focus of follow-on efforts [15].

5 SENSOR NETWORK PARADIGMS

The sensor systems discussed in the previous section have led to three types of sensor network paradigms that are either currently being used for SHM or are the focus of current research efforts in this field. These paradigms are described below. Note that the illustrations of these systems show them applied to a building structure. However, these paradigms can be applied to a wide variety of aerospace, civil, and mechanical systems, and the building structure is simply used for comparison purposes.

5.1 Sensor arrays directly connected to central processing hardware

Figure 1 shows a sensor network directly connected to the central processing hardware. Such a system is the most common one used for SHM studies. The advantage of this system is the wide variety of commercially available off-the-shelf systems that can

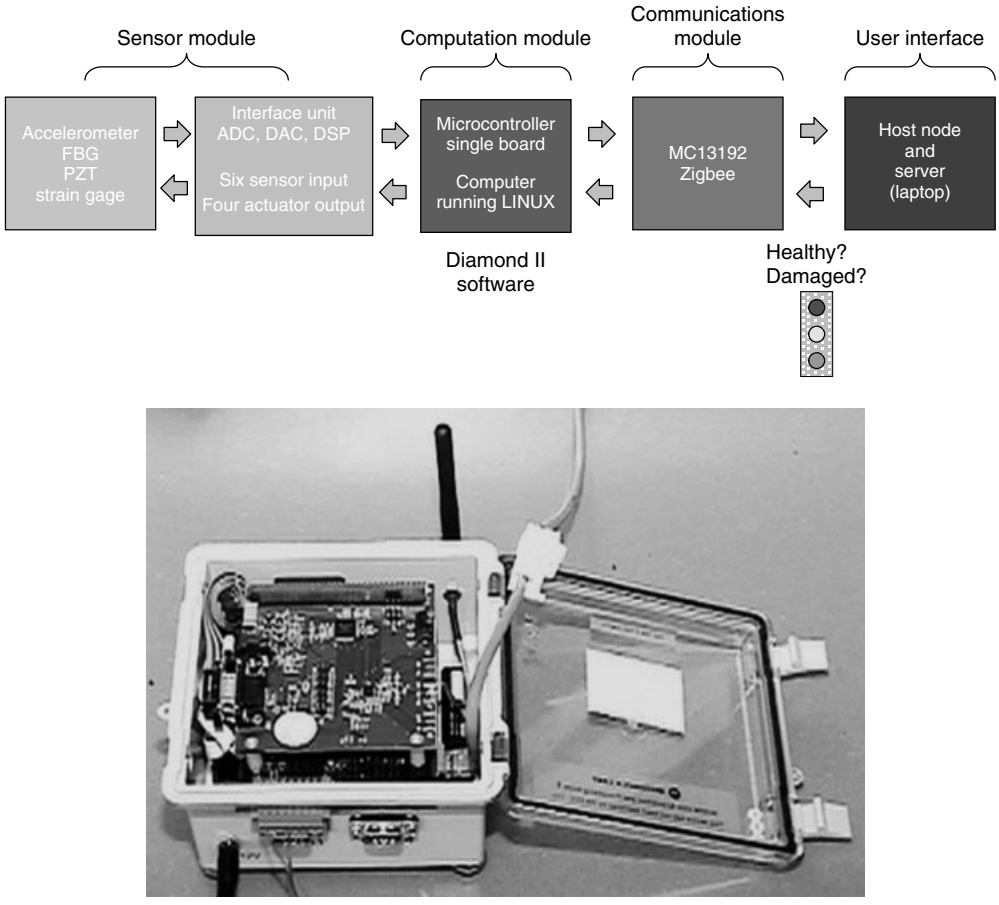


Figure 4. A sensor node incorporating a single board computer to increase processing power. Also included in the system is a digital signal-processing board with six A/D converters that interface to a variety of sensing modalities and a wireless network board that provides the ability to relay information to a central host, across a network, or through local hardware.

be used for this type of monitoring and the wide variety of transducers that can typically be interfaced with such a system. For SHM applications, these systems have been used in both a passive and active-sensing manner. Limitations of such systems are that they are difficult to deploy in a retrofit mode because they usually require ac power, which is not always available. Also, these systems are one-point failure sensitive as one wire can be as long as a few hundred meters. In addition, the deployment of such a system can be challenging with potentially over 75% of the installation time attributed to the installation of system wires and cables for large-scale structures such as those used for long-span bridges [16]. Furthermore, experience with field-deployed systems has shown that the wires can be costly to maintain because of general environmental degradation and damage caused by things such as rodents and vandals.

5.2 Decentralized processing with hopping connection

The integration of wireless communication technologies into SHM methods has been widely investigated to overcome the limitations of wired sensing networks. Wireless communication can remedy the cabling problem of the traditional monitoring system and significantly reduce the maintenance cost. The schematic of the decentralized wireless monitoring

system, which is summarized in detail by Spencer *et al.* [12], is shown in Figure 5.

From the large-scale SHM practice, however, several very serious issues arise with the current design and deployment scheme of the decentralized wireless sensing networks [12, 17]. First, the current wireless sensing design usually adopts *ad hoc* networking and hopping that result in a problem referred to as *data collision*. Data collision is a phenomenon that results from a network device receiving several simultaneous requests to store or retrieve data from other devices on the network. With increasing numbers of sensors, a sensor node located close to the base station will experience tremendous data transmission, possibly resulting in a significant bottleneck. Because the workload of each sensor node cannot be evenly distributed, the chances of data collision increase with expansion of the sensing networks. In addition, this decentralized wireless sensing network scales very poorly in active-sensing system deployment. Because active sensors can serve as actuators as well as sensors, the time synchronization between multiple sensor/actuator units is a challenging task. Because of the processor scheduling or sharing, the use of multiple channels on one sensor node would reduce the sampling rate, which provides neither a practical nor equitable solution for active-sensing techniques that typically interrogate higher frequency ranges. Therefore, for *in situ* applications, the current design scheme can potentially be a very expensive operation.

5.3 Decentralized processing with hybrid connection

The hybrid connection network advantageously combines the previous two networks, as illustrated in Figure 6. At the first level, several sensors are connected to a relay-based piece of hardware, which can serve as both a multiplexer and general-purpose signal router, shown in Figure 6 as a black box. This device will manage the distributed sensing network, control the modes of sensing and actuation, and multiplex the measured signals. The device can also be expanded by means of daisy-chaining. At the next level, multiple pieces of this hardware are linked to a decentralized data control and processing station. This control station is equipped with data-acquisition boards, on-board computer processors, and wireless telemetry, which is similar to the architecture of current decentralized wireless sensors. This device will perform duties of a relay-based hardware control, data acquisition, local computing, and transmission of the necessary results of the computation to the central system. At the highest

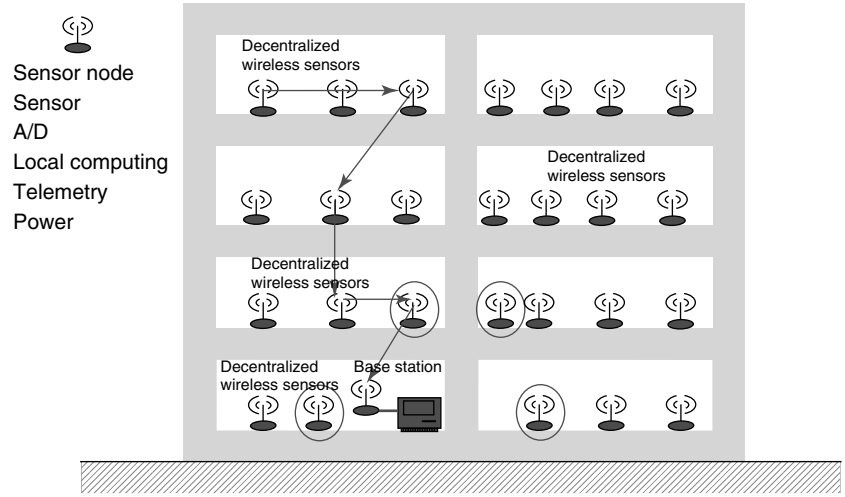


Figure 5. Paradigm II: decentralized processing with each sensor node running off battery power and utilizing a "hoping" telemetry protocol. The "mote" shown in Figure 3 is one such sensor node that can be deployed to form this type of sensor network.

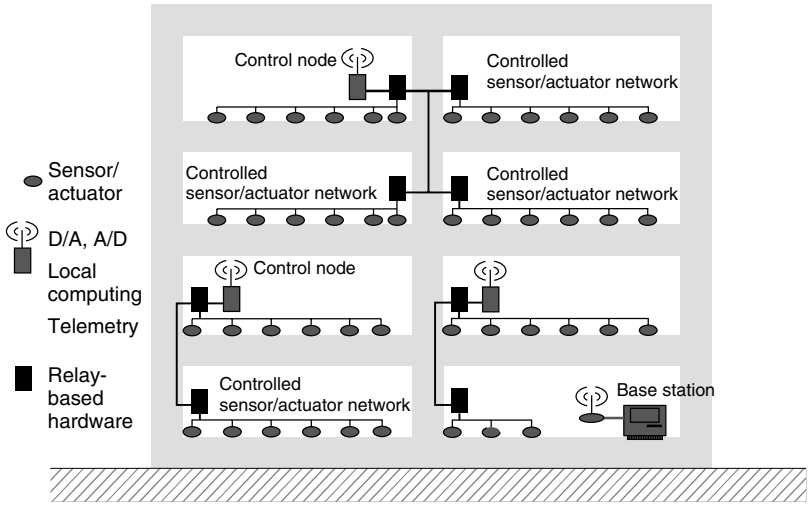


Figure 6. Paradigm III: a hybrid decentralized sensor system with local multiplexing and a hopping telemetry protocol.

level, multiple data processing stations are linked to a central monitoring station that delivers a damage report back to the user. Hierarchal in nature, this sensing network can efficiently interrogate large numbers of distributed sensors and active sensors while maintaining an excellent sensor–cost ratio because only a small number of data acquisition and telemetry units is necessary. This hierarchal sensing network is especially suitable for active-sensing SHM techniques, and is being investigated by Dove *et al.* [17]. In their study, the expandability or the sensing network was of most importance for significantly larger numbers of active sensors, as the number of channels on a decentralized wireless sensor is limited because of processor sharing and scheduling. The prototype of the relay-based hardware ("black box" shown in Figure 6) is illustrated in Figure 7.

6 FUTURE SENSING NETWORK PARADIGMS

The sensing network paradigms described in the previous section have one characteristic in common. The sensing system and associated power sources are installed at fixed locations of the structural system. As previously stated, the deployment of such sensing systems can be costly and the power source may not be always available. A new, energy-efficient future sensing network is currently being investigated collaboratively by Los Alamos National Laboratory and the University of California, San Diego, and is shown in Figure 8. This system couples energy-efficient embedded sensing technology and remote interrogation platforms based on either robots or unmanned aerial vehicles (UAV) to assess damage in structural systems [18]. This approach involves using an unmanned mobile host node (delivered via UAV or robot) to generate radio frequency (RF) signal near the receiving antennas connected to sensor nodes that have been embedded on the structure. Once a capacitor on the sensor node is charged by the RF energy emitted from the node on the UAV, the sensors measure the desired response (impedance, strain, etc.) at critical areas on the structure and transmit the signal back to a processor on the mobile host.

Figure 9 shows a sensor node that has been developed for this mode of remote powering and telemetry. This sensor node uses a low-power integrated circuit that can measure, control, and record an impedance measurement across a piezoelectric transducer. The sensor node integrates several components, including a microcontroller for local computing and sensor node control, a radio for wireless data transmission, multiplexers for managing up to seven piezoelectric transducers per node, energy storage mediums, and several triggering options including a wireless triggering into one package to realize a comprehensive, self-contained wireless active-sensor node for SHM applications. It was estimated that this sensor

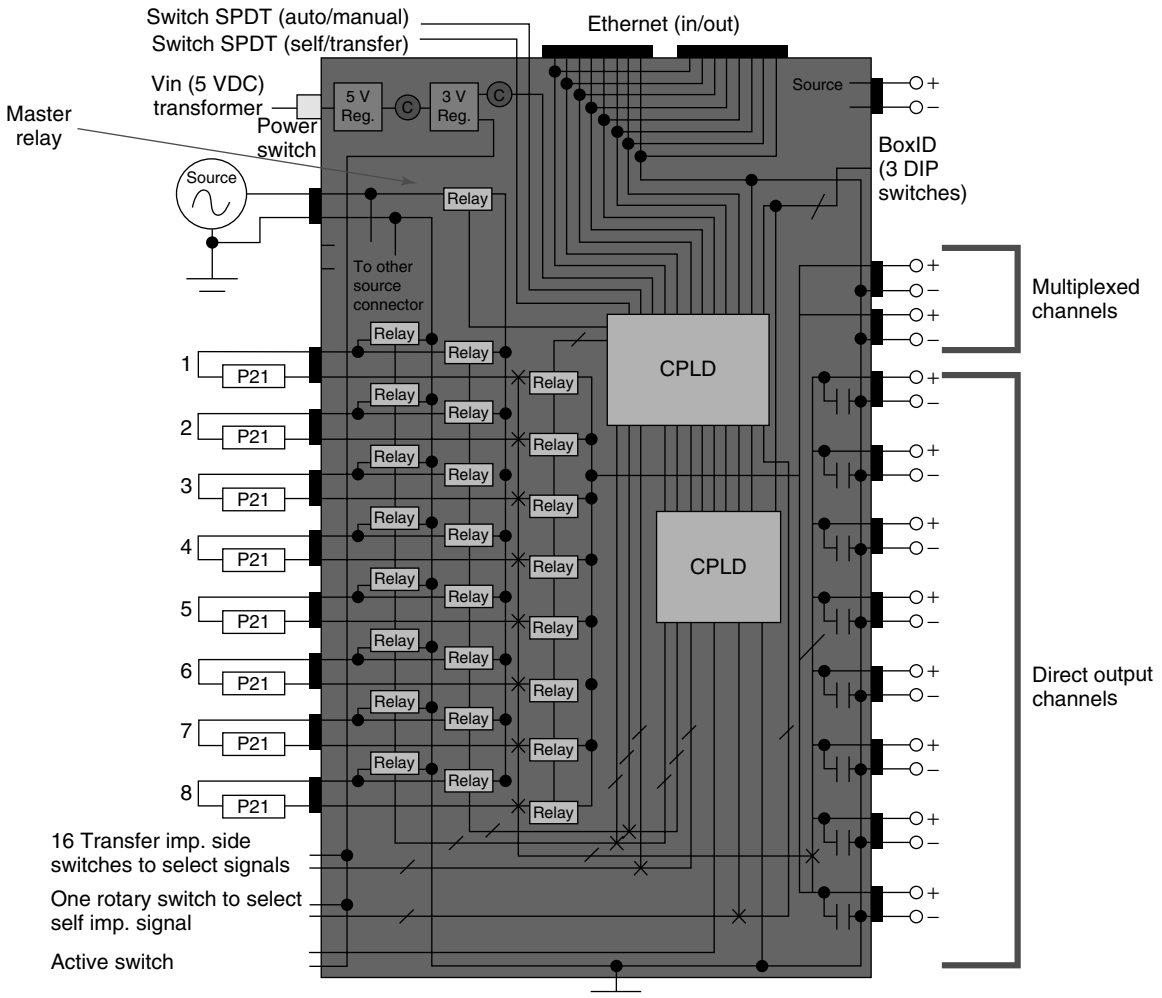


Figure 7. The relay-based hardware shown as a black box in Figure 6. This device manages the distributed sensing network, controls the modes of sensing and actuation, and multiplexes the measured signals.

node requires less than 60 mW of total power to operate, measure, compute, and transmit. Considering this amount of power consumption, the sensor node is within the range of the wireless energy transmission capabilities provided by the host node on the UAV as well as for energy harvesting devices such as small solar arrays.

One UAV with its power source, telemetry, and computing can be used to interrogate an entire sensor array placed on the structure and then can be used for other structures that have similar embedded sensor arrays. A recent field demonstration of this sensor network strategy is shown in Figure 10 where a remotely controlled helicopter was used to deliver power to sensor nodes mounted on a bridge structure. The next generation of this system will take traditional sensing networks to the next level of autonomy, as the mobile hosts (such as UAV), will fly to known critical infrastructure based upon a global positioning system (GPS) locator, deliver the required power, and then perform the SHM assessment without human intervention. This technology will be directly applicable to rapid structural condition assessment of buildings and bridges after an earthquake, where the sensor nodes may need to be deployed for decades during which conventional battery power will be

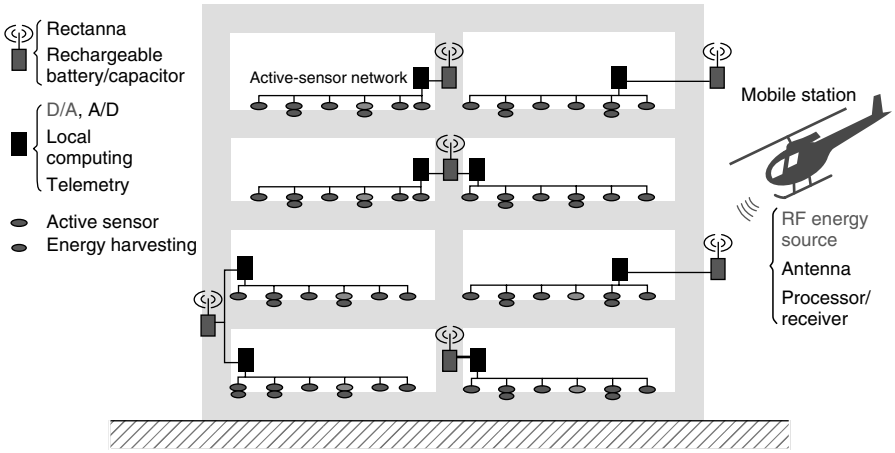


Figure 8. A new sensor network strategy where power and processing are brought to the sensor nodes on a robotic device. The sensor nodes are powered on demand by means of wireless energy transmission.

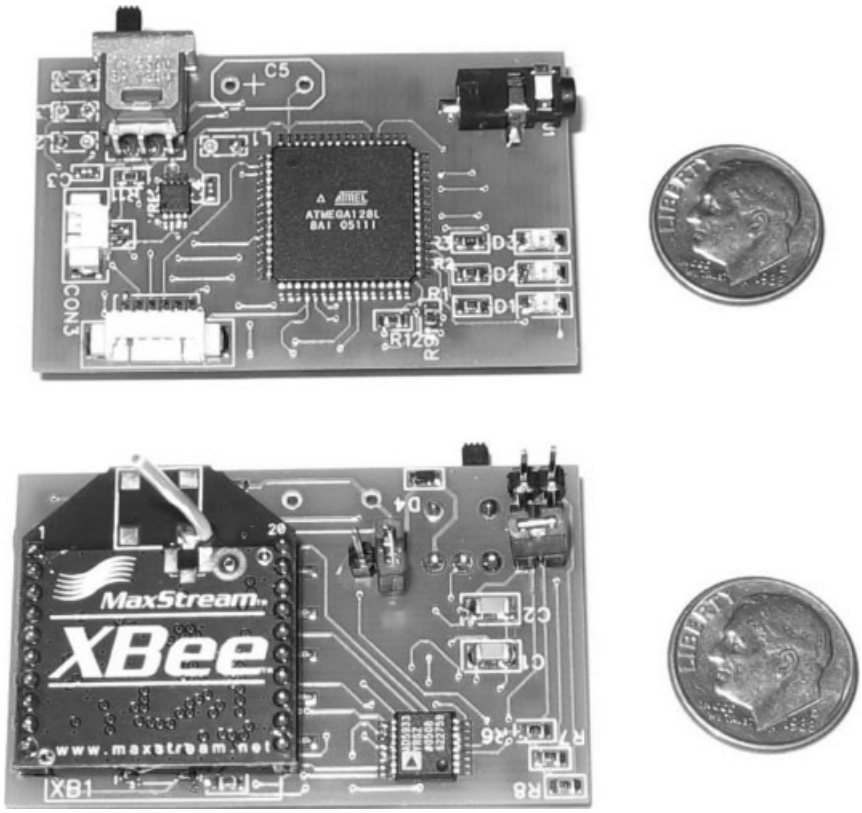


Figure 9. The sensor node that has been designed to receive power wirelessly from a remote host as depicted in Figure 8. This sensor node can measure impedance across up to seven piezoelectric sensors.

Figure 10. Field demonstration of wireless power delivery to a sensor node embedded on a bridge structure. The receiving antenna can be seen suspended from the bottom flange of the bridge girder.

depleted. Also, this technology may be adapted and applied to damage detection in a variety of other civilian and defense-related structures such as nuclear power plants where it is advantageous to minimize human exposure to hazardous environments during the inspection process. A review of other applications of robotic devices for SHM sensing can be found in [19].

7 PRACTICAL IMPLEMENTATION ISSUES FOR SHM SENSING NETWORKS

A major concern in the current sensing network development is the long-term reliability of the network and how to power the network. Other concerns are the abilities of the sensing systems to capture local and system level response, that is, the need to capture response on widely varying length and time scales, and to archive data in a consistent, retrievable manner for long-term analysis. These challenges are nontrivial because of the tendency for each technical discipline to work more or less in isolation. Therefore, an integrated systems engineering approach to the damage-detection process and regular, well-defined routes of information dissemination are essential. The subsequent portions of this article address specific sensing system issues associated with SHM.

7.1 Sensor properties

One of the major challenges with defining sensor properties is that these properties need to be defined *a priori* and typically cannot be changed easily once a sensor system is in place. These sensor properties include bandwidth, sensitivity (dynamic range), number, location, stability, reliability, cost, telemetry, etc. To address this challenge, a coupled analytical and experimental approach to the sensor system deployment should be used in contrast with the *ad hoc* procedures used for most current damage-detection studies. First, critical system failure modes should be well-defined in as quantifiable a manner as possible, using high-fidelity numerical simulations or from previous experiences, before the sensing system is designed. The high-fidelity numerical simulations/experiences can be used to define the required bandwidth, sensitivity, sensor location, and sensor

number. Additional sensing requirements can also be ascertained if changing operational and environmental conditions are included in the models so as to determine how these conditions affect the damage-detection process. The outcome of such analyses may indicate that additional sensors are needed to quantify the effects of varying operational and environmental conditions on the damage-detection process. As an example, it has been shown that the dynamic properties of a bridge structure vary significantly with temperature [20]; however, a measure of ambient air temperature does not correlate with the change in dynamic properties. Instead, it was found that the change in dynamic properties was correlated with the temperature differential across the bridge, which implies the need for multiple temperature sensors in the sensing system.

Another potential level of integration between modeling and sensing resides in the integration of software and hardware components. Once the actuation and sensing capability has been selected, their location has been optimized for damage observability, and the specifications of the data-acquisition system have been met, it may be advantageous to integrate model output and sensing information as much as possible. For example, surrogate models can be programmed on local digital signal-processing chips and their predictions can be compared to sensor output in real time. One obvious benefit would be to minimize the amount of communication by integrating the analysis capability with real-time sensing. In an integrated approach, features can be extracted from sensing data and numerical simulation. Test-analysis comparison and parameter estimation can then be performed locally, which would greatly increase the efficiency of the damage-detection process.

7.2 Power consideration

A major consideration in using a dense sensor array is the problem of providing power to the sensors. This demand leads to the concept of "information as a form of energy". Deriving information costs energy. If the only way to provide power is by direct connections, then the need for wireless communications protocols is eliminated, as the cabled power link can also be used for the data transmission. However, if a wireless communication protocol is used, the development of micropower generators will provide significant advantages over battery power sources as the concept of autonomous embedded sensing cannot be realized if one has to periodically replace batteries. A possible solution to the problem of localized power generation is technologies that enable harvesting ambient energy to power the sensor nodes [21, 22] (*see* **Energy Harvesting and Wireless Energy Transmission for SHM Sensor Nodes**, **On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System**; and **Energy Harvesting using Thermoelectric Materials**). Forms of energy that may be harvested include thermal, vibration, acoustic, and solar. Because such energy harvesting is somewhat new technology, the overriding consideration of reliability still exists, as it does with other components of the monitoring system. With two-way communication capability, the local sensing and processing units can also turn themselves off or go into a "sleep" mode for energy conservation and they can be resuscitated when a "wake-up" signal is broadcast. This approach to power management is discussed in detail in [22].

7.3 Sensor calibration, stability, and ruggedness

When discussing calibration, stability, and ruggedness, it must be clear that these concepts are applied to the entire sensor system and not just the sensor itself. Calibration is the process of determining the relationship between the field variable to be measured and the electrical signal generated by the sensor. Stability refers to how the calibration varies with time. Because SHM involves a comparison of measured response before and after a damaging event, stability is the more critical property for SHM sensor systems. Well-defined sensor calibration procedures exist, but approaches for establishing sensor stability are less well-defined. Most sensors are calibrated at a specialized calibration facility with well-established protocols and standards. This type of calibration is expected to endure, but for embedded sensor systems it needs to be supplemented by incorporating a self-diagnosing and self-calibrating capability into the sensors. In some cases, measurements are acceptable

with 20% error, as long as this error remains constant from one measurement to another one made at some future time. In other scenarios, absolute accuracies are necessary to ensure that the sensor has the fidelity to measure the changes in system response associated with the onset of damage.

Ruggedness of the sensors is a prime consideration for SHM. If part of the system is compromised, then the overall confidence in the system performance is undermined. For sensors implemented for SHM, several ruggedness considerations emerge:

1. The nontrivial problem of sensor selection for extreme environments (e.g. in-service turbine blades exposed to extreme temperatures, high-temperature components of an oil refinery and fluid systems of nuclear power plant that are exposed to radiation fields).
2. Sensors may be less reliable than the component they are monitoring—for example, reliable parts may have failure rates of 1 in 100 000 over several years time. Sensors are often small, complex assemblies with built-in microelectronics, so sensors subjected to the same operational and environmental loading conditions may fail more often than the component being monitored. Loss of sensor signal may then be falsely interpreted as component failure, not sensor failure.
3. Sensors may fail through outright sensor destruction while the component being monitored endures.

False indications of damage or damage precursors are extremely undesirable. If this occurs often, the sensor is either overtly or covertly ignored. Recently, several studies have focused on issues of sensor validation [23, 24]. However, in general, there is little data on the long-term stability and ruggedness of SHM sensor systems.

7.4 Multiscale sensing

Depending on the size and location of the structural damage and the loads applied to the system, the adverse effects of the damage can be either immediate or may take some time before it alters the system's performance. In terms of length scales, all damage begins at the material level and then under appropriate loading conditions progresses to component and system level damage at various rates. In terms of time scales, damage can accumulate incrementally over long periods of time such as that associated with fatigue or corrosion. Damage can also occur on much shorter time scales because of scheduled discrete events such as aircraft landings and from unscheduled discrete events such as enemy fire on a military vehicle. Therefore, the most fundamental issue that must be addressed when developing a sensing system for SHM is the need to capture the structural response on widely varying length and time scales.

The sensing systems that are able to capture the responses over varying length and time scales have not been substantially investigated by researchers, although it is quite possible to use the same piezoelectric patches in both an active, high-frequency mode and in a passive mode to capture the lower-frequency global response of the system. As an example, in the active mode, the piezoelectric sensors can be used to detect and locate damage on a local level using relatively higher frequency excitation and response measurements. This type of active sensing can be used to detect delamination in the composite skin on the wing of an unmanned aerial vehicle. In addition, the same sensors can be used in a passive mode to monitor the low-frequency global modal response of the wing when it is subjected to aerodynamic loading. This global response data can be used to assess the effect of the delamination on the flutter characteristics of that aircraft as determined by analysis of the coupling between the first bending and torsion mode of the wing.

7.5 Sensor–actuator optimization

Few researchers have addressed the issue of developing a systematic approach to the design of a sensor system for SHM. In very general terms, one approach is to consider the sensor system design as a constrained optimization problem. An example, of one such study that employed machine learning to optimize sensor number and location is given in [25]. In terms of a constrained optimization problem, the designer would like to maximize the "damage observability" subjected to a wide variety of possible constraints such as cost, weight, power (when active

sensing is used), and allowable locations. A challenge to actually implementing this approach is coming up with accurate mathematical definitions for damage observability and its relation to the various sensor system properties. This challenge is confounded by the fact that quite a few sensor system parameters may influence observability and the interactions between these various parameters may not be well understood.

One approach to solve the optimization problem is to determine (or assume) that a particular sensor to be employed has a certain damage-detection resolution (i.e., can detect a 1-mm crack through the thickness of a plate within 15-cm radius of the sensor). Then assume that you have an infinite number of sensors, which in turn maximizes observability. Next, optimization procedures such as genetic algorithms or gradient descent methods are used to maximize the observability while retaining some fraction of the infinite sensor array. This process produces a sensor layout with a minimum number of sensors placed at locations that maximize damage observability. Note that this optimization problem will become much more complicated when "real-world" issues such as operational and environmental variability have to be addressed. Also, one must consider the trade-offs between an optimal sensing system and a redundant sensing system. If one sensor or sensor node fails in an optimal system, it is most likely no longer optimal.

With the advent of active-sensing approaches, there can be SHM applications where the excitation is selectable, and, this excitation should be chosen to maximize damage observability. As a simple example, consider a beam or column with a crack that is nominally closed because of a preload. If the provided excitation is not sufficient to open and close the crack, the detectability of the crack in the measured output will be severely limited. Thus, if possible, it is important to answer the question: "Given ever-present physical limits on the level of excitation, and limited outputs that can be measured, what excitation should be provided to a system to make damage most detectable?" When one considers that an excitation may be viewed as a time series with hundreds or thousands of free parameters, optimization in this high-dimensional space might be a daunting task. However, as is demonstrated in [26, 27], a gradient-based technique may be used in which the gradient can be calculated very efficiently. This method does require a model of the system and the accuracy of that model will influence the results.

8 SUMMARY

In this article, the current sensor system design research that is being done to address the data-acquisition portion of the SHM problem is summarized. Several sensor systems that have been developed specifically for SHM are discussed. These sensor systems lead to the definition of several general SHM sensor network paradigms. All of these paradigms have relative advantages and disadvantages. Also, the paradigms described are not at the same level of maturity and, hence, some may require more development to obtain a field-deployable system, while others are readily available with commercial off-the-shelf solutions. At this time, no formal and accepted design methodology exists for the development of an SHM sensing system. As such, this article has also summarized practical implementation issues associated with the SHM sensor system in an effort to suggest the need for a more mathematically and physically rigorous approach to future SHM sensing system design. Finally, it should be noted that recently fundament axioms for SHM have been proposed [6] on the basis of the information published in the extensive amount of literature on SHM over the last 20 years. Of the eight axioms proposed in this article, seven are closely related to sensing aspects of the SHM problem and, therefore, should be considered when designing any SHM sensor network.

REFERENCES

[1] Worden K, Dulieu-Barton JM. An overview of intelligent fault detection in systems and structures. *International Journal of Structural Health Monitoring* 2004 **3**(1):85–98.

[2] Farrar CR, Worden K. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A* 2007 **365**:303–315.

[3] Farrar CR, Doebling SW, Nix DA. Vibration-based structural damage identification. *Philosophical Trans actions of the Royal Society: Mathematical, Physical and Engineering Sciences* 2001 **359**(1778):131–149.

[4] Doebling SW, Farrar CR, Prime MB, Shevitz DW. *Damage Identification and Health Monitoring of*

Structural and Mechanical Systems from Changes in their Vibration Characteristics: A literature Review, Los Alamos National Laboratory report LA-13070-MS, 1996.

[5] Sohn H, Farrar CR, Hemez FM, Czarnecki JJ, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature from 1996–2001*, Los Alamos National Laboratory report LA-13976-MS, 2004.

[6] Worden K, Farrar CR, Manson G, Park G. The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2007 **463**(2082):1639–1664.

[7] Farrar CR, Sohn H, Worden K. Data normalization: a key to structural health monitoring *Proceedings of the Third International Structural Health Monitoring Workshop*. Stanford, CA, 2001.

[8] Ni YQ, Wang BS, Ko JM. Simulation studies of damage location in Tsing Ma Bridge deck. *Proceedings of Nondestructive Evaluation of Highways, Utilities, and Pipelines IV*. SPIE: Bellingham, Washington, 2001, pp. 312–323.

[9] Lin M, Qing X, Kumar A, Beard S. SMART layer and SMART suitcase for structural health monitoring applications. *Proceedings of SPIE on Smart Structures and Materials*. The International Society for Optical Engineering: Newport Beach, CA, March 2001; Vol. 4332, pp. 98–106.

[10] Tanner NA, Wait JR, Farrar CR, Sohn H. Structural health monitoring using modular wireless sensors. *Journal of Intelligent Material systems and Structures* 2003 **14**(1):43–56.

[11] Lynch JP, Law KH, Kiremidjian AS, Carryer E, Kenny TW, Partridge A, Sundararajan A. Validation of a wireless modular monitoring system for structures. *Procedings of SPIE 9th Annual International Symposium on Smart Structures and Materials*. San Diego, CA, 17–21 March 2002.

[12] Spencer BF, Ruiz-Sandoval ME, Kurata N. Smart Sensing Technology: Opportunities and Challenges. *Structural Control and Health Monitoring* 2004 **11**(4):349–368.

[13] Zimmerman AT, Shiraishi M, Swartz A, Lynch JP. Automated modal parameter estimation by parallel processing within wireless monitoring systems. *ASCE Journal of Infrastructure* 2008 **14**(1): 102–113.

[14] Swartz RA, Lynch JP. A multirate recursive arx algorithm for energy efficient wireless structural monitoring. *4th World Conference on Structural Control and Monitoring*. San Diego, CA, 2006.

[15] Farrar CR, Allen DW, Ball S, Masquelier MP, Park G. Coupling sensing hardware with data interrogation software for structural health monitoring. *Proceedings of 11th International Symposium on Dynamic Problems of Mechanics*. Ouro Preto, Brazil, March 2005.

[16] Lynch JP, Partridge A, Law KH, Kenny TW, Kiremidjian AS, Carryer E. Design of a Piezoresistive MEMS-based accelerometer for integration with a wireless sensing unit for structural monitoring. *ASCE Journal of Aerospace Engineering* 2003 **16**:108–114.

[17] Dove JR, Park G, Farrar CR. Hardware design of hierarchal active-sensing networks for structural health monitoring. *Smart Materials and Structures* 2006 **15**:139–146.

[18] Todd M, *et al.* A different approach to sensor networking for shm: remote powering and interrogation with unmanned aerial vehicles. *Structural Health Monitoring 2007 Quantification, Validation and Implementation*. DEStech Publication: Lancaster, PA, 2007, Vol. 1, pp. 29–43.

[19] Huston DR. Robotic surveillance approaches for SHM. *Structural Health Monitoring 2005 Advancement and Challenges for Implementation*. DEStech Publication: Lancaster, PA, 2005, Vol. 2, pp. 1586–1593.

[20] Farrar CR, Cornwell PJ, Doebling SW, Prime MB. *Structural Health Monitoring Studies of the Alamosa Canyon and I-40 Bridges*. Los Alamos National Laboratory report, LA-13635-MS, 2000.

[21] Sodano HA, Inman DJ, Park G. A review of power harvesting from vibration using piezoelectric materials. *The Shock and Vibration Digest* 2004 **36**(3):197–205.

[22] Park G, Farrar CR, Todd MD, Hodgkiss W, Rosing T. *Power Harvesting for Embedded Structural Health Monitoring Sensing Systems*. Los Alamos National Laboratory report, LA-14314-MS, 2007.

[23] Park G, Farrar CR, Rutherford CA, Robertson AN. Piezoelectric active sensor self-diagnostics using electrical admittance measurements. *ASME Journal of Vibrations and Acoustics* 2006 **128**:469–476.

[24] Kerschen G, Boe PD, Golinval J, Worden K. Sensor validation using principal component analysis. *Smart Materials and Structures* 2005 **14**(1):36–42.

[25] Worden K, Burrows AP, Tomlinson GR. A combined neural and genetic approach to sensor

placement. *Proceedings of 13th International Modal Analysis Conference*, Nashville, TN, USA, 1995; pp. 1727–1736.

[26] Bement MT, Bewley T. Optimal excitation design for damage detection using adjoint based optimization Part 1 theoretical development. *Mechanical Systems and Signal Processing*, submitted for publication.

[27] Bement MT, Bewley T. Optimal excitation design for damage detection using adjoint based optimization Part 2 experimental verification. *Mechanical Systems and Signal Processing*, submitted for publication.

Chapter 73
Web-based SHM

Vistasp M. Karbhari[1] and Hong Guan[2]

[1] *University of Alabama in Huntsville, Huntsville, AL, USA*
[2] *HDR, Los Angeles, CA, USA*

1 INTRODUCTION

Structural health monitoring (SHM) is increasingly being considered as a means of not only obtaining data related to the response of components and structural systems but also as a means of providing an assessment of the "health" of the system. While there are still substantial differences of opinion regarding the scope and applicability of SHM (i.e., does the system merely relate to the collection of data and provision of a rudimentary knowledge of response, or does it actually serve as an autonomous or semiautonomous means of interpreting structural state in terms of characteristics such as capacity, remaining life, and necessity of repair), there is no doubt that the appropriate implementation of such a system can provide significant information of value to the owner, hitherto not available either in the same timescale or in quantity/depth, or both. In fact, recent developments in this field have led to the introduction of the term *civionics* to describe the "hardware and physical installation of the sensors, wires, conduits, termination, and control boxes" [1], which complements SHM in terms of a system that enables the acquisition and interpretation of data beyond that represented by traditional nondestructive evaluation (NDE).

Currently, civil infrastructure systems such as bridges, pipelines, waterways, and buildings are inspected at routine intervals till significant distress is noted, after which the period between inspections is decreased and the level of inspection is increased till such time that the distress has been corrected by replacement or repair. This time-based monitoring is inefficient not just in terms of resources but also since it does not minimize "downtime" of the structure. Since civil infrastructure forms a critical part of a nation's well-being, its deterioration has immense effect on the economies of the region, as it serves as the basis for the transfer of goods and services. The decrease in "downtime" due to maintenance, rehabilitation, or even rapid replacement is

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

a critical aspect that needs to be addressed, and hence a move to a condition-based assessment of structures is advantageous. The implementation of condition-based assessment, however, requires that changes in structural response be monitored as a result of both normal operation and extreme events. This can be done through the implementation of a true SHM system, which could enable autonomous and continuous recording and assessment of predetermined response parameters associated with materials, components, and/or the entire system.

The development of an SHM system is essentially based on the ability to acquire data, transmit it, interrogate it, and then make decisions based on the cumulative sets of data stored in the database. Thus, in effect, the SHM system is essentially a decision system that is fronted by sensors and backed by a knowledge base, the critical elements of which are represented in Figure 1.

Tremendous advances in technologies related to sensor technologies, data compression and transmission, and interrogation make the deployment of an SHM system substantially easier today than in the past. For example, advances in image capture and analysis have made it possible to track vehicles in traffic for purposes of transportation planning such that vehicles can be tracked and classified over long periods of time to accumulate large volumes of tracking data, which can then be used to build models consisting of the traffic flow parameters such as density, flow, and speed [2]. In addition, tracking data can be used for event detection and definition of normal motion paths for detection of abnormal events, as shown schematically in Figure 2.

In addition, systems have already been developed and implemented for the rapid assessment of pavement condition in terms of the presence and extent of pavement cracks, potholes, and other damage, which can disrupt the smooth and efficient flow of traffic and overall safety. For example, the Hanshin Expressway Public Corporation (HEPC) in Japan has implemented an intelligent inspection system based on the use of Charge Coupled Device (CCD) cameras, laser emitters, and image-processing techniques—all on board vehicles that continuously traverse the road network to obtain data, which is then transmitted and used from a central location [3]. In this, as in other SHM systems, the main components are those of (i) data acquisition, (ii) data transmission, (iii) data processing and interrogation, (iv) data storage, (v) data retrieval, and (vi) diagnostics of long-term response. While a number of these aspects can be performed through direct assessment systems without need for comparison of historical data or excessive diagnostics, this hinders the full development of an SHM system that would enable true diagnosis and prognosis of a structure and the system that it is a part of (for example, a single bridge within a network

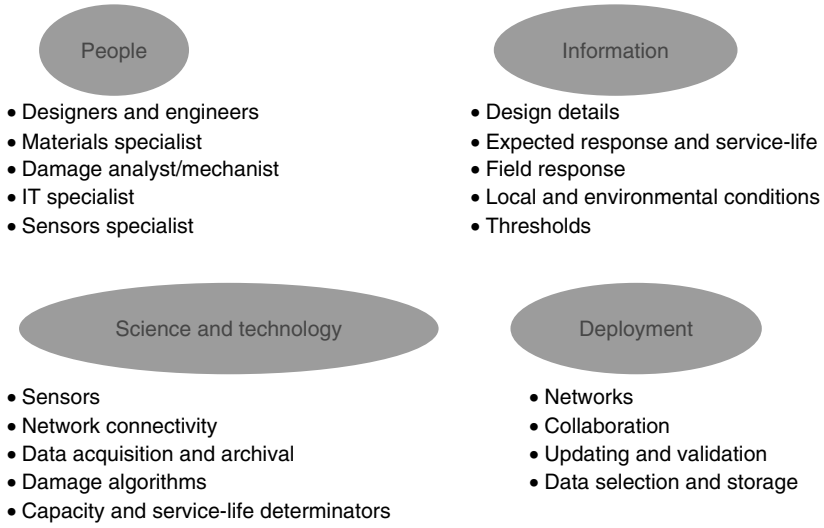


Figure 1. Critical elements of an SHM system.

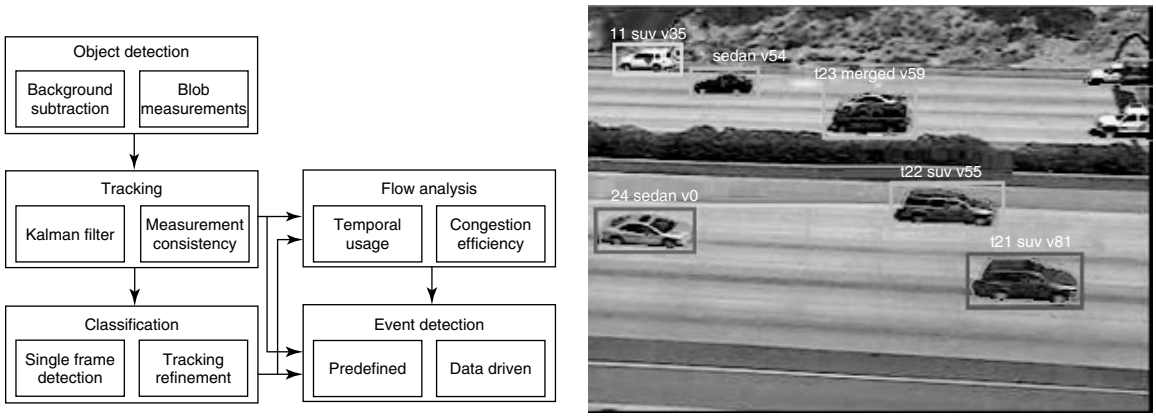


Figure 2. Schematic and basis for tracking of vehicles. [Courtesy of Mohan Trivedi, UCSD.]

of highways that have multiple bridges on them). For efficient use of such a system, it is incumbent that the SHM system be able to semiautonomously compare response at any point in time, not just with preset thresholds of performance but also with prior data to assess trends and develop self-learning models that can be used to predict future response. In addition, it is critical that the user be able to assess information rapidly about not just single structures but also about networks, so as to enable the assessment of the effect of deterioration in one element on the entire system and to thereby enable dynamic resource allocation. For example, in the case of earthquakes, it would be useful to be able to pull up, using regional maps, the location of critical structures, interrogate their health from a remote location, and reach decisions on which ones need immediate attention in terms of resources. All these aspects are best suited to be addressed through a web-based monitoring system.

2 COMPONENTS OF A WEB-BASED SHM SYSTEM

The essential components of a web-based SHM system are shown in Figure 3.

Data is collected from suites of sensors on individual structural components or systems and is brought back to a central location. While wireless systems are becoming ubiquitous, it must be remembered that vast amounts of data transmission, while possible, need large bandwidths. Depending on the location of the structures being monitored, it may be better to have packets of data sent at predetermined intervals, rather than on a continuous basis. Further, it may, in fact, be better to have the SHM itself be triggered to collect data only when specific thresholds are reached or exceeded (weight, deflection, acceleration, etc.). Once this data is brought into the system, it can be directly stored keeping in mind that storage can itself be a major issue. It is thus necessary to have a methodology in place by which data is assessed for storage through selection of "typical" sets for storage, events and novelties, and comparison with previously collected sets, and to store only the required values (such as maxima, minima, averages, etc.). The development and implementation of an appropriate selection and archival plan is essential, especially if the web-based SHM system is used for a large number of structures. A powerful backend database to manage raw data and analysis is essential in these cases. Data itself can be processed using both model-free and model-based approaches, which when compared to structural response models and thresholds can provide important characteristics for decision making. For a true level IV SHM system, it is important that the analysis tools be linked to predictive and "what-if" capabilities to provide the results that are of use to the owner/engineer in making decisions. A prototype system for web-based monitoring is used to demonstrate the requirements in the next section.

The value of the web-based system is threefold. First, it allows for direct assessment of data sets, both streaming and stored, by the user, using a forum

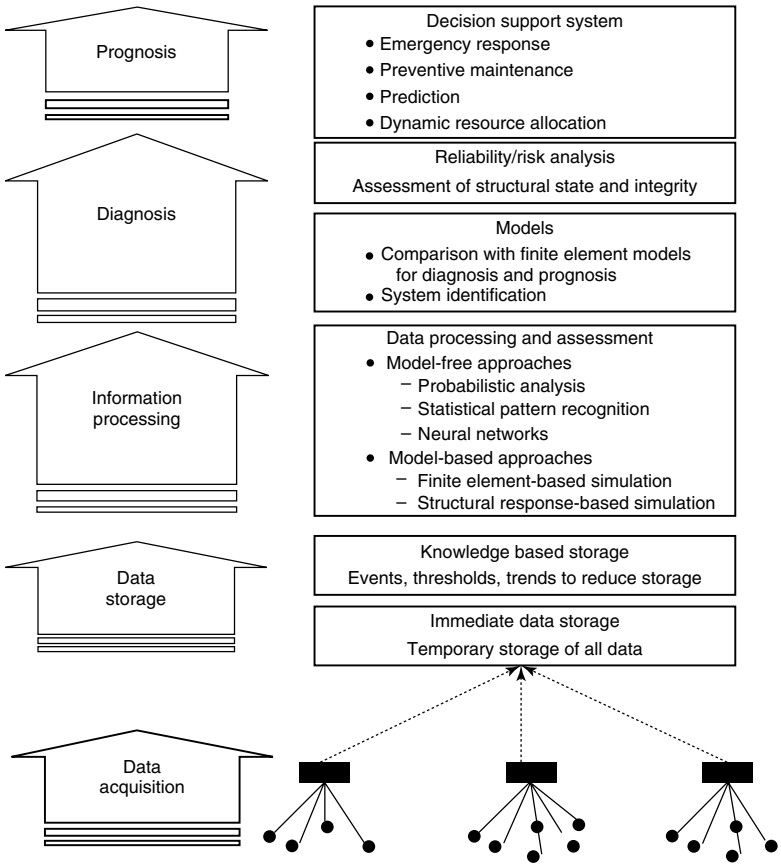


Figure 3. Components of a web-based SHM system.

that is now ubiquitous for rapid access of information. Secondly, it can be programmed to not only provide views of the data in sets chosen by the user (i.e., different users could simultaneously, from different sites, access different sets of sensors based on choices made from preset suites, or from the entire set) but to also interrogate the data, based on predetermined thresholds so as to enable rapid assessment of structural response. Thirdly, the use of a web-based schema allows for autonomous delivery of messages to personnel as needed. In addition, information from a set of bridges can be combined on a single system such that an entire network can be monitored remotely. It is essential, however, to recognize that the number of clients that can access the system at one time and the amount of data that can be accessed individually will depend intrinsically on bandwidth. In addition, the system has to be designed to both send and receive data, rather than the more commonly used schema of just receiving data, which is then accessed with a time delay. The use of a web-based interface also ensures that access is not restricted by location and that 24/7 communication and monitoring is available.

3 AN EXAMPLE FOR WEB-BASED SHM OF BRIDGES

3.1 Theoretical basis

Various health monitoring/NDE techniques have been proposed for continuous or time-based assessment of structural response, among which vibration-based damage detection (VDD) techniques have been widely accepted as a flexible choice for potential use

in long-term, real-time, remote monitoring systems. VDD can be described as a class of techniques that are capable of detecting the stiffness/mass change of the structure based on its dynamic response due to forced or ambient excitation. The change of stiffness/mass of the structure is often related to the damage of structural components and the consequent loss of performance. The general procedure of VDD can be roughly divided into four steps: (i) measurement of structural dynamic response in terms of accelerations or displacements; (ii) extraction of modal parameters, such as natural frequencies and mode shapes from the measurements obtained; (iii) calculation of stiffness/mass change based on modal parameters, used in conjunction with the same set of parameters from the as-built or "baseline" structure; and (iv) damage localization and severity estimation based on the results of the third step. In the present example, the time domain decomposition (TDD) method is used, which allows the extraction of mode shapes as the first step, followed by the determination of the corresponding natural frequencies as the second step. Although the TDD needs frequency information, it does not require use of discrete Fourier transforms, which result in substantial computational effort in most other modal parameter extraction algorithms. Once the modal parameters have been identified through this procedure, the next step in the modal-based damage detection procedure is localization of damage and severity estimation. This is done following the damage index method initially proposed by Stubbs and Osegueda [4]. The use of this methodology not only enables the monitoring of structural health but also assessment of damage and its progression as a function of time. In the present setup for a linear, undamaged structure, the ith modal stiffness of a linear, undamaged structure can be described as

$$K_i = \vec{\phi}_i^T \mathbf{C} \vec{\phi}_i \tag{1}$$

where $\vec{\phi}_i$ is the ith modal vector and $\mathbf{C}$ is the system stiffness matrix. The contribution of the jth member to the ith modal stiffness can then be given by

$$K_{ij} = \vec{\phi}_i^T C_j \vec{\phi}_i \tag{2}$$

The fraction of modal energy of the ith mode contributed by the jth member, also called *modal sensitivity*, is defined as

$$F_{ij} = K_{ij}/K_i \tag{3}$$

and correspondingly, using * to represent the damaged structure, the fraction of modal energy of a damaged structure can be defined as

$$F_{ij}^* = K_{ij}^*/K_i^* \tag{4}$$

in which

$$K_{ij}^* = \vec{\phi}_i^{*T} C_j^* \vec{\phi}_i^* \quad K^*{}_i = \vec{\phi}_i^{*T} C^* \vec{\phi}_i^* \tag{5}$$

and

$$C_j = E_j C_{jo} \quad C_j^* = E_j^* C_{jo} \tag{6}$$

A fundamental aspect in this method is that the modal sensitivity for the ith mode and jth member remain unchanged for both the undamaged and damaged structures, i.e.,

$$F_{ij}/F_{ij}^* = (K_{ij}^* K_i)/(K_i^* K_{ij}) = 1 \tag{7}$$

Therefore, a damage index β_j for the jth member, can be obtained as

$$\beta_j = \frac{E_j}{E_j^*} \quad \text{damage indicated when} \quad \beta_j > 1 \tag{8}$$

$$\beta_j = \frac{\gamma_{ij}^* K_i}{\gamma_{ij} K_i^*} = \frac{\phi_i^{*T} C_{jo} \phi_i^* K_i}{\phi_i^T C_{jo} \phi_i K_i^*} \tag{9}$$

and the severity of damage can be estimated by

$$E_j^* = E_j \left(1 + \frac{\mathrm{d}E_j}{E_j}\right) = E_j(1 + \alpha_j) \tag{10}$$

where

$$\alpha_j = \frac{\gamma_{ij} K_i^*}{\gamma_{ij}^* K_i} - 1 \tag{11}$$

in which

$$\gamma_{ij} = \vec{\phi}_i^T C_{jo} \vec{\phi}_i \quad \gamma_{ij}^* = \vec{\phi}_i^{*T} C_{jo} \vec{\phi}_i^* \tag{12}$$

3.2 Description of sensors and protocol

The bridge discussed in this article is a two-span highway bridge with a total length of approximately 20.1 m and a width of approximately 13.0 m. It carries two north-bound lanes on State Highway 86 in Riverside County, California. The superstructure is of slab-on-girder type, with two equal

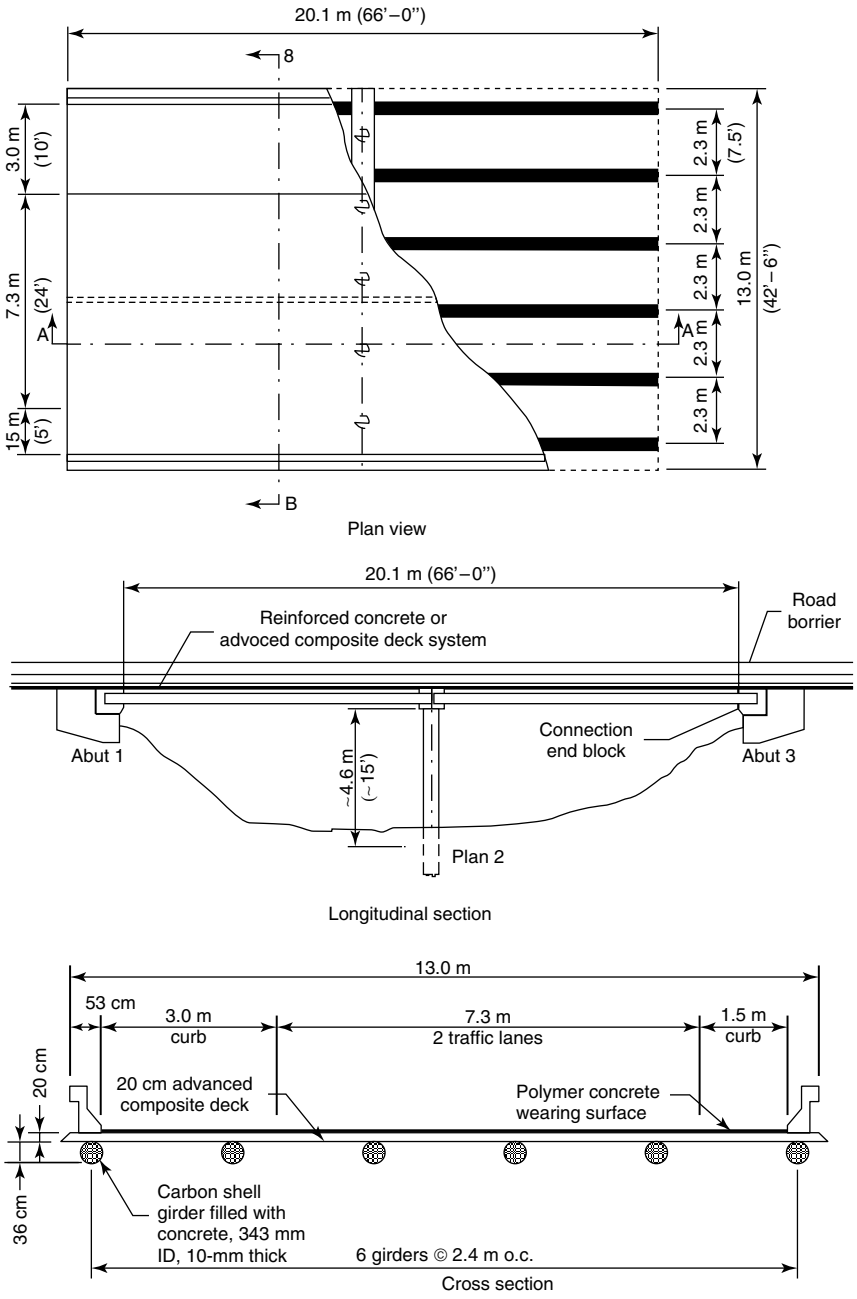


Figure 4. Schematic of the bridge.

spans of 10 m, and a cap beam connecting the two adjoining spans. The six main girders are composed of 10-mm-thick prefabricated filament-wound carbon/epoxy shells filled on-site with light-weight concrete, which support six modular E-glass fiber-reinforced polymer (GFRP) deck panels that serve as road surface, transfer vehicular loads to the girders, and also act as the transverse connections between girders. Schematics of the bridge are shown in Figure 4.

Details related to connections, proof testing at the component and systems levels, and design are reported in [5] and are hence not repeated herein. The bridge is open to heavy truck traffic and serves as a test bed for the implementation of SHM strategies, in addition to being one of the first fiber-reinforced polymer (FRP) vehicular bridges on a major route.

The response of the bridge is characterized through the placement of 63 accelerometers, 20 strain gauges, 4 linear potentiometers, 1 temperature sensor, and a pan-tilt-zoom (PTZ) camera. A total of 63 Model 3140 single axis accelerometers from IC Sensors, Inc. with a dynamic range of ± 2g, a sensitivity of $1\,\mathrm{V\,g^{-1}}$, and a frequency response of 0–200 Hz have been used. These provide an accurate measurement of the ambient-excitation-induced response of interest for this particular application. Forty-two accelerometers were placed at locations on the bottom of the composite deck, referred to as *nodes*, each protected by a sealed equipment housing. The nodes formed a 7×6 grid, with seven locations in the bridge longitudinal direction and six in the transverse direction (as shown in Figure 5), enabling the identification of several mode shapes required for damage detection, and for comparison with theoretical mode shapes already identified by finite-element analysis (FEA).

In addition to the 42 accelerometers measuring vertical acceleration, an additional horizontal accelerometer was installed in the same equipment housing at some of the nodes, to measure the horizontal accelerations that might be caused by an earthquake. A total of 20 bonded resistance gauges were also attached to critical locations on the bottom of the girder, the deck, and the middle section of the girder. Four linear potentiometers were used to measure the deflection of the composite girders at the midspan of two central girders, which experience the maximum deflection (as identified by previously conducted load tests).

A Campbell Scientific CR9000 modular high-speed data logger serves as the core of the data-acquisition system. The CR9000 is capable of making measurements at an aggregate sampling rate of 100 kHz for up to 200 channels. A Transmission Control Protocol/Internet Protocol (TCP/IP) interface enables communication and data collection via standard wired or wireless networks. The raw voltage data collected by the sensors are conditioned and digitized, and then stored into the cache memory of the CR9000 to be retrieved later. The High Performance Wireless Research and Education Network (HPWREN), funded by the National Science Foundation (NSF), which is operational in Southern California, is utilized as the wireless link for data transmission using a 900-MHz wireless antenna and a transmission rate of 45 Mbps. Figure 6 shows the bridge with the antenna used for data transmission.

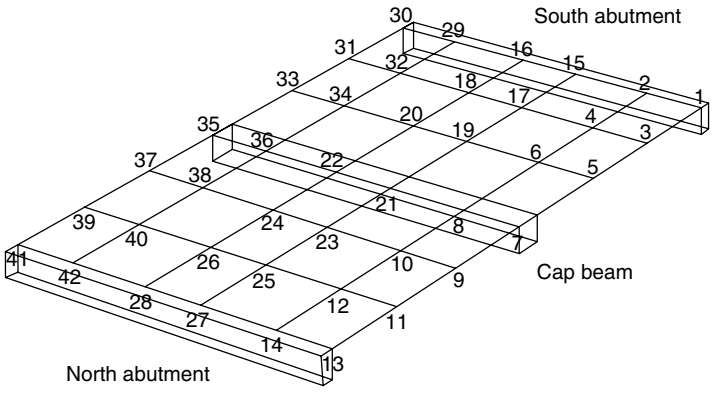


Figure 5. Location of accelerometers.

Figure 6. Bridge with antenna for data transmission.

The automated collection, digitization, and transmission of data are controlled by two collaborating programs. The first program, written in CRBASIC and running on a CR9000 DAQ system, serves the collection and digitization function. It enables data from all channels to be collected at preset intervals and when triggered either by an extreme event, such as an earthquake, or when a preset response threshold is exceeded, such as deflection or acceleration from extremely heavy traffic or permit loads. It also allows for data from a select number of channels, denoted as *streaming channels* to be collected and transmitted continuously in real time. The second program set is housed in the central server at the home node, and serves both data transmission and analysis functions. A schematic of flow is shown in Figure 7.

Continuous analyses performed on streaming data and results, such as peak displacement and strain, are then stored for retrieval on demand. Simultaneously, analyses are also performed on event-based data, as they become available, and the detailed results, such as estimates of damage localization and severity are also stored in the database. Event-based data are archived in an event data database. The server program also serves as a relay for streaming data to multiple end users.

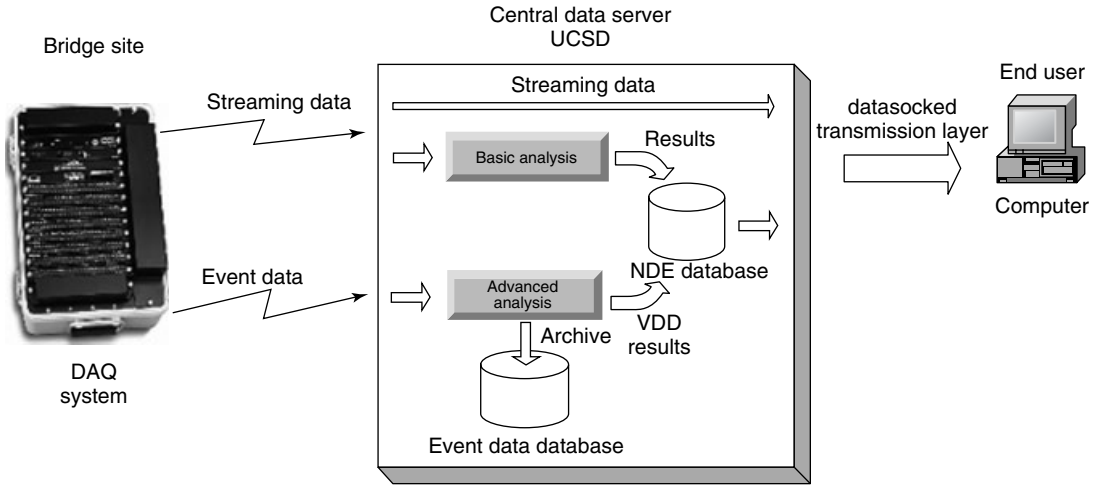


Figure 7. Schematic for information flow.

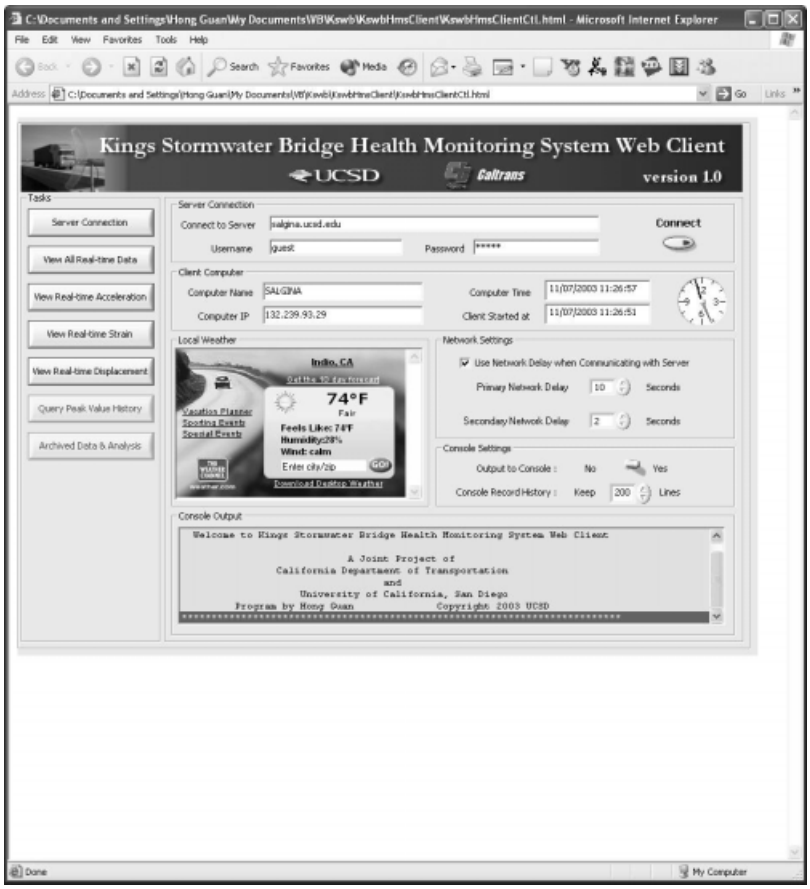


Figure 8. The web client page.

3.3 Web-based implementation

After the data is collected at the bridge site, controlled by a program running on the data logger, raw data is sent over the wireless network to a server located at the University of California, San Diego in La Jolla. A Matlab-based program then autonomously analyzes the raw data, generates results, and sends these to a web server. End users can then access these results using a standalone program or a specific web client, as shown in Figure 8.

The user can then select sensors for which data is desired using maps with sensor locations as in Figure 9.

The web-based user interface enables users with an appropriate connection to monitor the bridge's behavior and assess pertinent serviceability, reliability, and durability aspects related to the structure. By making use of the data stored on the central server, users can access historical data to make comparisons and/or to determine effects of specific events. For the purposes of the current investigation, an event is defined as a *significant excitation to the bridge*, usually caused by single or multiple vehicles crossing the bridge. In terms of measurement, an event is usually a time period over which the bridge is subject to excitation and measurements are taken. Both raw data and processed data are analyzed and stored with reference to the specific event through a date–time stamp. Users can either choose to plot sensor records pertaining to a single event, multiple events, or even records of

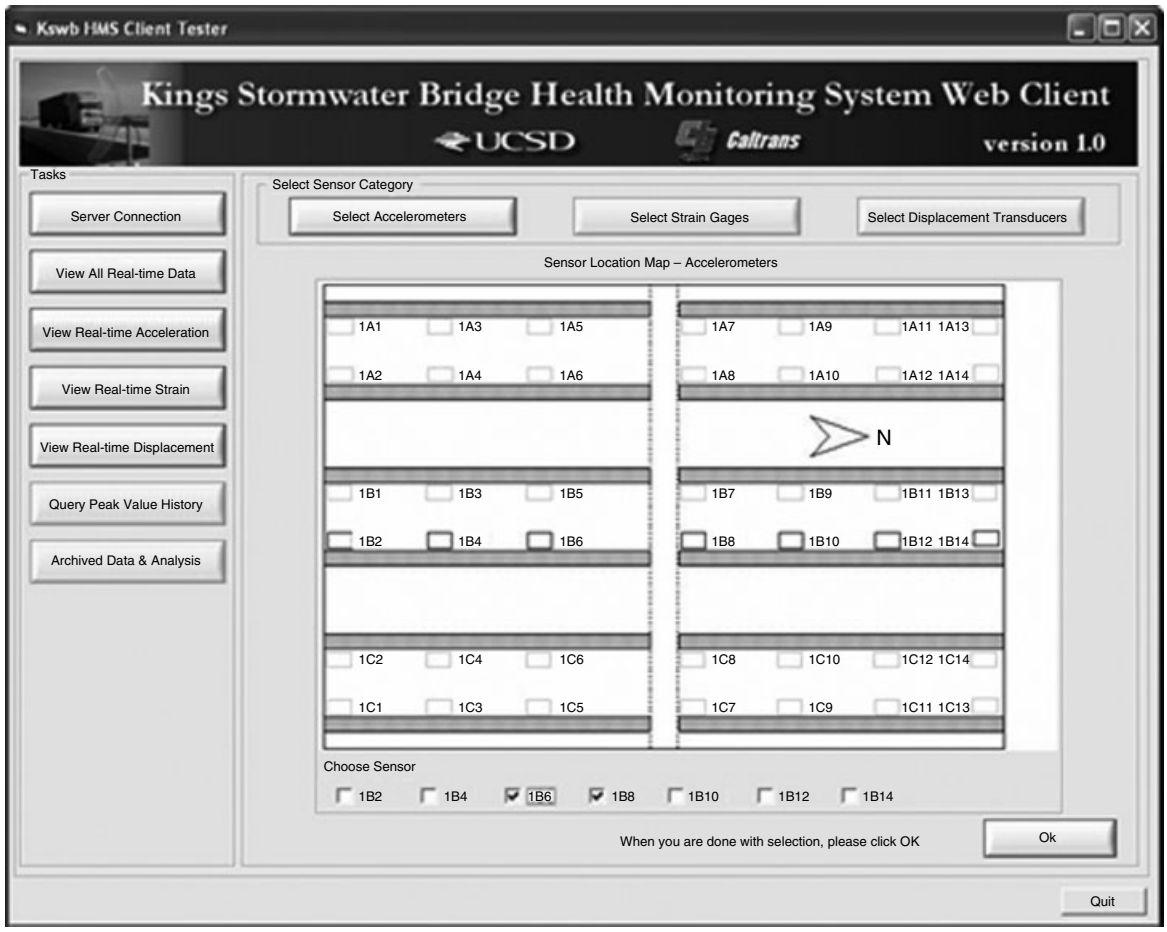


Figure 9. Example of map showing sensor locations.

multiple sensors selected by location. Results can then be viewed in real time (Figure 10a), with the capability of zooming in on details of any sensor (Figure 10b).

In addition, historical records can be assessed to compare peaks with current measurements, such as in Figure 11(a) for displacements, and in Figure 11(b) for strains.

3.4 Sample results of diagnostic analysis

At the current level of development, the vibration-based damage diagnostics information is supplemented by results from a finite-element model of the entire bridge (including deck panels, girders, deck-girder connections, cap beam, column/piers, and abutments), giving the user the capability of comparing structural response characteristics determined through the use of sensor measurements to FEA-based "ideal" or "updated" characteristics, as well as to predetermined response thresholds that signify bounds of acceptable structural behavior. The model was developed in ANSYS using design and construction drawings, and manufacturing specifications for the FRP composite components. The deck panels were modeled using elastic shell elements with equivalent properties, while the girders were modeled as beam elements. Connections between girders and deck panels were modeled as rigid constraints. The concrete barrier was also modeled using beam elements connected to the deck panels, while the connections to the abutments and the cap beam were modeled as rigid connections. The frequencies of the first five modes obtained from the initial FEA model are listed in Table 1 together with corresponding frequencies from a baseline forced

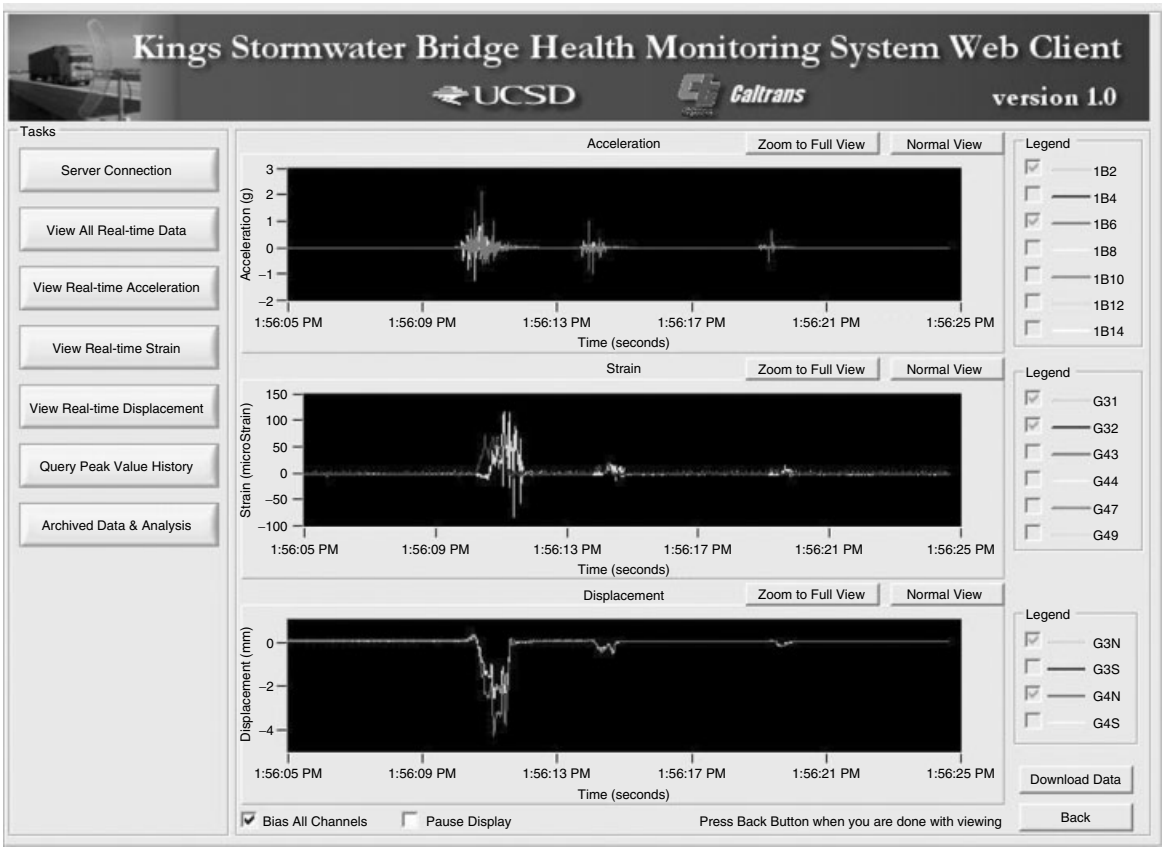


Figure 10a. (a) Real-time view of sensor-based response measurement. (b) "Zooming in" on data.

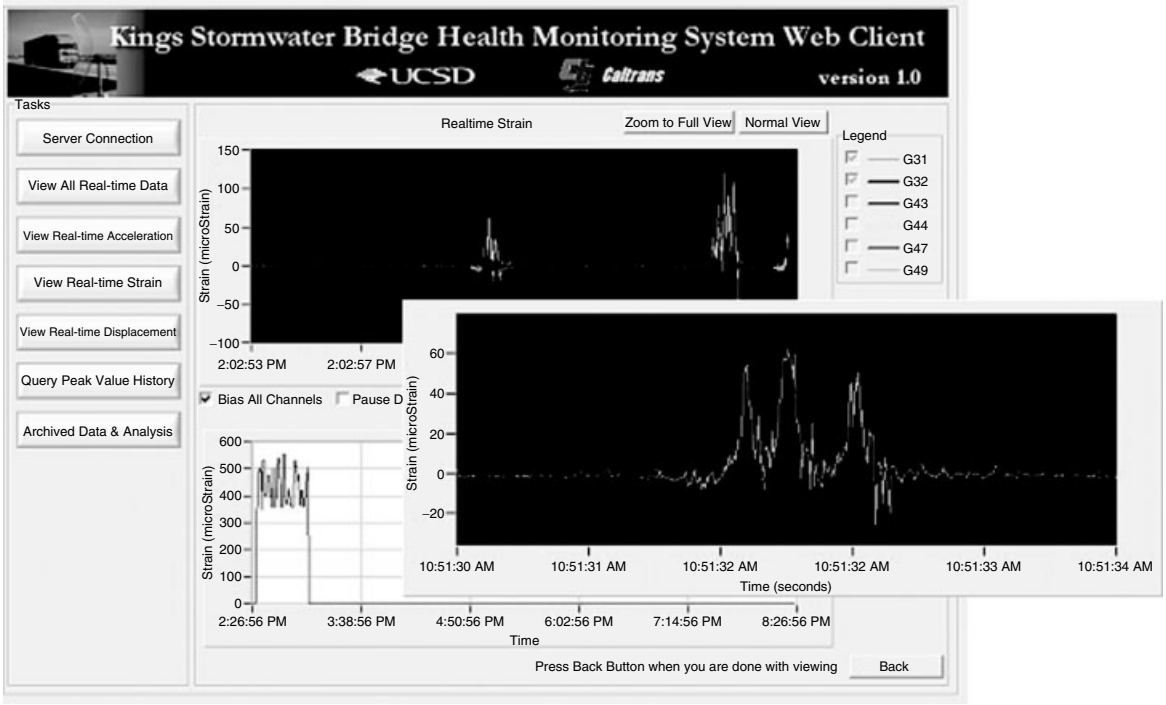


Figure 10b. *Continued.*

vibration test conducted just after the bridge was opened to traffic. A model updating process is then used to adjust the material properties used in the FEA model to better match the experiment results. The frequencies of the updated FEA model are also listed in Table 1 for comparison. This set was used as the initial baseline representative of the as-built nonaged structure. Figure 12 shows the mode shapes of the first two modes from the "baseline" finite-element model. Both the frequencies and the mode shapes are used hereafter for assessment of response as a function of time.

Shifts in natural frequency and the change of mode shapes provide an indication of stiffness changes in the structure. Four different data sets, collected in October 2001, June 2002, January 2003, and August 2003, respectively, are used as an example to show the time-related variations in modal parameters and the value of the methodology presented as a means of SHM. For each data set, the TDD algorithm was applied after the corresponding frequency ranges were determined. The change in natural frequencies is quite obvious when plotted out in Figure 13, which shows a decreasing trend of natural frequencies of both modes.

It should be noted, however, that there was a significant initial drop till June 2002 after which there appears to be very little change, especially considering the effect on response accruing from the changes in temperature. During this time period, some horizontal debonding at the saddle between the deck and girders was noticed, in addition to vertical cracking of the polymer concrete saddles in local areas with associated leakage of water through the deck joints and the cracks in the saddle, causing discoloration and streaking on the girder. This deterioration, in addition to the cracking and distress in the joints between decks, can be shown through FEA to lead to the change in frequencies, validating the experimental observations [6].

Although the changes may not be as clear when viewed in terms of mode shapes (Figure 14), a shift in peak location and magnitude can be noted on comparison of the August 2003 data with data from two other time periods.

Table 1. Comparison of frequencies of initial FE model and experiment

Mode number	Frequency (Hz)		
	Experiment	Initial FE model	Updated FE model
1	11.03	12.32	11.12
2	13.11	13.97	12.98
3	15.36	15.92	15.01
4	16.92	17.82	17.11
5	19.01	19.65	19.34

It is noteworthy that the analytically simulated mode shapes based on FEA models compare well with the experimental results, again showing the value of FEA use for both comparison at the threshold levels and for purposes of prediction. Translating the results into structural stiffness, this means the structure is becoming more flexible, i.e., decreasing structural stiffness with respect to time.

Use of the damage index method allows for determination of magnitude of damage indicators, which can provide assessment of damage in the structure. Changes in these indicators can be seen in Figure 15, and it is noted that the region with the highest damage indicator magnitude is concentrated around an area about 3.3 m (10 ft) on either side of the cap beam, which correlates with visual observation of damage through deterioration of the construction joint between deck panels. It is of interest to note that the increase in damage indicators in these

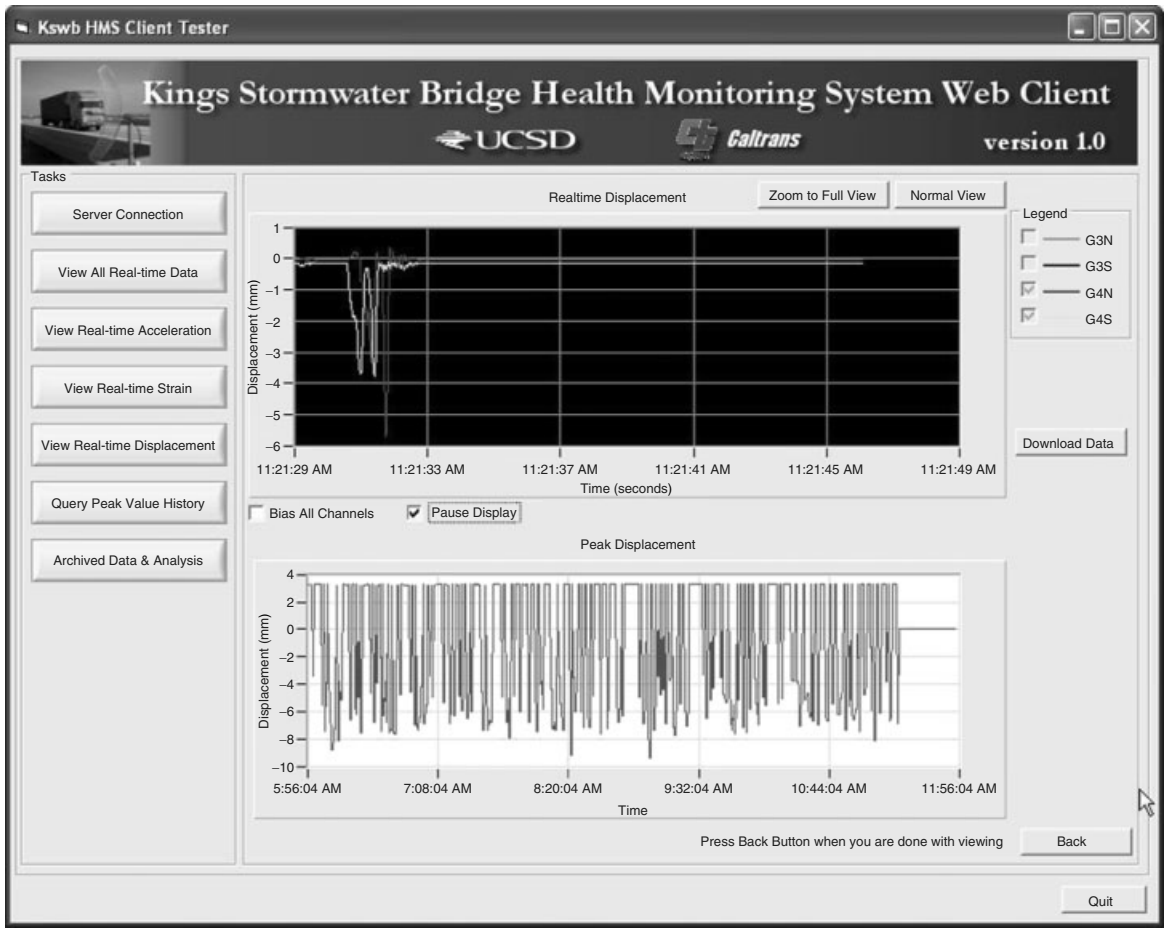


Figure 11a. (a) Access to peak displacements. (b) Access to peak strains from the overall data record.

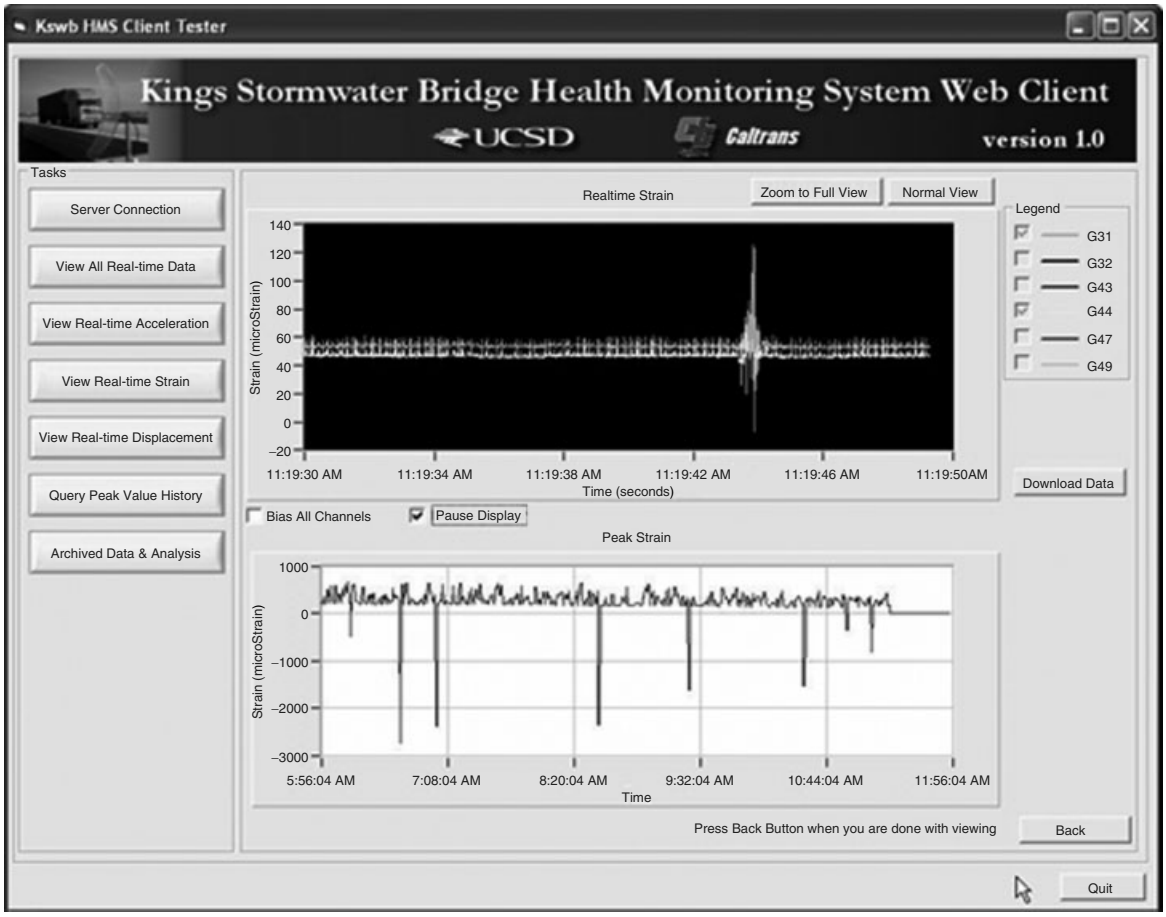


Figure 11b. *Continued.*

regions was noticed ahead of visual identification of the distress, emphasizing the value of such a health monitoring system in not just monitoring response but also in providing advance warning of deterioration.

4 SUMMARY

The design of components and structures is predicated on uncertainties related to a number of factors including those related to load, materials, and service environment. The lack of detailed knowledge in some of these has led conventionally to the use of intrinsically high factors of safety. While the use of an SHM system can provide, at the minimum, data that can be used to assess structural integrity in a manner allowing for provision of early warning of impending failure, it can also serve as a means of assessing reliability and, perhaps, even remaining characterizing capacity and remaining life on a continuous basis. In its development as a web-based system, it thus not only serves as a means of NDE but also as a means of collecting detailed records of response that will enable not just better designs of new structures but also the basis of better assessment and prediction of reliability of components and entire systems enabling better, and more timely, allocation of resources based on age, deterioration, and level of importance of the structure in terms of a local and regional context. Web-based SHM thus provides for the development of an *in situ* system similar to the system used by the federal aviation administration

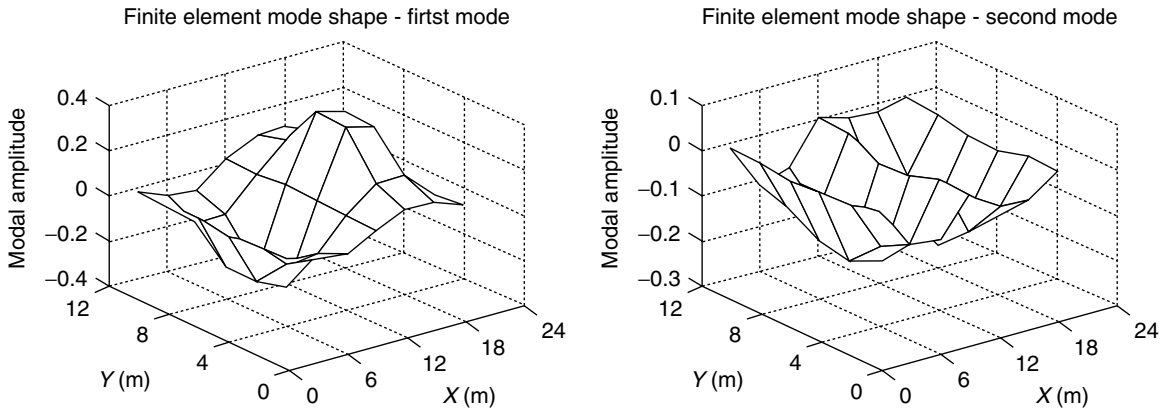


Figure 12. Mode shapes from the FEA model.

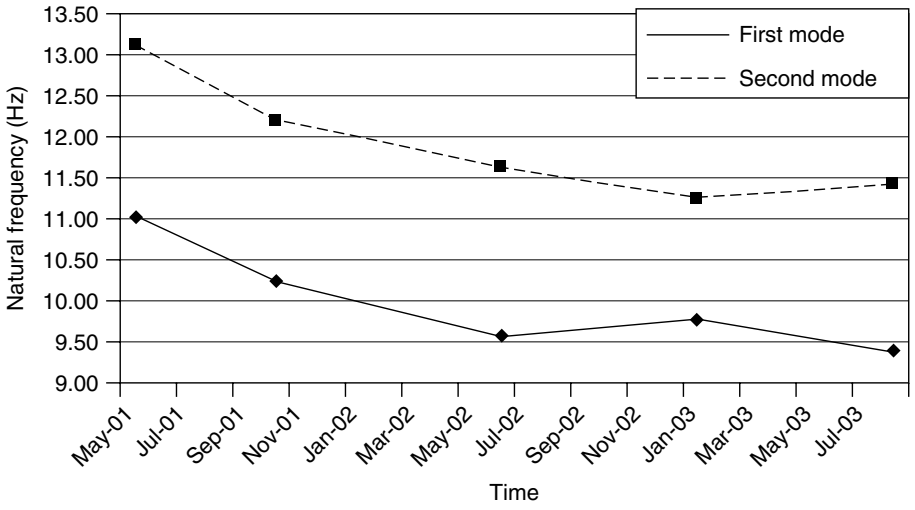


Figure 13. Change in frequency as a function of time.

(FAA) in tracking flights in the air, but based on fixed objects whose "health" varies and can be tracked in time.

The SHM system presented in this article provides an example of implementation of a system amenable to the use of autonomous monitoring. The system not only enables the collection and presentation of data from a suite of sensors but also provides for the assessment of the data in terms of structural characteristics, which can be compared to analytical results derived from an FEA model, which is calibrated to the initial response. This allows for the setting of response thresholds that would provide warnings to the user/owner when exceeded. In the current form of implementation, the web-based system has data transferred both in real time and in packages at preset intervals across a wireless network. Data is interrogated using a damage prognosis algorithm that provides indicators of damage severity and allows for comparison between response and damage state at various points in time. Thus subtle changes in bridge response, which may not be seen easily through just inspection of modal parameters, such as deterioration in

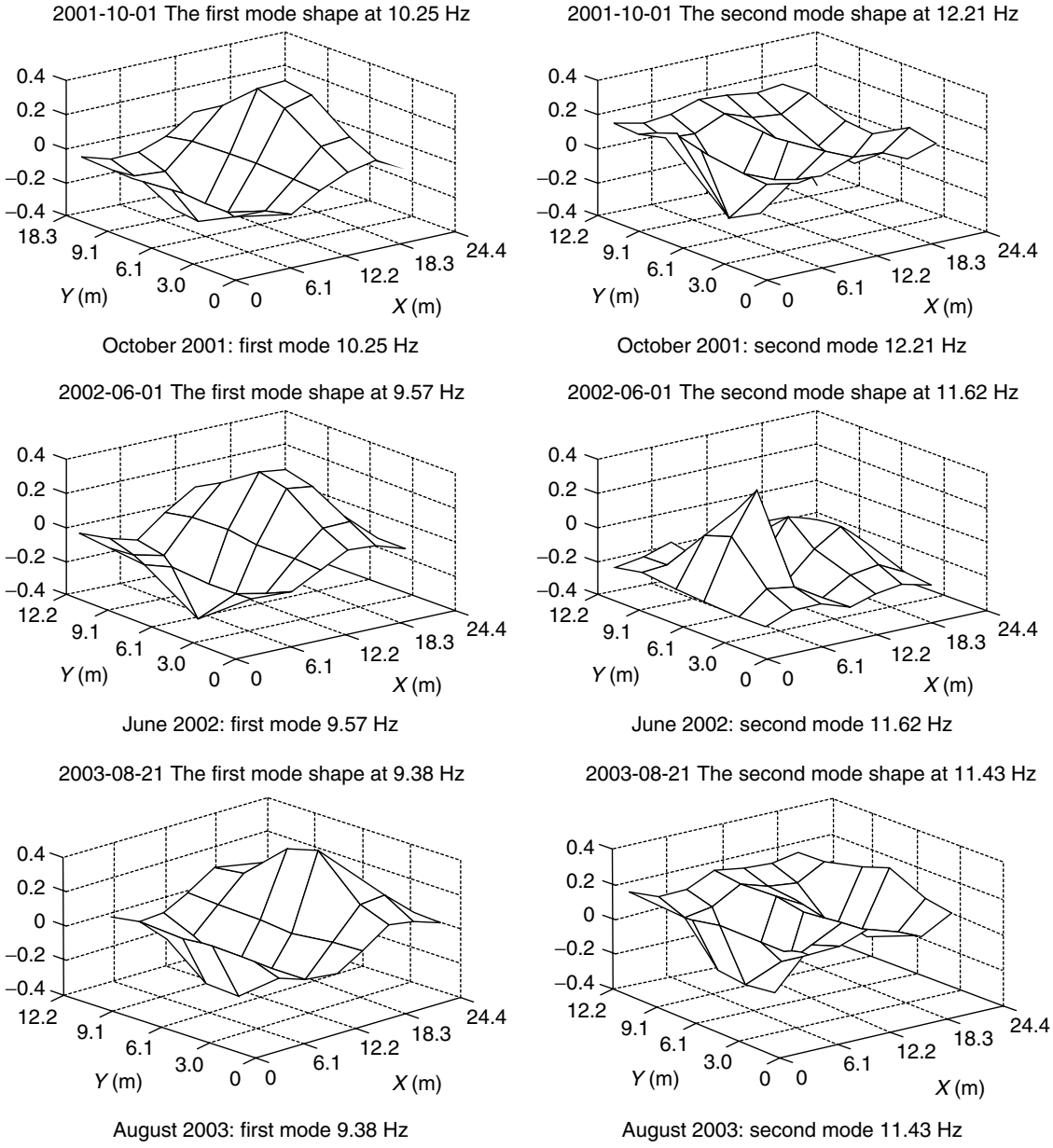


Figure 14. Changes in mode shapes over time.

expansion joints over time, can be easily determined. It is expected that the further development of such systems will lead to the establishment of a comprehensive methodology for autonomous health monitoring of structural systems to the point where true condition-based physical inspection and monitoring would become a reality. The integration of damage identification and finite-element-based tools further provides assistance to the engineer in assessing health immediately rather than having to resort to expensive closures while assessments are made off-line.

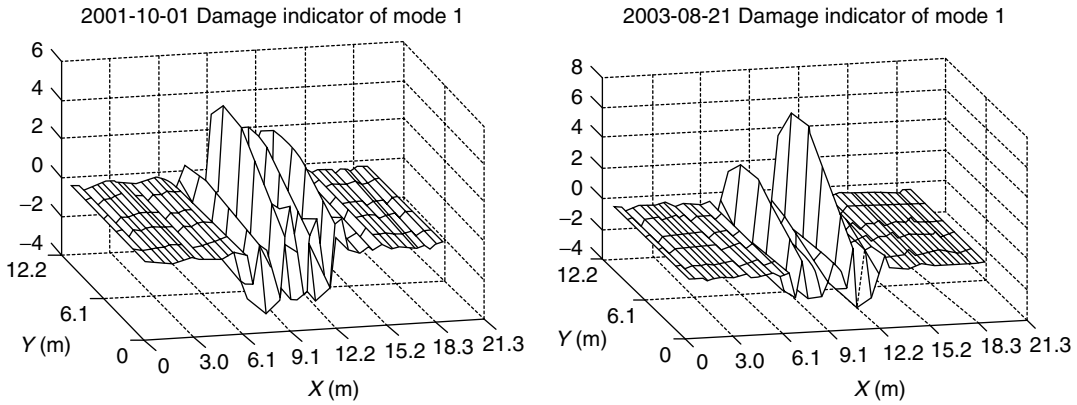


Figure 15. Examples of change in damage indicators.

REFERENCES

[1] Mufti AA, Bakht B, Tadros G, Horosko AT, Sparks G. Civionics—a new paradigm in design, evaluation, and risk analysis of civil structures. *Journal of Intelligent Material Systems and Structures* 2007 **18**:757–763.

[2] Morris B, Trivedi MM. Robust classification and tracking of vehicles in traffic video streams. *IEEE International Intelligent Transportation Systems Conference*, Toronto, Canada, September 2006.

[3] Adachi Y. Monitoring technologies for maintenance and management of urban highways in Japan. In *Sensing Issues in Civil Structural Health Monitoring*, Ansari F (ed). Springer: Dodrect, Netherlands, 2005, pp. 13–22.

[4] Stubbs N, Osegueda R. Global non-destructive damage evaluation in solids. *International Journal of Analytical and Experimental Modal Analysis* 1990 **5**(2):67–79.

[5] Karbhari VM, Seible F, Burgueño R, Davol A, Wernlli M, Zhao L. Structural characterization of fiber-reinforced composite short- and medium-span bridge systems. *Applied Composite Materials* 2000 **7**:151–182.

[6] Karbhari VM, Guan H, Sikorsky C. Web-based structural health monitoring of a FRP composite bridge. *Proceedings of the 1st International Conference on Structural Health Monitoring and Intelligent Infrastructure*, Tokyo, November 2003, pp. 217–226.

Chapter 74

Design of Active Sensor Network and Multilevel Data Fusion

Xiaoming Wang[1] and Zhongqing Su[2]

[1] *CSIRO Sustainable Ecosystems, Commonwealth Scientific and Industrial Research Organisation, Melbourne, VIC, Australia*
[2] *Department of Mechanical Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong, China*

1 INTRODUCTION

While traditional nondestruction evaluation (NDE), ranging from simple visual inspection to methods based on electrical properties, electromagnetic properties and acoustic or ultrasonic wave, has provided effective means in evaluating structural performance [1], it is understood that the application of structural health monitoring (SHM) has led to a paradigm shift in thinking from event-based structural evaluation to continuous on-line *in situ* monitoring. Such a philosophy may potentially reduce cost significantly without compromising performance and reliability of engineering structures such as airframes [2, 3] and bridges [4, 5]. It provides critical information of long-term structural performance for life cycle maintenance and management [6]. As one example, SHM has been introduced into a maintenance policy by the US Department of Defence, known as *Condition-Based Maintenance Plus* (*CBM+*), to reduce the cost in maintenance by a schedule-based approach [7].

SHM is a technique that uses sensors, on site or remotely, to collect information on structural, mechanical, or physical behavior, either continuously or periodically, for the diagnosis and prognosis of structural integrity and performance. It was comprehensively reviewed by many researchers [8–10]. Generally speaking, an SHM system may be specified

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

in terms of its functionality, sensing system design, and information acquisition and interpretation.

1.1 Functionality

An ideal SHM system should be capable of indicating structural health at four levels hieratically:

- level one: the occurrence of an adverse event to structure, affecting structural performance and functionality;
- level two: the location where an adverse event to structure occurs;
- level three: scale or severity of the adverse event to structure;
- level four: residual service life of structures or structural components to maintain required minimum performance and functionality with such an adverse event;

where an adverse event to structure can generally be considered as a consequence arising from an interaction between structures and their environment, which may include mechanical (loading, wind, impacts, and shock, etc.), chemical and biological (chloride, water, bacteria, etc.), thermal (radiation, heating, thermal shock, etc.), and so on. The consequence can be any form of damage or degradation in relation to performance and functionality of structures, such as crack, fatigue, buckling, delamination, dislocation, disconnection, etc.

1.2 Sensing system design

An appropriate sensing system is of vital importance to capture authentic signals to quantify the structural performance and identify any degradation in the performance. Much akin to the nerve cell of human beings, the sensor is a device for detecting the variation in physical, chemical, or biological properties, and, by proper transduction, transforming the measurands into interpretable information. Sensor technology is seen as a basic element in SHM. Basically, an ideal sensor system for SHM meets the following requirements: (i) veridical acquisition of changes in local or global structural responses; (ii) faithful delivery of captured changes; (iii) possibly less intrusion to host structure; (iv) endurance for general structural service conditions with its service life not less than that of the host structure; and (v) ease in handling, attachment, integration, and operation. In some demanding situations, e.g., for aerospace applications, the sensor should also feature small size, light mass, extremely high sensitivity and reliability, low cost and power consumption, little deterioration with aging, remote data transmission, robustness for noise, reduced wire, or even wireless, etc. While it is difficult to design a sensing system strictly satisfying all the requirements indicated above, an optimal design may be reached for a specific application by considering its most desired needs.

There are many sensing options including ultrasonic probe, acoustic emission sensor, magnetic sensor, eddy-current transducer, accelerometer, strain gauge, laser interferometer, optical fiber, electromagnetic acoustic transducer (EMAT), etc. In particular, piezoelectric elements have been generating a lot of interest for SHM, especially because of their potential of being used as either sensors or actuators [1–15].

By virtue of the utility of piezoelectric elements, an active sensing scheme can be designed on the basis of the assessment of electromechanical impedance spectrums (*see* **Electromechanical Impedance Modeling**). It is implemented by measuring the electric current in response to the voltage imposed on the piezoelectric element integrated with structures. The electromechanical impedance is directly related to structural dynamics [16], and therefore any structural change will, in principle, lead to variation in the impedance. Applications of such a scheme include detection for bridge joint deterioration [17, 18], disbond in composite repair patch [19], loose joints in pipelines [20], and cracks in a concrete beam [21].

The active sensing can also be achieved by using Lamb waves. Lamb waves have been widely explored [22] since 1961 when they were introduced as a means of damage detection [23]. Lamb waves, made up of a superposition of longitudinal and shear modes, are available in a thin plate or shell structures. The mode of a Lamb wave can be either symmetric or antisymmetric, and can be excited by a variety of means [10]. Details can also be found in Section 2 (*see* **Fundamentals of Guided Elastic Waves in Solids**). Piezoelectric elements are capable of generating and acquiring Lamb waves,

and are particularly suitable for integration into a host structure as an *in situ* actuator and sensor, as a result of their low mass, good mechanical strength, wide frequency band, low power consumption and acoustic impedance, as well as low cost. Theoretically, a sensing system can be designed on the basis of the concept to relate damage-modulated Lamb waves signal to damage location, size, or severity, offering an SHM system featured with functionality at all four levels.

1.3 Signal acquisition and interpretation

Signal acquisition is designed to capture any information related to structural health status, which can further be interpreted at the four levels of functionality required by an SHM system. The concept to apply piezoelectric elements for identifying structural damages by assessing Lamb waves can be schematically described by Figure 1. As seen, one piezoelectric element is used as an actuator to generate Lamb waves, and the other as a sensor to collect signals. With the occurrence of damage, the original signal is considerably modulated by the damage, as in the example described in Figure 2(a) and (b). In comparison with Figure 2a, without the effect of structural damage, Figure 2(b) shows that a structural damage such as delamination in a composite laminate induces a *shear horizontal* (SH) mode wave [24], and there is a time lag between the SH wave and the symmetric mode (S) wave.

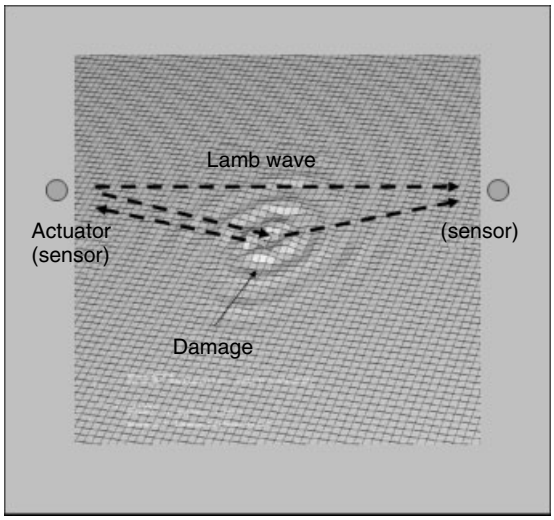


Figure 1. Active sensing scheme with Lamb waves.

Basically, there are two approaches to interpret the signals acquired by a sensor in an active sensing scheme based on Lamb waves:

- using *time of flight* (TOF) of damage-modulated Lamb waves between recipient wave and the damage-reflected wave, to identify a damage event and quantify the location of damage—the functionality at levels one and two required by a SHM system;
- correlating features of damage-modulated Lamb waves to identify a damage event, quantify

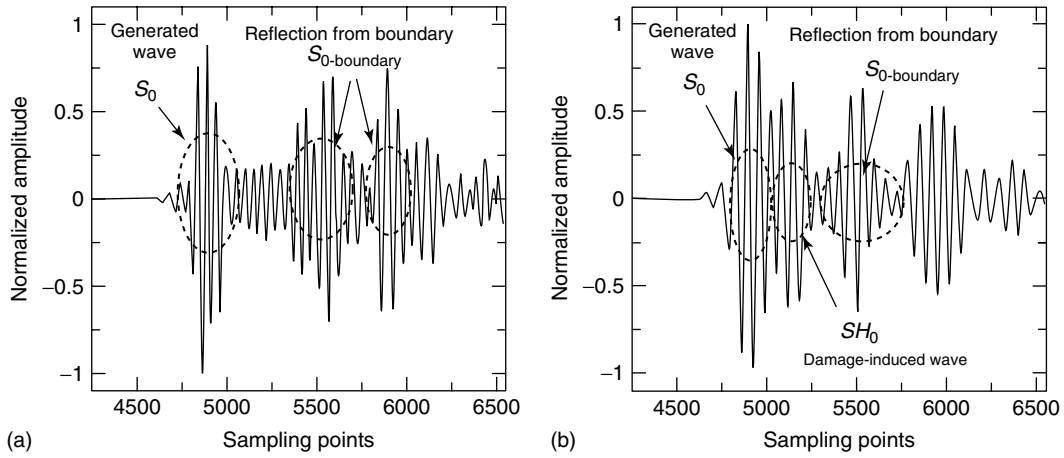


Figure 2. Acquired Lamb wave signals: (a) without damage modulation and (b) with damage modulation.

damage location, and size or severity—the functionality at levels one, two, and three required by an SHM system.

A proper sensing system design into network may also extend the first approach to achieve functionality at level four. The features mentioned in the second approach can be obtained in the time domain [25], but it is more common to analyze features in the frequency domain by approaches such as the fast Fourier transform (FFT) [26–28]. It is more reasonable to combine the analyses in both the time and the frequency domains so as to avoid any information loss in either of these domains, i.e., the time–frequency domain analysis. The joint time–frequency domain analysis is exemplified by the *short-term Fourier transform* (STFT) and the *Wigner–Ville distribution* (WVD) [29]. Recently, *wavelet transform* (WT) has become more popular in feature extraction for SHM [30–32], such as denoising a signal to facilitate feature extraction in the time domain [33].

2 ACTIVE SENSING

Active sensing design using piezoelectric elements has created great potential to advance the development of SHM techniques, especially in applications pertaining to plate and shell structures that largely exist in airframes, marine vehicles, containers or tanks, pipelines, and so on. An active piezoelectric sensing system is capable of both generating interrogative signals and capturing structural-modulated responses, which reflect the condition of structures, over *an area* instead of at *a point*. However, the identification of structural condition is normally shown as a typical inverse problem that represents the solution (structural condition) in terms of a *hidden* physical phenomenon to be obtained or estimated from *observed* signal from sensors. Mathematically, it often shows that the process is highly ill-conditioned or not well-posed that requires, (i) a solution to exist, (ii) the solution to be unique, and (iii) the solution to continuously depend on the observed data. Complication of structural damages, such as occurring at multiple locations, may impose even more difficulties to obtain the solution correctly. While a simple active piezoelectric sensing system may potentially monitor a broad structural area, it also creates uncertainties in signal interpretation caused by noises and signal scattering as a result of environment, structural boundary, and discontinuity interference. Distinguishing the damage-induced signal variation from noises, particularly in engineering practices, is very challenging.

Intrinsic problems in seeking an inverse solution as well as practical issues in view of system operations require a new path to increase the robustness and reliability of SHM. Artificial intelligence techniques can be adopted to reproduce human cognitive processes for the identification of structural abnormality and derive *beliefs* at the level of an individual sensor, which is essentially a part of a decentralized or distributed sensor network. An overall *decision* or consensus on structural condition becomes available by combining the *beliefs* within a group and then subsequently among groups of individual sensors in the network in multiple steps, known as *a multilevel fusion process*, which may considerably increase the robustness in obtaining inverse solutions and reduce uncertainties in damage identification for SHM. How to design a distributed active sensing system to suit the needs of a multilevel decision fusion process becomes important.

3 DISTRIBUTED SENSING SYSTEM AND DECISION FUSION

Neurological study [34] indicates that signals to different senses are initially segregated at the neural level and neurons do not interact with each other on what they sense until the signals are transmitted to the brain, where the sensory signals converge on the same target to supply perceptions and orient behavior. It is understood that there are three critical steps involved in such a process: (i) distributed sensing, (ii) information fusion, and (iii) establishing perception and orienting behavior. Extending the concept to SHM, sensors can be designed in a distributed way—not only in terms of a spatial network but also in terms of their functions—implying that they are able to sense different physical parameters at many locations in a specific pattern. Distributed sensors may firstly form their individual beliefs that represent their own interpretation of data or "the world" they sense.

Secondly, a fusion process then combines all their beliefs following conjunctive, disjunctive, or compromising rules to form a consensus about the view on "the world"-health status of the structures. The advantages of using a distributed sensor network include the combination of all kinds of information sources that can be the same type or completely different, the enhancement of the awareness of targets with superior system reliability and robustness [35], and the reduction in dependence of final perception/decision on the information from a single sensing source.

3.1 Distributed active sensing structure

Figure 3 gives a representative unit in a distributed active sensing structure, including the following four elements: (i) active sensor node, (ii) passive sensor node, (iii) sensor controller, and (iv) sensing path. The active sensor node plays an active role by generating an interrogative signal propagating through a sensing path and forming its belief about structural health on the basis of the response of the passive sensor node. The passive sensor node is used to receive a signal transmitted through structures from the active sensor and provide feedback to the active sensor node by the sensor controller. The sensor controller decides which sensor node takes an active role, relays and processes feedback to establish a belief for the active sensor node. The sensing path is the possible pathway of interrogative signal traveling from the active sensing node to the passive sensing nodes within a structure.

A distributed active sensing structure can be designed in a centralized format, as shown in Figure 4, where sensors may rotate to take an active role to generate interrogative signals and get feedback

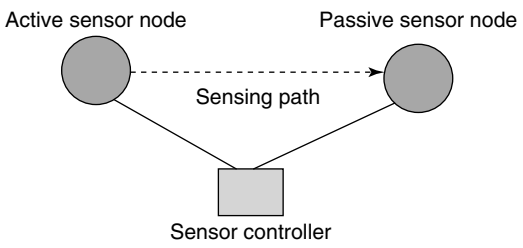


Figure 3. Sensor nodes, controller, and sensing path.

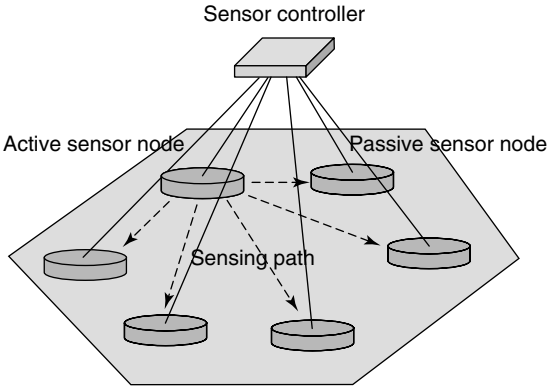


Figure 4. Centralized active sensing structure.

from all other (passive) sensors only through one sensor controller. All feedbacks then become inputs to form a belief of the active sensor group in regard to structural health status. It should be indicated that the entire process is controlled by one sensor controller, know as *a centralized structure*. Figure 5 shows a hierarchical structure, which has an extra sensor controller at a higher layer. As mentioned in the centralized structure, each sensor may have its own belief on the status of structural health. The sensor controller at the low layer may form a belief of a sensor group on the basis of the beliefs of individual sensors through the second level of the fusion process. In the hierarchical structure, the high-layer sensor control subsequently establishes a belief of entire sensor groups by fusing the beliefs formed at the low-layer sensor controllers.

Different from the hierarchical structure, the autonomous active sensing structure, as shown in Figure 6, does not possess any high-layer sensor controller. Sensor controllers are independent of each other, but maintain communication among themselves. Note that the belief of entire sensor groups is established at each sensor controller by acquiring information from all the others with their own rules, which are not necessarily the same.

More complex active distributed sensing structures can be established on the basis of the basic structures discussed above. One example is the hybrid structure, which is intended to integrate the features of centralized, hierarchical, and autonomous structures.

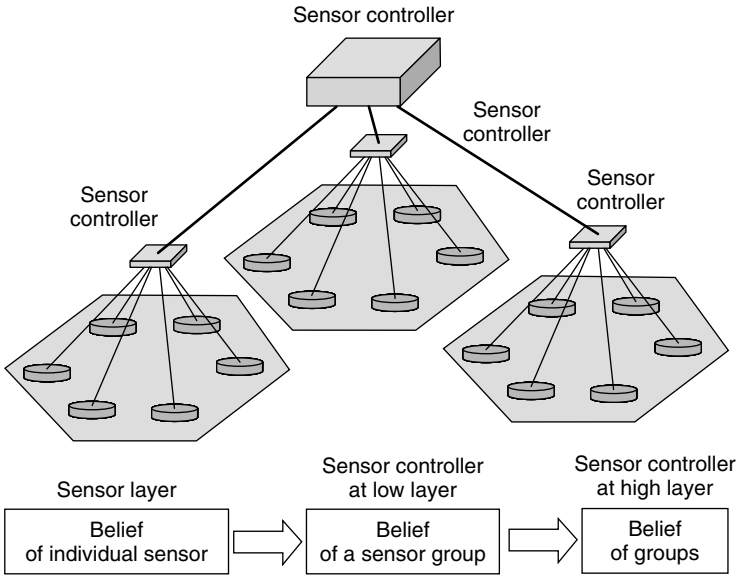


Figure 5. Hierarchical active sensing structure.

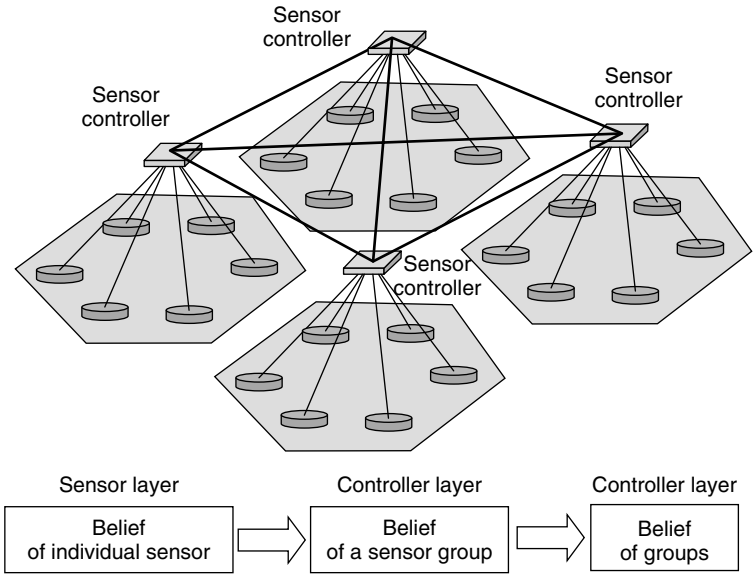


Figure 6. Autonomous active sensing structures.

3.2 Fusion in distributed active sensor network

As discussed previously, the establishment of a belief at each layer, i.e., sensor layer and sensor controller at a low or high layer, has to be implemented via fusion, which basically combines the beliefs at the lower layer to form a new belief or posterior belief at a higher layer. The fusion process is intended to make the posterior brief with less imprecision, uncertainty, and incomplete information, accurately characterizing the structural health status. More precisely, a brief

is described by a degree that represents how certain or possible the perception of a sensor or a sensor controller may enhance such a view or consensus as the occurrence of a series of potential adverse structural health events. A fusion process is described by a projection function, $\Im$, which projects a vector of the prior belief, $\vec{B}_{\text{prior}}$, at a lower layer to a posterior belief, $B_{\text{posterior}}$, at a higher layer presented by

$$\Im: \vec{B}_{\text{prior}} \longrightarrow B_{\text{posterior}}, \text{ for } \vec{B}_{\text{prior}} \in I^n \text{ and } B_{\text{posterior}} \in I \tag{1}$$

where I represents a measure set of the degree of the posterior belief as well as the degree of all subsets of the prior belief given in n dimension. It is normally defined as an interval, for example, [0, 1], where "0" indicates that the prior belief *cannot* enhance the posterior belief or provide information to strengthen the consensus, and "1" represents that the prior belief can enhance the posterior belief.

Character of fusion functions can be divided into three basics [36]. Assuming that there are two prior belief inputs with their degrees described by x and y, and a fusion function is given by $\Im$ with the posterior belief given by $\Im(x, y)$, then

- $\Im$ is conjunctive if $\Im(x, y) \leq \min(x, y)$;
- $\Im$ is disjunctive if $\Im(x, y) \geq \max(x, y)$;
- $\Im$ is a compromise if $\min(x, y) \leq \Im(x, y) \leq \max(x, y)$.

Conjunctive fusion gives the common part among the prior beliefs, reducing the less certain components and giving the smallest measure, as shown in Figure 7(a). Disjunctive fusion covers all possible beliefs, increasing the certainty and giving the greatest measure, as shown in Figure 7(b). Compromise fusion gives an intermediate measure, as in Figure 7(c).

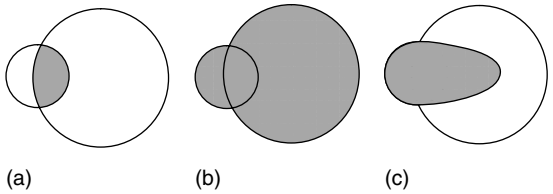


Figure 7. Fusion behavior: (a) conjunctive, (b) disjunctive, and (c) compromise.

In practice, a fusion function can have any one of the function character described above or their mixture, depending on the context of the fusion function. The selection of fusion schemes has to be considered in relation to the data sets and the fusion objectives. The following are a few examples of fusion functions [36]:

- conjunctive fusion functions

$$\Im(x, y) = \{\min(x, y); xy; \max(0, x + y - 1)\}, \quad (x, y) \in [0, 1] \times [0, 1] \tag{2}$$

- disjunctive fusion functions

$$\Im(x, y) = \{\max(x, y); x + y - xy; \min(1, x + y)\}, \quad (x, y) \in [0, 1] \times [0, 1] \tag{3}$$

- compromising fusion functions

$$\Im(x, y) = \left\{ \frac{2xy}{x + y}; \sqrt{xy}; \frac{x + y}{2}; \sqrt{\frac{x^2 + y^2}{2}} \right\}, \quad (x, y) \in [0, 1] \times [0, 1] \tag{4}$$

More fusion techniques include the voting scheme, Bayesian inference [37], Dempster–Schafer theory [38, 39], and fuzzy logic [40].

3.2.1 Fusion based on voting scheme

In a fusion based on voting scheme, the posterior belief is formed according to the voting index [37]:

$$\Im(\vec{Y}_i) = \sum_{j=1}^{n_i} w_{ij} Y_{ij} \tag{5}$$

where n_i is the total number of prior beliefs from all sensors used to assess the ith location ($i = 1, \ldots, N$), and Y_{ij} is the prior belief degree of the jth sensors on the structural condition at the ith location. Typically, it equals one when supporting a tested consensus and zero when against it. In equation (5), w_{ij} represents the voting weight of the jth sensor in relation to the ith location, and $\sum_j w_{ij} = 1$. The fusion based on voting scheme is compromising.

3.2.2 Fusion based on Bayesian theory

In Bayesian fusion, with multiple inputs, E_i, from a distributed sensor network, the posterior probability of the belief, $\Theta = \theta_i$, is given by [37]

$$P(\Theta = \theta_i | E_1, \ldots, E_j, \ldots, E_m) \sim P(\Theta = \theta_i) \prod_{j=1}^{m} P(E_j | \Theta = \theta_i) \quad (6)$$

where $P(\Theta = \theta_i)$ is the prior probability of the belief of $\Theta = \theta_i$; $P(\Theta = \theta_i | E_i)$, $(i = 1, \ldots, m)$ is the posterior probability of the belief of $\Theta = \theta_i$, given inputs $E_i (i = 1, \ldots, m)$; and $P(E_i | \Theta = \theta_i |)$ is the likelihood of the occurrence of E_i at the assumption of $\Theta = \theta_i$. The fusion based on Bayesian theory is conjunctive.

3.2.3 Fusion based on Dempster–Shafer rule

Assuming mass functions $m_j(B_j)(j = 1, \ldots, m)$ that represent the degree of belief of the sources (e.g., sensors) on the proposition B_j, the Dempster–Shafer rule for the fusion of belief degree on the proposition Θ, where $\Theta = B_1 \cap \ldots B_j \cap \ldots B_m$, is then given by [38, 39]

$$m_1 \oplus \ldots m_j \oplus \ldots m_m(\Theta) = \frac{\sum_{B_1 \cap \ldots B_j \cap \ldots B_m = \Theta} m_1(B_1) m_2(B_2) \ldots m_m(B_m)}{1 - k} \quad (7)$$

where

$$k = \sum_{B_1 \cap \ldots B_j \cap \ldots B_m = \emptyset} m_1(B_1) m_2(B_2) \ldots m_m(B_m) \quad (8)$$

where k is a measure of conflict between sources, reflecting the imprecision of information; and $B_1 \cap \ldots B_j \cap \ldots B_m \cap \ldots = \emptyset$ implies the scenario with null intersection among all $B_j (j = 1, \ldots, m)$. The fusion based on Dempster–Shafer rule is conjunctive.

3.2.4 Fusion based on fuzzy interference

Assuming membership functions, $\mu_{A_j}(x)(j = 1, \ldots, m)$ that define a fuzzy set of beliefs on the proposition, such as damage occurrence at a location, the fusion through fuzzy conjunction is given by [40]

$$\mu_{A_1 \cap A_2 \ldots \cap A_m}(x) = \min\{\mu_{A_1}(x), \mu_{A_2}(x), \ldots, \mu_{A_m}(x)\} \quad (9)$$

where x is the universe of discourse related to the degree of prior beliefs. To defuzzificate the combined fuzzy conclusion in equation (9), the centroid defuzzification method is applied by using the following equation:

$$\bar{x} = \frac{\sum_{j=1}^{m} \mu_{A_j}(x_j) x_j}{\sum_{j=1}^{m} \mu_{A_j}(x_j)} \quad (10)$$

where $\bar{x}$ is the posterior belief through fusing degree of prior beliefs, on the proposition of damage occurrence.

3.2.5 Fusion with neural network

A neural network consists of one input layer with α elements (im), two hidden neural processing layers respectively possessing λ and η computing elements (referred to as *neurons*), and one output layer containing β variables (ov). Mathematically, the ith output variable in the network is formularized as

$$ov_i = T_3\left(\left(\sum_{q=1}^{\eta} w_{q-i}^3 \cdot T_2\left(\left(\sum_{r=1}^{\lambda} w_{r-q}^2 \cdot T_1 \times \left(\left(\sum_{p=1}^{\alpha} w_{p-r}^1 \cdot im_p\right) + b_r^1\right)\right) + b_q^2\right)\right) + b_i^3\right) \quad (11)$$

where im_p denotes the pth input element. $w_{p-q}^r (r = 1, 2, 3)$, defined as *weight*, represents the linkage joining the pth input element/neuron in the rth layer with the qth neuron/output variable in the next layer. Similarly, b_q^i, called *bias*, is an offset constant for the qth element in ith layer, $(q = 1, 2, \ldots, \lambda/\eta/\beta$ for the first/second-processing/output layer). $T_s (s = 1, 2, 3)$ is the transfer function in the network [41]. In the fusion with a neural network, the inputs may be the features related to prior beliefs, while outputs are the parameters related to posterior beliefs.

4 DISTRIBUTED SENSOR NETWORK FOR DAMAGE DETECTION OF COMPOSITE STRUCTURES

The following gives a few examples to demonstrate the applications of distributed sensor network and one/multiple level of fusion in damage detection of composite structures, as shown in Figure 8. The simulation is concentrated on a representative active sensing unit (RASU), extracted from structures with a complex distributed sensor network.

In a distributed active sensor network, each sensor takes a role of Lamb wave generator (active sensor) or receiver (passive sensor). The information fed back to the active sensor from a passive sensor within a sensing path is represented using a set of digitalized signal features, termed digital damage fingerprint (DDF) in a study [42] pertaining to structural conditions. DDF is a kind of characteristic of a scattered Lamb waves signal that are extracted from the wavelet filter-processed signal and energy spectrum by gradually screening the noncharacteristic components and using principal signal components only, as described by a simple example shown in Figure 9.

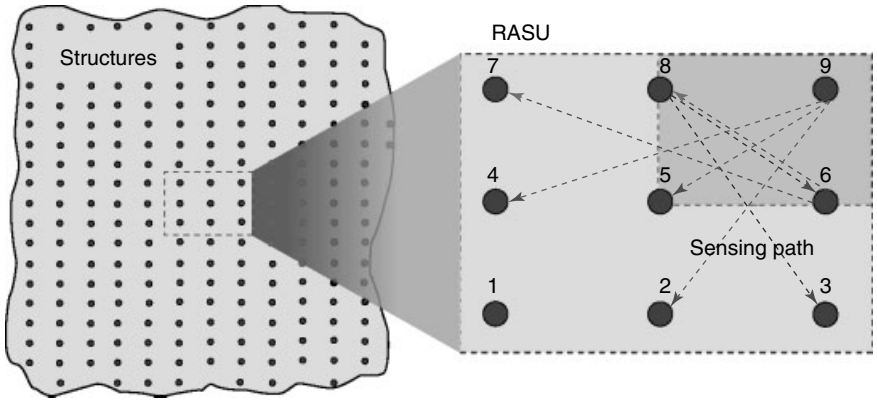


Figure 8. Representative active sensing unit (RASU) from structures.

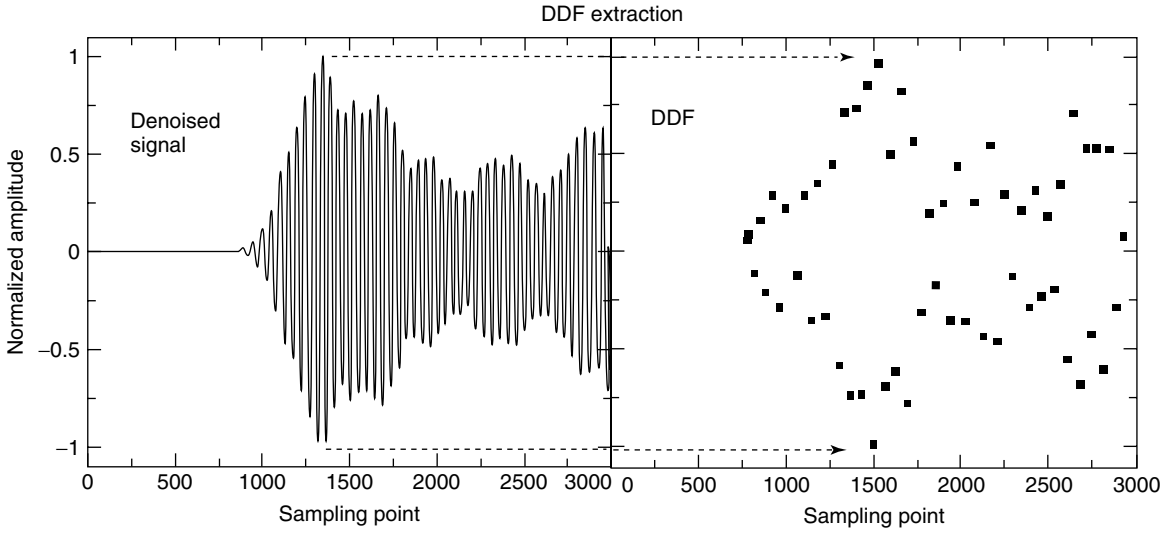


Figure 9. An example of DDF extracted from a denoised signal by a wavelet filter.

The one or multiple level fusion process is strongly related to DDF.

By way of illustration, a CF/EP (T650/F584) laminate was supported on all its four sides with a stacking sequence of $[45/-45/0/90]_s$ and measuring $500\,\text{mm} \times 500\,\text{mm} \times 1.275\,\text{mm}$ [43]. A sensor network of nine piezoelectric elements was surface-bonded on the laminate, sensors being numbered Pi $(i = 1, 2, \ldots, 9)$, similar to the one in Figure 8. The laminate was pretreated with damage (hole or delamination) during fabrication. Without losing generality, the damage was presumed elliptic, defined with six parameters: presence (0 or 1), location (ξ, ζ), semimajor/minor axes (α, β) and orientation of θ, although it is very often that not all of those details are required. Knowledge about digital damage fingerprints DDFs in relation to each sensing path under a specific damage was obtained through FEM simulation. Owing to the symmetric character in the discussed examples, only the area as shadowed in Figure 8, surrounded by sensors 5, 6, 9, and 8, was considered for the location of damage, though sensors 1 to 9 were all used in the damage identification process to identify damages. In establishing the knowledge, the cases of damage (location, shape, and orientation of holes/delamination) assumed are shown in Figure 10.

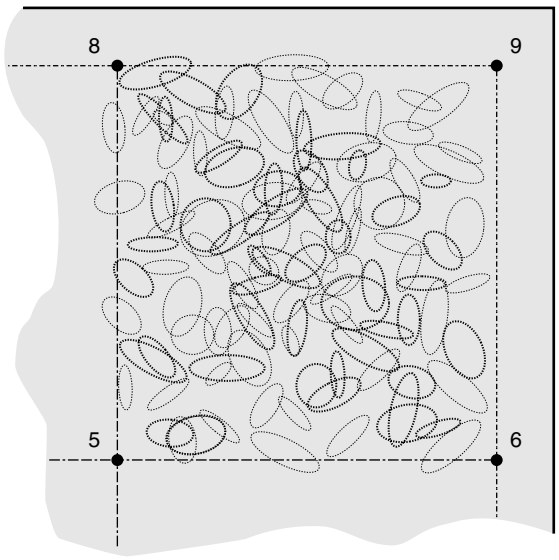


Figure 10. Samples of damages for acquiring knowledge of relevant DDFs.

4.1 Centralized sensing structure and one-level fusion process

In this example, the sensing structure was designed to be centralized, with one sensor controller to regulate one of sensors to generate Lamb waves and the others to collect signals, as shown in Figure 4. Feedback from all passive sensors became inputs to form a belief of structural condition status via a fusion process. Neural network was applied to implement the fusion process. The prior belief was described by the knowledge of delamination characteristics (location, shape, and orientation) in relation to DDFs obtained through FEM simulation. That knowledge was then employed to train a neural network and to establish the relation between DDFs and delamination.

In experiments, a delamination was pretreated as shown in Figure 11(a). DDFs were obtained on the same sensing paths as those utilized to train the neural network. A one-level fusion process was then carried out by combining all obtained DDFs as inputs of the trained neural network. The posterior belief about delamination was then established, as shown in Figure 11(b). The belief about the damage is reasonably close to the actual one.

4.2 Hierarchical distributed sensing structure and multilevel fusion

The belief of the sensors on structural damage status can also be quantified by correlating real-time acquired DDFs with preestablished knowledge about the DDFs, which are obtained from FEM simulation in association with specified damage patterns, illustrated in Figure 10. As a result, corresponding to each sensing path, there is one correlation coefficient C_{ij}^k, which can be considered as the degree of belief of an active sensor i within a sensing path $\vec{ij}$ on a specific damage pattern D_k defined in the preestablished knowledge. Theoretically, when there are N sensors, there exist $N-1$ sensing paths for one active senor i, or $\vec{ij}(j = 1, \ldots, i-1, i+1, \ldots, N)$. The level-one fusion is required to consolidate the correlation coefficients $C_{ij}^k(j = 1, \ldots, i-1, i+1, \ldots, N)$ to form a belief, $Bel(k; i)$, of the active sensor i on the damage patterns $D_k(k = 1, \ldots, M)$, or

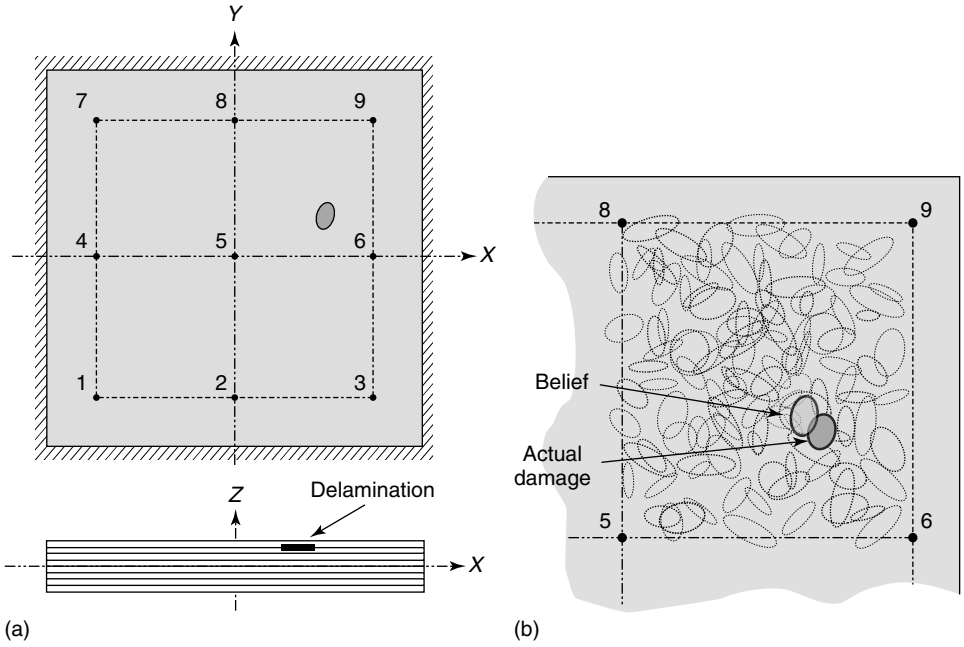


Figure 11. Experiment and identification: (a) A CF/EP laminate with a pretreated delamination and (b) comparison of the actual pretreated delamination with the belief on the damage formed on using an artificial neural network (ANN) [43].

described by

$$\Im_1(k) : C_{ij}^k \longrightarrow Bel(k;i), \text{ for } C_{ij}^k \in [0,1]^{N-1} \quad (j = 1, \ldots, i-1, i+1, \ldots, N) \quad \text{and } Bel(k;i) \in [0,1] \tag{12}$$

Meanwhile, all sensors may form their own beliefs, $Bel(k;i)(i = 1, \ldots, N)$, on each damage pattern. The level-two fusion is required to combine all beliefs of sensors to establish a belief of a group of sensors, which is normally at a sensor controller layer in the hierarchical active sensing structure as discussed previously. The fusion function for a sensor group p is described by

$$\Im_2(k) : Bel(k;i) \longrightarrow Bel(k;p), \text{ for } Bel(k;i) \in [0,1]^N (i = 1, \ldots, N) \quad \text{and } Bel(k;p) \in [0,1] \tag{13}$$

When there are groups of sensors, or there are groups of fusion approaches for one sensor group, that are used to form a posterior belief, the level-three fusion, $\Im_3(k)$, has to be implemented at a higher layer in the hierarchical structure, but very much similar to the one described in equations (12) and (13).

With fusion processes based on Bayesian theory and voting scheme [44], the belief about damage can be identified through level-one fusion, as shown in Figure 12(b) and (e), and level-two fusion, as shown in Figure 12(c) and (f). The actual damage is highlighted by a dotted circle. It is shown that the belief about the damage from one sensing path, as shown in Figure 12(a) and (d), is very vague, and actually cannot identify the damage. The belief after level-one fusion increases the confidence of belief about the damage, while the accuracy in identifying damage improves significantly after level-two fusion.

It is interesting to find that the contrast of the result from the Bayesian theory is much better than the one from the voting scheme. As discussed previously, the fusion based on the Bayesian theory is conjunctive, which may reduce uncertain components in the fusion process and provide a smallest measure. The fusion based on the voting scheme is compromising, which take an intermediate measure, implying that it could not reduce uncertain data as good as the fusion

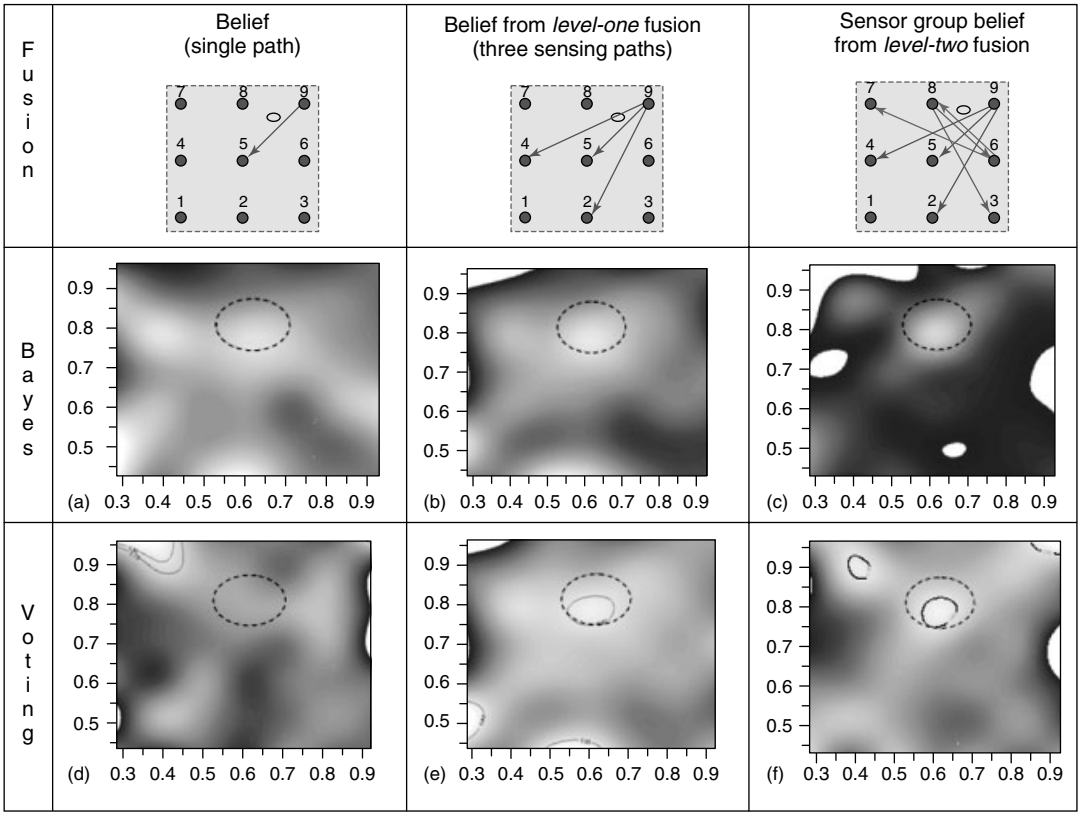


Figure 12. Damage identification by a fusion process, (a), (b), and (c) were obtained by fusion based on Bayesian theory, (d), (e), and (f) were obtained by fusion based on voting scheme (results of (c–f) extracted from [44]). The gray scale is equivalent to the probability of damage occurrence with the lighter corresponding to the higher possibility.

process based on the Bayesian theory. Therefore, the proper selection of fusion process is important.

5 SUMMARY

Design of a sensing network, and subsequently the interpretation of data acquired from sensors within the structure, are two critical issues in SHM for practical applications. It is understood that the process to identify structural health status is basically an inverse problem, which is often mathematically ill-posed or ill-conditioned creating considerable barriers to offer a robust and reliable identification algorithm, especially based on a simple sensor structure.

Active sensor network is designed to deal with the problems of health monitoring, capable of providing richer and sometime redundant information for structural condition identification by spatially distributed sensing structure. The sensing structure can be centralized, hierarchical, or even autonomous, in correspondence to the requirements of data interpretation.

The challenges faced by data interpretation for distributed active sensing network is how to consider information from all the sensors. The information from each sensor may be consistent or conflict with each other, and sometimes, meaningless. The combination of all information becomes crucial in the interpretation of data during the process of SHM. Data fusion can be utilized to implement the task. There have been a wide range of approaches in data fusion, including those algorithms based on voting scheme, Bayesian theory, Dempster–Schafer rules, fuzzy inference, and neural network. The behavior of data fusion operation can be conjunctive, disjunctive,

compromising, or a mix of them. It determines the quality of fusion process, which is often involved with imprecision, uncertainties, and incompleteness. An example has been demonstrated to see the difference of results from a fusion process based Bayesian theory and voting scheme.

Fusion can be considered as a process to establish a posterior belief about a set of propositions, such as structural damage events, on the basis of a set of prior beliefs that are possessed by physical elements such as sensors. Basically, a data fusion process is utilized to increase robustness and reliability of an SHM algorithm by reducing imprecision, uncertainties, and incompleteness as much as possible, through rich information from distributed sensor network.

ACKNOWLEDGMENTS

Z. Su thanks Grant A-PA8G from the Hong Kong Polytechnic University and supports from Prof. Lin Ye of the University of Sydney.

REFERENCES

[1] Hellier C. *Handbook of Nondestructive Evaluation*. McGraw-Hill: New York, 2001.

[2] Talbot D. Boeing's flight for survival. *Technology Review* 2003: 35–44.

[3] Beral B, Speckmann H. Structural health monitoring (SHM) for aircraft structures: a challenge for system developers and aircraft manufacturers. In *Proceedings of the 4th International Workshop on SHM*, Stanford, CA, Chang FK (ed). DEStech Publications: Lancaster, PA, 2003.

[4] Chase SB. Smarter bridges, why and how? *Smart Materials Bulletin* 2001 **2**(10):9–13.

[5] Pines D, Aktan A. Status of SHM of long-span bridges in the United States. *Progress in Structural Engineering and Materials* 2002 **4**(4):372–380.

[6] Ko JM, Ni YQ. Technology developments in SHM of large-scale bridges. *Engineering Structures* 2005 **27**(12):1715–1725.

[7] Derriso MM, Pratt DM, Homan DB, Schroeder JB, Bortner RA. Integrated vehicle health management: the key to future aerospace systems. In *Proceedings of the 4th International Workshop on SHM*, Stanford, CA, Chang FK (ed). DEStech Publications: Lancaster, PA, 2003.

[8] Sohn S, Farrar CR, Hemez FM, Czarnecki JJ, Shunk DD, Stinemates DW, Nadler BR. *A Review of SHM Literature: 1996–2001*. Los Alamos National Laboratory: Los Alamos, NM, 2001.

[9] Giurgiutiu V, Cuc A. Embedded non-destructive evaluation for SHM, damage detection, and failure prevention. *The Shock and Vibration Digest* 2005 **37**(2):83–105.

[10] Su Z, Ye L, Lu Y. Guided Lamb waves for identification of damage in composite structures: a review. *Journal of Sound and Vibration* 2006 **295**(3–5):753–780.

[11] Crawley EF, Luis J. Use of piezoelectric actuators as elements of intelligent structures'. *AIAA Journal* 1987 **25**:1373–1385.

[12] Liang C, Sun FP, Rogers CA. Electro-mechanical impedance modeling of active material systems. *Proceedings of Mathematics and Control in Smart Structures, SPIE 2192*. SPIE, Bellingham, WA, 1994; pp. 232–253.

[13] Wang X, Ehlers C, Neitzel M. Electro-mechanical dynamic analysis of the piezo-electric stack. *Smart Materials and Structures* 1996 **5**(4):492–500.

[14] Wang X, Ehlers C, Neitzel M. Analytical investigation on static models of piezo-electric patches attached on beams and plates. *Smart Materials and Structures* 1997 **6**(2):204–213.

[15] Wang X, Shen YP. On the characterization of piezoelectric actuators attached to structures. *Smart Materials and Structures* 1998 **7**:389–395.

[16] Giurgiutiu V. Embedded self-sensing piezoelectric active sensors for on-line structural identification. *Journal of Sound and Vibration* 2002 **124**(1):116–125.

[17] Ayres JW, Lalande F, Chaudhry Z, Rogers A. Qualitative impedance-based health monitoring of civil infrastructures. *Smart Materials and Structures* 1998 **7**(5):599–605.

[18] Ritdumrongkul S, Abe M, Fujino Y, Miyashita T. Qualitative health monitoring of bolted joints using a piezoelectric actuator-sensor. *Smart Materials and Structures* 2004 **13**(1):20–29.

[19] Chiu WK, Galea S, Koss LL, Najic N. Damage detection in bonded repairs using piezoceramics. *Smart Materials and Structures* 2000 **9**(4):466–475.

[20] Park G, Cudney HH, Inman DJ. Feasibility of using impedance-based damage assessment for pipeline structures. *Earthquake Engineering and Structural Dynamics* 2001 **30**(10):1463–1474.

[21] Tseng KK, Wang L. Smart piezoelectric transducers for *in situ* health monitoring of concrete. *Smart Materials and Structures* 2004 **13**(5):1017–1024.

[22] Rose JL. A vision of ultrasonic guided wave inspection potential. *Proceedings of the Seventh ASME NDE Topical Conference*. San Antonio, TX, 2001; NDE-Vol. 20(1–5).

[23] Worlton DC. Experimental confirmation of Lamb waves at megacycle frequencies. *Journal of Applied Physics* 1961 **32**:967–971.

[24] Rose JL. *Ultrasonic Waves in Solid Media*. Cambridge University Press (UK): New York, 1999.

[25] Sohn H, Farrar CR. Damage diagnosis using time series analysis of vibration signals. *Smart Materials and Structures* 2001 **10**:1–6.

[26] Heller K, Jacobs LJ, Qu J. Characterization of adhesive bond properties using Lamb waves. *NDT and E International* 2000 **33**:555–563.

[27] Koh Y, Chiu WK, Rajic N. Effects of local stiffness changes and delamination on Lamb waves transmission using surface mounted piezoelectric transducers. *Composite Structures* 2002 **57**:437–443.

[28] Youbi FEI, Grondel S, Assaad J. Signal processing for damage detection using two different array transducers. *Ultrasonics* 2004 **42**:803–806.

[29] Niethammer M, Jacobs LJ, Qu J, Jarzynski J. Time-frequency representations of Lamb waves. *The Journal of the Acoustical Society of America* 2001 **109**(5):1841–1847.

[30] Wang Q, Deng X. Damage detection with spatial wavelets. *International Journal of Solids and Structures* 1999 **36**(23):3443–3468.

[31] Kim H, Melhem H. Damage detection of structures by wavelet analysis. *Engineering Structures* 2004 **26**(3):347–362.

[32] Rucka M, Wilde K. Application of continuous wavelet transform in vibration based damage detection method for beams and plates. *Journal of Sound and Vibration* 2006 **297**(3–5):536–550.

[33] Smith C, Akujuobi CM, Hamory P, Kloesel K. An approach to vibration analysis using wavelets in an application of aircraft health monitoring. *Mechanical Systems and Signal Processing* 2007 **21**(3):1255–1272.

[34] Murphy RR. Biological and cognitive foundations of intelligent sensor fusion. *IEEE Transactions of Systems, Man, and Cybernetics—Part A: Systems and Humans* 1996 **26**(1):42–51.

[35] Xiong N, Svensson P. Multi-sensor management for information fusion: issues and approaches. *Information Fusion* 2002 **3**:163–186.

[36] Bloch I. Information combination operators for data fusion: a comparative review with classification. *IEEE Transactions of Systems, Man, and Cybernetics—Part A: Systems and Humans* 1996 **26**(1):52–67.

[37] Byington CS, Garga AK. Data fusion for developing predictive diagnostics for electromechanical systems. In *Handbook of Data Fusion*, Hall D, Llinas J (eds). CRC Press: Boca Raton, FL, 2001.

[38] Bloch I. Some aspects of Dempster–Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters* 1996 **17**:905–919.

[39] Schocken S, Hummel RA. On the use of the Dempster Shafer mode in information indexing and retrieval applications. *International Journal of Man-Machine Studies* 1993 **39**:843–879.

[40] Yen J, Langari R. *Fuzzy Logic: Intelligence, Control, and Information*. Prentice-Hall: Upper Saddle River, NJ, 1999.

[41] Haykin SS. *Neural Network: A Comprehensive Foundation*. Prentice-Hall, 1999.

[42] Su Z, Ye L. A damage identification technique for CF/EP composite laminates using distributed piezoelectric transducers. *Composite Structures* 2002 **57**:465–471.

[43] Su Z, Ye L. Digital damage fingerprints (DDF) and its application in quantitative damage identification. *Composite Structures* 2005 **67**:197–204.

[44] Wang X, Foliente G, Su Z, Ye L. Multilevel decision fusion in a distributed active sensor network for structural damage detection. *Structural Health Monitoring: An International Journal* 2006 **5**:45–58.

Chapter 75

Energy Harvesting and Wireless Energy Transmission for SHM Sensor Nodes

Kevin M. Farinholt, Gyuhae Park and Charles R. Farrar
Engineering Institute, Los Alamos National Laboratory, Los Alamos, NM, USA

1 INTRODUCTION

The management of energy resources is an essential component in the success of any structural health monitoring (SHM) system. With applications that span aerospace, civil, and mechanical engineering infrastructure, it is necessary for sensors and sensor nodes to be both physically robust and energy efficient. In many applications, a sensor network must be installed in locations that are difficult to access, and often these systems have a desired operation life span that exceeds the capabilities of conventional battery technologies. To augment or replace the need to manually recharge or replace batteries within sensors, it is desirable to utilize ambient energy sources that are present within or around the structure being monitored. This process of extracting energy from the environment or a surrounding system and converting it into usable electrical energy is known as *energy harvesting*. Recently, there has been a surge of research in this area, brought about by advances in wireless technology and low-power electronics such as microelectromechanical system (MEMS) devices.

In many SHM applications, there is a considerable amount of ambient energy present within the structure under analysis. This ambient energy is typically in the form of mechanical vibrations induced through environmental or operational conditions, or thermal gradients that develop throughout the day from solar heating or the operation of machinery. The purpose of this article is to provide an up-to-date assessment of available energy harvesting methods suitable for potential SHM sensing applications. This article is not intended to provide an exhaustive literature review, as this area is very broad and useful review articles are already available in

 ISBN: 978-0-470-05822-0.

the literature (*see* **On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System**; **Energy Harvesting using Thermoelectric Materials**). Instead, this article provides a concise introductory survey on the topic and outlines the current status of energy harvesting as applied to relevant themes in SHM. In the second part of the article, a recent alternative to energy harvesting based on vibrations and thermal gradients is presented with more details, using radio frequency (RF) signals to wirelessly deliver electrical energy to operate SHM sensing systems. In this approach, both energy and data interrogation commands are conveyed via a mobile host to each sensor node in order to perform the individual interrogation. Power does not have to reside at the sensor node, relaxing battery, or other such powering requirements. This article also discusses such a prototype system, which is used to interrogate piezoelectric impedance-based sensors on a full-scale bridge.

2 ENERGY HARVESTING

The source of ambient energy can take various forms, including sunlight, thermal gradients within a material, human motion and body heat, vibrations, and ambient RF energy. Several articles reviewing possible energy sources can be found in the literature [1–7]. Fry *et al.* [1] provide a survey of power supplies such as thermoelectric generators (TEGs), mechanical vibration devices using piezoelectric transducers, wind turbines, solar cells, and other exotic portable power sources that utilize ambient electromagnetic radiation, as well as traditional portable supplies such as batteries and fuel cells. While this report is concerned with sources for the US military special operations requirements, it provides insight on the future research trends in energy harvesting.

Roundy [2] compared the energy density of available and portable energy sources. He concludes that, for the device whose desired lifetime is in the range of 1 year or less, battery technology alone is sufficient to provide enough energy. However, if a device requires a longer service life, which is often the case, then the energy harvester can provide a better solution than the battery technologies. Paradiso and Starner [3] point out that the battery technology has evolved very slowly in mobile computing, with only a threefold increase in the battery energy density since 1990. At the same time, the disk storage density has been increased by 1300 times and the CPU speed by nearly 800 times. Park *et al.* [4] summarized several energy harvesting techniques that have been used in SHM applications. The summary also includes the issues associated with power-efficient hardware design and issues with energy harvesting system integration. Glynne-Jones and White [5], Qiwai *et al.* [6], and Mateu and Moll [7] also summarized the basic principles and components of energy harvesting techniques, including piezoelectric, electrostatic, magnetic induction, and thermal energy. A common suggestion listed in these articles is the combined use of several energy harvesting strategies in the same devices so that the harvesting capabilities in many different situations and applications can be increased. Furthermore, energy consumption can be minimized in an effort to close the gap between required and harvested energy [7].

2.1 Electrical energy from mechanical vibrations

One of the most effective methods of implementing an energy harvesting system is to use mechanical vibration to apply strain energy to the piezoelectric material or displace an electromagnetic coil. Energy generation from mechanical vibration usually uses ambient vibration around the energy harvesting device as an energy source, and then converts it into useful electrical energy. The research in this area has made use of mechanical vibration in order to quantify the efficiency and amount of energy capable of being generated and converted, as well as to power various electronic systems. Active materials, usually ceramic- or polymer-based, may be fabricated into a variety of shapes and sizes amenable to various applications. Some examples are shown in Figure 1.

The concept of utilizing piezoelectric material for energy generation has been studied by many researchers over the past few decades, which is well summarized in [8–10]. Piezoelectric materials form transducers that are able to interchange electrical energy and mechanical motion or force. These materials, therefore, can be used as mechanisms to transfer ambient vibration into electrical energy that

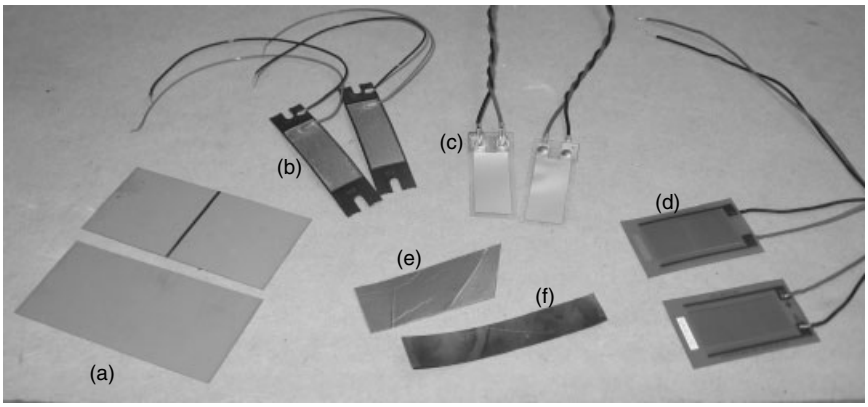


Figure 1. Some typical sensor/actuator materials: (a) PZT 5A4E from Piezo Systems; (b) thunder actuator from FACE International; (c) PVDF transducers from Measurement Specialties; (d) MFC actuators from Smart Material; (e) ionic polymer transducers from Discover Technologies; and (f) ionic polymer–metal composites from Environmental Robots.

may be stored and used to power other devices. A full description of the piezoelectric effect and the methods used to model the behavior of these materials is beyond the scope of this article. However, a significant number of journal articles and conference proceedings develop accurate models and discuss the fundamentals of these materials in great detail [11–13].

Basically, a variety of piezoceramic devices and piezo films have been examined in various settings to try and capture ambient vibration energy [14–16]. Most of them have been laboratory demonstrations and experiments. Some have focused on the electronics in an attempt to optimize the energy transferred from mechanical motion to electrical energy [17, 18]. Others have focused on using various different electrode patterns and packaging to optimize the electromechanical coupling [19]. Studies have also focused on using tuning to try and maximize the amount of energy transferred [20]; however, most of them have dealt directly with the random nature of mechanical energy. In all cases, the goal is to maximize the amount of energy flowing from the mechanical motion of vibration into usable electrical energy. The second feature addressed in the literature is, what should be done with the harvested energy. Some efforts examine storage through capacitors, immediate use, and storage through charging batteries [16, 21]. Other articles examine various applications of harvested power for powering the small electronics including SHM strain-sensing [22, 23] and environmental-sensing systems [24].

2.2 Electrical energy from thermal sources

A second method of obtaining energy from ambient sources is through the use of TEGs that capitalize on thermal gradients. TEGs use the Seebeck effect (Figure 2), which describes the current generated when the junction of two dissimilar metals experiences a temperature difference. Using this principle, numerous p-type and n-type junctions are arranged electrically in series and thermally in parallel to construct the TEG. Thus, when an electrical current is applied to the TEG, a thermal gradient is generated, allowing the device to function as a small solid-state heat pump. Inversely, if a thermal gradient is applied to the device, it will generate an electrical current that can be utilized to power other electronics.

TEGs have been used for capturing ambient energy in various applications. Lawrence and Snyder [25] suggest a potential method of retrieving electric energy from the temperature difference that exists between the soil and the air. To test their concept, a prototype was built without the TEG and the heat flow was measured to estimate the amount of power that could be obtained. The results showed that a maximum instantaneous power of approximately 0.4 mW could be generated by the thermoelectric device. Rowe *et al.* [26] investigate the ability to construct a large TEG capable of supplying 100 W of power from hot waste water. The system tested used numerous thermoelectric devices placed between two

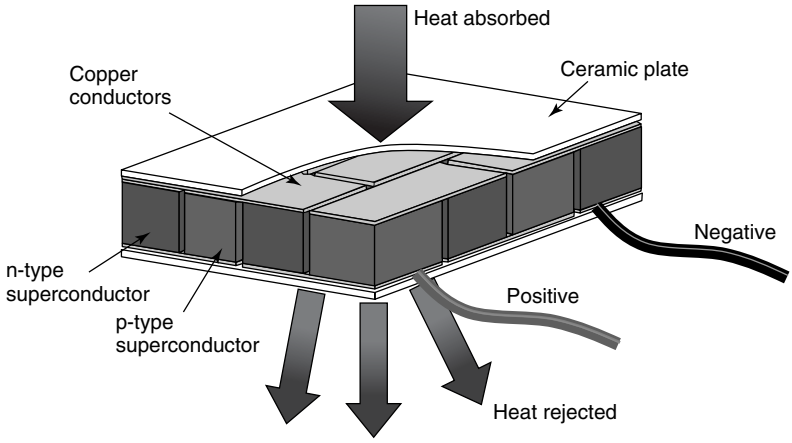


Figure 2. Schematic of the Seebeck effect.

chambers, one with flowing hot water and the other with cold water flowing in the opposite direction and thus maximizing the heat exchange. With a total of 36 modules, each with 31 thermocouples, 95 W of power could be generated.

As described, the idea of using thermoelectric devices to capture ambient energy from a system is not a new concept. However, in many cases, the research efforts utilize liquid heat exchangers or forced convection that significantly improve heat flow and power generation, but require complex cooling loops and systems. Therefore, Sodano *et al.* [27] investigated the use of TEGs as power harvesting devices that do not have an active heat exchanger, but function as a completely passive power scavenging system. Two potential applications are investigated, utilizing solar radiation and harvesting of waste heat. For each application, an experimental prototype was constructed and tested to determine the effectiveness in recharging a discharged nickel–metal hydride battery. The results showed that the TEG does produce significantly more power than a piezoelectric device and that the charge time needed to recharge a battery is significantly lower. This study is presented in more detail in **Energy Harvesting using Thermoelectric Materials**.

The TEG is a mature technology and a reliable energy converter with no moving parts compared to vibration-based harvesters. The TEG has been actively studied for the last three decades and the literature in this area is extensive. One of the drawbacks of this technology is low efficiency (<5%) if there is low temperature gradient present. Further, the fabrication cost is high, and the volume and weight are still too large for microscale sensing systems. Therefore, with the recent advances made in nanotechnologies, the fabrication of MEMS-scale TEG devices has been actively studied [28–30]. For instance, 0.5 cm^2 of a new thermoelectric thin film developed by Applied Digital Solution produces 1.5 μW of power with only a 5 $^{\circ}$C temperature gradient [31].

2.3 Current and future efforts in energy harvesting

Although extensive research work has been focused on energy harvesting, the amount of harvested energy appears to fall significantly behind that required by SHM sensing systems. The power requirement (not counting telemetry) for active-sensing SHM sensor nodes ranges in the order of tens to hundreds of milliwatts. Therefore, the major limitations facing researchers in the field of power harvesting revolve around the fact that the power generated by piezoelectric materials is far too small to power SHM devices. Therefore, methods of increasing the amount of energy generated by the power harvesting device or developing new and innovative methods of accumulating the energy are the key technologies that will allow energy harvesting to become a practical source of power for wireless SHM systems. Innovations in power storage must be developed before power

harvesting technology will see widespread use. The energy harvesting materials have typically been used to determine the extent of power capable of being generated rather than investigating applications and uses of the harvested energy. However, some limited research has focused on field applications of energy harvesters in SHM systems, one example being the intelligent repair application discussed in **On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System**. The practical applications for energy harvesting systems, such as wireless self-powered SHM sensing networks, must be clearly identified with emphasis on power management issues. Application-specific, design-oriented approaches are needed to help the practical use of this technology. Finally, the long-term reliability of energy harvesting devices under the field operating condition should be extensively studied and validated before the full-scale deployment can take place.

3 RADIO FREQUENCY WIRELESS ENERGY TRANSMISSION

Another potential solution for powering SHM sensor nodes and networks is the use of wireless energy transmission. This approach relies on the use of electromagnetic radiation to charge a capacitor or battery that is embedded within or near a sensor network. At short ranges (∼cm) this method is highly efficient owing to inductive or capacitive coupling through magnetic or electric fields [32]. At longer ranges, the efficiency decreases as the electromagnetic radiation is transmitted in the form of visible or near-infrared light, or microwave radiation in the gigahertz range [33, 34]. While the efficiency of short-range coupling mechanisms is ideal, the close proximity that is necessary between source and receiver makes this option difficult to implement in many SHM applications. At longer length scales, the use of visible or near-infrared light provides an effective method for transmitting energy from the source to the target because of the ability to focus light. Unfortunately, the inefficiencies in current photovoltaic cells are detrimental as the conversion capabilities are on the order of 12%. Some experimental triple-junction cells have been developed with efficiencies of 30–40%; however, they require a highly concentrated light source to operate effectively [35].

Another transmission mechanism that is being considered is microwave radiation. In this case, power is generated elsewhere and transmitted to a sensor node by some form of RF radiation. This concept can utilize two different RF energy sources, ambient or controlled RF sources. Previous studies showed that electronics can be used to efficiently capture the ambient radiation sources and convert them to useful electricity. Harrist [36] attempted to charge a cellular phone battery by capturing ambient 915 MHz RF energy. Although he was not able to fully charge the battery, he observed 4 mV s^{-1} charging time from a typical cellular phone battery. Although there are several electronics that may derive their required power from ambient RF sources, the amount of captured energy is extremely low, typically in the range of a few microwatts. Therefore, the technology that has received the most attention is the microwave transmission with a controlled or so-called beamed RF sources, and has been significantly improved in the last several decades. A source antenna transmits microwaves across the atmosphere or space to a receiver, which can either be a typical antenna with rectifying circuitry to convert the microwaves to dc power, or a rectenna (rectifying antenna) that integrates the technology to receive and directly convert the microwaves into dc power.

A pair of excellent survey articles was written to discuss the history of microwave power [37, 38]. With the use of rectennas, efficiencies in the 50–80% range of dc-to-dc conversion have been achieved. Significant testing has also been done across long distances and with kilowatts power levels [39]. The study showed the feasibility of the wireless energy delivery systems for actuating large devices, including dc motors and piezoelectric actuators. Briles *et al.* [40] invented an RF wireless energy delivery system for underground gas or oil recovery pipes. The RF energy is generated on the surface, and travels through the conductive pipe acting as an antenna or a waveguide. The sensor module in the bottom of the pipe captures this energy and uses it to power the electrical equipment. With a 100-W transmitted power from the surface, it was estimated that around 48 mW of instant power can be captured after traveling 1.6 km along the pipe.

The fundamental limitation in this approach is the dispersion of microwaves, since they cannot be focused as readily as light. However, this loss

from dispersion during transmission is made up for in the increased efficiency with which this RF energy can be converted into electrical energy. Relative to the photovoltaic cells, the conversion of RF energy is more efficient than the highest performing photovoltaics. To further improve the efficiency of the microwave transmission, larger arrays of receiving antennas can be assembled to harvest more of the transmitted energy. Current research efforts in RF wireless energy transmission focus on improving the conversion efficiency and attempt to maximize the output power by designing efficient antennas and rectennas. In particular, circular polarized antennas are being implemented in the rectenna design because it avoids the directionality of other antenna designs [41–43]. An array of rectennas is increasingly used to improve the output power [44] and several new rectenna design schemes are proposed [45, 46]. Different elements are also used for efficient rectification [47, 48] in an attempt to obtain optimum output power, and these research trends are similar to those typically pursued in the energy harvesting arena.

Originally considered for alleviating the wiring harness in space structures or microaerial vehicles or providing an extremely low power for those typically used in radio frequency identification (RFID) tags in the 1–100 μW range, the application of an RF wireless energy transmission system for powering electronics typically used in distributed sensing networks has not been studied substantially in the past. In particular, the application of this technology for SHM sensor nodes in order to alleviate the challenges associated with power supply issues has never been addressed in the literature. Therefore, recently, a new and efficient SHM sensing network is proposed, whereby the electric power and interrogation commands are wirelessly provided by a mobile agent [49–51]. This approach involves using an unmanned mobile host node to generate an RF signal near sensors that have been embedded on the structure. The sensors measure the desired response at critical areas on the structure and transmit the signal back to the mobile host again via the RF communications. This "wireless" communication capability draws power from the RF energy transmitted between the host and sensor node and uses it to both power the sensing circuit and transmit the signal back to the host, which is schematically described in Figure 3. This research takes traditional sensing networks to the next level, as the mobile hosts (such as UAV) will fly to known critical infrastructure based upon a GPS locator, deliver required power, and then begin to perform an inspection without human intervention. The mobile hosts will search for the sensors on the structure and gather critical data needed to perform the structural health evaluation. This integrated technology will be directly applicable to rapid structural condition assessment of buildings and bridges after an earthquake. Also, this technology may be adapted and applied to damage detection in a variety of other civilian and defense-related structures such as pipelines, naval vessels, hazardous waste disposal

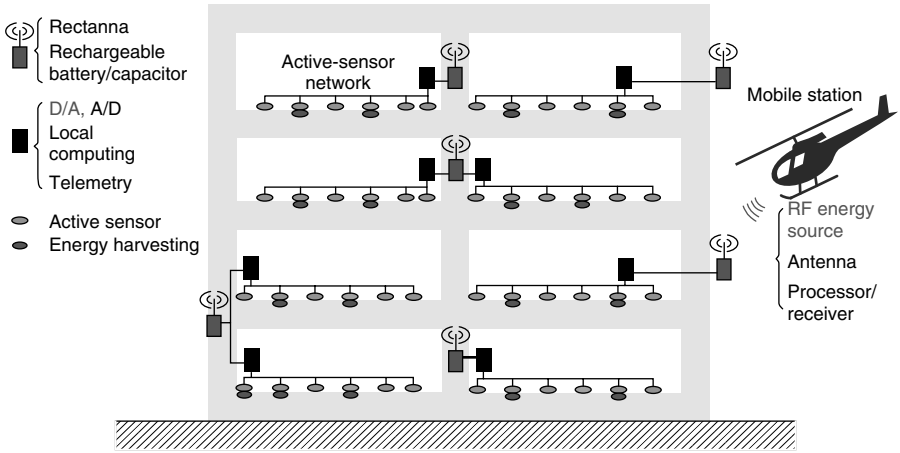


Figure 3. A new sensing network that includes wireless energy transmission and energy harvesting and is interrogated by an unmanned robotic vehicle.

containers, and commercial aircraft. It should be emphasized that this technology can be hybrid in that the sensor node is still equipped with energy harvesting devices and the mobile host would provide additional energy if the energy harvesting device is not able to provide enough power to operate the sensor nodes. Even if the energy harvesting device provides sufficient power, the mobile agent can wirelessly trigger the sensor nodes, collect the information, and/or provide computational resources, which significantly relax the power and computation demand at the sensor node level.

3.1 Patch rectenna design for SHM sensor node

There are two principal components in any wireless energy transmission system: the transmitting antenna and the receiving antenna. The transmitter is driven by an energy source such as a microwave generator and power amplifier, whereas the receiver is coupled with a rectifying circuit that converts the RF energy into usable electrical energy. A supercapacitor can serve as a storage device to collect the received energy and power the associated sensor node.

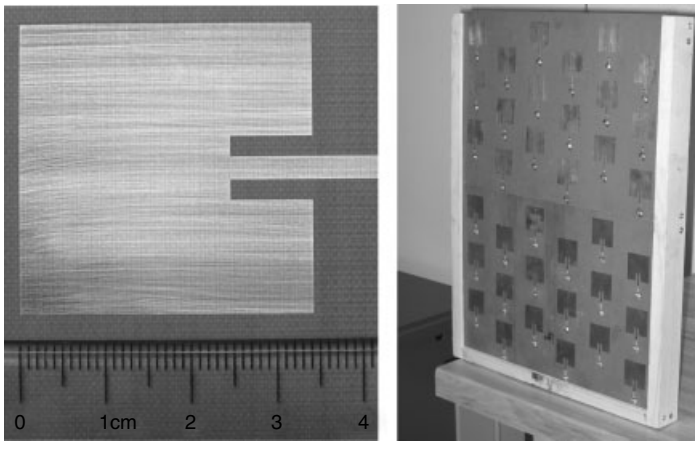
As stated previously, limitations in microwave transmission are associated with the dispersion of the wave as it travels through space. This dispersion can significantly reduce the amount of power received by the rectenna. This loss is seen to be a function of the square of the distance from the transmitting antenna, as seen in the Friis transmission equation:

$$P_{\mathrm{R}} = \frac{G_{\mathrm{T}} G_{\mathrm{R}} \lambda^2}{(4\pi R)^2} P_{\mathrm{T}} \tag{1}$$

where G_{T} and G_{R} are the transmitting and receiving gains, λ is the microwave wavelength in meters, R is the distance between antennas in meters, and P_{T} and P_{R} are the transmitted and received power in milliwatts. For a single microstrip patch antenna, such as the one shown in Figure 4, the power efficiency is calculated to be approximately 1.2% at 1 m spacing [52]. While this level of efficiency is relatively low, it can be increased considerably by assembling a larger array of these microstrip patch antennas as shown in Figure 4. This configuration greatly enhances performance as the array is capable of harvesting more of the incident wave from the source antenna. In this configuration, laboratory tests have shown that the antenna array is capable of charging a 0.1 F capacitor to 3.7 V in 28 s when located at 1.5 m from a source antenna that is emitting 900 mW of microwave power at 2.4 GHz.

3.2 SHM sensor node design for wireless energy transmission

This discussion of energy harvesting and wireless energy transmission is targeted toward one sensor



(a) (b)

Figure 4. (a) A single microstrip patch antenna used to receive microwave energy and (b) a rectenna array of 36 of these patch antennas.

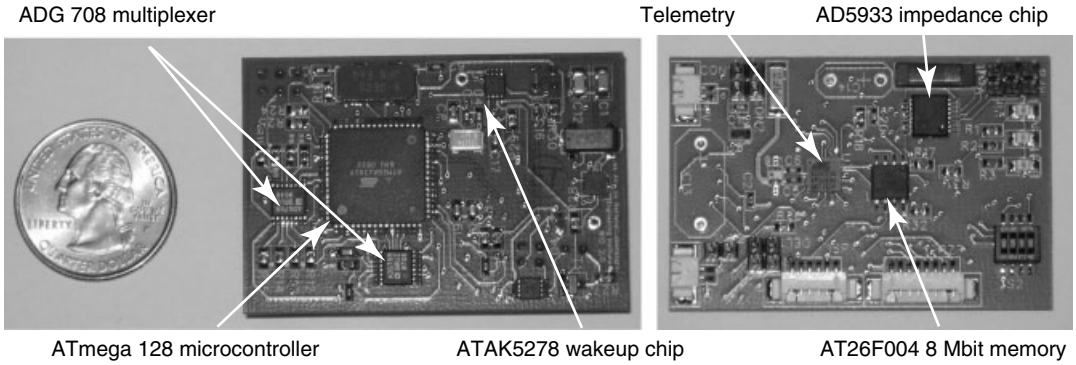


Figure 5. Wireless impedance device (WID2) developed by Los Alamos National Laboratory to measure, process, and transmit electrical impedance data for up to seven piezoelectric sensors.

node that has been developed by researchers at Los Alamos National Laboratory [53]. This system is designed as an active-sensor node that can interrogate the electrical impedance of seven piezoelectric patches, which utilizes impedance-based SHM [54], and is referred to as the *wireless impedance device* (WID2). The principal components of the design are shown in Figure 5, and include an ATmega1281V microcontroller, an AD5933 impedance chip with ADG708 multiplexers, an ATAK5278 wake-up chip, and an AT86RF230 telemetry module.

This sensor node is designed to provide onboard computing and data storage for a variety of SHM schemes. The standard operating condition for the WID2 is the idle mode in which the sensor node operates below 0.65 mA, consuming less than 1.82 mW of power. Proper configuration of the sleep modes should allow us to reduce current draw to 0.01 mA, providing a sensor node that is capable of extended operation at extremely low-power levels. The WID2 can be brought out of this idle mode at specific time intervals through an integrated real-time clock that is capable of waking the WID2 on intervals of 1 s to 1 year. Additionally, the sensor node can be activated through a low-frequency wakeup chip that monitors an integrated inductor for a magnetically coupled wakeup signal. Once the sensor node is activated, the microcontroller measures electrical impedance of each attached piezoelectric sensor, evaluates the status of the sensor, and either stores or transmits the result to a mobile host, which subsequently transmits the data back to the base station. Current and power consumption of the WID2 are outlined in Table 1 for each operational state.

Table 1. Current and power consumption for WID2 in different operational states

State	Current (mA)	Power (mW)
Measure	20	56
Transmit	22	61.6
Idle	0.65	1.82

The measure and transmit operations take a combined time of 10 s to complete. When considering the power consumption, the power requirement of WID2 is within the range of the capabilities of several energy harvesting or wireless energy transmission methods. Under laboratory testing, it was found that a 0.1 F supercapacitor provides sufficient energy for more than 10 s of operation, when it is charged to 3.7 V, which combines energy harvesting and transmission techniques to provide decades of sustainable operation.

3.3 Field testing

In August 2007, this method was field tested on the Alamosa Canyon Bridge near Truth or Consequences, New Mexico. Experiments were designed to test the effectiveness of this wireless energy transmission system when mounted on a mobile host. The source antenna, was mounted on the side of an automobile and the height was adjusted to match the height of the receiving antenna array. The microwave source and power amplifier were powered by a power inverter

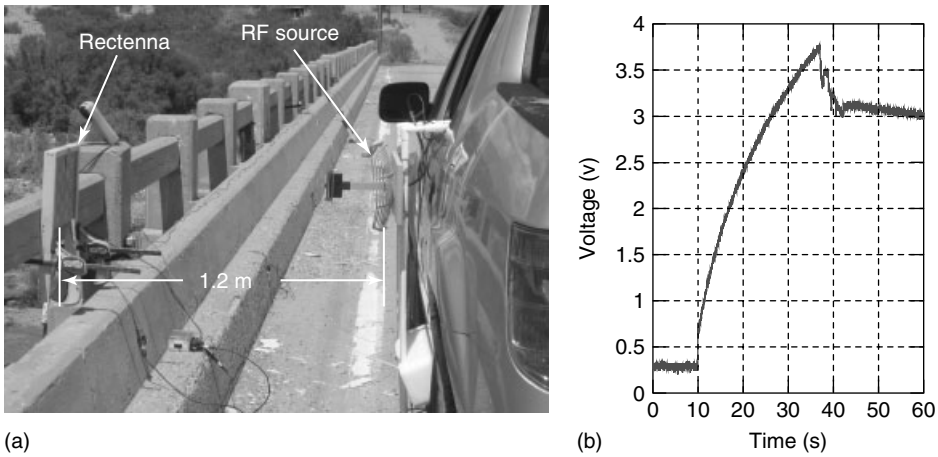


Figure 6. (a) Wireless energy delivery system field tested on Alamosa Canyon Bridge, NM. (b) The RF source was configured to emit 1 W of energy at 2.4 GHz, charging a 0.1-F capacitor to 3.7 V in 27 s.

that was plugged into the automobile's 12 V dc power supply. The receiving antenna was composed of 18 microstrip patch antennas clamped to the concrete side rail of the bridge as seen in Figure 6. This rectenna was used to charge a 0.1-F (P/N: PB-5R0V104) supercapacitor, which was connected to a WID2.0 sensor node and used to monitor the preload in two bolts on the bridge through piezoelectric sensors.

For this test, the automobile was stopped in parallel with the rectenna, as shown in Figure 6. The source antenna was activated and the voltage on the 0.1-F capacitor was visually monitored to prevent the capacitor from charging above 4 V (the WID2.0 operates on a nominal voltage between 2.7 and 3.3 V). During these experiments, the spacing between the transmitting and receiving antennas was maintained at 1.2 m, and the resulting charge time was measured to be 27 s to obtain a voltage of 3.7 V within the supercapacitor. Once charged, the sensor node was used to take measurements from two PZT active sensors, perform the local computing to determine bolt tightness, and then transmit 8 bytes of data back to the mobile host.

4 CONCLUSIONS

Energy harvesting is slowly coming into full view of the SHM and the more general sensing network communities. With continual advances in wireless sensor/actuator technology, improved signal-processing technique, and the continued development of power-efficient electronics, energy harvesting will continue to attract the attention of researchers and field engineers. However, it should be emphasized that energy harvesting still remains in its infancy and only a few successful examples are in practice. Still, a tremendous research effort is required to convert, optimize, and accumulate the necessary amount of energy to power such electronics.

One potential alternative that has shown successful performance in remotely powering sensor nodes is the use of wireless energy transmission. This technique builds upon an already existing technology that has been redirected to focus on low-power SHM systems. Laboratory and field tests have demonstrated that this method is highly effective in powering nodes that are designed to perform discrete measurements of a structure's health. Further development is needed to determine the optimal configuration of the receiving array and to couple this technique with charging electrochemical batteries to provide testing capabilities between charging cycles.

REFERENCES

[1] Fry D, Holcomb D, Munro J, Oakes L, Maston M. *Compact Portable Electric Power Sources*, ORNL/TM-13360. Oak Ridge National Laboratory Report, 1997.

[2] Roundy SJ. *Energy Scavenging for Wireless Sensor Nodes with a Focus on Vibration to Electricity Conversion*, Ph.D. Dissertation. Department of Mechanical Engineering, University of California: Berkeley, CA, 2003.

[3] Paradiso JA, Starner T. Energy scavenging for mobile and wireless electronics. *IEEE Pervasive Computing* 2005 **4**:18–27.

[4] Park G, Farrar CR, Todd MD, Hodgkiss W, Rosing T. *Energy Harvesting for Structural Health Monitoring Sensor Networks*, LA-14314-MS. Los Alamos National Laboratory Report, 2007.

[5] Glynne-Jones P, White N. Self-powered systems: a review of energy sources. *Sensor Review* 2001 **21**:91–97.

[6] Qiwai M, Thomas J, Kellogg J, Baucom J. Energy harvesting concepts for small electric unmanned systems. *Proceedings of the SPIE* 2004 **5387**:84–95.

[7] Mateu L, Moll F. Review of energy harvesting techniques and applications for microelectronics. *Proceedings of the SPIE* 2005 **5837**:359–373.

[8] Sodano H, Inman D, Park G. A review of power harvesting from vibration using piezoelectric materials. *The Shock and Vibration Digest* 2004 **35**:451–463.

[9] DuToit NE, Wardle BL, Kim SG. Design considerations for MEMS-scale piezoelectric mechanical vibration energy harvesters. *Integrated Ferroelectrics* 2005 **71**:121–160.

[10] Anton S, Sodano H. A review of power harvesting using piezoelectric materials (2003–2006). *Smart Materials and Structures* 2007 **16**:R1–R21.

[11] Niezrecki C, Brei D, Balakrishnam S, Moskalik A. Piezoelectric actuation technology: state of the art. *The Shock and Vibration Digest* 2001 **33**:269–280.

[12] Inman DJ, Cudney HH. *Structural and Machine Design Using Piezoceramic Materials: A Guide for Structural Design Engineers*, Final Report NASA Langley Grant NAG-1-1998, 2000.

[13] Sirohi J, Chopra I. Fundamental understanding of piezoelectric strain sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**:246–257.

[14] Goldfarb M, Jones LD. On the efficiency of electric power generation with piezoelectric ceramic. *ASME Journal of Dynamic Systems, Measurement, and Control* 1999 **121**:566–571.

[15] Clark W, Ramsay MJ. Smart material transducers as power sources for MEMS devices. *International Symposium on Smart Structures and Microsystems*. Hong Kong, 2000.

[16] Lesieutre GA, Hofmann HF, Ottman GK. Electric power generation from piezoelectric materials. *The 13th International Conference on Adaptive Structures and Technologies*. Potsdam, Berlin, 7–9 October 2002.

[17] Kasyap A, Lim J, Johnson D, Horowitz S, Nishida T, Ngo K, Sheplak M, Cattafesta L. Energy reclamation from a vibrating piezoceramic composite beam. *Proceedings of 9th International Congress on Sound and Vibration*, Paper No. 271. Orlando, FL, 2002.

[18] Han J, von Jouanne A, Le T, Mayaram K, Fiez TS. Novel power conditioning circuits for piezoelectric micro power generators. *IEEE Applied Power Electronics Conference and Exposition*, Anaheim, CA, USA, 2004; Vol. 3, pp. 1541–1546.

[19] Sodano HA, Inman DJ, Park G. Comparison of piezoelectric energy harvesting devices for recharging batteries. *Journal of Intelligent Material Systems and Structures* 2005 **16**:799–807.

[20] Cornwell PJ, Goethal J, Kowko J, Damianakis M. Enhancing power harvesting using a tuned auxiliary structure. *Journal of Intelligent Material Systems and Structures* 2005 **16**:825–834.

[21] Sodano HA, Inman DJ, Park G. Generation and storage of electricity from power harvesting devices. *Journal of Intelligent Material Systems and Structures* 2005 **16**:67–75.

[22] Elvin NG, Elvin AA, Spector M. A self-powered mechanical strain energy sensor. *Smart Materials and Structures* 2001 **10**:293–299.

[23] Inman DJ, Grisso BL. Towards autonomous sensing. *Proceedings of the SPIE* 2006 **6174**:T1740–T1749.

[24] Microstrain. http://www.microstrain.com, 2008.

[25] Lawrence EE, Snyder GJ. A study of heat sink performance in air and soil for use in a thermoelectric energy harvesting device. *Proceedings of the 21st International Conference on Thermoelectronics*. Portland, OR, 2002; pp. 446–449.

[26] Rowe MD, Min G, Williams SG, Aoune A, Matsuura K, Kuznetsov VL, Fu LW. Thermoelectric recovery of waste heat—case studies. *Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference*. Honolulu, HI, July 27–August 1 1997; pp. 1075–1079.

[27] Sodano HA, Dereux R, Simmers GE, Inman DJ. Power harvesting using thermal gradients for recharging batteries. *Proceedings of the 15th International Conference on Adaptive Structures and Technologies*. Bar Harbor, ME, 25–27 October 2004.

[28] Bottner H. Thermoelectric micro devices: current state, recent developments and future aspects for technological progress and applications. *Proceedings of 21st International Conference on Thermoelectric*, La Grande Motte, France, 2003; pp. 511–518.

[29] Snyder GJ, Lim JR, Huang CK, Fleurial JP. Thermoelectric microdevice fabricated by a MEMS-like electrochemical process. *Nature Materials* 2003 **2**:528–531.

[30] Jovanovic V, Ghamaty S. Design, fabrication and testing of energy-harvesting thermoelectric generators. *Proceedings of the SPIE* 2006 **6173**:G–1–G–8.

[31] http://www.asdx.com, 2006.

[32] Jang J, Liu JF, Yue CP, Sohn H. Development of self-contained sensor skin for highway bridge monitoring. *Smart Structures and Materials, Proceedings of the SPIE*, San Diego, CA, USA, 2006; Vol. 6174, pp. 1291–1300.

[33] Blackwell T. Recent demonstrations of laser power beaming at DFRC and MSFC. *BEAMED ENERGY PROPULTION: Third International Symposium on Beamed Energy Propultion. AIP Conference Proceedings*. American Institute of Physics, 2005; Vol. 766, pp. 73–85.

[34] Choi H, Song K, Golembiewskii W, Chu S, King G. Microwave power of smart material actuators. *Smart Materials and Structures* 2004 **13**:38–48.

[35] Green M, Zhao J, Wang A, Wenham S. Progress and outlook for high-efficiency crystalline silicone solar cells. *Solar Energy Materials and Solar Cells* 2001 **65**:9–16.

[36] Harrist DW. *Wireless Battery Charging System Using Radio Frequency Energy Harvesting*, M.S. thesis. Department of Electrical Engineering, University of Pittsburgh: Pittsburgh, PA, 2004.

[37] Brown WC. The history of wireless power transmission. *Solar Energy* 1996 **56**:3–21.

[38] Maryniak GE. Status of international experimentation in wireless power transmission. *Solar Energy* 1996 **56**:87–91.

[39] Choi S, Song K, Golembiewskii W, Chu SH, King G. Microwave powers for smart material actuators. *Smart Materials and Structures* 2004 **13**:38–48.

[40] Briles SD, Neagley DL, Coates DM, Freund SM. *Remote Down-Hole Well Telemetry*, US Patent # 6,766,141, 2004.

[41] Strassner B, Chang K. 5.8 GHz circularly polarized dual-rhombic-loop traveling-wave rectifying antenna for low power-density wireless power transmission applications. *IEEE Transactions on Microwave Theory and Techniques* 2003 **51**:1548–1553.

[42] Ali M, Yang G, Dougal R. A new circularly polarized rectenna for wireless power transmission and data communication. *IEEE Antennas Wireless Propagation Letters* 2005 **4**:205–208.

[43] Ren YJ, Chang K. 5.8 GHz circularly polarized dual-diode rectenna and rectenna array for microwave power transmission. *IEEE Transactions on Microwave Theory and Technique* 2006 **54**:1495–1502.

[44] Kim J, Yang SY, Song DD, Jones S, Choi SH. Performance characterization of flexible dipole rectennas for smart actuator use. *Smart Materials and Structures* 2006 **15**:809–815.

[45] Park JY, Han SM, Itoh T. A rectenna design with harmonic-rejecting circular-sector antenna. *IEEE Antennas and Wireless Propagation Letters* 2004 **3**:52–54.

[46] Chin CH, Xue Q, Chan CH. Design of a 5.8 GHz rectenna incorporating a new patch antenna. *IEEE Antennas and Wireless Propagation Letters* 2005 **4**:175–178.

[47] Epp LW, Khan AR, Smith HK, Smith RP. A compact dual-polarized 8.51 GHz rectenna for high-voltage actuator applications. *IEEE Transactions on Microwave Theory and Technique* 2000 **48**:111–119.

[48] Zbitou J, Latrach M, Toutain S. Hybrid rectenna and monolithic integrated zero-bias microwave rectifier. *IEEE Transactions on Microwave Theory and Technique* 2006 **54**:147–152.

[49] Farrar CR, Park G, Allen DW, Todd MD. Sensor network paradigms for structural health monitoring. *Structural Control and Health Monitoring* 2006 **13**(1):210–225.

[50] Todd MD, *et al.* A different approach to sensor networking for SHM: remote powering and interrogation with unmanned aerial vehicles. *Proceedings of 6th International Workshop on Structural Health Monitoring*. Stanford, CA, 11–13 September 2007.

[51] Park G, Overly TG, Nathnagel M, Farrar CR, Mascarenas DL, Todd MD. A wireless active-sensor node for impedance-based structural health monitoring. *Proceedings of US-Korea Smart Structures Technology for Steel Structures*. Seoul, 16–18 November 2006.

[52] Nothnagel M, Park G, Farrar CR. Wireless energy transmission for structural health monitoring embedded sensor nodes. *Proceedings of the SPIE*, San Diego, CA, USA, 2007; Vol. 6532, pp. 653216.

[53] Overly T, Park G, Farrar CR. Development of impedance-based wireless active-sensor node for structural health monitoring. *The 6th International Workshop on Structural Health Monitoring*, Stanford, CA, USA, 2007.

[54] Park G, Sohn H, Farrar CR, Inman DJ. Overview of piezoelectric impedance-based health monitoring and path forward. *The Shock and Vibration Digest* 2003 **35**(6):451–463.

Chapter 72

Nondestructive Evaluation of Cooperative Structures (NDECS)

Daniel L. Balageas

Structure and Damage Mechanics Department, ONERA (The French Aerospace Lab), Châtillon, France

1 INTRODUCTION

The recent evolution of nondestructive evaluation (NDE) techniques used for maintenance of structures is characterized by rapid and dramatic changes. These changes also concern the general philosophy of structure maintenance and monitoring.

Three major evolutions can be highlighted: (i) in the pure NDE field, the ever growing importance of full-field, real-time, noncontact imaging techniques (*see* **Full-field Sensing: Three-dimensional Computer Vision and Digital Image Correlation for Noncontacting Shape and Deformation Measurements**); (ii) the birth and the impressive development of structural health monitoring (SHM), which could be superficially considered as an avatar of NDE, if only seen as a fully integrated NDE; and (iii) the importance of Lamb waves to elaborate SHM systems.

An analysis of these evolutions leads us to consider a possible third way, intermediate between NDE and SHM, called *nondestructive evaluation of cooperative structures* (NDECS). To illustrate this concept, three possible NDECS are presented. The first two techniques result from the combination of localized and embedded Lamb wave generation and noncontact optical detection; the last one combines laser ultrasonics and embedded piezoelectric detection.

2 RECENT EVOLUTION IN NDE, SHM, AND MAINTENANCE PHILOSOPHY

Presently, the maintenance of aerospace structures is based on the use of NDE techniques. One

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

of the characteristic evolutions of NDE and of experimental mechanics techniques is the importance gained by the development of full-field, real-time, noncontact imaging techniques. Such techniques, thanks to charge-coupled device (CCD) cameras, are performing a parallel acquisition of information coming from a very large number of locations. This is achieved by cameras working in various spectral domains: ultraviolet, visible, near, or far infrared, associated with coherent light illumination (interferometric techniques such as electronic speckle pattern interferometry (ESPI) [1]) or incoherent sources (stimulated thermography [2], visible image correlation techniques [3], etc.). These techniques can image fields of displacements, deformation, temperature, etc. of structures submitted to various types of solicitations: loads [4], fatigue [5], vibrations [6, 7], radiant heat fluxes [8], eddy currents [9], electromagnetic fields [10], etc. A high-frequency periodical excitation often used in NDE consists in propagating ultrasounds in the structure. Nevertheless, it has been only very recently coupled to full-field imaging techniques, and more particularly with shearography and thermography.

SHM, which belongs to the domain of smart materials and structures, can be considered as a new form of NDE, characterized by the full integration of sensors, actuators, and intelligence inside the structure during its manufacturing process. It is the reason why the recent growth of SHM community is partially explained by the progressive aggregation of members of the NDE community, while SHM, at its beginning, was mainly done by the mechanical engineering community.

SHM techniques were introduced in the 1990s and show an impressive development [4]. If we look at the SHM literature, we see an increasing importance of Lamb wave-based work and, for this purpose, a prevalent use of piezoelectric patches (*see* **Ultrasonic Methods**; **Piezoelectric Wafer Active Sensors**; **Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection**). This is illustrated by a statistical analysis of the recent publications given in [4, 11].

These three evolutions (growing importance of full-field real-time imaging techniques, impressive development of SHM, and importance gained by the use of Lamb waves to elaborate SHM systems) are mainly driven by a cost-reduction strategy of maintenance operations (*short-term objective*). The second evolution — development of SHM — aims at the replacement of scheduled maintenance inspections by performance-based or condition-based inspections (*long-term objective*).

3 BETWEEN NDE AND SHM COULD WE IMAGINE AN INTERMEDIATE WAY?

If we consider the progress achieved in both NDE and SHM fields, a third approach seems possible now. This third way could consist in only embedding at well-chosen locations into the structure the stimulation function (actuators or emitters) and leaving the full-field noncontact detection (sensors or receivers) outside, or in adopting an opposite configuration with a full-field noncontact stimulation outside and a localized detection inside. To optimize such systems, Lamb waves could be used in conjunction with full-field noncontact optical techniques. So, using the more recent and promising techniques of both SHM and NDE domains, the proposed approach could be applied in a *short term*, with existing technologies, and would permit time and money saving for maintenance operations. A well-suited appellation for this type of technique could be NDECS.

Such an idea is not totally new. A similar approach has been made by Walsh [12] with a more limited field of application. Walsh proposed to replace conventional ultrasonic testing using surface contact probes by a semiembedded system in which the emitter is inside the structure and the detector is outside, which allows an improved resolution and a larger depth of penetration. Walsh called this concept as *nondestructive evaluation ready material* (*NDERM*) technology.

The NDECS concept proposed here is more efficient and more ambitious. Figure 1 presents a schematic view of what could be an NDECS system. Three possible systems are described using (i) lock-in thermography to monitor the thermal effects resulting from the interaction of the waves with possible damages like delaminations, cracks, or corrosions; (ii) stroboscopic shearography to monitor the full field of the surface displacements created by the Lamb waves; (iii) ultrasounds generated by a

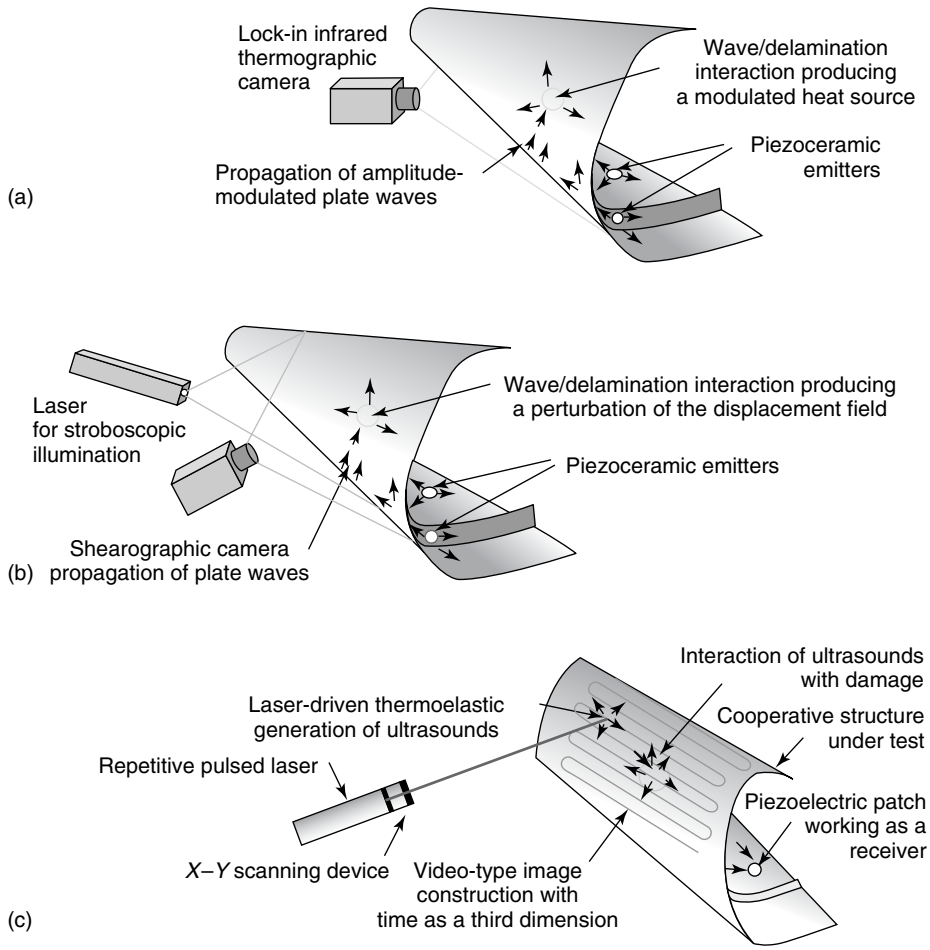


Figure 1. Three possible NDECS: (a) lock-in ultrasonic vibrothermography with Lamb waves generated by embedded piezoelectric emitters [Reproduced with permission from Ref. 11. © 2007.], (b) shearographic visualization of Lamb wave interaction with damage [11], and (c) piezoelectric detection of laser-generated ultrasounds. [Reproduced with permission from Ref. 13. © Takatsubo, 2006.]

pulse repetitive laser and detected by an embedded piezoelectric receiver. These three imaging techniques are chosen because they seem more promising and they are almost ready to be used, since their feasibility has been proved in laboratory. The two first systems are based on work done at Office National d'Etudes et de Recherches Aérospatiales (ONERA) and were presented as NDECS systems in [11]. The third system is studied by a group at National Institute of Advanced Industrial Science and Technology (AIST) (Japan) [13]. The following sections present the principle and some laboratory results already obtained with these techniques.

4 IMAGING INTERACTIONS OF LAMB WAVES WITH DAMAGES USING LOCK-IN THERMOGRAPHY

4.1 Principle of lock-in ultrasonic vibrothermography

Ultrasonic vibrothermography is an NDE technique based on the application of a modulated mechanical stress on the tested structure, while an IR camera

maps the surface temperature [14]. Thermomechanical coupling (*see* **Thermal Imaging Methods**) is responsible for heat production, which, in turn, is partly responsible for damping. In particular, damaged regions convert energy into heat through enhanced viscoelastic dissipation, collisions, and/or rubbing of internal free surfaces present in delaminations and cracks. Surface defects and internal defects appear hotter when the surface temperature is mapped, a fact that provides the basis for the use of joint mechanical excitation and thermography as a nondestructive technique.

Most of the experimenters have used high-power ultrasound emitters producing simultaneously several uncontrolled waves: mechanical shakers [15], a 500-W ultrasound cleaner [16], ultrasound generators such as sonotrodes (generally designed for plastic welding) with power up to 2 kW [17, 18], etc. The use of such powerful devices may cast some doubt on the nondestructive aspect of the control. The first danger is thus, in the case of polymer composite testing, of overheating the material in the contact area. The second danger is that small tolerable cracks may grow unexpectedly fast when the defect surfaces are submitted to energetic rubbing and/or "clapping".

Ultrasonic vibrothermography can also be performed through video lock-in processing [15] (*see* **Thermal Imaging Methods**) if the high-frequency mechanical excitation is amplitude modulated at a low frequency. In this case, the heat generated by the *high-frequency vibrations* is modulated at this *low frequency* and the thermographic lock-in system is synchronized with this thermal wave frequency. At ONERA it has been proposed to evaluate whether lock-in ultrasonic vibrothermography could be applied by means of small, low-energy piezoelectric transducers embedded or surface bonded to the tested structure, and generating Lamb waves (Figure 2). These transducers are currently used in acousto-ultrasonic SHM systems. This has been successfully demonstrated in the case of C/epoxy plates [14]. This solution presents three main advantages: (i) low-power actuators (less than 1 W) can be used, inducing a low dissipation at the defect location, nevertheless detected, thanks to the lock-in procedure; (ii) embedding or bonding the piezoelectric patches guarantee a coupling constant in time and ascertainable by measuring the electromechanical impedance [19, 20]; and (iii) the use

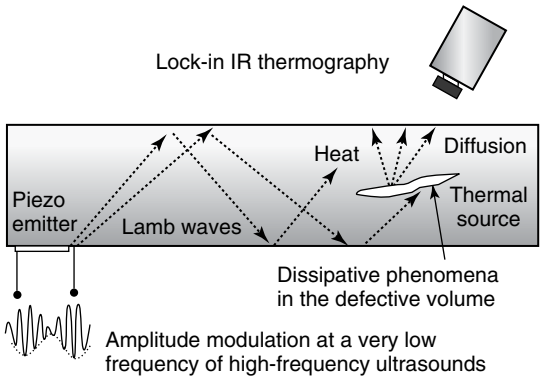


Figure 2. Principle of lock-in ultrasonic vibrothermography applied to the visualization of the interaction of Lamb waves with a damage.

of Lamb wave permits long-distance propagation with moderate attenuation, thus giving an almost uniform sensitivity to the defect detection in the full structure.

The choice of the frequency and the nature of the wave to excite the defects are very important. This has been demonstrated by several experimenters [21–23].

4.2 Low-energy detection of delaminations in composite panels

Following the principle presented in Figure 2, results have been obtained using Lamb waves and very low energies. Temperature mapping was performed using an IR focal plane array camera (Amber 4128, 128 × 128 pixels), with a noise equivalent temperature difference (NETD) of 7 mK. The lock-in technique was applied to demodulate the thermal signal. The ultrasound actuator was fed with a high-frequency electric modulation, between a few tens and a few hundreds of kilohertz, and the amplitude was modulated with a sinus function or a square function at a low frequency (33–300 mHz). By processing 500 images, a temperature modulation amplitude image with an NETD lower than about 0.5–0.6 mK was obtained, making the detection of defects such as delaminations, debonding, microcracking, and macrocracks possible.

The results presented here were obtained with a $[0_4/45_4/90_4/-45_4]_S$ C/epoxy plate of dimensions $70 \times 70\,cm^2$ with a delamination produced by

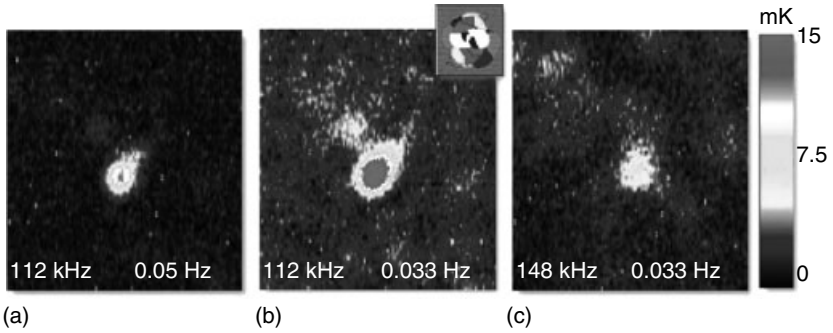


Figure 3. Temperature amplitude map on the front side of the impacted area of a C/epoxy coupon. Influence of the low-frequency modulation for a Lamb wave frequency of 112 kHz: (a) low amplitude modulation frequency of 50 mHz, (b) same as in (a) but with 33 mHz. Influence of the wave frequency (same Lamb mode) with amplitude modulation at 33 mHz: (b) 112 kHz, (c) 148 kHz.

a 5-J impact. The electric power injected in the piezoelectric emitter (disc-shaped, 30 mm in diameter and 200 micron thick) was 1 W. The amplitude of the Lamb waves is modulated at a low frequency (30–300 mHz) to cope with the thermal wave attenuation. In the present case, the delaminations can be detected from front face (impacted face) provided the modulation frequency is lower than 100 mHz. Figure 3(a–b) presents the thermal amplitude images obtained for two modulation frequencies: 50 and 33 mHz. Decreasing the modulation frequency has a beneficial effect on the sensitivity, but results in more blurring. These results highlight the well-known depth probing property of "thermal waves".

With the modulation frequency fixed at 33 mHz to get a good signal-to-noise ratio, several wave frequencies between 50 and 150 kHz were used. The temperature amplitude images obtained after synchronous demodulation are presented in Figure 3(b–c) for 112 and 148 kHz. The highest contrast, i.e. 22 mK, is observed for a frequency of 112 kHz. When the actuator is excited at 148 kHz, the defect is still detectable, but the contrast dropped to about 8 mK. Decreasing the wave frequency has the same detrimental effect. This shows how important the choice of the ultrasound frequency for a safe detection by thermography. For this optimum frequency, the apparent size of the defect is comparable to the image obtained by classical D-scan (see upper right part of Figure 3b).

These experiments show that defects can be detected whatever their depth be, provided that dissipation is high enough and that the modulation frequency is chosen sufficiently low (thick plates need higher mechanical energy and deep defects need lower lock-in modulation frequency to be revealed). The high selectivity of the ultrasound frequency regarding the thermomechanical coupling at the defect location has, however, to be taken into account. This requires some theoretical work to predict, for a given structure and a given defect geometry, which kind of ultrasound wave would lead to maximum dissipation at the defect locus.

4.3 Practical applications

The experimental results just presented concerned delaminations in composite structures. In fact, the field of application of the technique is much wider and ultrasonic vibrothermography has already been applied to real structures presenting realistic damages. These works have been essentially performed with sonotrodes or high-energy lead zirconate titanate (PZT) patches. The types of structures and damages detected until now in these conditions are varied. Let us mention, taken from [17, 22, 23] (i) delaminations in skin and stringer and stringer debondings in composite aircraft panels; (ii) cracks and debonds in metallic structures, hidden corrosions between skin and riveted stringers in aluminum structures; and (iii) heterogeneities in a multilayer C/C–SiC ceramic composite. We can suppose that by choosing tailored Lamb waves, similar results could be obtained with low or moderate energies.

5 IMAGING INTERACTIONS OF LAMB WAVES WITH DAMAGES USING STROBOSCOPIC SHEAROGRAPHY

5.1 Principle of the shearographic imaging of ultrasounds

Shearography is a speckle interferometric technique that appeared in the 1970s [24–26], in which the interfering photons are issued from two closed points of the structure thanks to an optical "shearing" device. This prevents the system from being sensitive to vibrations and solid motions of the structure. The measurement is performed for two successive states of deformation of the structure. By electronic data reduction, a subtraction of the two speckle images produces a resulting image containing information linked to the gradient of deformation between the two states in the direction of the shear. To make the technique quantitative, a phase stepping is required and, from four intensity images corresponding to four phase lags introduced thanks to a controllable mobile mirror, a phase image is obtained, which can be directly graded in gradient of deformation. This differential image is different from the images of deformation given by more classical interferometric techniques like ESPI [1].

To produce two different states of deformation of the structure, it is possible to use depressurization [27], photothermal stimulation [28], or vibrations. To define two states in the case of vibrations, it is necessary to produce a stroboscopic illumination with an acousto–optic modulator in the laser beam [29] as seen in Figure 4(a).

Gordon and Bard [30] were the first to propose and apply this technique to the visualization of ultrasounds. As shown in Figure 4(b), the two states of deformation of the structure correspond to two laser illuminations with a phase difference of 180° with respect to the ultrasounds. The phase stepping

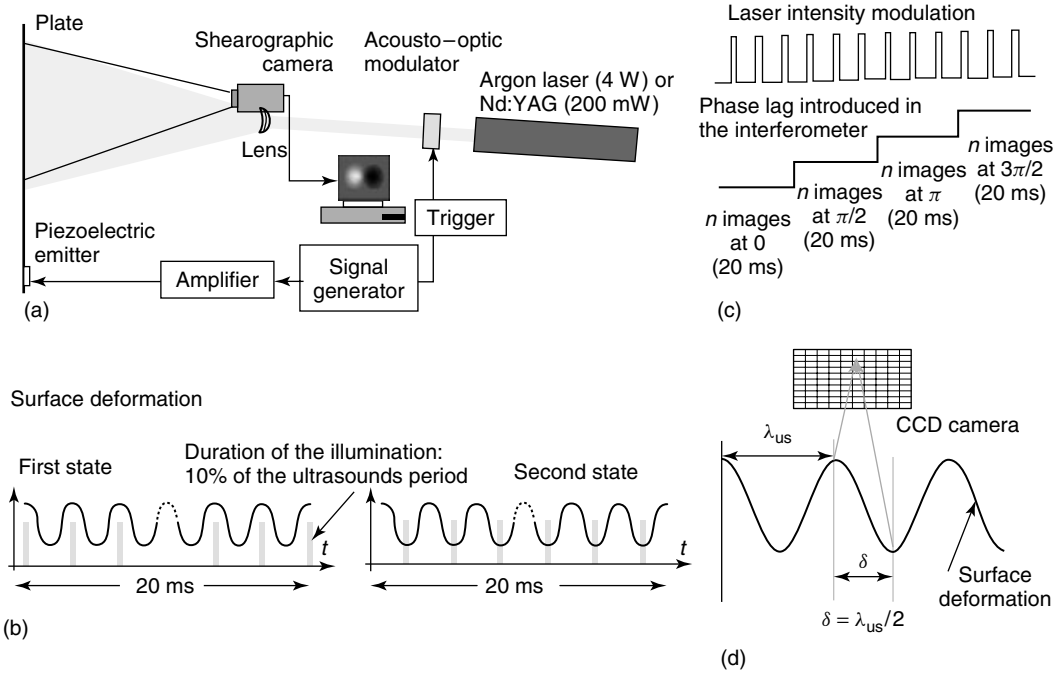


Figure 4. Visualization of Lamb waves by shearography: (a) experimental setup (the shearographic camera contains the three elements of the Michelson interferometer: shearing, phase stepping, and shear control devices). (b) synchronization between the signal feeding the emitter and the laser illumination, producing the stroboscopic effect; (c) synchronization between the laser illumination and the phase lag introduced in the arm of the interferometer (valid for the two states); and (d) shear adjustment (shear distance equal to half the wavelength).

technique is used (Figure 4c) and, to increase the signal-to-noise ratio, for every phase shift, image accumulation is performed for 20 ms. Finally, the better sensitivity is obtained with a shearing distance equal to half the ultrasound wavelength (Figure 4d), producing a difference of displacement between the two interfering points that is maximum. Under these conditions, the shearographic image level is strictly equal to four times the ultrasound amplitude, with a sensitivity of the order of 1 nm. If the shear is equal to the wavelength, then the wave is not seen, the deformation of the interfering points being identical, which can be interesting to visualize the waves diffracted by a damage [31, 32].

The use of continuous Lamb waves leads to results, which can be hard to interpret whether real structural parts are likely to induce reflections, mode conversions etc. Such elements can be fasteners, stringers, rivets, etc. or simply the plate edges. In particular, these effects are important in metallic structures for which the attenuation of Lamb waves is very low. To avoid this inconvenience, it is more convenient to generate short bursts and to follow their propagation [32]. Some modifications have to be introduced relative to the wave imaging mode: only one laser stroke is used per burst and its firing is delayed according to the propagation stage one wants to image. The second surface state is simply obtained by reversing the transducer input voltage.

In its simplest way of application, the shearographic camera views the structure perpendicularly to its surface, and then only the out-of-plane displacement is detected (present results). Nevertheless, it is possible to visualize the in-plane displacement, and this can be useful for Lamb wave imaging [33, 34].

5.2 Illustrative examples of shearographic imaging of Lamb waves

Figure 5 presents the visualization of the interaction of Lamb waves with a delamination in a C/epoxy plate [31, 32]. The coupon, equipped with a PZT emitter, is the one already used for the vibrothermographic experiment. The detection and localization are unambiguous and feasible on both front and rear faces of the plate. The interaction between the incident wave and the delamination is a diffraction phenomenon. It is possible to detect, although not so easily, this diffraction (emergent wave) around the damaged area (in particular, in Figure 5c).

Figure 6 presents an interaction in a more complex configuration: an artificial defect in a composite sandwich structure of a radome [35]. The defect is a lack of material, representative of a debond between the outer skin (glass/epoxy composite) and the core made of low-density foam. There is an interaction only in the outer skin, which allows us to delimit the extent of damage. No interaction is produced in the inner skin, which is thicker, because its dispersion curves are almost identical to that of the full sandwich.

The demonstration of the possibility of imaging the propagation of Lamb wave bursts has been achieved too [32]. Figure 7 presents the images of

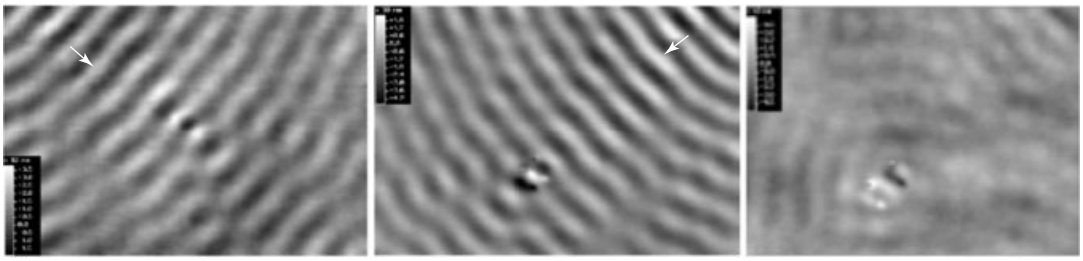


Figure 5. Interaction of Lamb waves (S_0 mode, 68 kHz) propagating in the C/epoxy plate described in the section on vibrothermography. The dimensions of images are $17 \times 12\,cm^2$ and their dynamic range is near to 50 nm. The electric power injected in piezoelectric emitter is 1 W. In figure (a) the shearographic image of the front face obtained with a shear distance equal to half the wavelength shows both incoming and diffracted waves. This is also the case for the image of the rear face given in figure (b). In figure (c) the image of the rear face obtained with a shear distance equal to the Lamb wavelength only shows the diffracted waves. [Reproduced with permission from Ref. 32. © EDP Sciences, 2006.]

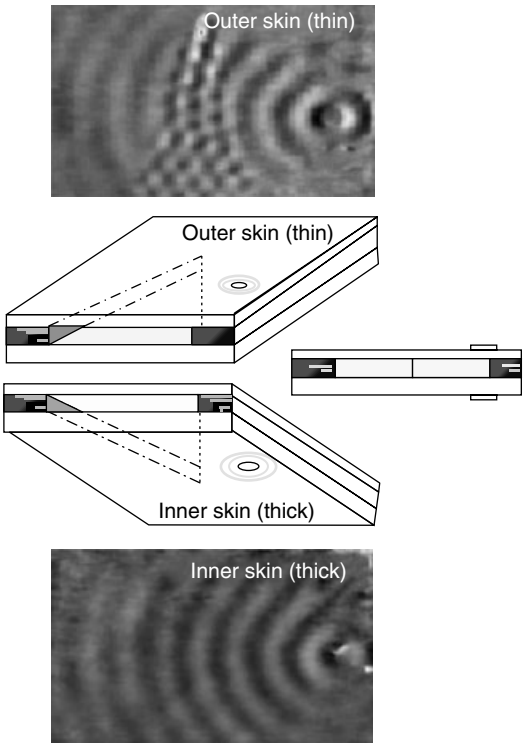


Figure 6. Shearographic visualization of the interaction of Lamb waves (A_0 mode, $f = 20\,\text{kHz}$) with a defect in the core (lack of foam) in a composite sandwich of a radome.

the displacements recorded on the impacted sample previously monitored (Figure 5) for three different delays after the burst emission. In the first image, the Lamb wave front is still upstream from the impact defect and in the second image the burst has already passed the defect. Again a 30° wide wake with a 180° phase delay appears downstream. In the last image, a circular diffracted wave is seen around the defect. Five-period bursts have been used, with windowing, and this unavoidably induced higher frequency content for the mechanical excitation. Modes other than S_0 with different velocities and wavelengths could thus be present. Together with the fact that the bursts are repetitively emitted, this could explain why the defect already appears before the burst reaches it. Indeed small amplitude ripples can be seen in front of and behind the burst.

6 IMAGING INTERACTIONS OF LAMB WAVES WITH DAMAGES USING SCANNED LASER ULTRASONICS AND EMBEDDED PZT RECEIVERS

6.1 Principle

The concept of the third technique (Figure 1c) proposed by Takatsubo and colleagues from AIST (Japan) [13], is based on the fact that when ultrasonic waves propagate between two points using one emitter and one receiver having the same frequency characteristics, the same waveform would be detected if the emitter and the receiver replaced each other. The technique uses a pulsed laser generating ultrasounds by thermoelastic effect and a piezoelectric receiver. The waves detected by the piezoelectric sensor at a point B and generated by the laser at a point A would be almost the same as the waves detected at point A issued from the laser generation at point B. This visualization is based on the reversibility of the propagation of ultrasonic waves.

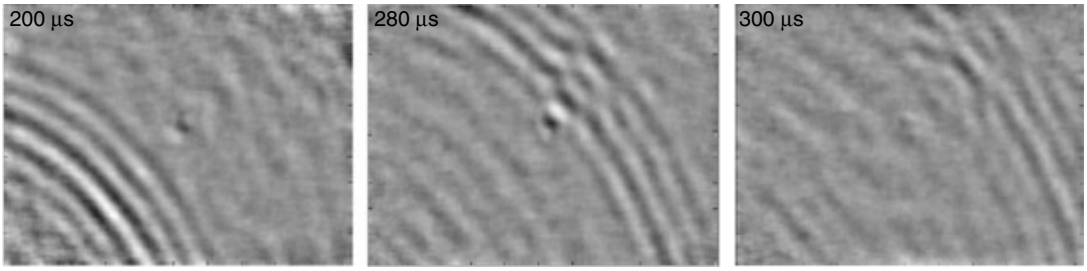


Figure 7. Shearograms of burst propagation in a delaminated C/epoxy plate (5-J impact) obtained for three delays after the burst. Transducer driven with repetitive five-period bursts (68 kHz). [Reproduced with permission from Ref. 32. © EDP Sciences, 2006.]

Let us consider now that the laser impact is scanned on the structure surface. Then, the train of waveforms detected by the fixed-position piezoelectric sensor during the laser scanning will be the same as those obtained by detecting the waves generated by directing the laser beam toward the position of the piezoelectric sensor while scanning the sensor on the structure surface. Therefore, images may be created by the waveforms detected during scanning of the laser. Displaying these images consecutively in the order of the measurement times produces an animation showing the propagation pattern that would be generated by a laser generation of ultrasounds at the location of the piezoelectric receiver. These authors have experimentally demonstrated that this principle is still valid if a defect is present in the propagation path [13].

The technique has several remarkable advantages: (i) it allows visualization of ultrasonic waves propagating in a complex-shaped 3D structure with curved surfaces, steps, and dents; (ii) there is no need to adjust the laser incidence angle and the focal distance; (iii) high detection sensitivity is obtained; and (iv) noncontact wide-field imaging of complex structures is made possible.

6.2 Practical applications

The technique has been applied to the detection of various types of structures and defects: holes and flaws in curved metallic pipes, slits in metallic plates, and debonding in a carbon fiber reinforced plastic (CFRP) structure. Figure 8 presents an illustration of the possibilities of the technique for detecting debonding between the skin and a hat stringer in a CFRP structure (airplane wing). The pulsed laser was a yttrium aluminum garnet (YAG) (1064 nm) delivering 5-mJ, 10-ns pulses with a repetition rate of 20 Hz. The laser is fixed on two rotation axis stages controlled by a PC. The image results from the impact of the 2-mm-diameter laser beam onto

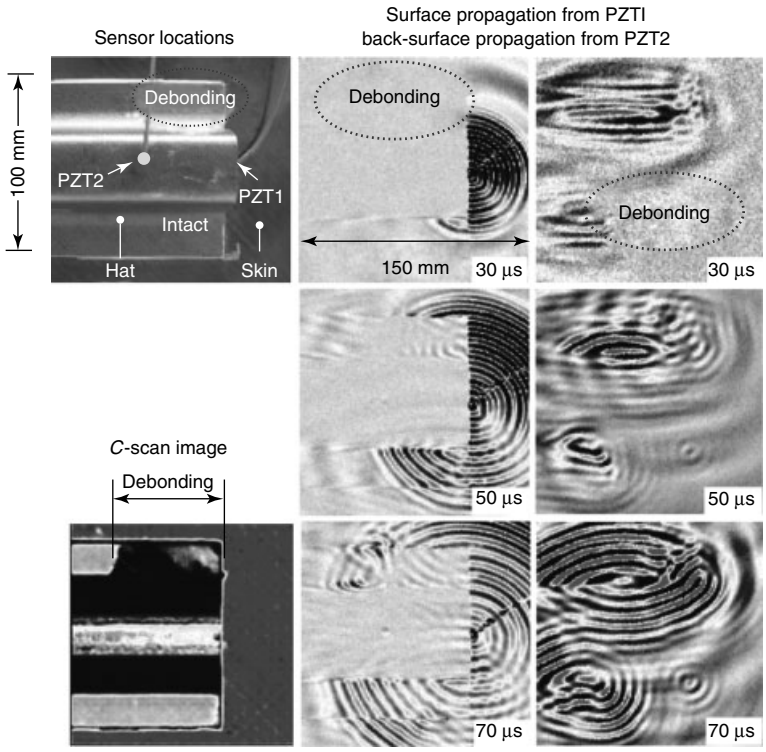


Figure 8. Detection of a debonding between skin and hat stringer of a CFRP structure. [Reproduced with permission from Ref. 13. © Takatsubo, 2006.]

100×100 points on the structure. For each impact, the PZT receiver signal is recorded. Thanks to the reversibility, the amplitudes of the recorded trains of waveforms at a given time correspond to the displacements of the ultrasounds at the irradiation points. Images can be created whose intensities are determined by these amplitudes. When displayed consecutively, the propagation at the surface of the structure of the ultrasonic waves can be observed, allowing the detection of damaged areas. Using Fourier analysis, it is possible to construct monofrequency images. The visualized signal in Figure 8 is a single-frequency (300 kHz) ultrasound signal, so there is very little velocity dispersion in the propagation image since such a dispersion depends on frequency. The debonding forms a thin air layer, and ultrasonic waves hardly propagate across it. The debonded part and the intact part can thus be clearly distinguished.

7 CONCLUSIONS

The concept of NDECS has been defined and proposed as a third possible way between the classical NDE and SHM. From the analysis of the present evolution of techniques in NDE and SHM fields, it has been deduced that the best NDECS solutions could be the combined use of Lamb waves and optical noncontact detection or emission systems. Three NDECS configurations have been described on the basis of the combined use of embedded PZT transducers with the following techniques: stroboscopic shearography, lock-in infrared thermography, and scanned laser ultrasonics. To demonstrate that these solutions are applicable in the short term, promising experimental results have been presented.

REFERENCES

[1] Rastogi PK. *Holographic Interferometry: Principles and Methods*, *Springer Series in Optical Sciences*. Springer-Verlag: Berlin, 1994; Vol. 68.

[2] Maldague X. *Theory and Practice of Infrared Technology for Nondestructive Testing*, *Wiley Series in Microwave and Optical Engineering*. John Wiley & Sons, 2001.

[3] Schmidt T, Tyson J. Full-field dynamic displacement and strain measurement using advanced 3D image correlation photogrammetry. *Experimental Techniques* 2003, Part I: **27**(3):47–50, Part II: **27**(4):44–47.

[4] Balageas D. Introduction to structural health monitoring. In *Structural Health Monitoring*, Balageas D, Fritzen CP, Güemes A (eds). ISTE: London, 2006, pp. 13–43.

[5] Arnould O, Hild F, Brémond P. Thermal evaluation of the mean fatigue limit of a complex structure. *Thermosense XXVII, SPIE Proceedings 5782*. SPIE: Bellingham, WA, 2005; pp. 255–263.

[6] Potet P, Bathias C, Degrigny B. Quantitative characterization of impact damage in composite materials: a comparison of computerized vibrothermography and X-ray tomography. *Materials Evaluation* 1988 **46**(8):1050–1054.

[7] Tenek LH, Henneke EGII, Gunzburger MD. Flaw dynamics and vibro-thermography thermoelastic NDE of advanced composite materials. *Thermosense XIII, SPIE Proc. 1467*. SPIE: Bellingham, WA, 1991; pp. 252–263.

[8] Balageas DL, Déom AA, Boscher DM. Characterization and nondestructive testing of carbon-epoxy composites by a pulsed photothermal method. *Materials Evaluation* 1987 **45**(4):461–465.

[9] Riegert G, Zweschper Th, Busse G. Lock-in thermography with eddy-current excitation. *QIRT Journal* 2004 **1**(1):21–31.

[10] Balageas DL, Levesque P, Nacitas M, Krapez JC, Gardette G, Lemistre M. Microwaves holography revealed by photothermal films and lock-in IR thermography: application to electromagnetic materials NDE. *Proceedings of SPIE 2944*. SPIE: Bellingham, WA, 1996; pp. 55–66.

[11] Balageas D. Non-destructive evaluation of cooperative structures (NDECS): a third way? *First Asia-Pacific Workshop on SHM*. Yokohama, December 2007.

[12] Walsh SM. Practical issues in the development and deployment of intelligent systems and structures. In *Proceedings of the 2nd International Workshop on SHM: SHM 2000*, Chang F-K (ed). Technomic Publishing: Lancaster-Basel, 1999, pp. 612–621.

[13] Takatsubo J, Yashiro S, Wang B, Tsuda H, Toyama N. Laser ultrasonics imaging technique for nondestructive inspection of defects in three-dimensional objects. *First Asia-Pacific Workshop on SHM*. Yokohama, December 2007 *see also, in Japanese* 高坪純治、王波、津田浩、遠山暢之, 発振レーザ走査法による三次元任意形状物体を伝わる超音波の可視化、日本機械学会論文集 2006 72-718A～: 945–950.

[14] Krapez JC, Taillade F, Balageas D. Ultrasound-lock-in thermography NDE of composite plates with low power actuators. Experimental investigation of the influence of the Lambwave frequency. *QIRT Journal* 2005 **2**(2):191–206.

[15] Rantala J, Wu D, Busse G. Amplitude modulated lockin vibrothermography for NDE of polymers and composites. *Research in Nondestructive Evaluation* 1996 **7**:215–218.

[16] Rantala J, Wu D, Busse G. NDT of polymer materials using lock-in thermography with water-coupled ultrasonic excitation. *NDT&E International* 1998 **31**(1):43–49.

[17] Busse G, Dillenz A, Zweschper T. Defect-selective imaging of aerospace structures with elastic-wave-activated thermography. *Thermosense XXIII, Proceedings of SPIE 4360*. SPIE: Bellingham, WA, 2001; pp. 580–586.

[18] Favro LD, Han X, Zhong O, Sun G, Sui H, Thomas RL. Infrared imaging of defects heated by a sonic pulse. *Review of Scientific Instruments* 2000 **71**(6):2418–2421.

[19] Giurgiutiu V, Zagrai A, Bao JJ. Piezoelectric wafer embedded active sensor for aging aircraft structural health monitoring. *Structural Health Monitoring—An International Journal* 2002 **1**(1):41–62.

[20] Pacou D, Pernice M, Dupont M, Osmont D. Study of the interaction between bonded piezo-electric devices and plates. In *Proceedings of First European Workshop on SHM: Structural Health Monitoring 2002*, Balageas DL. DEStech Publishing: Lancaster, PA, 2002, pp. 406–413.

[21] Han X. Frequency dependence of the thermosonic effect. *Review of Scientific Instruments* 2003 **74**(7):414–416.

[22] Dillenz A, Zweschper T, Riegert G, Busse G. Progress in phase angle thermography. *Review of Scientific Instruments* 2003 **74**(7):417–419.

[23] Zweschper T, Riegert G, Dillenz A, Busse G. Frequency modulated elastic wave thermography. *Thermosense XXV, Proceedings of SPIE 5073*. SPIE: Bellingham, WA, 2003; pp. 386–391.

[24] Leendertz J, Butters J. An image shearing speckle pattern interferometer for measuring bending moments. *Journal of Physics E: Scientific Instruments* 1973 **6**:1107–1110.

[25] Hung YY. A speckle-shearing interferometer: a tool for measuring derivatives of surface displacements. *Optics Communications* 1974 **11**(2):132–135.

[26] Hung YY. Shearography: a novel and practical approach for non-destructive inspection. *Journal of Nondestructive Evaluation* 1989 **8**(2):55–67.

[27] Clarady JF, Summers M. Electronic holography and shearography NDE for inspection of modern materials and structures. *Review of Progress in QNDE* 1993 **12A**:381–386.

[28] Paoletti D, Schirripa Spagnolo G, Zanetta P, Facchini M, Albrecht D. Manipulation of speckle fringes for non destructive testing of defects in composites. *Optics and Laser Technology* 1994 **26**(2):99–104.

[29] Hariharan P, Oreb B. Stroboscopic holographic interferometry: application of digital. *Optics Communications* 1986 **59**(2):83–86.

[30] Bard BA, Gordon GA, Wu S. Laser modulated phase-stepping digital shearography for quantitative full-field imaging of ultrasonic waves. *Journal of the Acoustical Society of America* 1998 **103**:3327.

[31] Krapez JC, Taillade F, Lamarque T, Balageas D. Shearography: a tool for imaging Lamb waves in composites and their interaction with delaminations. *Review of Progress in QNDE* 1999 **18A**:905–912.

[32] Taillade F, Krapez JC, Lepoutre F, Balageas D. Shearographic visualization of lamb waves in carbon epoxy plates: interaction with delaminations. *The European Physical Journal—Applied Physics* 2000 **AP9**:69–73.

[33] Rastogi PK. Measurement of in-plane strains using electronic speckle and electronic speckle-shearing pattern interferometry. *Journal of Modern Optics* 1996 **43**(8):1577–1581.

[34] Moulin E, Assaad J, Delebarre C, Kaczmarek H, Balageas D. Study of a piezoelectric transducer embedded in composite plate: application to Lamb waves generation. *Journal of Applied Physics* 1997 **82**(5):2049–2055.

[35] Devillers D, Taillade F, Osmont D, Krapez JC, Lemistre M, Lepoutre F. Shearographic imaging of the interaction of ultrasonic waves and defects in plates. *Proceedings of SPIE 3993*. SPIE: Bellingham, WA, 2000; pp. 142–149.

Chapter 76

On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System

Stephen C. Galea, Stephen Van der Velden, Scott Moss and Ian Powlesland

Air Vehicles Division, Defence Science and Technology Organisation (DSTO), Melbourne, VIC, Australia

1 INTRODUCTION

The application of bonded composite patches or doublers to repair or reinforce defective (secondary) metallic structures is becoming recognized as an effective and versatile repair procedure for many types of problems [1]. However, the application of bonded composite repairs to cracked aircraft primary structure is generally acceptable only on the basis that a margin on design limit-load (DLL) capability is retained in the event of loss (total absence) of the repair [2]. However, assuming that the static requirements and quality-assurance processes are satisfied, one approach to certify full credit for the patch in slowing crack growth could be justified by a *continuous safety-by-inspection approach.* This approach is based on the continuous self-assessment of the patch system integrity using a *smart patch* approach [2, 3], by incorporating *in situ* sensors to continuously monitor the structural condition of the patch system and associated remaining damage in the parent structure. The need to follow approved patch design, fabrication, and quality-assurance procedures is unchanged; this approach simply allows a relaxation of the probability of failure requirements, particularly in relation to environmental degradation. However, the viability of any *smart patch* or *in situ* structural health monitoring (SHM) approach now depends on establishing its reliability or probability of damage detection, which is similar to

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 Commonwealth of Australia. ISBN: 978-0-470-05822-0.

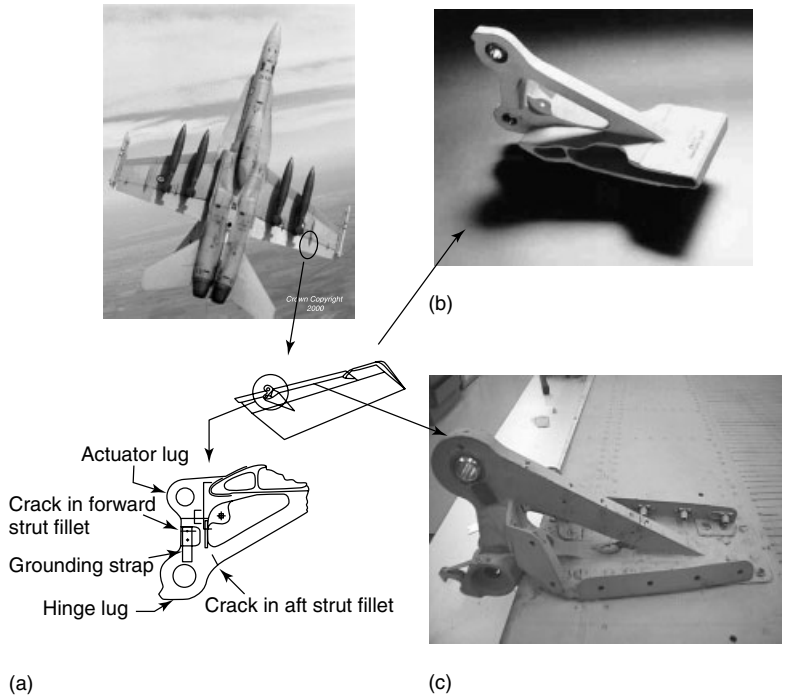


Figure 1. (a) Schematic of the aileron, hinge lug, and hinge aft strut with likely crack locations (taken from [4]). Photograph of F/A-18 (b) aluminum, and (c) titanium aileron hinges.

the problem of probability of detection in Non-Destructive Inspection (NDI), and should include system self-checking and redundancy to provide the required level of confidence.

To demonstrate and evaluate the feasibility of the *smart patch* concept, the Australian Defence Science and Technology Organisation (DSTO) has developed two SHM systems to interrogate the structural health of a boron/epoxy (b/ep) doubler (or reinforcement) on an F/A-18 inboard aileron hinge, as shown in Figure 1(a). One system, the battery-powered *smart patch* (BPSP), uses a lithium ion-based battery as the power source and measures load transfer from the structure to the patch, using conventional strain gauges, to monitor patch health. The patch health data is uploaded by the operator via an infrared (IR) link. The other concept, the self-powered (wireless) *smart patch* (SPSP), involves more technical risk and consists of a piezoelement-based self-powered sensing system powered by an array of piezotransducers, which convert structural dynamic strain to electrical energy, and monitors damage in the patch via piezoelectric film strain sensors. In this system, the patch health data is uploaded by the operator using a magnetic transceiver. The instrumented aileron was installed on a Royal Australian Air Force (RAAF), Aircraft Operation Support Group (AOSG) F/A-18 in early 2006 and was flown for 12 months.

This article describes some of the critical issues associated with the development, evaluation and implementation of the self-powered *smart patch* system, including system design, functionality and certification testing, and installation. The article gives a brief overview of the various available power-harvesting techniques and discusses several key issues when designing and developing the self-powered *smart patch* system, viz., (i) demand—power requirements, (ii) supply—energy generation from vibration or strain-based sources, and (iii) conversion—issues and efficiencies associated with energy conversion from mechanical to electrical energies. Flight data from the SHM system and lessons learned during the program are also presented.

2 BACKGROUND

2.1 General system specifications

The initial idea was to evaluate the *smart patch* concept on a bonded composite patch/reinforcement designed for an F/A-18 aluminum aileron hinge with a propensity to cracking [5] (Figure 1a). During the initial system design phase, the RAAF decided to refurbish the ailerons and, in so doing, replaced the aluminum hinges (Figure 1b) with redesigned titanium hinges (Figure 1c), thus eliminating the cracking problem. However, in consultation with senior RAAF engineering managers, it was decided to proceed with the *smart patch* demonstrator on a titanium aileron hinge.

The first phase of the project was to establish the operating envelope of the system, including temperature, strain, and vibration at the aileron hinge and, flying time and elapsed time between system installation and removal from the aircraft. All these issues are critical when designing and implementing such a system. These general specifications are listed below [6]:

- The expected operational temperature range was -40 to $+70\,^{\circ}$C.
- No strain time history data was available for the titanium aileron hinge; however, flight loads and design data for the aluminum aileron hinge indicated that the operational peak strains were of the order of $1500\,\mu\varepsilon$ at an excitation frequency of between 8 and 42 Hz.
- Flight times were expected to vary between 40 and 60 min per sortie.
- The system was expected to be operational for an elapsed time of one year, which would entail between 100 and 200 sorties.

2.2 Concept development and high-level system design

In general, the *smart patch* concept needs to continually inspect regions of the reinforcement likely to suffer disbond damage [2]. There are several methods of achieving this goal and providing an indication of structural health. One method involves subjecting the structure to a known excitation and then measuring a response at the "interrogation" time (*see* **Development of an Active Smart Patch for Aircraft Repair**; **Design, Analysis, and SHM of Bonded Composite Repair and Substructure**). An alternative would be to use some form of "in-service" excitation (*see* **Design, Analysis, and SHM of Bonded Composite Repair and Substructure**). In this case, in-flight loading is used as the excitation mechanism and in-flight strain response as the damage indicator. A finite element (FE) analysis of the hinge is undertaken to determine the most likely regions of damage and to assist in developing the most appropriate damage indicator (see Section 3.1) [6].

An autonomous remotely accessed patch health monitoring (*smart patch*) system was the final objective. As discussed previously, the full *smart patch* demonstrator consisted of two, developed concurrent, wireless SHM systems, namely, the BPSP and the SPSP. However, even though the BPSP system is a lower risk solution it does suffer from the limitation of a finite power supply and would require occasional maintenance to replace the battery. The SPSP offers significant advantages over the BPSP since it is maintenance-free and therefore facilitates rapid acceptance by operators, maintainers, and certification authorities and would incur minimum additional through-life-support costs. This latter approach would enable minimum interruption to relatively rigid and well-established logistic and maintenance procedures associated with aging aircraft. In addition to monitoring the health of bonded patches, this approach would facilitate SHM of structural "hot spots" in "difficult-to-access" locations. Therefore, to encourage acceptance by operators and maintainers, a fully autonomous system design was chosen with no batteries. This self-powered system had the following major hardware components; (i) power-harvesting elements, (ii) electronics to manage power harvesting, interrogate the sensors, calculate/store patch health, and allow data download, (iii) sensing elements, (iv) system wireless interrogation, and (v) handheld data gathering unit. Issues associated with the selection and design of these components are discussed in Section 3.

2.2.1 Energy harvesting

DSTO is exploring structural power-harvesting techniques for use on defence platforms [7]. The

main driver for this research is the development of powering systems for autonomous distributed sensor networks (DSN) applied to airframes for SHM. Using "smart" sensor concepts, damage and damage growth in the airframe, and other structural life-related problems, would be continuously monitored on board the aircraft to provide real-time damage assessment. This technology could potentially permit a safe reduction in inspection and regular maintenance costs with substantial impact on the through-life costs. Using a DSN for SHM has only recently become feasible because of advances in micropower electronics and microelectromechanical system (MEMS); however, there are still a number of issues to overcome.

One of the main hurdles to the use of a DSN is powering. It is unsatisfactory to use the conventional approach of running power through copper wires on a central power/data bus because of

1. certification issues;
2. the added mass of wiring;
3. the drain on the limited electrical power available on an aircraft;
4. the wiring occupying significantly more space on the aircraft than the sensors themselves;
5. the fact that the DSN is no longer unobtrusive on the airframe and would interfere with maintenance procedures;
6. installation of the DSN itself would be complicated, time consuming, and expensive;
7. the significant reliability and durability issues with excessive wiring.

An alternative to the central power/databus is to have a ubiquitous network of autonomous sensors with independent power supplies. One option is to use batteries as the primary power source; however, replacing batteries in a DSN would be a considerable maintenance overhead and therefore, alternative approaches for powering such systems are required. Because of the low-power nature of MEMS systems, the concept of self-powering a DSN via harvesting power from the local environment becomes an intriguing possibility. On an airframe, in particular, there can be significant accelerations and dynamic strains available that lend themselves to the power-harvesting concept. Power harvesting has the potential to

- reduce the logistical burden and costs incurred by replacing a number of battery-powered systems;
- increase system availability due to the longevity of a device that is "self-powered" (compared with a battery-powered device); and
- power large numbers of tiny sensors in a network that may be difficult (or impossible) to power any other way.

Power harvesting itself is the process of energy scavenging from the local environment. Depending on the operational environment, various energy sources may be available [8] (*see* **Energy Harvesting and Wireless Energy Transmission for SHM Sensor Nodes**), for example: vibration, kinetic, solar, thermal (*see* **Energy Harvesting using Thermoelectric Materials**), chemical, and biological. On an aircraft, some or all of the first four sources could be available, depending on the location of the sensors. Within DSTO, the focus has been on the use of piezoelectric elements (both ceramic and polymer) for power-harvesting applications [9] using vibrational energy.

There is significant worldwide activity in the area of power harvesting, and a number of reviews of the literature are available [10]. Eggborn [11] discusses optimization of power harvesting from the piezoelectric/cantilever combination. Defense Advanced Research Projects Agency (DARPA) supported a "micro power generation" program that funded a number of energy harvesting investigations, for example, the "piezoelectric eel" project [12] and the "thermoelectric generator" [13]. Perhaps the best-known example of power harvesting is the Massachusetts Institute of Technology (MIT) piezoelectric shoe [14]. In 1996, Starner [15] discussed the possibility of using the everyday actions of the human body as a source of potentially harvestable energy. Starner calculated that one of the best prospects for power harvesting from the human body is walking—an average 68-kg person, walking with an average human gait of two steps per second, generates 67 W of "heel strike" power. Paradiso, at the MIT Media Lab [16], investigated the possibility of parasitic "shoe scavenging" and the result was a piezoelectric shoe-based power-harvesting system [17]. Both PVDF (polyvinylidene fluoride) piezoelectric films and PZT (lead zirconate titanate) piezoelectric ceramics and the associated harvesting electronics were placed in a running shoe. As the shoe wearer walked, the piezoelectric

deformed, generating piezoelectric voltage that was then (through the power-harvesting electronics) used to charge a storage capacitor. The storage capacitor was then used to drive a radio frequency (RF) transmitter.

There are a number of other MEMS power harvesting examples available in the literature, for example, El-hami *et al.* [18] describe a micromachined magnet/coil power harvesting device. The so called smart dust being developed at the University of California, Berkeley is a significant MEMS-based project, which involves the development of a "self-powered" device [19]. More recently, Beeby unveiled a MEMS-based energy harvester with practical volume $0.15\,\mathrm{cm}^3$ capable of producing $46\,\mu\mathrm{W}$ in a $4\,\mathrm{k}\Omega$ load from vibration levels of $0.59\,\mathrm{m\,s}^{-2}$ at the device resonance frequency of 52 Hz [20].

However, for the self-powered system that was implemented on the F/A-18 aileron hinge, time and resource limitations led to the development of a relatively simple power-harvesting system based on the conversion of strain energy to electrical energy (charge) via two different electromechanical coupling devices, PVDF piezoelectric films, and PZT piezoelectric ceramics. The combination provided a degree of robustness in terms of mechanical and thermal durability. Also, since the operational strains in the titanium aileron hinge were not well understood, the combination of PVDF and PZT power harvesting elements ensured that enough electrical power would be generated to power the system. PVDF elements are durable, have an adequate operating temperature range, up to about $80\,^{\circ}\mathrm{C}$, and are chemically inert. The PZT wafers have better strain/electrical energy conversion efficiency than PVDF and have a better operating temperature range, up to $\sim 180\,^{\circ}\mathrm{C}$, but are not as mechanically durable as PVDF (i.e., low strain to failure of 0.1–0.2% strain).

2.2.2 System operation and information flow

The basic operation and information flow issues can be summarized as follows:

- When in flight, a combination of piezoelectric film (chosen for its durability) and piezoelectric ceramic (chosen for its higher output) elements are used to power up the system.
- A patch health measurement is taken by the piezoelectric film sensors when power is available and the temperature and strain are within a selected range. As damage growth is slow, the relatively sparse sampling is considered adequate.
- The reading is used to refine the current health estimate stored in a nonvolatile memory.
- A handheld interrogator, placed in close proximity to the hinge, powers the circuitry of the self-powered system via inductive coupling and allows readings to be downloaded via the same inductive link. Inductive coupling was chosen as it is able to carry power and information through a thin barrier (in this case, the plastic aileron hinge cowl, although transmission through thin aluminum is also possible).

3 DETAILED DESIGN

3.1 Damage detection scheme

The FE model of the hinge with a b/ep reinforcement, which is shown in Figure 2(a), was used to assess the proposed damage detection techniques [21]. Two load cases were considered in the FE analysis, viz., load case WO39 a tensile design ultimate load condition and WO42 a compressive design ultimate load condition [5]. Figure 2(b) shows the peel and shear stress distribution in the adhesive of the reinforcement, respectively, for load case WO39, and indicates that two critical regions exist, namely, the tapered region at the end of the patch on the hinge strut ($x \sim 200\,\mathrm{mm}$) and the region of high adhesive peel stresses in the concave portion of the reinforcement ($x \sim 70\,\mathrm{mm}$). The BPSP monitored damage in the former and the SPSP monitored damage in the latter. In both cases, damage was detected by monitoring the change in ratio of critical region strains to far-field strains (referred to here as the *patch health indicator*) [22]. The discussion below focuses on the analysis associated with the damage detection approach for the SPSP system.

In this case, the damage was simulated by a 10 mm-long disbond in the adhesive layer at the high peel stress region (at $x \sim 70\,\mathrm{mm}$) of the patch. The surface longitudinal strains (ε_x) plotted in Figure 2(c) for the tensile load case, WO39, show significant increases in surface strains when the damage is introduced,

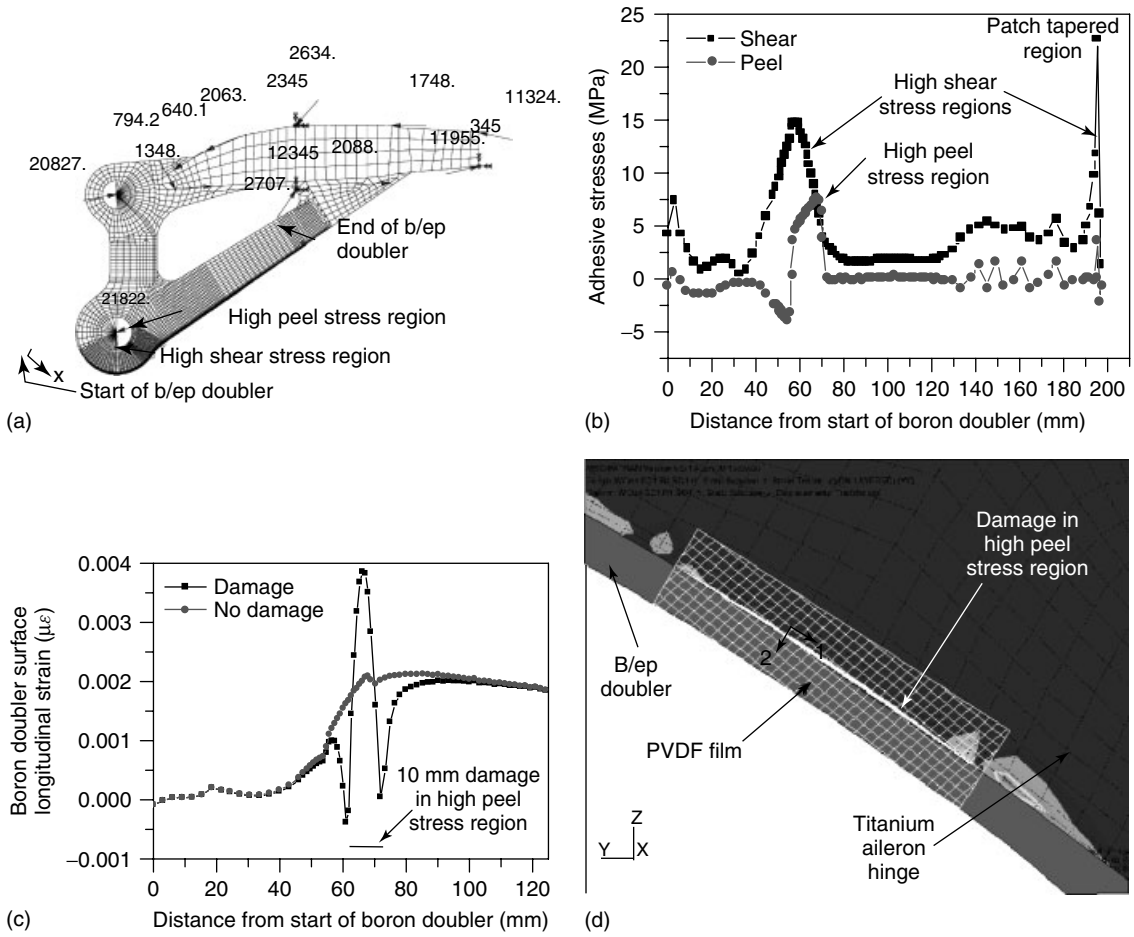


Figure 2. (a) Finite element model of titanium aileron hinge with b/ep doubler. (b) Predicted adhesive stresses in the reinforcement system with no damage present (load case WO39). (c) Longitudinal surface strain profile on b/ep doubler due to a 10 mm disbond in the adhesive high peel stress region. (d) FE displacement plot and damage detection scheme for damage in the high peel stress region.

indicating that damage detection is possible using surface strain sensors. However, the distinctive sharp peak/trough in the surface strains (over a 10 mm length) means that the probability of detection is very sensitive to the sensor size and proximity to the damage site. Thus sensors will only detect the damage if the sensing length is small, compared to the strain gradient length, and if the sensors are positioned over or very close to the damage or when the damage reaches a substantial size. Also, it is possible that the damage will initiate on one side of the strut before growing, through the width, to the other side of the strut. Therefore, to have a reasonable chance of detecting damage, one approach would be to use an array of short gauge length strain sensors on both sides of the b/ep reinforcement. However, this would require several dual-channel devices or one multichannel device to monitor a reasonable length of about 10–20 mm. A more simplistic (in terms of the electronics and number of sensors required) approach is to make use of the disbond opening under load (Figure 2d).

In this application, piezoelectric film sensors were proposed for measuring strain, or, more specifically, to measure strain produced in the film bridging the disbond as it opens. Piezoelectric film sensors

Table 1. Voltage outputs for two piezoelectric film sensors

Result case	$\int \varepsilon_x \, dS$ (mm^2)	$\int \varepsilon_y \, dS$ (mm^2)	$\int \delta u \, dy$ (mm^2)	PVDF output (V)	Ratio V_{dam}/V_{nodam}
No damage case (4 mm × 15 mm sensor covering the anticipated damage)	0.0899	0.0345	0	2.7	2.5
Damage case (4 mm × 15 mm sensor covering the damage)	0.0899	0.0511	0.147	6.8	
No damage case (2 mm × 15 mm sensor covering the anticipated damage)	0.0470	0.0226	0	1.5	3.7
Damage case (2 mm × 15 mm sensor covering the damage)	0.0463	0.0394	0.147	5.5	
No damage case (5 mm of the 4 mm × 15 mm sensor covering the anticipated damage)	0.0836	0.0294	0	2.5	1.6
Damage case (5 mm of the 4 mm × 15 mm sensor covering the damage)	0.0835	0.0360	0.0506	3.9	
No damage case (5 mm of the 2 mm × 15 mm sensor covering the anticipated damage)	0.0449	0.0182	0	1.4	1.9
Damage case (5 mm of the 2 mm × 15 mm sensor covering the damage)	0.0443	0.0176	0.0506	2.7	

generate a charge when strained rather than conventional electrical-resistance foil strain gauges, which *require* electrical power to measure strain. Thus the use of such sensors also reduces the system power requirements. In this case, a piezoelectric film sensor covering the edge of the aileron hinge strut (over the edge of the patch, adhesive layer, and the titanium strut), as shown in Figure 2(d), was used to detect disbond opening. FE studies for two sensor configurations, one 2 mm × 15 mm long and the other 4 mm × 15 mm long, were undertaken [23]. The area integral of the longitudinal (ε_x) and transverse (ε_y) strains of the sensing area is tabulated in Table 1 for various sensing conditions, i.e., no damage, the entire sensor covering the damage, and only half the sensor covering the damage. Table 1 also includes the area integrals of the crack opening ($\int \delta u \, dy$) and the predicted electrical voltage calculated from the overall area integral of strain [24]. The results show that the 2 and 4 mm wide (by 15 mm long) sensors will have a 2–4 and 1.5–2.5 fold increase in response, respectively, due to a 10 mm disbond, when compared with the undamaged case. The piezoelectric film sensor (see Figure 3) was manufactured from 28 μm thick PVDF sheet where the silver ink electrodes were removed using a grit blasting process and a mask was placed over the sheet to protect the sensing regions during grit blasting.

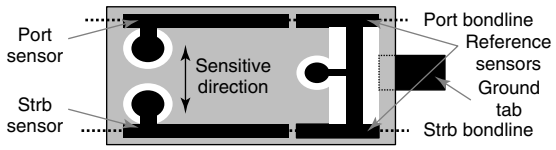
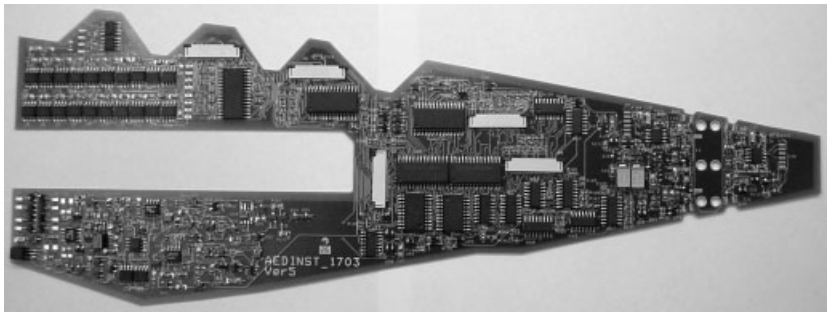


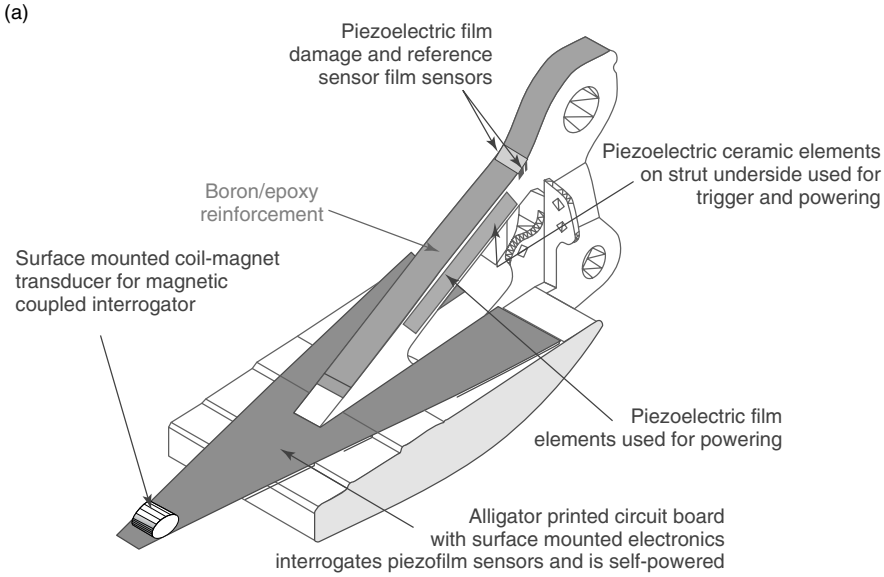
Figure 3. PVDF sensor configuration (hatched area indicates silver ink electrode on top surface and gray is the silver electrode on the lower surface).

3.2 Electronic design and power (demand) requirements

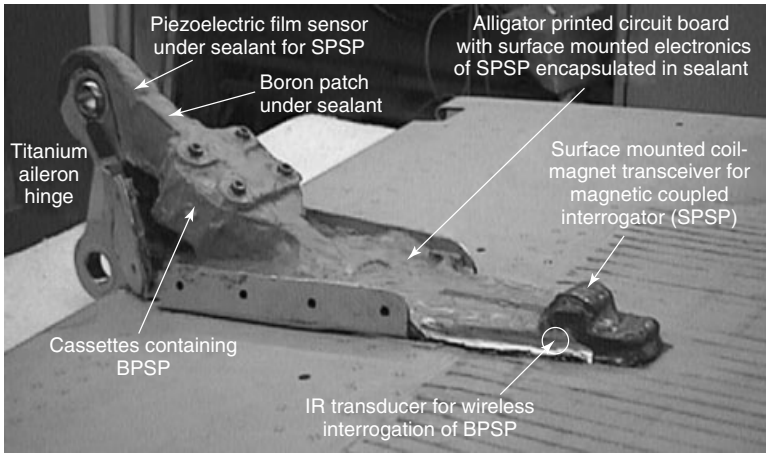
Geometric constraints and limitations in strain level and frequency response limit the amount of power available while using piezoelectric elements to fractions of a milliwatt. Since it is difficult to get a reasonable amount of signal processing done on this kind of power budget, the harvested power was stored in a capacitor. Consequently, a reading was taken only once the following conditions were met: (i) sufficient energy was available for a complete reading/computation/store cycle, (ii) the structure was experiencing significant strain (within a given window), and (iii) the temperature was also within a given window. Budget and time constraints necessitated a fairly large A-shaped printed circuit board (APCB) footprint with surface-mounted commercial off the shelf (COTS) electronic components, shown in Figure 4(a).



(a)



(b)



(c)

Figure 4. (a) Photo of the SPSP printed circuit board, (b) schematic of self-powered *smart patch* system, and (c) photograph of the installed *smart patch* system on an F/A-18 aileron hinge. (SPSP and BPSP are self-powered and battery-powered *smart patch* configurations, respectively.)

Figure 4(b) and (c) show the mounting arrangement of the *smart patch*, including the BPSP, and the location of the transducers on the aileron hinge. However, the design is such that it should be a relatively straightforward process to reduce the unit to very small proportions using standard complementary metal oxide semiconductor (CMOS) technology—this would have the added benefit of quite substantially reducing the energy requirements and overall weight.

To keep the energy requirements of the circuit as low as possible, it was necessary to minimize the number of components. This is illustrated in Figure 5 for the analog section and Figure 6 for the digital section. Figure 5 shows that the analog section consists of a small number of building blocks: (i) a power harvester (ii) a power controller (with temperature lockout), (iii) a trigger, (iv) a signal conditioner, (v) two logarithmic analog to digital converters, and (vi) an interrogation transceiver. It can be noted from the actual circuit, inserted under the block diagram, that there are, indeed, only a few components.

To minimize energy usage, the digital section, depicted in Figure 6, uses as few components as is practicable. The building blocks are (i) a flash to binary converter, (ii) a computation unit, (iii) data storage, (iv) a finite state machine, (v) a clock control, (vi) an interrogation control, (vii) a data output unit, and (viii) programming/testing ports. Items (i)–(iv) are all implemented in ferroelectric nonvolatile random access memory (FRAM) to reduce power requirements and to reprogram the circuit function; this proved extremely valuable in the test and verification stage of the system design.

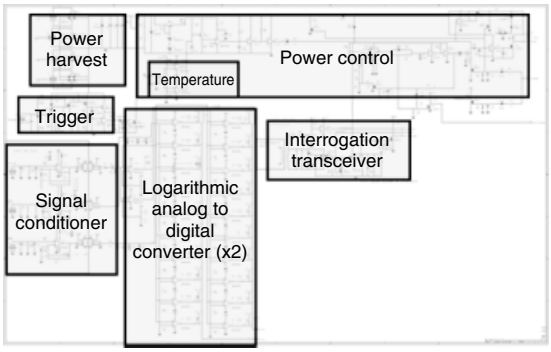


Figure 5. Electrical schematic block diagram—analog section.

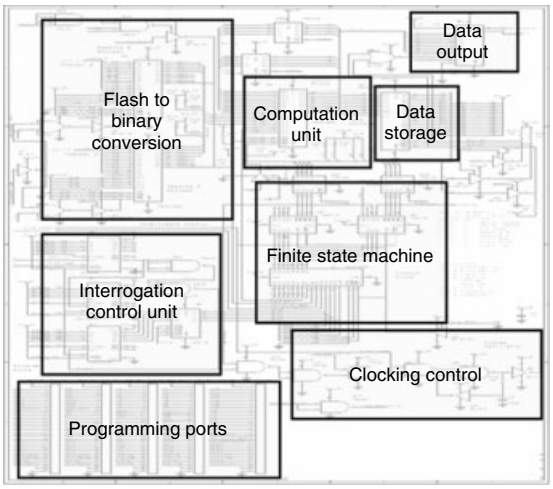


Figure 6. Electrical schematic block diagram—digital section.

As discussed previously, PVDF film was proposed as the sensing material because of the fact that it requires no excitation power. Other advantages with piezoelectric film sensors are their very good fatigue life and strain to failure properties, simple signal conditioning requirements, good high-frequency response; they are also light weight and conform easily to complex shapes. Some limitations with PVDF sensors that need to be considered are their limited operating temperature, poor thermal sensitivity, poor response in the low-frequency regime, and also the extreme care that is required to ensure the correct sensor polarity during installation. As there is a high probability of a substantial amount of strain being generated at low frequencies, it was considered important to ensure that the frequency roll-off characteristics were fairly accurately matched, especially in adverse environments. This in itself can be tedious, but as high impedances are involved, the greater challenge is ensuring that they remain matched in adverse environments. To assist with this problem, buffer amplifiers with low input leakage were incorporated, but the need for thorough protection against moisture ingress remained a critical issue. Considerable care was taken to reduce the risk of moisture ingress creating a shunting effect that would adversely affect this matching.

To minimize system power requirements, the patch health measurement was a ratio obtained using an

analog technique, where the signal from the "reference" transducer was fed through a divider chain and compared with the signal from the active transducer (i.e., sensor over the damage region) via a chain of 10 comparators (with 10% resolution). Each digitizer/divider "reading" uses about 100 nJ (i.e., typically 2.5 μW per comparator, with a 1.2 ms power-up time for 10 comparators plus resistive chain). One advantage of this technique is that it has a frequency response in the kilohertz region, and minimal possibility of time skew problems between the two readings forming the ratio. A disadvantage is the relatively long (1.2 ms) "switch on" settling time of the comparators (and amplifiers). This necessitated a two-stage power supply, where the analog and digital circuits were powered for milliseconds and a few microseconds, respectively. Also, since the PVDF sensors could not be calibrated before installation, owing to time and resource constraints, it was likely that the sensor readings from the undamaged critical region might not exactly match those from the reference; consequently, perfect patch health readings might not result in a ratio value of unity.

The device was designed to measure two ratios, one on the starboard and the other on the port side of the strut, with respect to a common reference and to a 10% resolution. Each reading, taken when the conditions previously described were met, was recorded in FRAM, wrapping back to the initial position in storage after approximately 16 million readings. FRAM was chosen primarily because of its relatively low access energy requirement, which is in the region of 1 nJ/bit compared typically with a few microjoules per bit for electrically erasable programmable read-only memory (EEPROM) devices, and its tolerance of extended temperatures. The parallel variant was chosen, as the higher access speed helped in minimizing system energy requirements. Unfortunately, at the time of design, only the 5 V supply variant of these parts was available, thus to keep the design simple, most of the circuit operates at this relatively high voltage. The high-level choice of FRAM as the storage medium dictated that the device required some digital circuitry, and the choice of input transducers dictated analog input circuitry.

Besides the storage function, the other aspects of the electronic design that were chosen to be implemented in digital technology were sections of the computation and most of the interrogation functions. A programmable logic device (PLD) was originally chosen on the grounds it provided low operating power with maximum flexibility. However, further investigation revealed a large and potentially problematic start-up energy requirement. As standard cell and custom integrated circuits (IC) were beyond our budget it was decided to use medium scale integration (MSI) CMOS as the most appropriate option. The information stored was two (4 bit) ratios and one (24 bit) integer indicating the number of reading cycles. A temperature-dependent lockout was fitted to the supply circuit to prevent any operations being attempted if the operating temperature was outside specified temperature limits. The limits set were quite tight, within a window range of 10–40 °C, to ensure that readings were not influenced by thermal effects. This was considered an acceptable option in this application as the data required was a very small number of samples and not considered to be particularly dependent on temperature.

In the most elementary case, the only value needed would be the last worst patch health indicator reading. However this was considered not to be sufficiently robust, since a single noise "spike" could cause a misleading reading. This led to the conclusion that filtering and some other confidence-building indicators were required. A time stamp was considered, but this would necessitate a permanent power source to keep the clock running. Also, since the aim was for "indefinite" operation, a reading counter was chosen as a compromise. Unfortunately, the only place to maintain the count is in the nonvolatile memory, thus necessitating a "measurement cycle" to consist of (i) reading the memory, (ii) taking the measurement, and (iii) writing the updated information back into memory. In an attempt to reduce the effects of spurious spikes, a simple slew rate filter algorithm was employed. This was thought to be appropriate as the damage is expected to grow relatively slowly with respect to the reading rate, and even if the patch suddenly failed this would be indicated after only 10 reading cycles at most. During a reading cycle, the system is synchronized to an internal ~1 MHz clock, the speed being basically chosen to minimize the energy required.

Measurements of the overall system indicated the prototype consumed about 210 nJ of energy to perform the digitizing and dividing function, 60 nJ to perform the storage and retrieval function, and 10 nJ

for clocking. Hence, the total energy required of the system to make a reading, retrieve and store data, and undertake simple real-time data processing is approximately 280 nJ.

3.3 Power harvesting—energy harvesting and conversion issues

On the supply side, there were two PVDF stacks, each stack having three PVDF elements electrically in series, with poling directions perpendicular to the loading direction. The individual PVDF elements had a nominal thickness of 52 μm, and a manufacturer-rated capacitance of 5.7 nF [25] from which the PVDF relative permittivity $\varepsilon_R \sim 11.3$ was calculated. The PVDF elements were laminated prior to installation, bonded together into a stack using Hysol EA9309 adhesive, and found to have a fairly consistent measured series capacitance of ~1.9 nF. Care was taken to ensure that PVDF electrodes of similar potential (while under cyclic loading) were colocated to minimize capacitive coupling. Insulation resistance was measured to be in excess of 50 GΩ, so both dc and low ac frequency dielectric loss appeared to be negligible for PVDF.

Using the typical published piezoelectric coefficient for PVDF (d_{31}) of 23 pC/N and assuming a simple one-dimensional piezoelectric case with no losses, the predicted peak voltage due to a peak strain of 300 με was ~32.1 V across a stack of three PVDF elements in series. Each PVDF stack fed power through a separate bridge rectifier, which incorporated BAS70 [26] diodes to minimize electrical loss. Each diode had a typical forward voltage drop of 0.22 V (at relevant current levels) and 3 nA of leakage current. The PVDF stacks were directly bonded to the aircraft structure and had a calculated output of ~0.1 V/με (open circuit). The PVDF stack capacitance of 1.9 nF meant that about 0.19 nC (0.1 V × 1.9 nF)/1 με of charge could potentially be generated.

The energy harvesting power supply was a simple diode bridge (Figure 7a) and capacitor, which can be thought of as having two states, diode conduction "ON" and "OFF" as shown in Figure 7(c). The circuit was designed for a nominal operating voltage of about 5.2 V. Thus the first 5.2 V/(0.1 V/1 με) = 52 με of mechanical load excursion did no work on the electrical load (assuming the previous excursion reached the conducting state) and, in fact, incurred a slight leakage current loss while it was occurring. At the nominal operating point, the storage capacitor holds an electrical potential energy of $E = 1/2\ \text{C}\ \text{V}^2 = 0.5 \times 1\,\mu\text{F} \times (5.2\ \text{V})^2 \approx 14\,\mu\text{J}$.

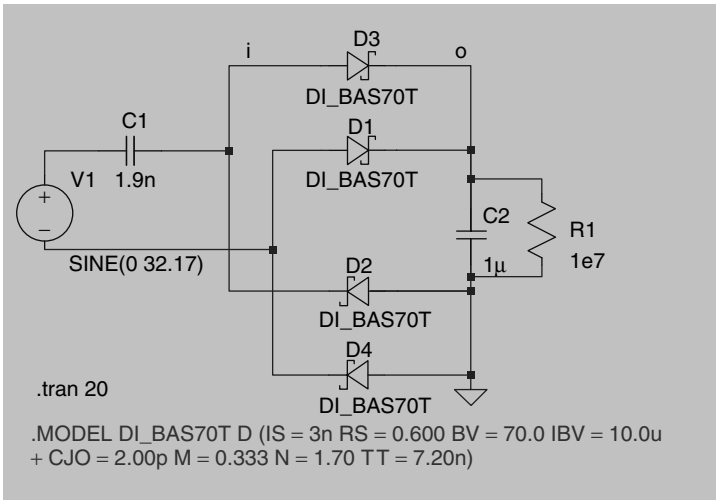
Once the diodes turn "ON", the voltage can be thought of as being approximately constant (assuming the storage capacitance is much larger than the PVDF stack capacitance, 1 μF vs. 0.0019 μF in our case) and each 1 με would add energy of $1/2 \times 1.9\ \text{nF} \times [(5.2\ \text{V} + 0.1\ \text{V})^2 - (5.2\ \text{V})^2] \approx 1$ nJ (not allowing for losses).

With this type of system the primary "losses" associated with the diodes is because of leakage when the diode is reversed biased, and forward drop when the diode is conducting. The power "loss" associated with leakage can be approximated by multiplying the leakage current by the (mean) reverse voltage and by the fraction of time the device is reverse biased. The forward loss can be approximated by the ratio of forward voltage drop to load voltage.

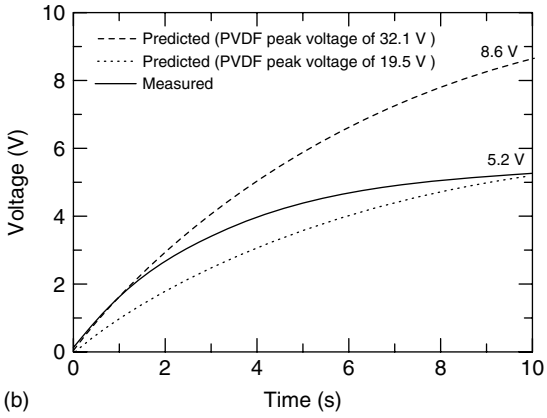
As mentioned, the energy harvesting circuit implemented had a working load voltage of about 5.2 V and since the diodes had a forward voltage drop of about 0.22 V, the diodes of the full bridge (two diodes on at a time) would have dissipated about $100 \times (2 \times 0.22\ \text{V}/5.2\ \text{V}) \sim 8\%$ of the incoming power while in the conducting state.

If the input was truly sinusoidal, then it is expected that the diodes would be reverse biased for about half the time. Given that the leakage of the rectifying diodes was about 3 nA at the operating voltage of 5.2 V and with at least one pair being reverse biased at any point in time, then the expected drain would be about $(2 \times 3\,\text{nA}) \times 5.2\text{V} \approx 31.2\,\text{nW}$ while in operation. Should the excitation amplitude reduce for a period then this value is expected to double, since all four diodes would be reverse biased. Hence, a minimum of 31.2 nW/1 nW ≈ 31 με of "active" strain excursion would be required per second to replenish diode losses. This figure was kept low by the use of low-leakage diodes. If BAT54 diodes with 600 nA leakage [27] had been chosen, then an "active" strain excursion level of $\sim 6200\,\mu\varepsilon\,\text{s}^{-1}$ would be required. Here the term *active* refers to the part of the strain excursion that occurs while the diodes are "ON".

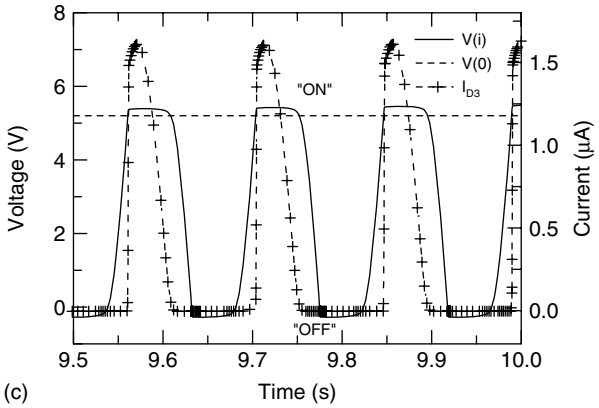
The energy harvesting process was modeled via LTSpice [28] simulation, as shown in Figure 7(a),



(a)



(b)



(c)

Figure 7. (a) Model of PVDF/diode rectifier energy harvesting circuit. (b) Measured versus predicted storage capacitor voltage. Predictions were made using two PVDF peak voltages viz., 32.1 and 19.5 V. The measured voltage was for a 1 µF storage capacitor being charged using a PVDF stack subject to a sinusoidal strain with a peak value of 300 $\mu\varepsilon$ at a frequency of 7 Hz. Storage capacitor voltage was ~5.2 V after 10 s, and ~5.36 V after 14 s. (c) The current for diode D3 (I_{D3}) versus time for a PVDF peak voltage of 19.5 V and a storage capacitor voltage near the operation point of 5.2 V.

which incorporates the BAS70 full bridge diode rectifier, a 1.9 nF PVDF stack, and a 1 μF storage capacitor with a 10 MΩ parallel resistance to model a 10X oscilloscope probe (at low frequency, the probe capacitance of ~13 pF was considered irrelevant). The model assumes that the energy harvester and aircraft structure are effectively decoupled, i.e., actions of the harvester have negligible effect on both the bulk and local strains in the aircraft structure. The PVDF is modeled as a 1.9 nF capacitor in series with an ac voltage source of 7 Hz with peak voltage 32.1 V. The manufacturer states that the BAS70 diode leakage is 3 nA (at 25 °C and 5 V reverse voltage); this was confirmed experimentally (the diode measured was from the same diode batch that was later flown) and hence the saturation current "I_s" in the LTSpice diode model was adjusted to 3 nA. LTSpice modeling showed that after 10 s of charging, from a sinusoidal strain excitation of 300 με peak strain at 7 Hz, the storage capacitor dc voltage should be ~8.6 V assuming only diode losses.

Figure 7(b) also shows a comparison of the LTSpice simulation with the measured voltage of a 1 μF storage capacitor during the charge up phase, using a PVDF stack (viz., one stack of three 52 μm thick by 156 mm long by 19 mm wide PVDF elements) subjected to a sinusoidal strain with a peak value of 300 με at a frequency of 7 Hz. In this case, the PVDF stack is a similar configuration to those flown on the flight demonstrator. The storage capacitor voltage achieved was ~5.2 V after 10 s, and after 14 s the final storage capacitor voltage was 5.36 V.

Given the stated mechanical loading conditions (7 Hz at a peak strain of 300 με), a comparison between the final simulated storage capacitor voltage (8.6 V) and the measured storage capacitor voltage (5.2 V), as shown in Figure 7(b), indicates that the inefficiencies (losses) totaled about $100 \times (32.1 - 19.5/32.1) \sim 39\%$. Three mechanisms have been postulated for this energy harvesting loses. In decreasing order of importance they are

1. the effect of Poisson's ratio, a tensile stress in the 1 direction will produce ~1/3 compression in the 2 direction, which will reduce the piezoelectric output voltage by ~1/3 (assuming $d_{31} \sim d_{32}$);
2. a "shear lag" effect that is the combination of bondline mechanical compliance issues and, more importantly, the reduction of strain in the PVDF elements that are located furthest from the metallic substructure due to the compliance of the piezoelectric film; since the PVDF elements are quite compliant, not all the strain is transferred from the metallic substructure through these elements to the ones above—the larger the stack the less strain will be transferred to the top elements;
3. the effect of capacitive coupling between the PVDF metallization layers.

It appears that Poisson's ratio effect alone cannot explain the 39% efficiency decrease measured. Using the measured storage capacitor voltage of 5.2 V (at $t \sim 10$ s) generated by a single PVDF stack, the energy stored in the 1 μF capacitor after 10 s was approximately,

$$E = \frac{1}{2} C\ V^2 = 0.5 \times 1\ \mu F \times (5.2\ V)^2 \sim 14\ \mu J \quad (1)$$

This means 14 μJ of energy could be harvested every 10 s from each side of the strut, resulting in 1.4 μW, which was sufficient energy for the device to operate when attached to the aluminum hinge configuration.

The limited flight strain data available for the aluminum aileron hinge configuration, indicated that the strain loading scenario (i.e., 300 με peak at 7 Hz) should be easily achieved at the aileron aft hinge strut. Note also that the 7 Hz used in this experimental study is quite conservative since the in-flight strain and acceleration data indicate that structural resonances occurred at about 14, 30, and 55 Hz, as shown in Figure 8. These higher frequencies would produce more energy for harvesting within a given time period.

The replacement of the aluminum aileron hinge with the titanium hinge meant that there was a significant reduction in the operational strain levels on the hinge. FE studies indicated that the strains were reduced by about a factor of about 2. The maximum surface strains on the b/ep doubler, for load condition WO39, were reduced from 5000 με for the aluminum to 2100 με for the titanium hinge [21, 23]. Therefore, it was decided to add additional power-harvesting PZT elements on the lower side of the hinge strut, as shown in Figure 4(b). The PZT power-harvesting elements consisted of two 32 mm long by 12.5 mm

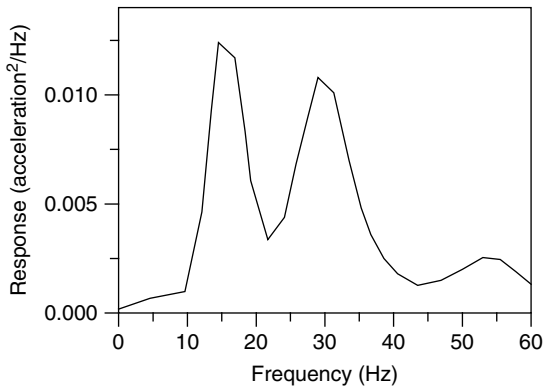


Figure 8. Typical acceleration frequency spectra measured on in the aileron hinge region.

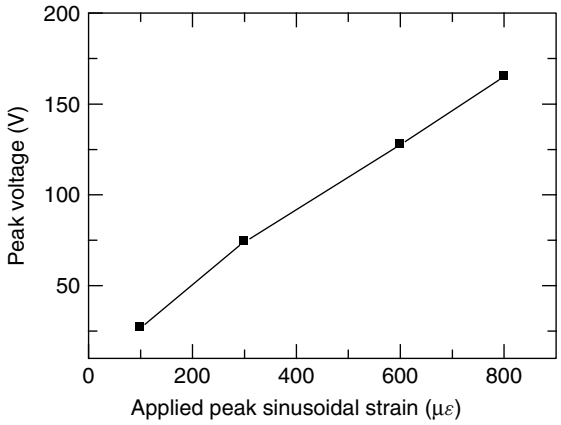


Figure 9. Peak strain versus open-circuit voltage for a T226-A4-303Y PZT-5A bimorph subjected to a sinusoidal peak strain (10–30 Hz).

wide Piezo Systems T226-A4-303Y bimorph transducers, consisting of a 0.12 mm thick brass shim sandwiched between dual parallel poled 0.27 mm thick PZT wafers [29]. The calculated capacitance of the PZT-5A bimorph was 12.4 nF compared with the measured capacitance of approximately 9.2 nF.

When a load equaling 300 $\mu\varepsilon$ was applied to the PZT-5A bimorph the calculated open-circuit voltage was ~115 V, assuming no losses or inefficiencies. To experimentally confirm that the voltage produced by the PZT-5A bimorph was as expected, a specimen was manufactured from 4340 steel plate with dimensions 260 mm × 75 mm × 4 mm. The PZT-5A bimorph element was bonded to the plate using silver-loaded epoxy with a resultant bondline thickness of less than 200 μm. The top surface of the bimorph was also coated with silver-loaded epoxy to approximate the bonding conditions that the bimorph would experience in service.

As shown in Figure 9, when the PZT-5A bimorph was subjected to a sinusoidal 300 $\mu\varepsilon$ peak strain (at excitation frequencies between 10 and 30 Hz), the recorded peak voltage produced by the bimorph was about ~75 V across a 10 MΩ oscilloscope probe, which is an $100 \times (115\,\text{V} - 75\,\text{V}/115\,\text{V}) \approx$ 35% efficiency decrease. The difference between the calculated and measured PZT-5A bimorph voltages appears to be almost entirely explainable by the Poisson's ratio effect.

For the purpose of energy harvesting, the output from each of the PZT bimorphs was fed into a separate bridge rectifier that was effectively working in a half-wave manner. The main benefit of the half-bridge approach was that it allowed one electrode of each of the PZT-5A bimorph to be bonded to, and hence, electrically grounded with, the aircraft structure.

To summarize, assuming no mechanical or electrical losses, the individual PVDF elements appeared to be rated at ~2.5 nJ/(m^2·$\mu\varepsilon$) and the PZT-5A bimorphs at ~4.4 μJ/(m^2·$\mu\varepsilon$). Maintaining the assumption of no mechanical or electrical losses, and also assuming 100% strain transfer efficiency and with no maximum placed on voltages, with a constant sinusoidal mechanical loading of 7 Hz at 300 $\mu\varepsilon$ peak strain, it was calculated that

1. ~0.96 μW of power could be harvested from a single PVDF element, resulting in a total of ~5.8 μW harvestable from the two PVDF stacks;
2. ~2.2 mW from a single PZT bimorph, resulted in a total of 4.4 mW from the two PZT-5A bimorphs attached to the aileron strut; and
3. the overall minimum harvestable power from the PVDF stacks and the PZT bimorphs was ~4.406 mW.

As can be seen, the two PZT-5A bimorphs were capable of producing significantly more power than the two PVDF stacks; however, there was a risk that the ceramic PZT-5A bimorphs might not have been robust enough to cope with the stresses induced by operational flight loads.

Further expanding upon the energy harvesting efficiency and loss factors discussed above, we arrive at the following results:

1. Dielectric loss: this appears to be sufficiently small as to be considered negligible for both for both PVDF and PZT-5A bimorphs. For example, a PVDF charged to 5 V, with 50 GΩ insulation resistance, has a continuous dc power loss of $\sim(5\,\mathrm{V})^2/50 \times 10^{12}\ \Omega = 0.5$ pW, and assuming low-frequency structural vibration, ac power loss to the dielectric will also be negligible.
2. Geometrical strain compliance between the underlying structure and the piezoactive material: this is an efficiency loss due to the single sided mechanical coupling that is normally used to attach a piezoelectric element to the structure.
3. Diode loss: the diode loss in the bridge rectifier is caused by both the forward voltage drop across the diode and the diode leakage current, both of which produce real energy losses via heat production. Generally speaking, the peak currents produced by the energy harvesting process are small, and assuming no losses, the peak currents will be $<3\,\mu$A for the PVDF stack and $<50\,\mu$A for the PZT-5A bimorph. The Schottky diodes used in the rectifiers must be carefully chosen; however, it is not a trivial task to determine the best diode because the LTSpice diode models provided by manufacturers are generally quite inaccurate when currents are small (i.e., 100 μA or less), so accurate diode losses cannot be simulated. For energy harvesting from low-frequency structural excitations the authors consider diode leakage to be the main diode loss component. A laboratory testing program appears to be the most efficient method of determining the best rectification diodes for the particular energy harvesting task.

At this point it should be noted that the simple energy harvesting circuits implemented on the *smart patch* may not have been optimal since the voltage across the piezoelements, and hence the electrical component of the force was rather low. On the other hand, having very small forces feeding back from the energy harvester in the aircraft structure is conservative from a structural dynamics point of view. In any case, circuit board space and design time limitations prevented us from using a more elaborate circuit.

It is difficult to assess the exact mechanical and electrical losses in the harvesting system during any particular flight because of the

1. lack of low current fidelity in the diode spice models provided by manufacturers;
2. action of the diode bridge rectifiers during the energy harvesting process being nonlinear; and
3. mechanical loading of the aileron hinge during flight being unknown.

However, for a specific sinusoidal loading condition of $300\,\mu\varepsilon$ peak strain at 7 Hz, measurements have shown (Figure 7b) that a PVDF stack can charge up a 1 μF capacitor to the operational load voltage of 5.2 V in ∼10 s. Under the same loading conditions, taking into account the inefficiencies and loss mechanisms discussed above, LTSpice simulation calculated that a PZT-5A bimorph will charge up a 1 μF capacitor to the operating load voltage of 5.2 V in ∼0.57 s. Given that the time to charge the 1 μF storage capacitor to the operating load voltage 5.2 V is known, the expected power-harvesting level for the combination of the two PVDF stacks $(2 \times [0.5 \times 1\ \mu\mathrm{F} \times (5.2\ \mathrm{V})^2/10\ \mathrm{s}] \approx 2.8\,\mu\mathrm{W})$ and the two PZT-5A bimorphs $(2 \times [0.5 \times 1\ \mu\mathrm{F} \times (5.2\ \mathrm{V})^2/0.57\ \mathrm{s}] \approx 90\ \mu\mathrm{W})$ was estimated to be ∼93 μW, meaning that ∼0.55 s of time was required to generate enough energy to power the electronics outlined in Section 3.2.

3.4 Interrogation

During the information uploading operation, the circuit was both powered and clocked by an alternating magnetic field provided by the interrogator. This field was modulated with a serial representation of the reading information, which was continually repeated until the interrogator received several identical samples at which time it accepted the data.

The antenna consists of a coil of copper wire wound on a ferrite core. The device is energized by the current produced in the coil in response to a low-frequency magnetic field produced by the interrogator. The data is modulated on this carrier by changing the reflected impedance. The data rate

is derived as a submultiple (1/16th) of the carrier frequency.

4 SAFETY-OF-FLIGHT AND FUNCTIONAL TESTING

A number of testing activities were undertaken, during the development of the *smart patch* system, to cover two main aspects:

1. safety-of-flight issues, which required testing on a prototype *smart patch* system to characterize the system response from specified mechanical vibration/shock, electromagnetic and pressure/temperature environments;
2. system functional testing consisting of both bench-top testing of the electronic components (including interrogator) and a prototype *smart patch* system on an "original" aluminum aileron hinge component (removed from an F/A-18 aileron).

4.1 Safety-of-flight testing

The vibration test on a system mock up consisted of a resonance search test, and a low-frequency swept sine, a low-frequency sine dwell and a high-frequency random vibration test, and finally a shock test in each of the three principal axes. Testing requirements are based on specifications given in [30, 31]. Visual inspection of the test article after testing did not reveal any deformation, delamination, fragmentation, or breakage [32]. Results of the tests showed that the system survived all vibration and shock testing without failure and that the equipment performed normally after the test.

Temperature and altitude testing included temperature cycles between -40 and $+80\,^{\circ}$C with concurrent altitude pressure cycles from sea level to 9000 m. No adverse structural or chemical problems were encountered; however, significant variations in the coefficient of thermal expansion between the potting compound and the wiring caused solder joints to fail—these issues meant that a more compliant potting compound needed to be used for the in-flight demonstrator.

The electromagnetic test consisted of a radiated emissions test in accordance with MIL-STD-461E [33] and Boeing requirements [34] for RF susceptibility from 1 MHz to 18 GHz. No emissions were detected from the *smart patch* when it was optically isolated from the interrogator and monitor computer [35].

4.2 Functional testing

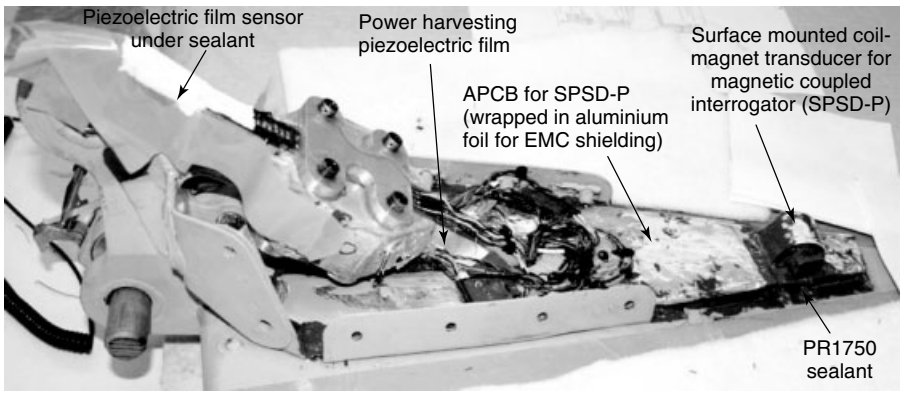
Electronic components (such as the APCB and handheld interrogator) were bench tested using simulated voltage inputs for the piezoelectric film sensor and power-harvesting elements. A trial installation was also performed on a subcomponent "original" aluminum aileron hinge taken from an F/A-18 aileron, similar to the component shown in Figure 1(b) and then subjected to functional and environmental testing. That was undertaken to "fine tune" and validate the installation procedure and evaluate the performance of the *smart patch* after adverse environmental (i.e., hot/wet/cold) loading. The environmental testing involved placing the subcomponent in an environmental chamber and subjecting it to several moderate and severe thermal cycles of -30 to $40\,^{\circ}$C and -40 to $70\,^{\circ}$C, respectively. All thermal cycling was undertaken at the maximum relative humidity of the chamber (which was about 100%), and each temperature extreme was maintained for at least 30 min, to ensure condensation would occur on the subcomponent during the environmental cycling; the aim was to ensure that the environmental cycling would comprehensively test the protective coating process. The specimen was then installed in the testing machine to perform a number of functional tests. This testing program was extremely useful and outlined several electronic and installation deficiencies, involving the environmental protection [6]. The main issues were problems associated with the FRAM units, as well as issues with the protective coating process associated with the electromagnetic shielding layer bonding to the protective polysulfide PR1750 sealant. Polysulfide PR1750 was used for the environmental protection for the electrical components and piezoelectric transducers, as well as to bond the APCB to the aileron hinge surface. Electromagnetic shielding was achieved by wrapping aluminum foil around the APCB, after applying the PR1750, and wiring. Care needed to be taken to ensure that air pockets were removed between the foil and the

PR1750. A coating of conductive paint provided the final electromagnetic protective layer.

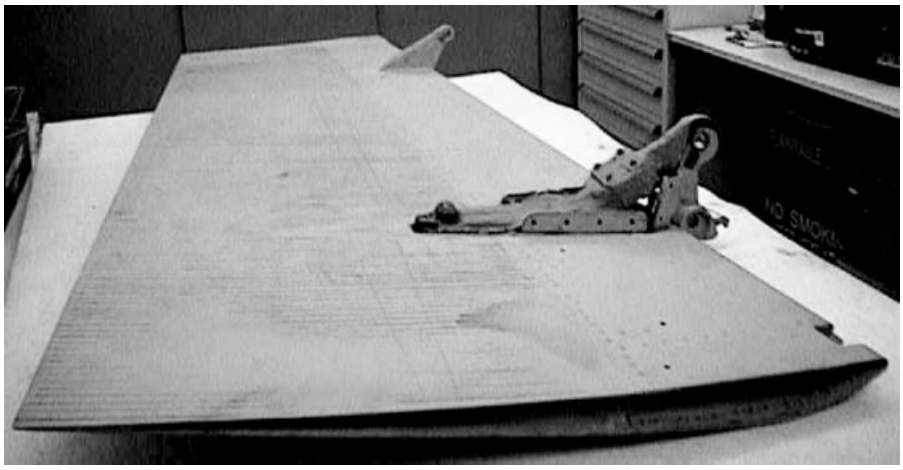
5 SYSTEM IMPLEMENTATION AND FLIGHT TEST RESULTS

A comprehensive installation procedure was written and approved by the appropriate airworthiness authorities [6]. The system was then installed on an Aircraft Research and Development Unit (ARDU), Flight Test Squadron (AFTS) aileron, over about a two-week period in November 2005. The aileron with the installed *smart patch* is shown in Figure 10.

The instrumented aileron hinge was installed on Hornet A21-101 in February 2006. The patch health information was to be downloaded by ground personnel from the instrumented aileron using a handheld unit, as shown in Figure 11, at the end of every week. No downloads were attempted if the aircraft did not fly during the week or if operational requirements meant that the aircraft was located at another base. After each download, the handheld unit was connected to a nearby docking station, which ensured that the handheld device was fully charged and allowed the patch health data, in the handheld unit, to be downloaded via a GSM mobile (cell) phone link for analysis. A plot of this data is shown in Figure 12.



(a)



(b)

Figure 10. (a) *Smart patch*, without protective coating and electromagnetic shielding, installed on aileron hinge; (b) photograph of *smart patch* on the F/A-18 aileron hinge.

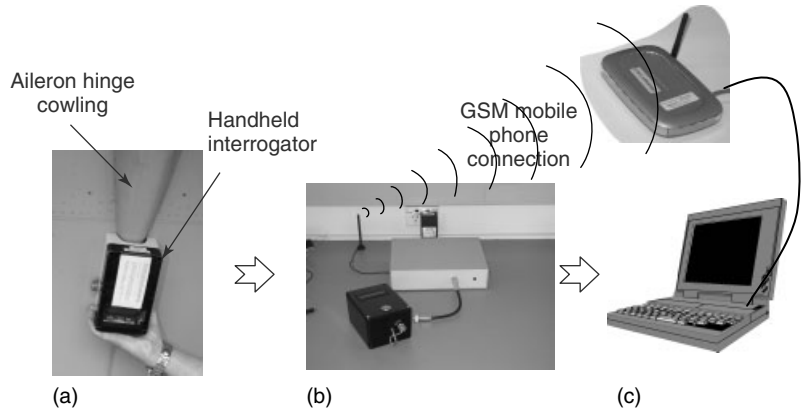


Figure 11. Patch health data download from *smart patch* (a) using handheld interrogator, (b) connecting handheld interrogator to docking station, and (c) download data to base station via GSM mobile phone link.

The trial ended in late February 2007 after 110 flights had been completed. During this time, the self-powered unit had registered 178 valid readings. The patch health indicator readings stored in FRAM, from the starboard and port piezoelectric film sensors, were initially set to an arbitrary low health value of about 0.8 and 0.1, respectively, as shown in Figure 12(a). Both patch health readings exhibited similar trends toward their respective perfect patch health ratio. As discussed previously, the PVDF sensors were not calibrated before installation and, as a result, the port and starboard ratios may settle to a nonunity ratio. The starboard sensor converged to a ratio of about 1.0 after 14 flights (37 readings) and then maintained a consistent value until the end of the trial. Therefore, any divergence from this value would indicate damage in the patch on the starboard side. The port sensor converged to a different ratio of about 0.63 after 97 flights (134 readings). As mentioned above, 178 valid readings were recorded after a total of 110 flights, as shown in Figure 12(b). The figure shows a steady increase in valid readings, of between one and two readings per flight, which was another indication that the device was working correctly.

6 DISCUSSIONS—LESSONS LEARNED

Considerable experience in low-power electronic design, design and implementation of packaging, as well as understanding issues associated with the design and implementation of a strain-based power-harvesting techniques has been achieved during the development of this demonstrator. One significant lesson learnt was to be cautious in the choice of electronic components that have only recently been released. This is because the project suffered considerable delays in the early part of the program due to the early adoption of immature COTS electronics, in particular, early versions of the Ti430 and a nonoperational FRAM chip. The use of COTS surface-mounted electronic components necessitated a large printed circuit board (PCB) footprint, making the installation complicated, cumbersome, and time consuming. Also, working with polysulfide PR1750 sealant to provide environmental protection, provided an additional level of complexity mainly due to the fact that the sealant was difficult to work with and required 24 h to cure, besides posing health and safety concerns. Hence, it would be desirable to eliminate the use of this sealant during any future *smart patch* installations.

Overall, there is a need to simplify the installation process into as few steps as possible. If the electronics were miniaturized (on a single chip), then this would greatly assist with packaging and installation, as well as reducing energy requirements, improving robustness, and reducing weight. A vibration-based energy harvesting approach would have also been feasible in this situation and would have facilitated in simplifying packaging and installation. This is because the system could be installed as a self-contained device thus eliminating the need

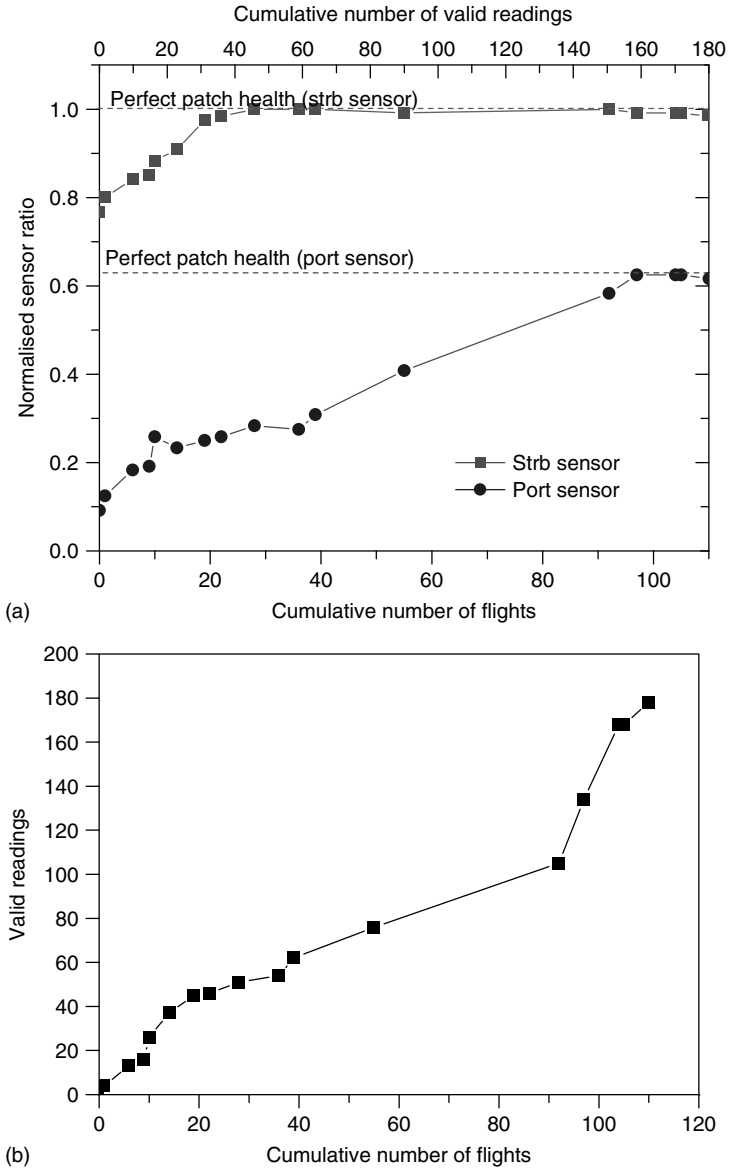


Figure 12. (a) Variation of *smart patch* normalized sensor ratios (patch health indicator readings) with increasing number of flights (from Mar 2006 to Feb 2007). (b) Plot shows the number of valid readings for the self-powered system with increasing number of flights.

to install separate power-harvesting piezoelectric elements, which required a significant footprint, to the structure.

After the instrumented aileron was removed from the aircraft, the *smart patch*, including the b/ep reinforcement, was removed from the aileron. Before removing the b/ep reinforcement from the aileron hinge, a standard tap test was undertaken, indicating no disbonds were present in the patch. Removal of the sensors, power-harvesting elements, PCB, and electromagnetic shielding indicated no significant adverse environmental effects on the various components.

However, it was observed that, in some regions, the PR1750 had not adhered to the shielding, suggesting that care needs to be taken when handling the shielding. Also, some of the PVDF layers in the power-harvesting stacks had disbonded. This observation reinforced the fact that more durability and robustness testing of the various sensing and power-harvesting elements would be required if the system were to be installed for a longer period.

7 CONCLUSIONS

The *smart patch* approach, which, in this case, is based on a self-powered *in situ* SHM system, would assist in alleviating certification concerns associated with implementing bonded composite repairs to primary aircraft structures. This approach relies on the ability to autonomously detect disbonding in the patch and is basically a continuous safety-by-inspection approach for the bonded repair. The main aspects of a self-powered *in situ* patch health monitoring system, applied to a composite bonded reinforcement on an operational F/A-18 aircraft, were outlined in this article. The system incorporates piezoelectric film PVDF sensors and is designed to operate using the electrical power generated by an array of PVDF and PZT power-harvesting piezoelements, which convert structural dynamic strain to electrical energy. The system was successfully installed on an operational RAAF F/A-18 aircraft. Patch health information downloaded from the *smart patch* system after 12 months of operational flying has indicated that the device is performing as expected and that the patch has "good" health. This was confirmed by applying a standard tap test to the composite bonded patch.

The system was designed with the focus on minimizing power requirements. This meant that issues such as the type of sensor, the amount and type of patch health information stored, electronic design/components including memory type and the interrogation technique were chosen from the standpoint of minimizing overall energy consumption. Owing to time constraints, the self-powered approach was strain-based, rather then vibration-based, since this required minimal development time even though it significantly complicated the installation process. To further lower risk, two types of power-harvesting piezoelectric elements, polymer and ceramic, were used in the system. This ensured adequate durability while ensuring enough power was generated at the reduced strain levels anticipated in the refurbished titanium hinges. The *smart patch* approach also demonstrated to the operators and maintainers of the aircraft that the interrogation of such systems required minimal interruption to their current maintenance routine.

ACKNOWLEDGMENTS

This program was undertaken under a RAAF sponsored task with support from Aircraft Structural Integrity-Directorate General Technical Airworthiness (ASI-DGTA) especially Officer-In-Charge ASI WGCDR Jason Agius and Dr Madabhushi Janardhana. Also, significant support was provided by AOSG, Aerospace Systems Engineering Squadron, especially Chief Engineer WGCDR Greg Young, Mr Des Cass, SGT Gavin Jones and SGT Andrew Schutz. The authors would like to also acknowledge Andrew Rider for development of the surface treatment procedure for the titanium aileron hinge and Ivan Stoyanovski for the application of the composite patch to the hinge; David Rowlands for assistance in installation of the *smart patch* and fabrication of the handheld interrogation unit; Keith Muller, Rodney Gray, Carlos Rey and Bruce Crosbie for assistance in mechanical testing; Peter Virtue and Howard Morton for undertaking the NDI of the aileron; Anthony Walley for assistance in developing the installation procedure and testing the PZT elements; Mike Konak, Quang Nguyen, Peter Smith and Sami Weinberg for assistance in circuit design and software development; and Richard Callinan for undertaking FE analysis. Support from Richard Chester and Alan Baker during the program is also gratefully acknowledged. Significant technical support from Aerostructures Technologies Pty Ltd is also gratefully acknowledged—Bryan Stade and Karmal Gill for designing the attachment and mounting arrangements; Mathew Goldstraw, Greg Rowlinson, Simon Maan and Ron Westcott for undertaking the FE analyses.

REFERENCES

[1] Baker AA. Introduction and overview. In *Advances in the Bonded Composite Repair of Metallic*

Airframe Structures, ISBN 0080429939, Baker AA, Rose LRF, Jones R (eds). Elsevier, 2002; Chapter 1, pp. 1–17.

[2] Baker AA. Certification issues for critical repairs. In *Advances in the Bonded Composite Repair of Metallic Airframe Structures*, ISBN 0080429939, Baker AA, Rose LRF, Jones R (eds). Elsevier, 2002; Chapter 22, pp. 643–656.

[3] Galea SC. Smart patch systems. In *Advances in the Bonded Composite Repair of Metallic Airframe Structures*, ISBN 0080429939, Baker AA, Rose LRF, Jones R (eds). Elsevier, 2002; Chapter 20, pp. 571–612.

[4] *F/A-18 A/B/C/D Trailing Edge Flap and Aileron Hot Spot Analysis—Final Report*, McDonnell Douglas Aerospace Report MDA 96A0138, Revision A, March 1997.

[5] Chester R (ed). *Life Extension of F/A-18 Inboard Aileron Hinges by Shape Optimization and Composite Reinforcement*, Defence Science and Technology Organisation (DSTO), Technical Report DSTO-TR-0699, January 1999.

[6] Galea S, *et al. Smart Patch Flight Demonstrator—System Implementation and Lessons Learned*, Defence Science and Technology Organisation (DSTO), Technical Report DSTO-RR in draft, 2008.

[7] Galea SC, Moss SD, Powlesland IG, Baker AA. Application of a smart patch on an F/A-18 aileron hinge. *Proceedings of ACUN-4 Composite Systems: Macrocomposites, Microcomposites, Nanocomposites*. University of New South Wales, Sydney, 2002.

[8] Roundy S, Wright PK, Rabaey JM. *Energy Scavenging for Wireless Sensor Networks with a Special Focus Vibrations*, ISBN: 1-4020-7663-0. Kluwer Academic Publishers, 2004.

[9] Konak MJ, Powlesland IG, van der Velden SP, Galea SC. A self powered discrete time piezoelectric vibration damper. *Proceedings of the Far East and Pacific Rim Symposium on Smart Materials Structures and MEMS*. SPIE, 1997; Vol. 3241, pp. 270–279.

[10] Sodano G, Inman DJ. A review of power harvesting using piezoelectric materials. *The Shock and Vibration Digest* 2004 **36**(3):197–205.

[11] Eggborn T. *Analytical Models to Predict Power Harvesting with Piezoelectric Materials*, Master's thesis. Virginia Polytechnic Institute and State University, May 2003.

[12] Taylor GW, Burns JR, Kammann SM, Powers WB, Welsh TR. The energy harvesting eel: a small subsurface ocean/river power generator. *IEEE Journal of Oceanic Engineering* 2001 **26**(4):539–547.

[13] Fleurial JP. Thermoelectric energy conversion: future directions and technology development needs. *Indo-US Workshop on Emerging Trends in Energy Technology*. New Delhi, 11–16 March 2007.

[14] Shenck NS, Paradiso JA. Energy scavenging with shoe-mounted piezoelectrics. *IEEE Micro* 2001 **21**(3):30–42.

[15] Starner T. Human powered wearable computing. *IBM Systems Journal* 1996 **35**(3–4):618–629.

[16] Starner T, Paradiso JA. Human-generated power for mobile electronics. In *Low Power Electronics Design*, ISBN-10: 0849319412, Piguet C (ed). CRC Press, 2004, pp. 45-1–45-26.

[17] Shenck NS. *Generation from Piezoceramics in a Shoe*, Master's thesis. MIT, May 1999.

[18] El-hami M, Glynne-Jones P, White NM, Hill M, Beeby S, James E, Brown AD, Ross JN. Design and fabrication of a new vibration based electromechanical power generator. *Sensors and Actuators, A* 2001 **92**:335–342.

[19] Warneke B, Last M, Liebowitz B, Pister KSJ. Smart dust: communicating with a cubic-millimeter computer. *Computer* 2001 **34**(1):44–51.

[20] Beeby SP, Torah RN, Tudor MJ, Glynne-Jones P, O'Donnell T, Saha CR, Roy S. A micro electromagnetic generator for vibration energy harvesting. *Journal of Micromechanics and Microengineering* 2007 **17**:1257–1265.

[21] Armitage RP. *FE Analysis of a Boron Reinforcement on an F/A–18 Aileron Hinge*, Aerostructures Letter Report, Ref. 4-13-12-6. PM1058, May 2002.

[22] Moss SD, Galea SC, Powlesland IG, Konak M, Baker AA. In-situ health monitoring of a bonded composite patch using the strain ratio technique. *SPIE's 2000 Symposium on Smart Materials and MEMS, Smart Structures and Devices Conference*, Paper 4235-41. SPIE: Melbourne, 2000; Vol. 4325.

[23] Rowlinson GR. *FE Analysis of F/A–18 Aileron Hinge (Blueprint Profile) for Active Sensor Design*, Aerostructures Letter Report, Ref. 4-13-12-3. PM661, March 2001.

[24] Zhang H, Galea SC, Chiu WK, Lam YC. An investigation of thin PVDF films as fluctuating strain measuring and damage monitoring devices. *Smart Materials and Structures* 1993 **2**:208–216.

[25] Measurement Specialities (MEAS), *Piezo Film Product Guide and Price List*, 2007, http://www.meas-spec.com.

[26] Diodes Incorporated, *BAS70 Diode Specification Sheet*, 2007, http://www.diodes.com/datasheets/ds11007.pdf.

[27] Diodes Incorporated, *BAT54 Diode Specification Sheet*, 2007, http://www.diodes.com/datasheets/ds11005.pdf.

[28] Engelhardt M. *LTspice/SwitcherCAD III*. Linear Technology Corporation, 2007, http://www.linear.com.

[29] Piezo Systems, *Catalog 7B*, 2007, http://www.piezo.com.

[30] MIL-STD-810E, *Environmental Test Methods and Engineering Guidelines*, July 1989.

[31] *F-18 Vibration, Shock, and Acoustic Noise Design Requirements and Test Procedures for Aircraft Equipment*, McDonald-Douglas Report MDC A2276, March 1978.

[32] Accredited Test Services, *Vibration Testing oF F/A18 Smart Patch Aileron Hinge*, Report TS1167, April 2002.

[33] MIL-STD-461E, *Requirements for the Control of Electromagnetic Interference Characteristics of Subsystems and Equipment*, 1999.

[34] O'Byrne MA. *Electromagnetic Interference Control Requirements*, Boeing Document Number D6-16050-4, July 1991.

[35] Accredited Test Services, *EMC Testing on F/A 18 Smart Patch Aileron Hinge*, Report TS1150, March 2002.

Chapter 78

Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications

Xinlin P. Qing[1], Shawn J. Beard[1], Amrita Kumar[1], Irene Li[1], Mark Lin[2] and Fu-Kuo Chang[2]

[1] *Acellent Technologies, Inc., Sunnyvale, CA, USA*
[2] *Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

A typical structural health monitoring (SHM) system consists of three basic components: (i) actuators and/or sensors, (ii) diagnostic software, and (iii) integrated hardware to monitor the "health state" of in-service structures [1–4]. Many research activities have been done to evaluate new sensor technologies for health monitoring of structures, such as optical fibers [5], piezoelectric materials [6], magnetostrictive materials [7, 8], and microelectromechanical system (MEMS) [9]. Among various types of transducers, piezoelectric materials are being widely used for SHM because they can be used as either actuators or sensors due to their piezoelectric effect and vice versa (*see* **Piezoelectric Wafer Active Sensors**; **Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection**). In the active-sensing mode, both wave propagation and impedance-based SHM methods have been developed [10–15]. In the passive sensing mode, the piezoelectric sensors are used as sensors that continuously monitor external impact events [16]. In the wave propagation-based SHM, guided waves, such

as Lamb and Rayleigh waves, are most widely used for damage detection in metallic and composite structures. Guided waves used for damage detection are introduced into a structure at one point by a piezoelectric actuator and sensed by another piezoelectric sensor at a different position. A key advantage of using piezoelectric elements is that a larger area of the structure can be monitored with fewer transducers, which is vitally important for the monitoring of large-scale structures. Other sensors, like optical fiber-based types, can only scrutinize smaller, specific areas, thus leaving larger areas of a structure unmonitored.

However, it is well understood in the field of SHM that the network of transducers plays a key role in the performance of the SHM system. The ability of sensors and actuators in the network to communicate with each other establishes the intelligence of the system. The type, location, and number of sensors and actuators critically affect the sensitivity of the SHM system.

As the number of sensors increases, the integration of such a network with a structure can be very challenging or become impractical. Obviously, for SHM, the network sensors must be able to

1. integrate/adapt easily within/onto the structure;
2. accommodate any structural configuration;
3. carry a minimal weight; and
4. operate under variable environments.

Stanford multiactuator–receiver transduction (SMART) Layer technology, originally developed by Lin and Chang [17], can overcome these major challenges listed above. In this article, the progress of the SMART Layer technology is summarized in terms of its adaptability for practical applications.

2 SMART LAYER

2.1 SMART Layer design

An important part of the SHM system is the proper integration/adaptation of the sensors with the structure. Sensors permanently mounted onto structures provide the capability to monitor the condition of these structures throughout their service life. The SMART Layer is a unique and cost-effective method for integrating a network of piezoelectric elements with a structure [17, 18]. The layer is made of a thin dielectric film with an embedded network of distributed piezoelectric elements that can be used as either actuators or sensors. The novelty of the SMART Layer lies in its networking capabilities with any type of sensor that enhances its monitoring capabilities and eliminates the need to place each type of sensor individually on the structure. The major features of the SMART Layer include

- actuating and sensing capabilities;
- built-in sensor network for area sensing;
- signal consistency and sensor reliability;
- multiple wires from every transducer to improve reliability of the circuit;
- shielded layer to reduce electromagnetic (EM) noise;
- "hardwired" sensors for direct hardware diagnostic;
- ease of installation.

As shown in Figure 1, the SMART Layer utilizes a layered construction: a circuit layer, an insulation layer, and a sensor layer. In the sensor layer, PZT (lead zirconate titanate) discs with a diameter of 6.35 mm and a thickness of 0.25 mm are most often used, while other PZTs of different shapes and sizes can also be included in the layer as needed. Typical properties of the SMART Layer are given in Table 1.

2.2 Double/multiwire SMART Layer

The SMART Layer can be designed and manufactured in different configurations based on the requirements of the application. To increase the reliability

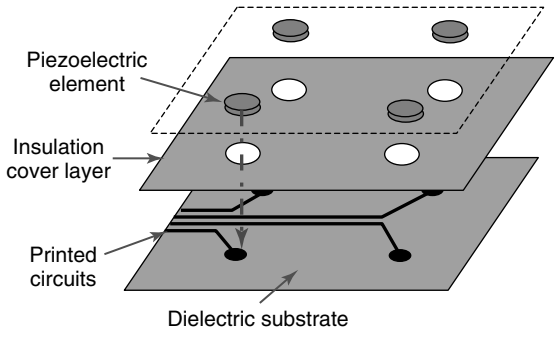


Figure 1. Overview of SMART Layer.

Table 1. Typical properties of SMART Layer

Network sensors	Dielectric material	Thickness of the layer	Temperature range	Environment
PZT-5A disc $D = 6.35$ mm $T = 0.25$ mm	Polyimide	~0.15 mm thick at the area without PZT	−54 to 121 °C for continuous operation	Coated SMART Layer passed salt fog and humidity tests

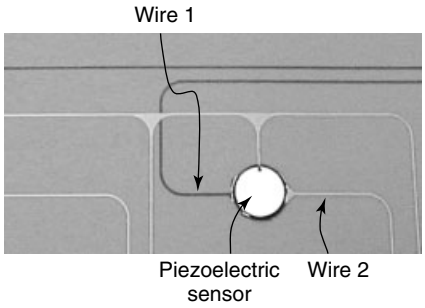


Figure 2. A double/multiwire version of SMART Layer.

of the layer, multiple printed circuitry has been developed for the layer design. With a single-wired layer, a small cut caused by impact or damage in the circuit can break the communication between the sensor and the diagnostic hardware system, rendering it worthless. The method of placing multiple circuits on the layer has been developed so that if one circuit breaks, the other could still be used. Figure 2 shows a picture of the sample of a layer with double wires for each sensor that has been manufactured in this way. Each sensor/actuator is connected via double wires to the final terminals.

This wire redundancy design in the layer eliminates the "fear of breakage of the wires" and makes the layer more reliable. This concept can be extended to any number of wires emanating from a single piezoelectric sensor. The circuit is designed so that the wires emanating from the same piezoelectric sensor are as far away from each other as possible. This design aspect ensures that any damage to one circuit does not damage the other circuit as well.

2.3 Layer for 3D structures

The SMART Layers can be fabricated in different shapes for integrating with different contours of structures. They vary in complexity ranging from a simple flat strip with one or two sensors to a complex 3-D "shell" with more than 50 sensors [18]. For structures with multiple curvatures and complex geometry, SMART Layers can be custom designed with special shapes and cutouts to provide a perfect fit. Two innovative manufacturing methods are described below.

One method of fabricating three-dimensional complex-shaped SMART Layers is to use mechanical locks at preselected locations and then shape the layer based on the required geometry. Upon curing, the layer will hold its shape. A schematic of this concept is presented in Figure 3.

A two-dimensional SMART Layer can also be fabricated through a compressed molding process into a layer with a three-dimensional configuration. A typical fabrication process for a three-dimensional SMART Layer was developed [19] and is briefly outlined as follows (Figure 4):

1. A two-dimensional SMART Layer is first fabricated based on a sensor location map from the 3-D structure.
2. Upon completion of the manufacturing process of the 2-D layer, appropriate cutting, trimming, and scissoring may be applied to conform the layer into the shape of the targeted structure.
3. The 2-D layer is then placed between the molds and subjected to a required processing temperature and pressure.
4. The pressure is maintained during the cooling cycle. After cooling down to room temperature, the layer can be removed from the molds and maintains its 3-D shape.

Figure 5 shows a 3-D layer that was fabricated on the basis of the procedures from items 1 to 4 outlined above, to conform to a composite mold. The 3-D layer was then embedded into a composite frame through

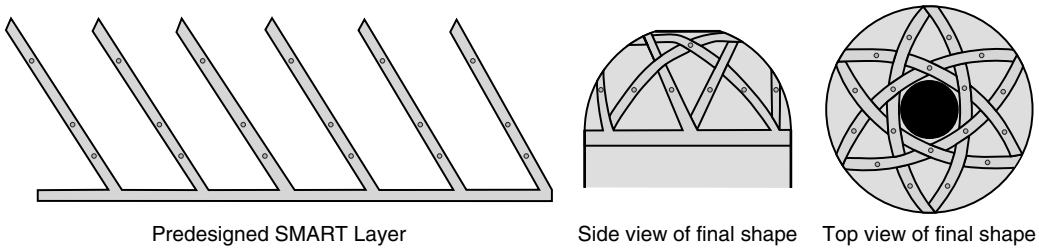


Figure 3. Fabrication method for 3-D SMART Layer.

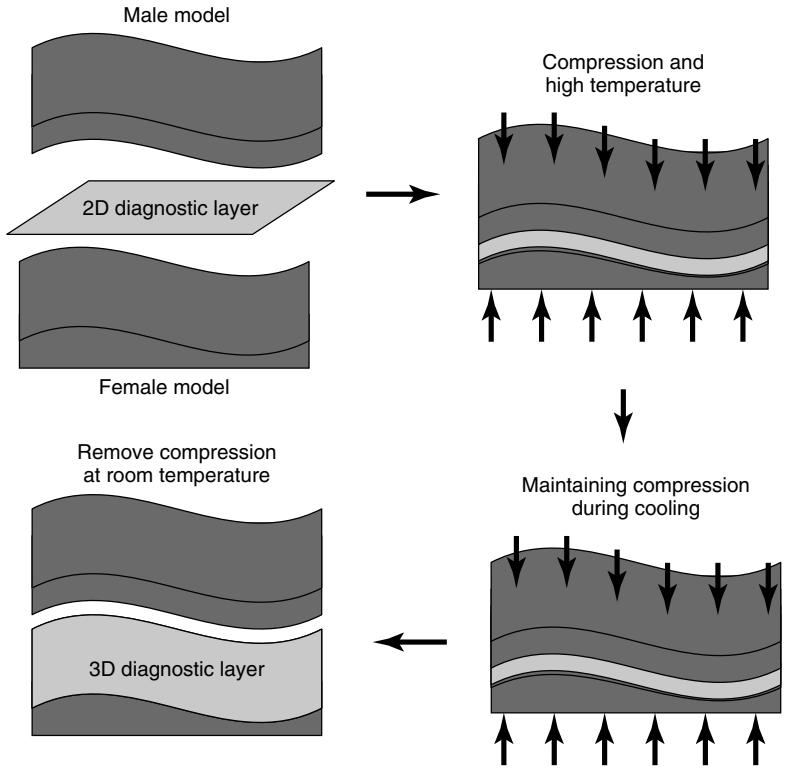


Figure 4. Conversion process of 2-D SMART Layer to 3-D.

a resin transfer molding process, which is described in the next section.

2.4 Hybrid SMART Layer

The SMART Layer can also accommodate other types of transducers along with piezoelectric elements. For instance, techniques have been developed to incorporate optical fibers into the layer. As shown in Figure 6(a), fiber Bragg grating (FBG) sensors are incorporated in the same layer with piezoelectric elements to create a hybrid piezoelectric/fiber-optic SMART Layer. Typical procedures for manufacturing such a diagnostic layer and the physical principles of the hybrid system can be found in the literature [20]. There is another article in the same section of this Encyclopedia that describes the hybrid SHM system (*see* **Hybrid PZT/FBG Sensor System**). Similarly, other types of sensors, such as strain gauges, MEMS, and temperature sensors can also be incorporated in

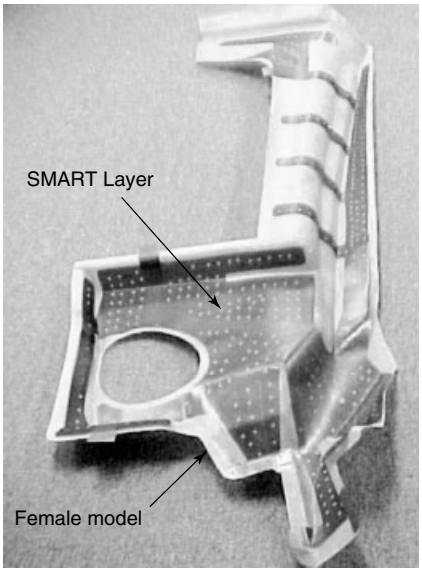


Figure 5. 3-D SMART Layer with 50 piezoelectric elements embedded.

the SMART Layer. Figure 6(b) shows a PZT/strain gauge hybrid layer mounted on a composite panel.

3 INTEGRATION OF SMART LAYER WITH A HOST STRUCTURE

In an SHM system, a sensor network needs to be permanently mounted onto the host structures. As shown in Figure 7, the SMART Layer can either be surface-mounted onto an existing structure or embedded into a composite structure during fabrication.

3.1 Surface Mounting the SMART Layer

As an example of a surface mounting the SMART Layer, the design and installation of the layers on bonded repairs on an F-16 airplane [21] (*see* **Flight Demonstration of a SHM System on a USAF Fighter Airplane**) is briefly described here. The center keel area of the F-16 fuselage station 341 bulkhead is susceptible to fatigue crack growth due to a maintenance-induced initiation site. The bulkhead is a large single-piece machined structure, and replacement of a damaged bulkhead is time consuming and costly. A system of adhesively bonded repairs was developed by Southwest Research Institute and applied to the keel structure as a potential alternative to replacing the bulkhead. To assess the health of the bonded structure as the airplane continues to be used, a health monitoring system was developed to inspect the bonded repairs.

Structure in the keel area of the F-16 station 341 bulkhead is complex, with material interfaces, fastened joints, and system penetrations. Figure 8 shows two views of the area in the main landing gear wheel well of the airplane that will be repaired and monitored. To ensure that the sensors would fit on the airplane and still provide adequate coverage of the repair area, a full-scale paper mock-up of the sensor layer, including positioning the transducers, routing

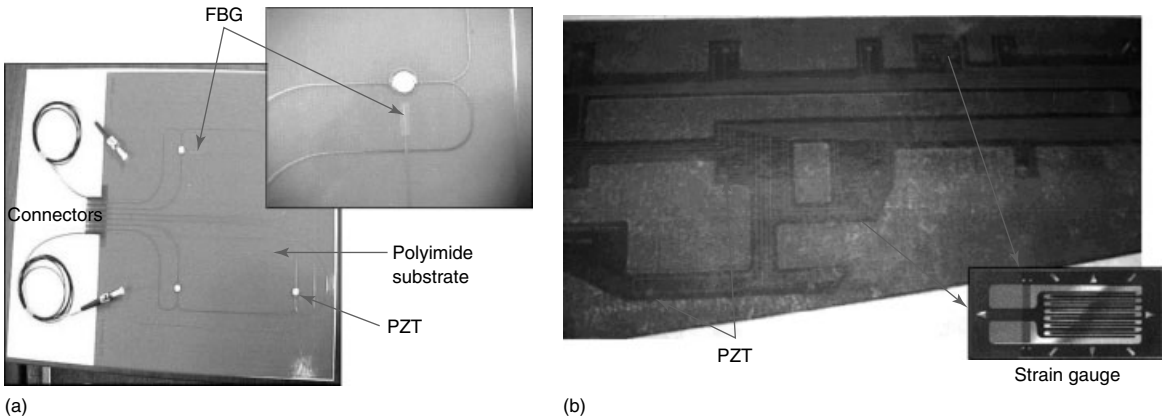


Figure 6. Hybrid SMART Layer.

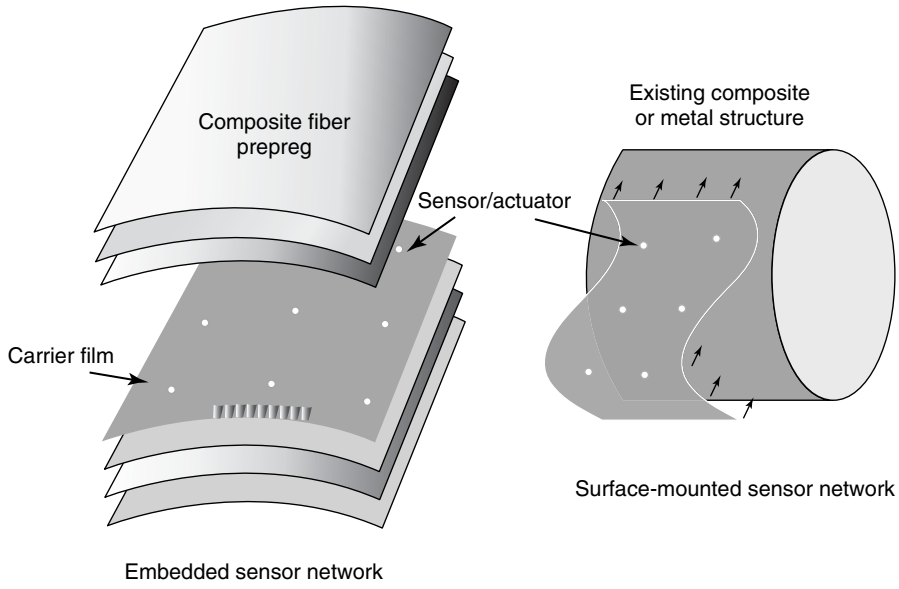


Figure 7. Diagnostic layer integrated with structure.

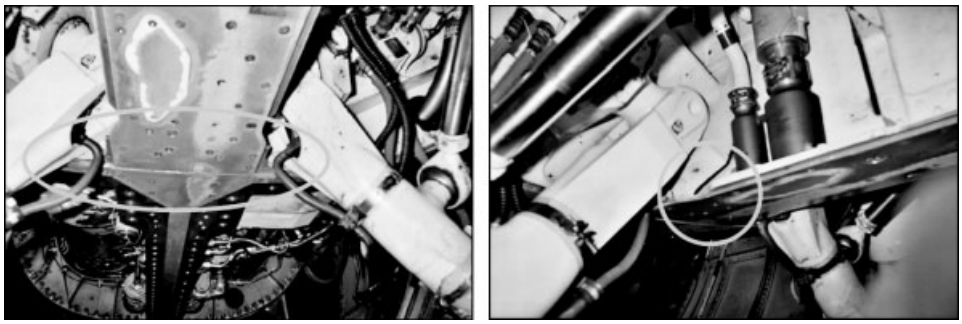


Figure 8. The main landing gear wheel well of an F-16 contains complex structures and multiple system installations [21].

the wiring, and finding an accessible but protected location for the connector, was created by using the F-16 at the US Air Force Museum. A SMART Layer was fabricated to match the paper mock-up. Figure 9 shows the layer prior to installation. The completed system consists of three separate layers and an integral connector.

The installation was performed at Hill Air Force Base in Ogden, Utah. Hysol EA 9394 epoxy adhesive was used to bond the sensor layer on the structure. As shown in Figure 10(b), the layer was adhered to the structure and fixed temporarily with clamps and Kapton tape. After room temperature curing, the tape and clamps were removed, and an overcoat of epoxy adhesive was brushed on the exterior of the sensor layer.

3.2 Embedding SMART Layer within composite

Methods for integrating a SMART Layer into a composite structure during different fabrication processes have been developed, including hand layup of prepregs or wet layup of fiber cloth, the resin transfer molding (RTM) process, and the filament-winding process. For hand layup of prepregs, the layer is simply treated as an additional ply laid

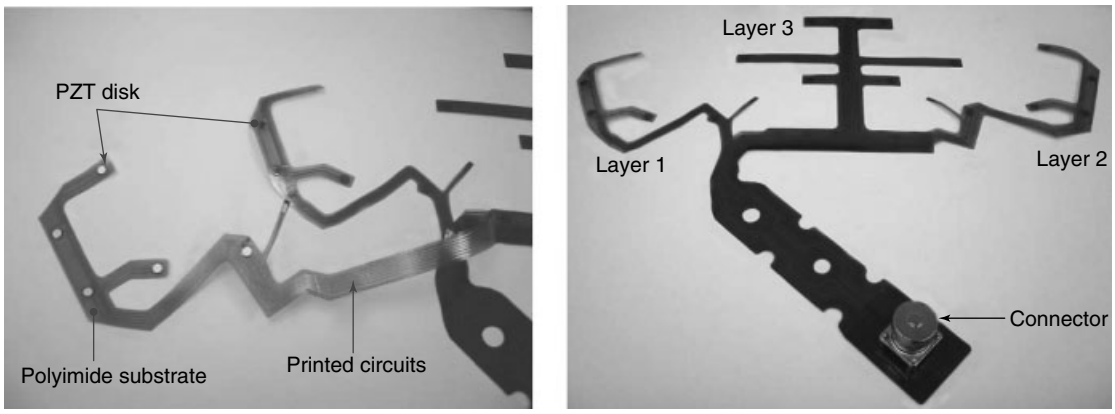


Figure 9. SMART Layer designed for monitoring the bonded repairs of F-16 [21].

Figure 10. The SMART Layers were mounted on the three repairs installed in the keel area of the station 341 bulkhead. (a) In-service F-16, (b) installation, and (c) mounted SMART Layer with surface coating [21].

down during layup. When inserting the layer inside composites, the sensor side preferably should face the thicker side of the laminate to account for the extra thickness of the sensors. For most of the cases, no other adhesive is needed when cocured with composite. Examples for integrating SMART Layers inside composites with RTM process and filament-winding process are given below.

3.2.1 Diagnostic layer integrated into composite during the RTM process

In this section, the study for integrating a sensor network into a three-dimensional composite foam core structure during the RTM process is presented [19]. RTM, which is a popular composite manufacturing process, is a low-pressure, closed molding process, where a mixed resin and catalyst are injected into a closed mold containing a fiber pack or preform. After the resin has cured, the mold can be opened and the finished component can be removed. The inclusion of a SMART Layer inside composite structure during RTM process is shown in Figure 11. The layer is inserted in the mold as an extra preform.

As an example, the 3-D layer shown in Figure 5 was inserted into the interface of the composite and foam core during the RTM manufacturing process. The open sides of piezoceramic disks on the diagnostic layer were oriented toward the composite. The cured composite sandwich structure with sensor network integrated is shown in Figure 12.

3.2.2 Sensor network integrated into filament-wound structure

Filament winding is the process of winding resin-impregnated fiber or tape on a mandrel surface in a precise geometric pattern. This is accomplished by rotating the mandrel while a delivery head precisely positions fibers on the mandrel surface. It is used for the manufacture of parts with high fiber volume

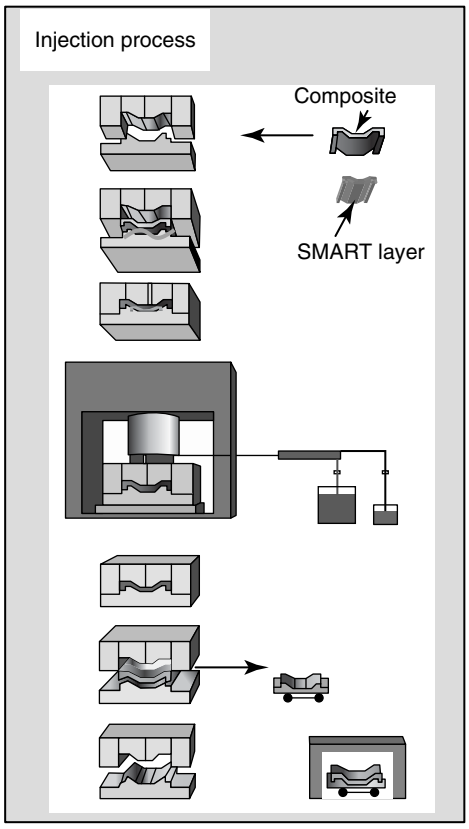


Figure 11. Composite RTM process and inclusion of SMART Layer.

fractions and controlled fiber orientation. To monitor the integrity of filament-wound composite structures, such as solid rocket motors and liquid fuel bottles, a way to embed a sensor network into a filament-wound composite structure was investigated [19, 22].

The composite bottle measures 500 mm in length and 381 mm in diameter. For the composite filament-winding process, the layer was designed in the shape of a thin flat strip with a row of piezoelectric elements. The PZT is 6.35 mm in diameter and 0.25-mm thick. There were five piezoceramic disks on each strip. To reduce EM noise, the connection circuits were shielded on one side by a solid copper foil. As shown in Figure 13, eight strips were used for the composite-wound bottle. This provided a sensor spacing of slightly less than 153 mm in the hoop direction. The result is an approximate square grid of sensors distributed over the bottle. The design of the bottle has five hoop composite layers and two helical composite layers in the cylindrical section, but only two helical layers in the dome sections.

Figure 14 shows some pictures of how the strips of SMART Layer were incorporated into the composite bottle during the filament-winding process. The main process involves winding composite prepreg tows onto an aluminum liner mandrel in either a hoop direction or a helical pattern. First, a hoop layer was wound onto the cylindrical section of the bottle. Four sensor strips were then placed on the top of the hoop layer and held in place by Teflon tape affixed on the dome part. Then two hoop layers were wound over the cylindrical section again. Next, the tapes used to fix the location of the four strips were removed before winding two helical layers over the entire bottle. Then, another hoop layer was wound over the cylindrical section. Next, the four outer sensor strips were placed on the bottle, held in place by Teflon tape. Finally, a hoop layer was wound on the top of the cylindrical section, leaving the outer strips exposed on the dome part. All the open

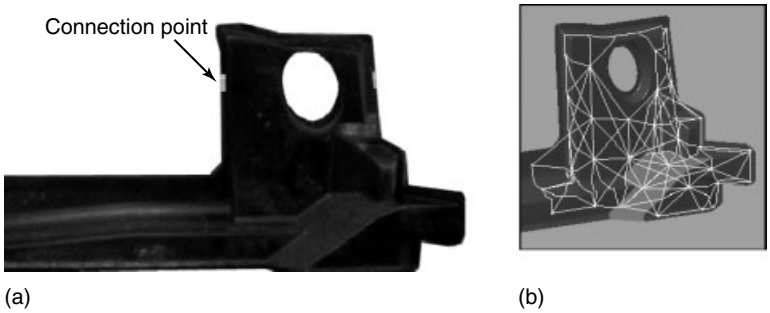


Figure 12. SMART Layer embedded inside a composite sandwich structure in RTM process: (a) photo of the structure and (b) diagram of the sensor network paths.

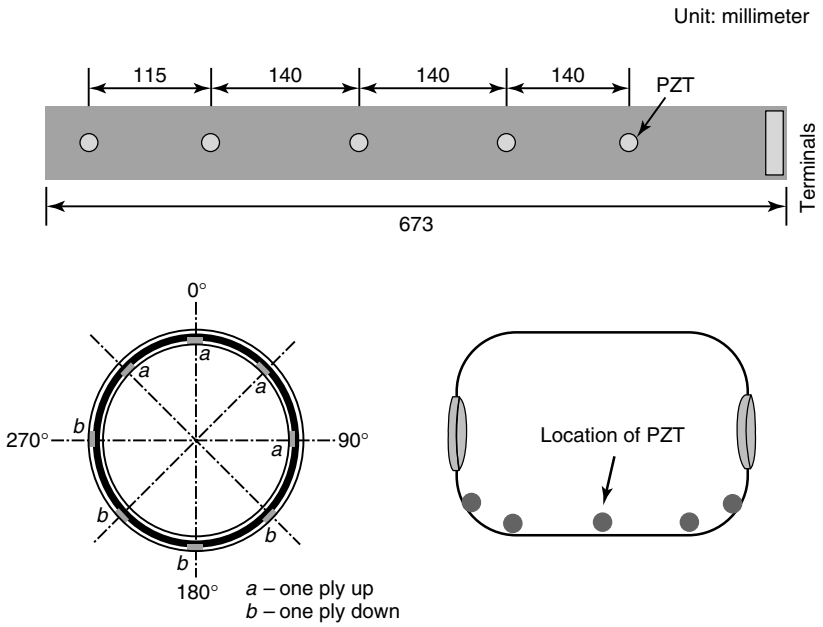


Figure 13. Piezoelectric sensors and their locations in filament-wound bottle.

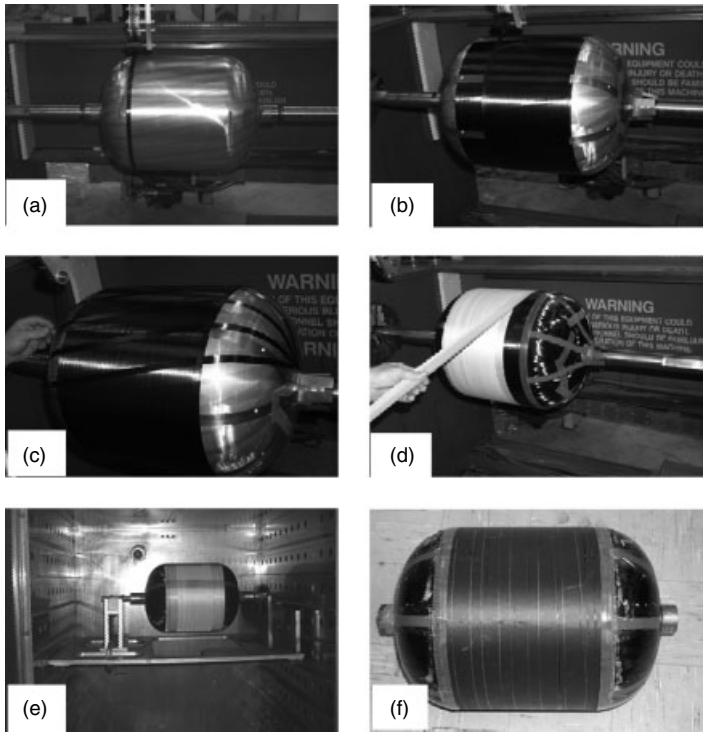


Figure 14. Embedding the sensor network in the filament-wound bottle. (a) The liner, (b) SMART Layers applied after winding of the hoop section, (c) helical winding, (d) preparation of bottle for curing, (e) bottle ready to be cured in the oven, and (f) complete filament-wound bottle with embedded sensor network.

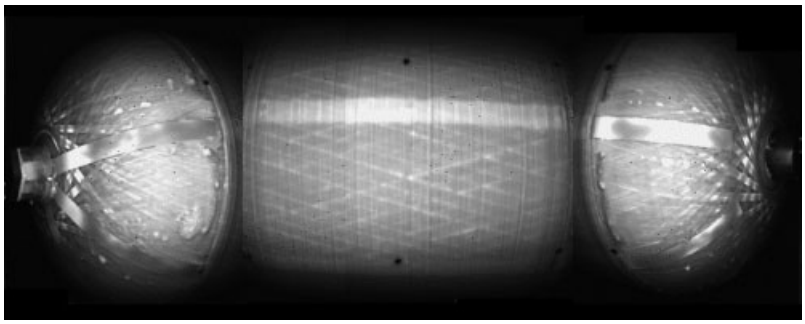


Figure 15. Finished bottle and thermographic image showing embedded SMART Layer.

sides of piezoelectric elements on the strips were oriented toward the thicker side of the composite: the piezoelectric elements on the inner strips face outward and the piezoelectric elements on the outer strips face inward. Since the composite material used to fabricate the bottle is a prepreg, i.e., has both fiber and resin in it, no epoxy resin was transferred into the bottle prior to cure. Since a vacuum bag was not used during curing, Teflon tape was used to push the part of four sensor strips on the outer surface of the dome to stick with the dome surface. The composite bottle was then cured at 177 °C (350 °F) for 2 h in the oven. During curing, the bottle was rotated continuously to obtain an even flow of resin and to eliminate any resin pools.

Figure 15 shows the finished bottle and the corresponding image from a thermographic scan (for thermal imaging techniques, *see* **Thermal Imaging Methods**). The imager used was an Indigo Merlin running under Thermal Wave Imaging software. The heating source was a TWI Flash System. Prior to scanning, the outer surface of the bottle was painted flat black [23]. From the thermographic image, the strips looked well bonded to the composite. This clearly demonstrated the compatibility of the SMART Layer with the filament-winding process typically used for the manufacturing of composite rocket motor cases.

4 SURVIVABILITY AND RELIABILITY OF BUILT-IN SENSOR NETWORK

The reliability of the sensor network for SHM must be entirely guaranteed through the life of structures. But the environments, such as mechanical loading, temperature, and chemicals, can significantly degrade the performances of sensors due to the nature of sensor materials [24] (*see* **Integrated Sensor Durability and Reliability**). A series of tests have been conducted to determine the survivability and functionality of the SMART Layer and its sensors under the different environment conditions. These tests include

- static and fatigue tests;
- temperature tests;
- vibration tests;
- salt fog and humidity tests.

4.1 Characterization of loading effect on the performance of PZT

4.1.1 Mechanical tests on coupon specimens

Tensile tests were conducted on three types of specimens: (i) aluminum specimens with SMART Layer mounted on the surface; (ii) composite specimens with SMART Layer mounted on the surface; and (iii) composite specimens with the layer embedded in the plies [25]. There were two PZT disks on each SMART Layer. For the composite specimens, the layer was cocured one ply down or on the surface of the specimen at 180 °C in an autoclave, while for the aluminum specimen the layer was bonded with an epoxy adhesive film Hysol EA 9696 and cured at 120 °C. During the test, the actuating signal was a five-cycle sinusoidal tone burst with an amplitude of 50 V and a frequency of 50 kHz. The performance of

PZT was studied by the difference between the amplitudes of sensing signals taken at different load histories. The piezoelectric property of PZT was evaluated at intervals of certain strain with and without loading.

Figure 16 shows the results of monotonic tensile tests. As shown in Figure 16(b) and (c), for the PZTs bonded on the composite specimens, the amplitude, η/η_o (η_o is the amplitude of the signal before loading), was almost constant up to the failure strain of PZT, $\varepsilon = 0.1\%$. The difference between the surface-mounted and embedded sensors was negligible except for the region of high strain. However, as shown in Figure 16(a), the amplitude, η/η_o, was almost constant up to about $\varepsilon = 0.3\%$, which is much larger than the failure strain of PZT.

The PZTs mounted on the aluminum specimen could survive much higher tensile strain because of the compressive prestress applied to the PZT during the adhesive mounting process of PZTs [25]. It is clear that the performance of PZT remains unchanged when the applied strain does not exceed the static failure strain of PZT, whereas the degradation of PZT occurs when the applied strain exceeds the static failure strain of PZT.

In addition, it can be concluded that the performance of PZT depends not on the materials of the host structure but on the thermal residual strain induced during the cure cycle. It is important to note that the PZT embedded in the structure can be used without any degradation even if the applied strain to the host structure exceeds the failure strain of PZT by introducing compressive prestress via a specific cure cycle.

The fatigue loading tests were carried out by controlling the strain applied to the specimen on the servo-hydraulic testing machine. The frequency of loading was 10 Hz and the ratio of minimum strain to maximum strain was 0.1. Figure 17 shows the results of fatigue loading tests. For the aluminum specimens, the amplitude, η/η_o, remained unchanged after 10^6 cycles when the maximum strain did not exceed 0.3%. The critical point of the onset of degradation in the fatigue loading tests of aluminum specimens, $\varepsilon = 0.3\%$, agreed with that in the monotonic tensile

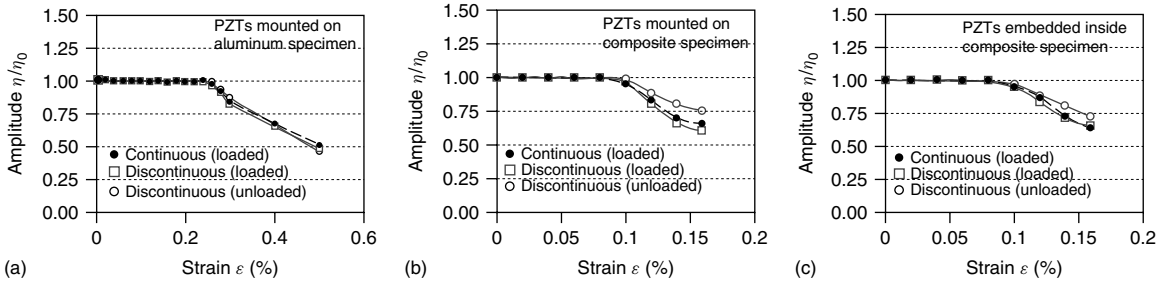


Figure 16. Results of tensile test for (a) aluminum specimen, (b) composite specimen with surface-mounted layer, and (c) composite with embedded layer.

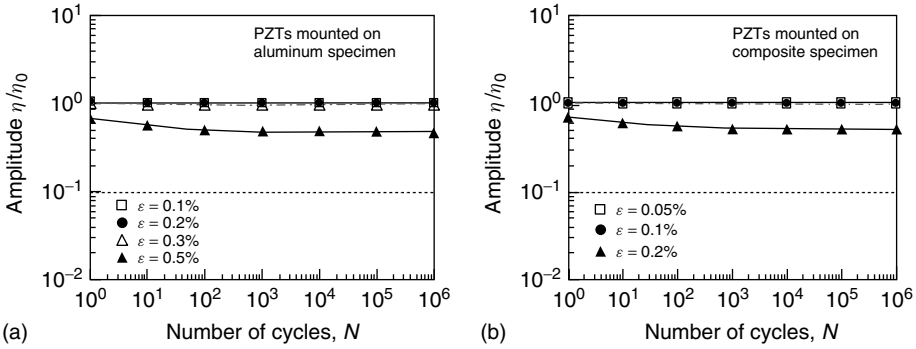


Figure 17. Effect of fatigue loading on the performance of PZT: (a) aluminum and (b) composite.

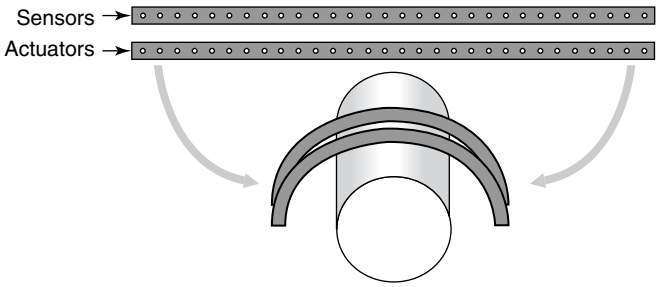


Figure 18. Thirty actuators and 30 sensors on two SMART Layers.

tests. For composite specimens, the amplitude, η/η_o, was almost constant for maximum $\varepsilon < 0.1\%$ when increasing the number of cycles up to 10^6. From the results, it can be concluded that the performance of PZT is not degraded under fatigue loading when the applied strain does not exceed the static failure strain of PZT, whereas the degradation of PZT is considerably stable after the first or several cycles of loading when the applied strain exceeds the static failure strain of PZT.

4.1.2 Fatigue test on a Thunder Horse steel pipe sample

The purpose of the test was to demonstrate the ability of SMART technology to detect and monitor the growth of a fatigue crack at the girth weld of a steel pipe used in offshore industry, and also to investigate the reliability and survivability of PZTs under fatigue load for real-world applications.

Test article

A Thunder Horse steel pipe sample was used in the test. The pipe, consisted of two 3-m-long segments joined together with a girth weld, has a 324-mm outer diameter with a wall thickness of 40 mm. Two 30-sensor SMART Layer strips were mounted on either side of the girth weld as shown in Figures 18 and 19. The sensors used in the strips were PZTs, each having a diameter of 6.35 mm and a thickness of 0.76 mm.

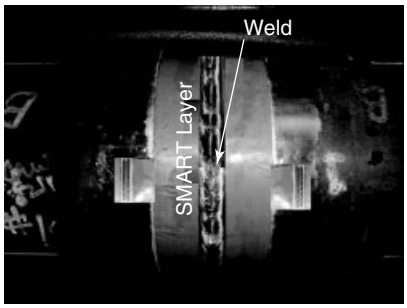


Figure 19. SMART Layers were mounted right up against the cap of the girth weld.

Test procedures and results

The testing was conducted with 2.8 MPa (400 psi) internal pressure and the fatigue cycling ran at ±206 MPa (30 ksi) at 27 Hz. For this test, the 30 piezos on one strip were used as actuators while the 30 piezos on the other strip were used as sensors. Each actuator, in turn, was excited with a five-peak modulated sine wave burst. The excitation generates stress waves that propagate through the structure and these are recorded by the sensor directly across from the actuator. Data was collected at 25 different intervals during the test. The test was stopped at about 12 300 000 cycles because a through-wall crack was observed in the pipe. All sensors were functioning when the tests were stopped.

The effect of cycling electric loading on the performance of PZTs was also investigated. The maximum voltage, which a 0.25-mm PZT can survive, is 70 VAC. Both surface-mounted PZTs and embedded PZTs inside the composite were tested. In each cycle, five-peak modulated sine wave bursts with maximum amplitude of 50 V were input to the PZT actuators at 10 different frequencies (50, 100, 150, 200, 250, 300, 350, 400, 450, and 500 kHz), respectively. The amplitude of sensor response to the excitation generated by the actuator is employed to evaluate the performance of the actuator. Test results that showed sensor signals for their actuator–sensor paths remained unchanged after 10^7 cycles for all PZTs on both specimens.

4.2 Effect of temperature and vibration

4.2.1 Temperature effect

Owing to the temperature dependence of material properties of the PZT, adhesive, and host structure, the sensor signals responding to the same voltage input to the actuator are temperature dependent. Figure 20 shows the temperature dependence of piezoelectric coefficients of PZT. It is expected that the sensor signal will change when the temperature changes. Tests on both metal and composite structures were conducted to investigate the temperature effect on the performance of SMART Layer. The SMART Layers were mounted on the structures with Hysol EA 9396. Hysol EA 9396 is a low-viscosity room temperature curing adhesive system with excellent strength properties at temperatures from −55 to 177 °C (−67 to 350 °F). Before mounting the SMART Layers on the structures, the surfaces of the structures were well prepared. The adhesive was cured at room temperature, followed by a post-curing of 1 h at 93 °C (200 °F).

Sensor signals from the SMART Layers on both metal and composite structures were recorded at different temperatures and cycles within the temperature range from −51 to 93 °C (−60 to 200 °F). Typical sensor signals from the SMART Layers on an Alloy 718 duct are shown in Figure 21. On the basis of the results, it is clear that the sensor signals for all frequencies tested here are repeatable at the same temperature after several temperature cycles. However, both amplitude and phase of the signals are different at different temperatures. According to the signals recorded from the SMART Layers on the metal structures, it can be seen that the amplitude of the signals slightly changed when the temperature changed. Also the phase of the signals shifted. However, the amplitude of signals from the layer on the composite panels significantly reduced when the temperature increased from −51 to 93 °C (−60 to 200 °F), depending on the resin used in the

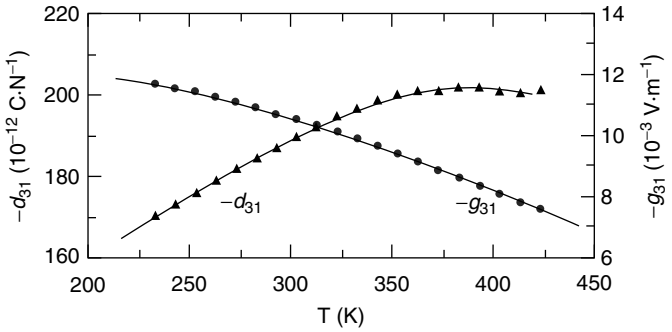


Figure 20. Temperature dependence of piezoelectric coefficients (d_{31}, g_{31}) of PZT (APC 850).

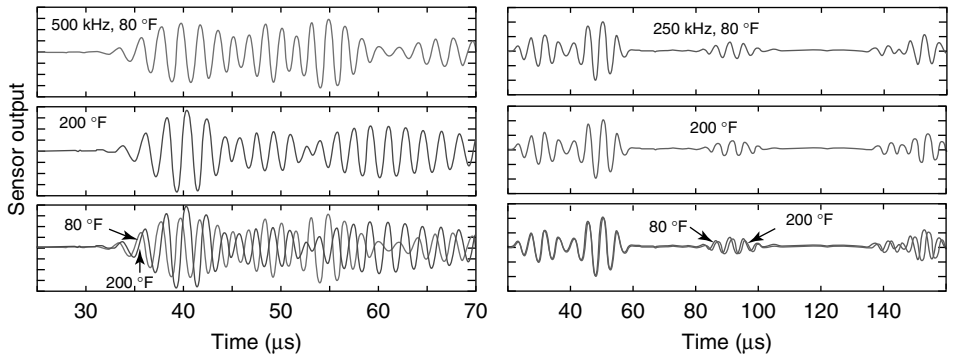


Figure 21. Typical sensor signals recorded at different temperatures.

composites. It can be concluded that the SMART Layer works well within the temperature range from −51 to 93 °C if the host structure can survive the environment [26].

For some specific applications, such as monitoring crack growth in a liquid rocket engine pipe, the sensor network must be able to survive a much harsher environment. Acellent has conducted the tests to determine the ability of the current SMART Layer to survive the extreme cold temperatures and high pressures of rocket engines. The test procedures and results are described in the following section [27]. From the successful application of composite manufacturing process monitoring, it demonstrated that the SMART Layer indeed survive high temperatures up to 200 °C [18].

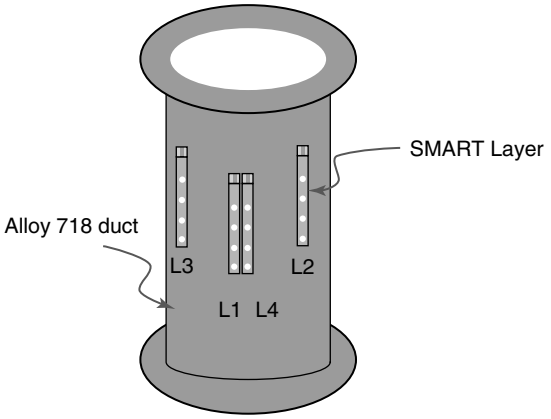


Figure 22. Schematic of the test specimen with SMART Tapes mounted.

4.3 Performance of SMART Layer under combined cryogenic temperature and vibration environment

This section describes how the physical robustness of the SMART Layer as well as operational survivability and functionality were verified with a duct simulator, conditioned to flight vibration and shock environments on a simulated large booster LOX-H_2 engine [27].

4.3.1 Test specimen and setup

Two layers (L1 and L3) with four piezoelectric single crystals on each were fabricated and mounted on the surface of an Alloy 718 duct with a diameter of 152 mm and length of 406 mm, as shown in Figure 22. The other two layers (L2 and L4) with PZTs were also manufactured and mounted on the duct to compare the performance of different piezoelectric materials. The low-temperature adhesive EP29LPSP was used to bond the layers on the Alloy 718 duct.

As shown in Figure 23, the duct with flanges was bolted between two aluminum "bookend" fixtures on the shaker machine. Liquid nitrogen (LN_2) flowed into one flange, through the duct, and out of the other flange. The vibration was controlled via an arithmetic average between accelerometers on each bookend. The test environments were derived from operational measurements on large booster LOX-H_2 engines during engine hot-fire testing. The defined random and sine environments represent an average of the data from the large propellant duct with the highest overall vibration level.

4.3.2 Test procedure

The duct assembly was installed in the test setup and a functional checkout of the sensors was conducted to verify their pretest operational condition. Test instrumentation was installed, including the triaxial response accelerometer and two temperature-monitoring type-K thermocouples. After completion of the flat random characterization run and instrumentation checks, the duct assembly was chilled down and attempt made to perform the same characterization at cryogenic temperatures. Following the completion of 30 min at 0 dB for the cryogenic test, four shock pulses were applied to the test setup at 0 dB.

Pitch–catch signals for both types of piezoelectric sensors were taken at ambient and cryogenic temperatures prior to applying the dynamic environments, during full-level vibration testing, and after completion of full-level testing.

4.3.3 Measurement results

Figure 24 shows some typical sensor signals at different stages. The sensor signals for piezoelectric sensors after the full-level test are the same as the signals before the test. It demonstrated that the piezoelectric sensors could withstand the operational

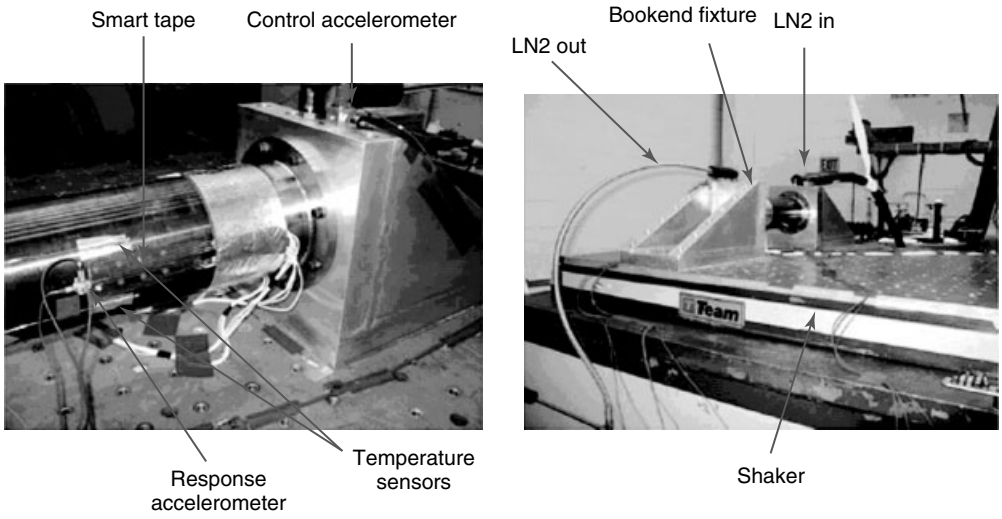


Figure 23. Setup for the combined cryogenic temperature and vibration test.

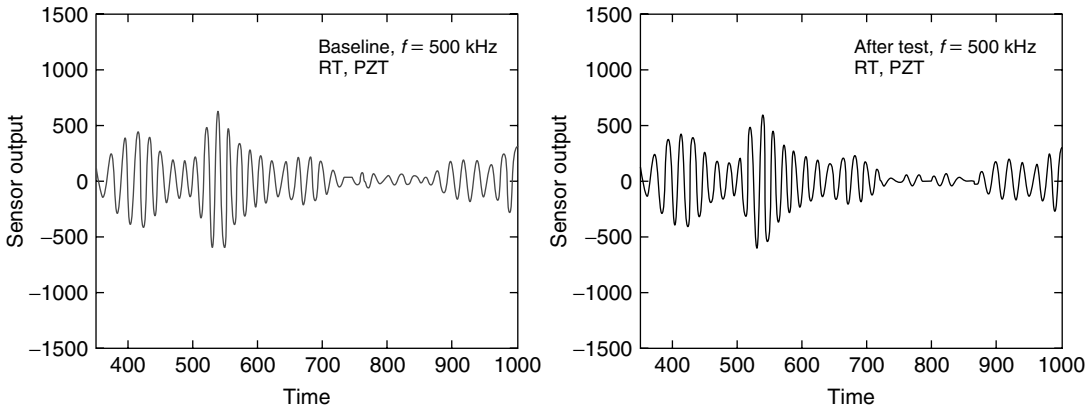


Figure 24. Comparison of 500-kHz signals from PZTs at room temperature before and after the full-level test.

levels of vibratory and shock energy on a representative rocket engine duct assembly within a restricted frequency band in laboratory testing.

4.4 Effect of moisture and aggressive chemical environment

Generally, the piezoelectric materials and polyimide used in the SMART Layer do not have a very good durability in high humidity or aggressive chemical environments. Protective coatings or encapsulation is recommended if the layer is used in high humidity or aggressive chemical environments. To evaluate the effectiveness and quality of protective coatings for the SMART Layer, both humidity testing and salt fog testing were conducted in a humidity chamber and a salt fog chamber per MIL-STD-810F, respectively. The coating tested is an epoxy primer per MIL-PRF-23377, Class C, which is then top coated per MIL-C-46168, Class H. The thickness of the epoxy primer is 15–25 μm, while the thickness of the top coating is 45–75 μm.

The comparisons of signals for the specimens before and after the salt fog test are shown in Figure 25. Test results showed that the SMART

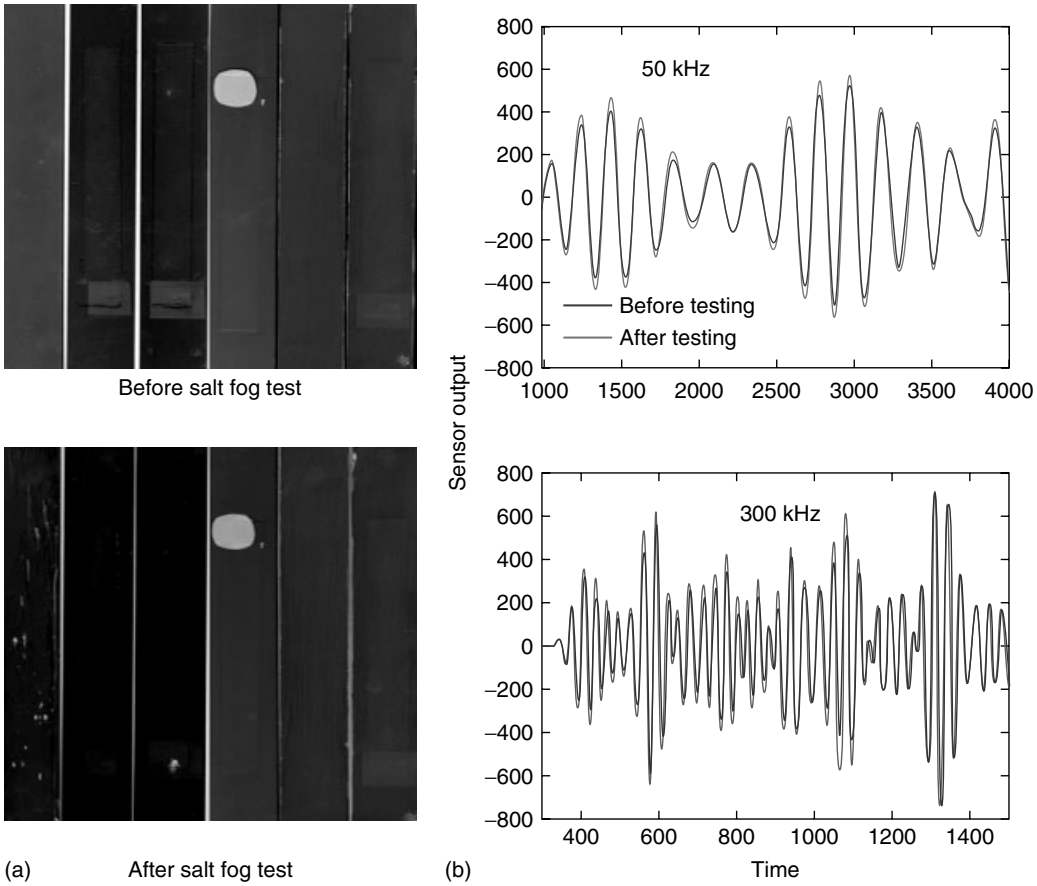


Figure 25. Salt fog test results: (a) aluminum specimens with protective coatings and (b) sensor signals.

Layers can survive humidity and salt fog exposure with proper surface treatment (sanding or chemical treatment), bonding the layers on the surface with Hysol EA 9396, and use proper coatings (epoxy primer per MIL-PRF-23377, class C, which is then top coated per MIL-C-46168, class H). Note that there is no protective coating necessary if the layer is embedded inside a composite.

5 EFFECT OF SMART LAYER ON STRUCTURAL INTEGRITY

When the SMART Layer is embedded inside a composite structure, the effect of the layer on the structural integrity is a concern. It is important to investigate the performance of the composite structure with embedded SMART Layer. Mechanical tests on composite coupon specimens with and without embedded polyimide layer were conducted to assess the change in structural integrity due to inclusion of the SMART Layer.

5.1 Quasi-static impact test

To determine the effects of embedding a SMART Layer into a composite, measurements of the damage tolerance of composites to transverse impact are needed. Quasi-static impact tests were used to simulate a low velocity impact and create a delamination. The tests were performed on both woven graphite/epoxy composites and toughened prepreg carbon–fiber/epoxy composites.

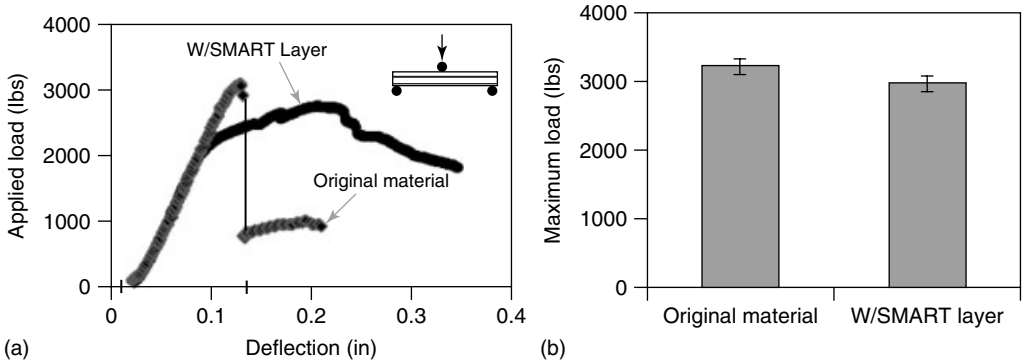


Figure 26. Mechanical test results of specimens with and without embedded SMART Layer (1 lb = 0.45 kg, 1 in. = 25.4 mm). (a) Load–deflection curve and (b) strength comparison.

5.1.1 Woven graphite/epoxy composite

The woven graphite/epoxy coupon specimens tested have a $[0_4/90_4/0_4]$ stacking sequence, specifically designed to promote delamination at the two ply-group interfaces. The specimens measured $140 \times 76 \times 5.2\,\text{mm}^3$. A SMART Layer was placed at the lower 0/90 interface. The thickness of the layer was 0.15 mm, while the thickness of the coupon specimens was 5.2 mm. For testing, the specimens were simply supported on two sides by steel rods spaced 90.0 mm apart and loaded in the center by a 12.7-mm diameter spherical indenter. The specimens were loaded at a displacement rate of $0.127\,\text{mm}\,\text{min}^{-1}$ ($0.05\,\text{in.}\,\text{min}^{-1}$). Test results on these specimens are presented in Figure 26. The test results indicate that the presence of the SMART Layer does not noticeably affect the strength of the host composite structure.

An examination of the cross section of the specimens corroborates these findings. Magnified views of the cross sections of the specimens are shown in Figure 27. It is clear that delamination in the specimens without the embedded SMART Layer occurs at the lower 0/90 interface, as expected, because of the high interfacial shear stress at the ply-group interface. However, in the specimens with an embedded SMART Layer at the lower 0/90 interface, there is no delamination. The actual delamination occurs one or two plies away from the interface, indicating that the SMART Layer does not promote delamination.

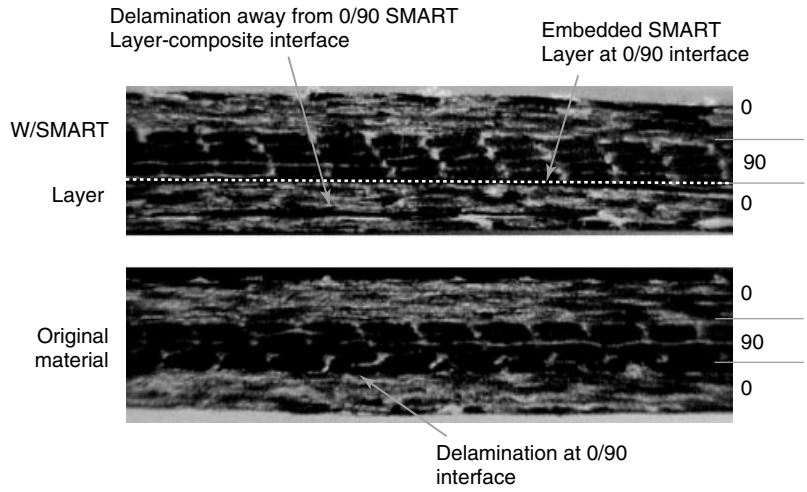


Figure 27. Cross-sectional view of test specimens.

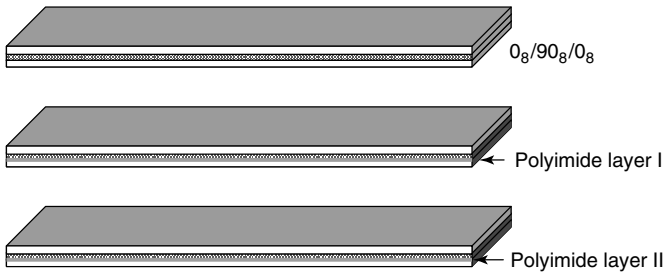


Figure 28. EX-1522 composite coupon specimens with or without polyimide layer inserted at the 0/90 interface.

5.1.2 Toughened carbon–fiber/epoxy composite

The toughened carbon–fiber/epoxy (EX-1522) composite coupon specimens used for quasi-static impact test have a layup $[0_8/90_8/0_8]$. For the embedded specimens, a 0.076-mm-thick polyimide layer I or a 0.152-mm-thick polyimide layer II is inserted at the interface to determine the effect of the thickness of SMART Layer on delamination growth, as shown in Figure 28. The specimens measured $130.0 \times 24 \times 4.0\,\text{mm}^3$. Similar to the woven graphite/epoxy composite, the specimens were supported on two sides by steel rods spaced 75.0 mm apart and loaded in the center by a 12.7-mm diameter spherical indenter. The specimens were loaded at a displacement rate of $01.27\,\text{mm}\,\text{min}^{-1}$. Test results on these specimens are presented in Figure 29.

The effect of embedding a SMART Layer on the damage tolerance of T300 and T800 composites to transverse impact was also studied by Lin and Chang [17]. Typical X-ray photographs for the tested specimens with $[0_8/90_8/0_8]$ stacking sequence were shown in Figure 30. As indicated by the X-ray photographs, the delamination area is actually smaller for the specimens with an embedded polyimide bondply; except in the case of the Toray T800H/3900–2 material, which is a resin system with a very high toughness, that it showed a slightly larger delamination. The improvement in impact resistance for most materials is attributed to the fact that the ductile polyimide layer toughens the interface and reduces matrix cracking, thus suppresses delamination.

5.2 Short beam shear test

Short beam shear tests on toughened carbon–fiber/epoxy (EX-1522) composite coupon specimens with and without polyimide layer were conducted. As shown in Figure 31, three types of specimens with unidirectional layup were tested: regular specimen

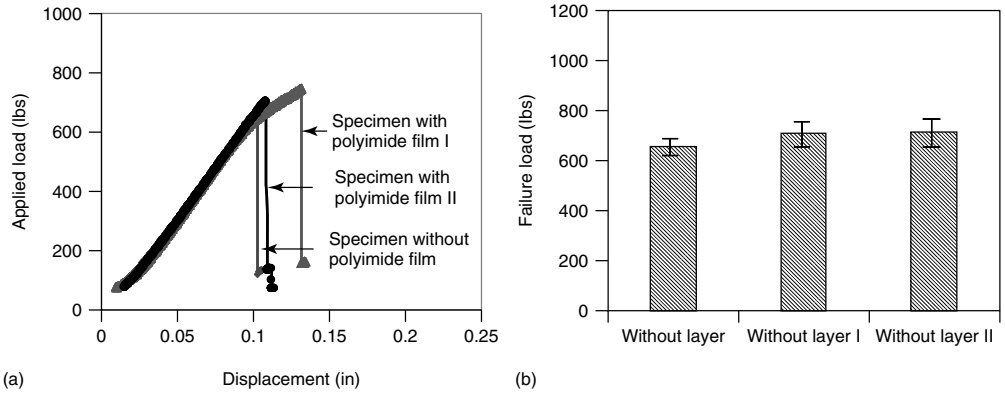


Figure 29. Three-point bending test results of layup $[0_8/90_8/0_8]$ with and without polyimide layer embedded at 0/90 interface. (a) Typical load–deflection curve and (b) strength comparison.

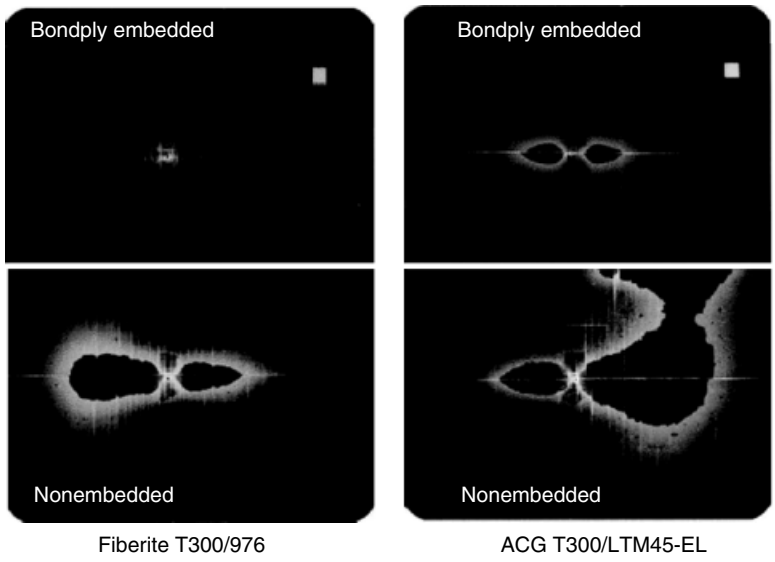


Figure 30. X-ray photographs of delaminations created by the impact test.

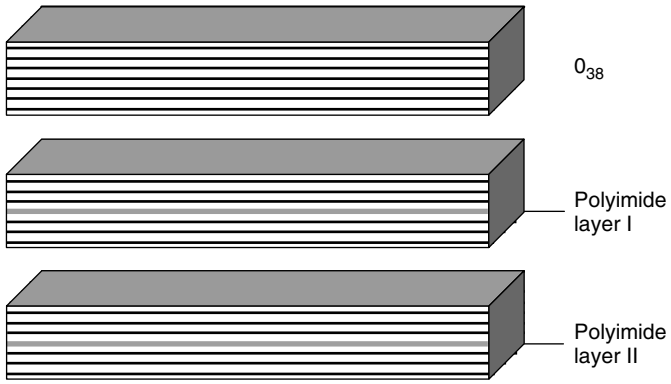


Figure 31. Short beam shear test specimen configurations. Polyimide layer was inserted in the middle plane of some specimens.

without polyimide embedded (Beam 1), specimen with 0.076-mm-thick polyimide layer I embedded (Beam 2), and specimen with 0.152-mm-thick polyimide layer II embedded in the middle plane (Beam 3). The specimens with a layup $[0_{38}]$ measured $40 \times 6.35 \times 6.35\,\text{mm}^3$. The specimens were loaded under three-point bending with a 25.4-mm span. A loading rate of $0.127\,\text{mm}\,\text{min}^{-1}$ was applied. Test results on these specimens are presented in Figure 32.

As shown in Figure 32, there is not much difference between the short beam shear strength of composite with and without polyimide layer. All types of specimens failed in a similar fashion—multiple simultaneous shear cracks between plies, i.e., delamination.

Once again, the test results indicate that the presence of the SMART Layer does neither noticeably affect the strength of the host composite structure nor promote delamination. Besides the studies mentioned above, more investigations on the change in structural integrity due to inclusion of the layer can be found in the literature [17, 28–30]. Results showed that the embedded layer would not decrease either the stiffness or strength of a composite structure.

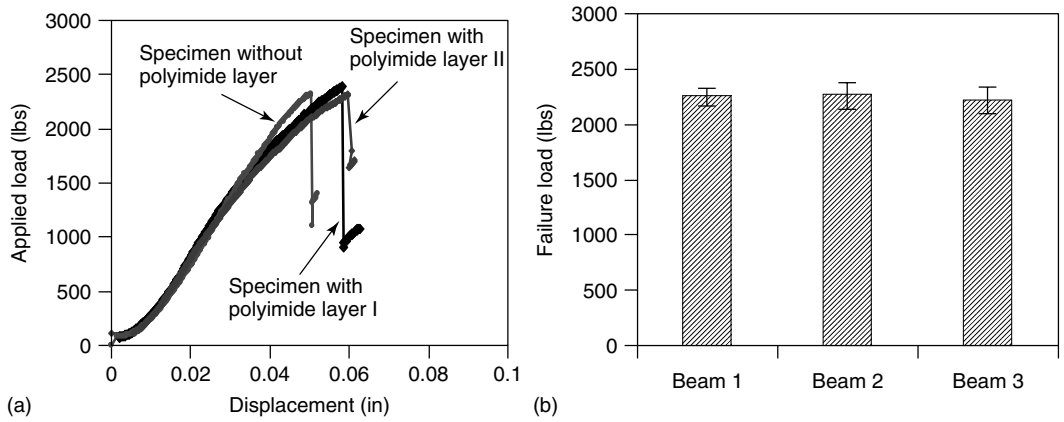


Figure 32. Short beam shear test results of unidirectional layup specimens with and without polyimide layer embedded in the middle plane. (a) Typical load–deflection curve and (b) strength comparison.

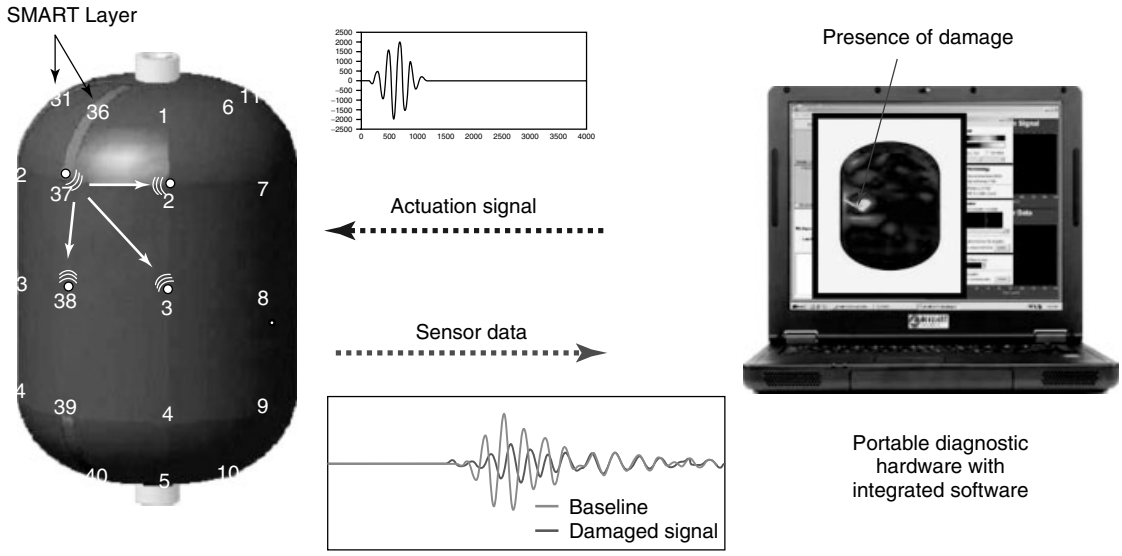


Figure 33. Diagram of Acellent's active-sensing system.

6 SENSOR NETWORK-BASED STRUCTURAL HEALTH MONITORING SYSTEMS

The SMART Layer integrated with a structure can be used for both active and passive sensing. The functioning of the active-sensing system is analogous to that of a built-in acousto-ultrasonic nondestructive evaluation (NDE) with a network of miniaturized piezoelectric transducers. The signal generator of the diagnostic hardware produces a diagnostic signal capable of propagating in structures with minimum distortion while being sensitive to any induced damage. Figure 33 shows the principles of Acellent's portable SHM system. The signal wave generated by the diagnostic hardware drives the actuator to produce a stress wave that propagates across the structure and is measured by the sensors.

When a propagating stress wave encounters a discontinuity in the geometry or material property of the structure, the wave is reflected or scattered.

The methodology used in the diagnostic process is based on the comparison of the current sensor responses with previously recorded sensor responses (baselines) from the undamaged structure. The differences between the two sets of signals are what contain the information about any existing damage or other anomalies. Some applications of the active sensing system can be found in the literature [31–34].

The passive sensing can be used in real time to detect impact events including both impact location and energy. The passive sensing system is set up to continuously listen for impacts at all times. When a sufficiently large force strikes the structure, the piezoceramic sensors that are bonded onto the structure pick up the stress waves traveling through the surface of the structure. A trigger mechanism enables software control to determine the trigger condition, which could be related to a specific set of features of the sensor measurements. As an example, to monitor the safety of a thermal protection system (TPS), a built-in passive sensing system is being developed for detecting the impact location and impact force in real time on TPS panels supplied by Lockheed Martin Space Systems [35].

7 DAMAGE DETECTION IN COMPOSITE STRUCTURES

A SmartComposite system based on SMART Layer technology has been developed to handle some practical issues for the application of SHM on composite structures in real world [36]. Regardless of the application, there are a number of elements that are essential to the practical usage and implementation of any SHM system. These elements can be grouped into three main categories: (i) the system must be easy to use, (ii) it must provide a well-defined resolution, and (iii) it must be very reliable. The essential elements for each category are listed below:

- Easy to use
  - the sensor installation should be straightforward with minimal training;
  - the system should have simple calibration procedures—preferably automated;
  - all data analysis shall be automated;
  - results should be output in standard formats.
- Well-defined resolution
  - the system must provide a quantifiable probability of detection (POD);
  - when damage is detected, the system must indicate size/severity along with the measure of uncertainty.
- High reliability
  - the system must be able to compensate for environmental changes;
  - the system must have built-in test for hardware and sensor self-diagnostics;
  - the sensors must survive extreme environmental conditions;
  - in the event of damage to a sensor, the system must be repairable.

Each element listed above has, by itself, been studied by many researchers and developers for a variety of applications. But there has been little work conducted to combine all the essential elements into a common framework. On the basis of the SMART Layer technology, Acellent has addressed this issue by focusing on all of the above elements, customizing each one for composite structures, and then unifying all of them into an integrated system called *the SmartComposite System*.

The key features of the system include sensor self-diagnostics and an adaptive algorithm to automatically compensate for damaged sensors, reliable damage detection under different environmental conditions, and generation of POD curves. In addition, state-of-the-art techniques to optimize sensor placement, automated calibration, and automated damage detection with no user interpretation of data make it an efficient and user-friendly system. A few of the features are discussed in the following sections.

7.1 Self-diagnostics

If one or more sensors are degraded, damaged, or missing, the SHM system may not function properly and can give false indications of structural damage. Measuring the impedance of each channel can be used to find an open or short circuit. This can indicate a missing sensor or damaged connection/wiring.

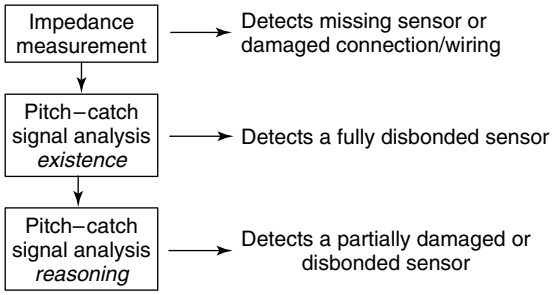


Figure 34. Integrated three-step method to automatically detect faulty sensors.

But a degraded or damaged sensor may go undetected using the impedance method. To resolve this, a reasoning process using the active sensor signals has been developed and implemented to detect degraded or damaged sensors that the impedance method may miss.

An integrated three-step method shown in Figure 34 has been developed to automatically detect

- faulty sensors caused by a missing sensor or damaged connection/wiring;
- a sensor that is still connected to the electronics, but is disbonded from the structure; and
- a partially damaged or disbonded sensor.

The impedance measurement of each channel can be used to find an open or short circuit. If a sensor is flagged as degraded, damaged, or missing, all signal data from the faulty sensor are removed from the analysis routines.

7.2 Temperature compensation

Current state-of-the-art damage detection methodologies rely on the use of baseline data collected from the structure in the undamaged state. The methodologies are based on comparing the current sensor responses to the previously recorded baseline sensor responses, and using the differences to glean information about structural damage. However, it is known that environmental effects, such as temperature differences, will also cause changes in the sensor signals, and will thus interfere with most damage detection schemes.

Acellent has developed a calibration technique utilizing multiple baselines that can be employed to mitigate the effects of environmental changes. The technique has been tested and verified for both global and local changes in temperature and can be used to compensate for global temperature changes in the structure as well as temperature gradients [36].

With this method, data is collected from the healthy structure at various temperatures and stored in a so-called baseline space. The baseline signals from each temperature can be used to create a baseline surface for each actuator–sensor path. At a later time, when a sensor scan is performed to search for damage, the newly recorded signals are compared with the

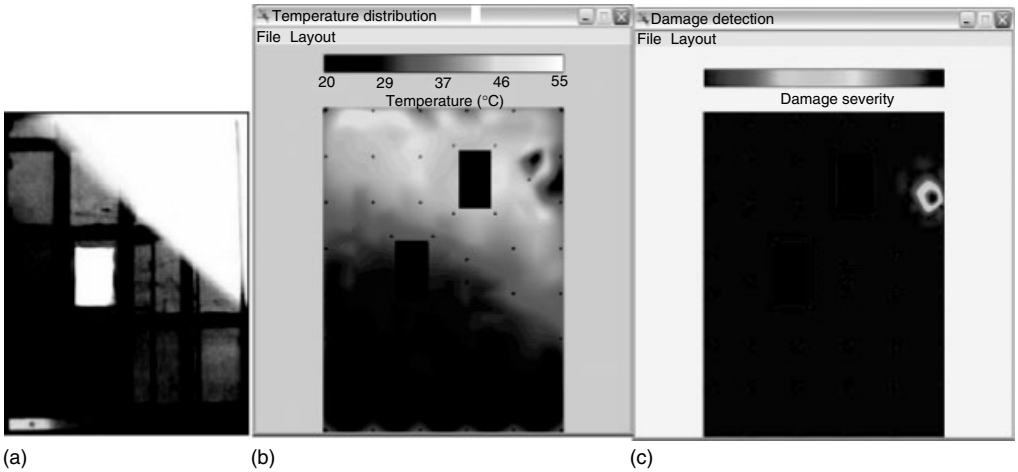


Figure 35. Environment compensation to eliminate temperature effect on a composite panel. (a) Structure in sun, (b) temperature distribution, and (c) detected damage.

corresponding baseline surfaces to determine a best fit along the temperature axis. An example of the temperature compensation is shown in Figure 35. The specimen shown in Figure 35 is a 1.5 m × 1.5 m fiberglass stiffener panel with surface-mounted SMART Layers containing a total of 50 actuators/sensors.

8 CONCLUSION

SHM offers the promise of a paradigm shift from schedule-driven maintenance to condition-based maintenance of assets. SMART Layer technology is a viable and cost-effective means of monitoring the structural condition and detecting damage while structures are in service. The techniques for fabricating the diagnostic layers and the methods for integrating the layer into structures are developed. The performance of the piezoelectric elements embedded in the SMART Layer under different loading and environment conditions was investigated. Other practical issues, such as the self-diagnostics of the embedded network and environmental compensation, are also discussed. On the basis of the study, the following remarks can be made:

1. Using SMART Layer, large sensor (and actuator) networks can be easily integrated with metal/composite structures. The characteristic features of the diagnostic layer include (i) ease of installation; (ii) adaptation to any structure with complex geometry; (iii) use of an area sensing network; (iv) actuation and sensing capabilities; (v) signal consistency and sensor reliability; and (vi) shielding to reduce EM noise.
2. The performance of PZT remains unchanged when the applied strain does not exceed the failure strain of PZT and the electric cycling input is below the maximum voltage.
3. The SMART Layer is functional within a wide temperature range. Proper protective coatings can effectively isolate the layer from the moisture and aggressive chemical environment when the layer is mounted on the surface of a structure.
4. Test results showed that the presence of the SMART Layer does neither noticeably affect the strength of the host composite structure nor promote delamination.
5. A SmartComposite system based on SMART Layer technology has been developed to handle some practical issues for the application of SHM in real world. The major features of the system include sensor network self-diagnostics, environment compensation, and POD curves with quantified damage size.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the National Institute of Standards and Technology (NIST), Army, Air Force, Missile Defense Agency, Defense Advanced Research Projects Agency (DARPA), Federal Aviation Administration (FAA), and National Aeronautics and Space Administration (NASA) of the United States for sponsoring some of the developments presented in this article. The authors also would like to thank Dr. Roy Ikegami, Dr. David Zhang, and other colleagues at Acellent Technologies Inc. for the technical consultation and assistance in the development of the technology.

RELATED ARTICLES

Acoustic Emission

REFERENCES

[1] Boller C. Identification of life cycle cost reductions in structures with self-diagnostic devices. *Proceedings of the NATO RTO Symposium on Design Issues*. Ottawa, 1999; pp. 1–8.

[2] Chang F-K. Ultra reliable and super safe structure for the new century. In *Proceedings of the First European Workshop on SHM, Structural Health Monitoring 2002*, Balageas D (ed). Cachan, July 2002. DEStech Publications: Lancaster, PA, 2002, pp. 3–12.

[3] Beral B, Speckmann H. Structural health monitoring (SHM) for aircraft structures: a challenge for system developers and aircraft manufactures. In *Proceedings of the 4th International Workshop on SHM, SHM 2005: From Diagnostics and Prognostics to Structural Health Management*, Chang F-K (ed). Stanford University, September 2003. DEStech Publications: Lancaster, PA, 2003, pp. 12–29.

[4] Trego A, Akdeniz A, Haugse E. Proceedings of the Second European Workshop on SHM, Structural Health Monitoring 2004. In *Structural Health Management Technology on Commercial Airplanes*, Boller C, Staszewski W (eds). DEStech Publications: Munich, Lancaster, PA, 2004, pp. 317–323.

[5] Ansari F. Fiber optic health monitoring of civil structures using long gage and acoustic sensors. *Smart Materials and Structures* 2005 **14**:S1–S7.

[6] Giurgiutiu V, Zagrai A. Characterization of piezoelectric wafer active sensors. *Journal of Intelligent Material Systems and Structures* 2000 **11**: 959–975.

[7] Kwun H, Kim S-Y, Light GM. Magnetostrictive sensor guided-wave probes for structural health monitoring of pipelines and pressure vessels. In *Proceedings of the 5th International Workshop on SHM, SHM 2005: Advancements and Challenges for Implementation*, Chang F-K (ed). Stanford University, September 2005. DEStech Publications: Lancaster, PA, 2005, pp. 694–701.

[8] Calkin FT, Flatau AB, Dapino MJ. Overview of magnetostrictive sensor technology. *Journal of Intelligent Material Systems and Structures* 2007 **18**:1057–1066.

[9] Varadan VK, Varadan VV. Microsensors, microelectromechanical systems (MEMS), and electronics for smart structures and systems. *Smart Materials and Structures* 2000 **9**:953–972.

[10] Giurgiutiu V, Zagral A, Bao JJ. Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring. *Structural Health Monitoring* 2002 **1**(1):41–61.

[11] Lee BC, Staszewski WJ. Modeling of Lamb waves for damage detection in metallic structures: part II. Wave interactions with damage. *Smart Materials and Structures* 2003 **12**:815–824.

[12] Kessler S, Spearing S, Soutis C. Damage detection in composite materials using Lamb wave methods. *Smart Materials and Structures* 2002 **11**: 269–278.

[13] Qing X, Chan H, Beard S, Kumar A. An active diagnostic system for structural health monitoring of rocket engines. *Journal of Intelligent Material Systems and Structures* 2006 **17**: 619–628.

[14] Ihn J, Chang F-K. Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: I. Diagnostics. *Smart Materials and Structures* 2004 **13**(3): 609–620.

[15] Park G, Sohn H, Farrar CR, Inman D. Overview of piezoelectric impedance-based health monitoring and path forward. *The Shock and Vibration Digest* 2003 **35**(6):451–463.

[16] Seydel R, Chang F-K. Impact identification of stiffened composite panel: I. System development. *Smart Materials and Structures* 2001 **10**:354–369.

[17] Lin M, Chang F-K. The manufacture of composite structures with a built-in network of piezoceramics. *Composite Science and Technology* 2002 **62**:919–939.

[18] Lin M, Qing X, Kumar A, Beard S. SMART layer and SMART suitcase for structural health monitoring applications. In *Proceedings of SPIE on Smart Structures and Materials 2001: Industrial and Commercial Applications of Smart Structures Technologies*, McGowan A-MR (ed). Newport Beach, CA, March 2001. SPIE, 2001; Vol. 4332, pp. 98–106.

[19] Qing X, Beard B, Kumar A, Ooi T, Chang F-K. Built-in sensor network for structural health monitoring of composite structure. *Journal of Intelligent Material Systems and Structures* 2007 **18**:39–49.

[20] Qing X, Kumar A, Zhang C, Gonzalez IF, Guo G, Chang F-K. Hybrid piezoelectric/fiber optic diagnostic system for structural health monitoring. *Smart Materials and Structures* 2005 **14**(3):S98–S103.

[21] Malkin M, Qing X, Leonard M, Derriso M. Flight demonstration: health monitoring for bonded structural repairs. In *Proceedings of the Third European Workshop on SHM, Structural Health Monitoring 2006*, Güemes A (ed). Granada, July 2006. DEStech Publications: Lancaster, PA, 2006, pp. 167–175.

[22] Qing X, Beard S, Kumar A, Chan H, Ikegami R. Advances in the development of built-in diagnostic system for filament wound composite structures. *Composite Science and Technology* 2006 **66**:1694–1702.

[23] Russell S, Walker J, Workman G. Efficient nondestructive evaluation of prototype carbon fiber reinforced structures. *Proceedings of the 10th US-Japan Conference on Composite Materials*. Stanford University, Stanford, CA, 16–18 September 2002.

[24] Blackshire JL, Jata KV. *Integrated sensor durability and reliability*. Air Force Research Laboratory, Wright-Patterson Air Force Base: Ohio, 2008.

[25] Kusaka T, Qing X. Characterization of loading effect on the performance of SMART layer embedded or surface mounted on structures. In *Proceedings of the 4th International Workshop on SHM, SHM 2005:*

From Diagnostics and Prognostics to Structural Health Management, Chang F-K (ed). Stanford University, September 2003. DEStech Publications: Lancaster, PA, 2003, pp. 1539–1546.

[26] Qing XP, Kumar A, Beard S, Yu P, Zhang D, Liu C, Hannum R. Advanced self-sufficient structural health monitoring system. In *Proceedings of the Third European Workshop on SHM, Structural Health Monitoring 2006*, Güemes A (ed). Granada, July 2006. DEStech Publications: Lancaster, PA, 2006, pp. 807–814.

[27] Qing XP, Beard SJ, Kumar A, Sullivan K, Aguilar R, Merchant M, Taniguchi M. Performance of piezoelectric sensors based SHM system under combined cryogenic temperature and vibration environment. *Smart Materials and Structures* 2008 **17**(5):055010.

[28] Yang SM, Hung CC, Chen KH. Design and fabrication of a smart layer module in composite laminated structures. *Smart Materials and Structures* 2005 **14**(2):315–320.

[29] Tang S, Xiong K, Liang D, Li D. The development of SMART layer used in structural health monitoring. *Journal of Experimental Mechanics (in Chinese)* 2005 **20**(2):226–234.

[30] Qi B, Bannister M. Mechanical performance of carbon/epoxy composites with embedded polymeric films. *Key Engineering Materials* 2007 **334–335**:469–472.

[31] Qing X, Beard S, Kumar A, Hannum R. A real-time active smart patch system for monitoring the integrity of bonded repair on an aircraft structure. *Smart Materials and Structures* 2006 **15**:N66–N73.

[32] Qing X, Wu Z, Chang F-K, Ghosh K, Karbhari V, Sikorsky C. Monitoring the disbond of externally bonded CFRP composite strips for rehabilitation of bridges. In *Proceedings of the Third European Workshop on SHM, Structural Health Monitoring 2006*, Güemes A (ed). Granada, July 2006. DEStech Publications: Lancaster, PA, 2006, pp. 463–470. Also in *FRP International* 2006 **3**(3):11–14.

[33] Qing X, Kumar A. Integrated active-passive "SMART layer" system monitoring structural defects. *Technology Advances, MRS Bulletin* 2005 **30**(7):506.

[34] Qing X, Beard S, Kumar A, Yu P, Chan HL, Zhang D, Ooi T, Marotta SA. Practical requirements for implementation and usage of SHM systems on aerospace structures. In *Proceedings of the 5th International Workshop on SHM, SHM 2005: Advancements and Challenges for Implementation*, Chang F-K (ed). Stanford University, September 2005. DEStech Publications: Lancaster, PA, 2005, pp. 1502–1509.

[35] Yu P. Real time impact detection system for thermal protection system. In *Proceedings of the 7th International Workshop on SHM, SHM 2007: Quantification, Validation, and Implementation*, Chang F-K (ed). Stanford University, September 2007. DEStech Publications: Lancaster, PA, 2007, pp. 153–159.

[36] Beard S, Liu B, Qing P, Zhang D. Challenges in implementation of SHM. In *Proceedings of the 7th International Workshop on SHM, SHM 2007: Quantification, Validation, and Implementation*, Chang F-K (ed). Stanford University, September 2007. DEStech Publications: Lancaster, PA, 2007, pp. 65–84.

Chapter 80
The HELP-Layer® System

Michel B. Lemistre
Laboratoire SATIE/CNRS, Ecole Normale Supérieure de Cachan, Cachan, France

1 INTRODUCTION

The major risk for composite materials is the creation of delaminations resulting from impacts being often associated by fiber breaking. Besides, other types of damages can be caused by thermal aggressions, inducing pyrolysis of the polymeric matrix and by liquid ingress linked to aggressive environments. The most current structural health monitoring (SHM) system (based on fiber-optic sensors and acousto-ultrasonics), well suited to damages of mechanical origin, are unfortunately poorly sensitive to these last two types of damages. Assuming that all damages can affect the main electrical properties of structures such as electrical conductivity and dielectric permittivity, a measurement of these two parameters may allow detecting all kinds of damages. This is the reason why a new SHM system has been developed. This system, based on the interaction between a low-frequency electromagnetic field and the composite structure, allows the electrical characterization of materials. As shown here, this method detects and characterizes all kind of damages with a good sensitivity.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 THE HELP-Layer® SYSTEM

2.1 Short recall of the principle

The HELP-Layer® system [1, 2] is a low-frequency electromagnetic technique (*see* **Electric and Electromagnetic Properties Sensing**) applied to an SHM concept for composite structures. Its principle is based on a simple concept: a carbon fiber reinforced plastic (CFRP) structure is a double medium made, on the one hand, of a conductive medium (carbon fibers of conductivity $\sigma \approx 10^4\,\mathrm{S\cdot m^{-1}}$) and, on the other hand, of a perfectly dielectric medium (the polymeric matrix). The electromagnetic behavior of the first medium is sensitive to conductivity variations and the second medium to the orientational polarization phenomenon. So, one can define a polarization vector $\mathbf{P}$ linked with the local electric field E_1 by the following relation:

$$\mathbf{P} = \chi_e \varepsilon_0 E_1 \tag{1}$$

where ε_0 is the dielectric permittivity of the free space ($8.84 \times 10^{-12}\,\mathrm{F{\cdot}m^{-1}}$) and χ_e is the electric susceptibility, itself linked with the relative dielectric permittivity ε_r by the following expression:

$$\varepsilon_r = 1 + \chi_e \tag{2}$$

After induction of eddy currents in the structure, the measurement of the reflected electric field E_r gives access to both electrical parameters σ and ε_r, by the following relation:

$$E_r = \frac{J}{\sigma} + E_l\left(\varepsilon_r - 1\right) \tag{3}$$

with J being the eddy current density. This method allows for detection of a possible local variation of σ and/or ε_r. Therefore, all damages inducing a local variation of one or both these parameters will be detected.

In the case of a glass fiber reinforced plastic (GFRP) structure, the induction of eddy currents being impossible, one excites the structure with an electric field. So, only one parameter is measured, the relative dielectric permittivity ε_r.

2.2 Technology used

The HELP-Layer® system is a complete system having two parts. The first part is the sensitive layer itself. The second part is the associated electronics having in charge the excitation for inducing eddy currents (or electric field) in the structure, and the data reduction process to extract the relevant information and to build an image of the structure where the damages clearly appear.

The sensitive layer is made of a printed circuit on a 200-μm-thick dielectric substrate including a double network of crossed wires; this layer is bonded or embedded in the structure under test. Figures 1 and 2 respectively present the principle of the layer and a photo of a carbon/epoxy plate equipped with such a layer. By scanning these networks, one can perform local measurements of the electric field, with each analyzed zone having a dimension of 20 mm × 20 mm corresponding to the distance between two successive wires of each network (Figure 1). The first network, called the *inductive network*, is short-circuited at one end (or in open circuit in the case of a GFRP structure);

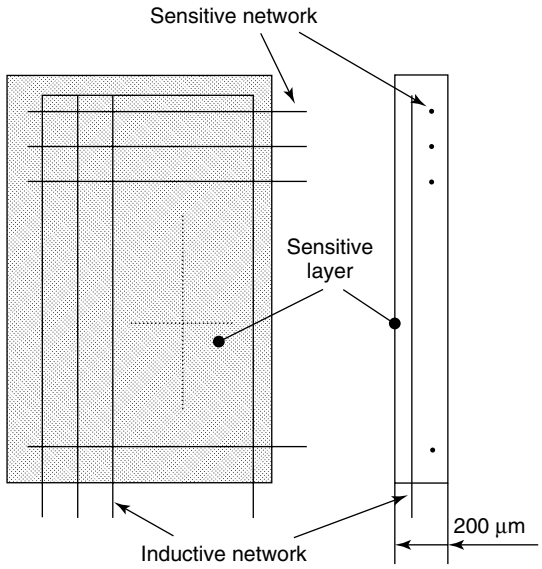


Figure 1. Geometry of the sensitive layer.

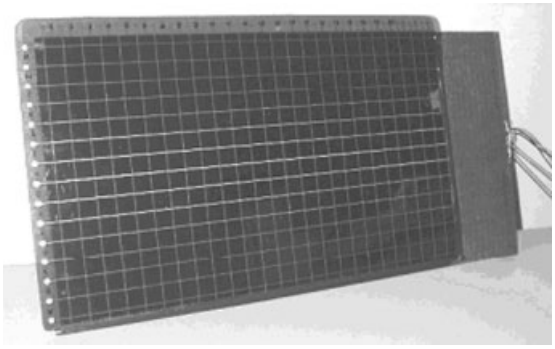


Figure 2. Instrumented structure.

each pair of wires represents a loop of induction (or a capacitance in the case of a GFRP structure). The second network, called the *sensitive network*, is not short-circuited, and each pair of successive wires can be considered as a capacitance allowing the in-plane component of the electric field perpendicular to the wires to be measured. Furthermore, each wire of this network can be considered as an elementary antenna, measuring the in-plane component of the electric field parallel to the wires (Figure 3). There are two possible exploitation methods of this technique. The first one consists of using only the modulus of the in-plane electric field to build an image of the damaged structure. The second one exploits separately the two in-plane components of the electric

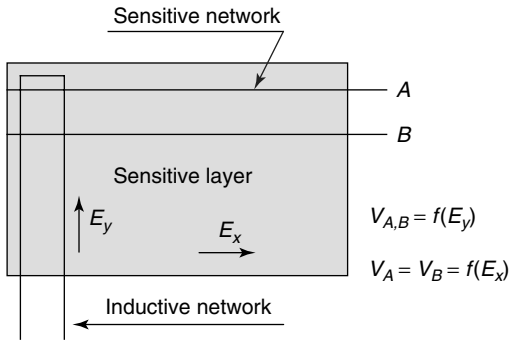


Figure 3. Measurement of the electric field components.

field and allows computation of the new values of the electric parameters to fully characterize the damages and to estimate their severity. This second method of exploitation is explained in the following subsections.

2.3 Example of application

An example of results obtained with the first method of exploitation is shown in Figure 5. Figures 2, 4, and 5 respectively present a photograph of an instrumented structure (a 16-ply orthotropic carbon/epoxy plate $[0_2, 90_2]_{2s}$ of dimensions: 610 mm × 305 mm × 2 mm), the types and locations of the damages in the structure (as seen from the opposite face of the plate), and the electromagnetic image obtained. The plate has been damaged by six different defects: a 4 J impact (I2) inducing a severe delamination with fiber breakage, a 2 J impact (I1) inducing a light delamination, and four local burns produced by "high energy sparks" (30 V, 5 A) of various duration, delivering energies of 40 J (B2), 80 J (B3), 120 J (B1), and 400 J (B4) (Figure 4).

In Figure 5, one can see that all damages are perfectly detected except the lower 2 J impact, probably because of the fact that there is no fiber breakage, so there is no variation of electric properties inside the structure under test. The inductive network has been excited by a continuous signal having a frequency of 700 kHz. The data reduction process is based on a multiresolution processing, using wavelet transform [3–5].

3 NUMERICAL SIMULATION

3.1 Basic principle

The numerical simulation method used is a concept developed at the Ecole Normale Supérieure of Cachan/France, called *distributed point source method* (DPSM) [6–8]. The main originality of this method is that, contrary to a classical finite

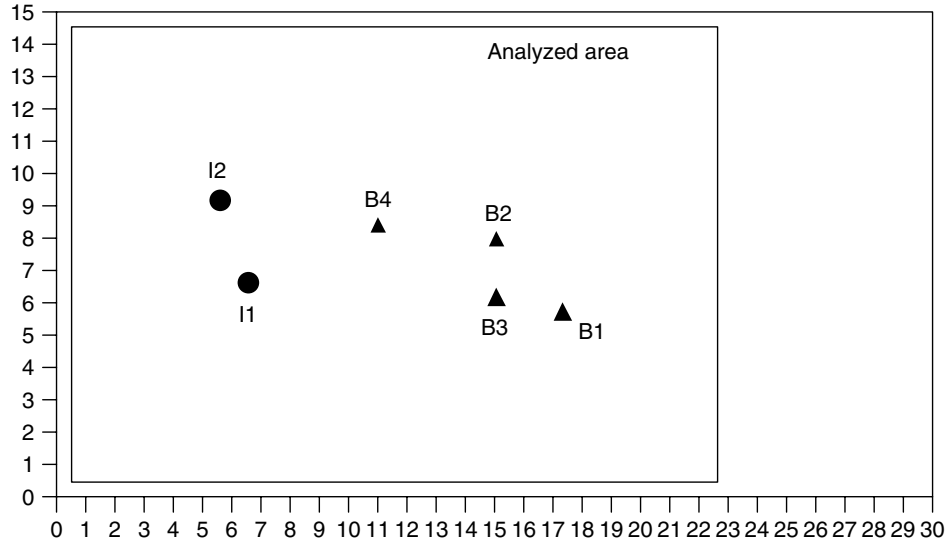


Figure 4. Damaged structure.

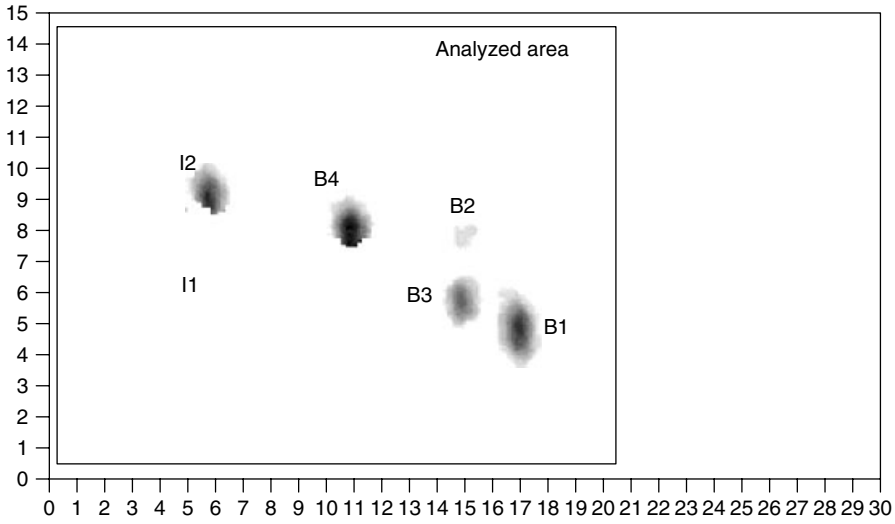


Figure 5. Electromagnetic image of the damaged structure shown in Figure 4.

elements method, it is not necessary to mesh the totality of the computation volume, but only the surface of interest. The implementation of the model simply requires discretization of the active surface of the transducer or the interfaces to obtain an array of point sources, so that the initial complexity is changed into a superposition of elementary problems. The active surfaces like transducers, emitters, or interfaces reflecting a part of an incident field, are discretized into a finite number of elementary surfaces, a point source being placed at the centroid of every elementary surface. It is interesting to note that the energy (or the power) radiated by such a system is the product of a scalar quantity by the flux of a vector (or the time derivative of the flux, for power). Let us call the scalar quantity θ_k the scalar potential, and ϕ_k the flux emitted by the source k, the vector being the field $\boldsymbol{V}$. For magnetic systems, θ_k and ϕ_k represent the magnetic potential and the flux of magnetic induction for the N elementary sources (Figure 6). For instance, one can calculate the potential θ at point M by superposition of elementary charges as follows:

$$\theta(M) = k\left(\sum_{n=1}^{N} \frac{q_n}{R_n}\right) \tag{4}$$

where R_n represents the distance between the source k and the point M, and q_n is the elementary charge at the point n (i.e., the charge generating the source k).

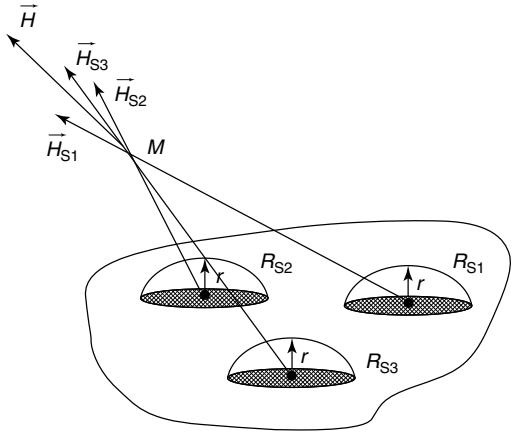


Figure 6. Discretization of a sensor surface into a finite number of hemispherical surfaces.

However, it is necessary to calculate the potential θ at the top point of each hemispherical surface (contribution of each elementary source toward others) by the following equation:

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} = k \begin{bmatrix} F_{1,1} & F_{1,2} & :: & F_{1,n} \\ F_{2,1} & F_{2,2} & :: & F_{2,n} \\ :: & :: & :: & :: \\ F_{n,2} & :: & :: & F_{n,n} \end{bmatrix} \cdot \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \tag{5}$$

where the function F represents the inverse of the distance R between the charge q_n and the calculation

point. From matrix $[F]$, which is a regular matrix, it is possible to calculate the elementary charges by the following relation:

$$[q] = \frac{1}{k}[G] \cdot [\theta] \tag{6}$$

with

$$[G] = [F]^{-1} \tag{7}$$

One can thus calculate the three quantities θ, V, and ϕ in each point of the space by using the knowledge of the elementary sources. If we place a target at point M, one can calculate the same quantities on its surface by using the matrix of reflection coefficients. A major issue of this method is that the surface of interest is meshed uniquely and the thickness of the target is taken into account by the matrix of reflection coefficients. This method allows the behavior of the HELP-Layer® system to be simulated.

3.2 Simulation conditions

As shown before, the HELP-Layer® system includes two crossed conductive networks. One of these networks is in charge of inducing eddy currents inside the carbon structure and appears in the form of parallel lines, with a spacing of 20 mm, and is short-circuited at one end. To induce significant eddy currents inside the structure (i.e., a significant electric field), it is necessary to sequentially inject a current of 1 A into each line of the inductive network. The frequency of excitation is chosen taking into account the depth of the structure (i.e., the skin effect). To simplify the problem, only one element of the HELP-Layer® of 60 mm × 60 mm is modeled, including only one inductive line. The structure is set at 0.1 mm above the HELP-Layer®. The current source is an inductive line of 60-mm length located in the xy plane at $x = 30$ mm and $y = 0$–60 mm, meshed by 60 current elements *Idl* (Figures 7 and 8). This current source is called *DPSM primary sources JAp*. The formulation used is the DPSM/Green's formulation [9].

The first computation consists of calculating the electromagnetic field on the surface of the structure ($\vec{E}$ and $\vec{H}$) or, more precisely, the magnetic vector potential $\vec{A}_1$ in medium 1 (free space) computed by superimposing the effect of current sources *JAp* and *JA*1 (DPSM virtual sources; see Figure 8. The magnetic vector potential in medium 2 (the structure) $\vec{A}_2$ is only defined by the current sources *JAs* and called *DPSM secondary sources*, the boundary

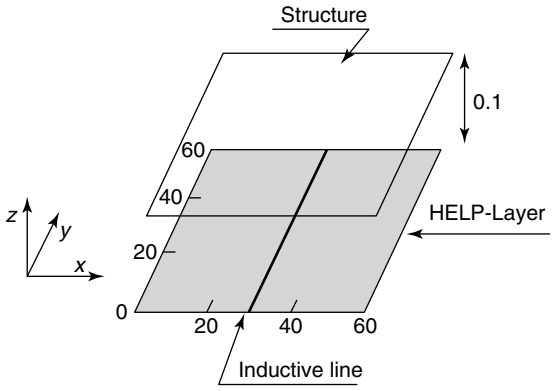


Figure 7. Geometry of the simulation (dimensions are given in millimeters).

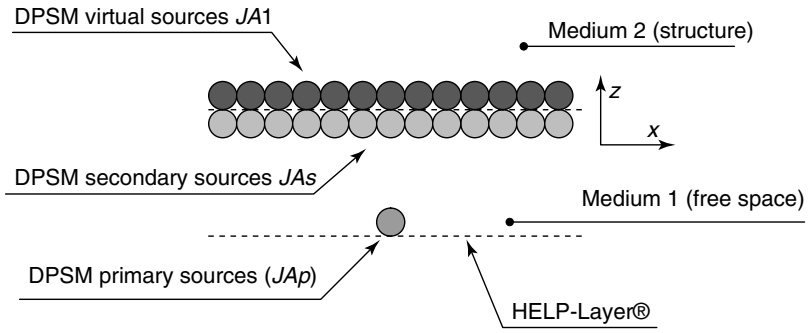


Figure 8. DPSM current sources.

conditions being the continuity on the vector potential and its first derivative along the z axis:

$$\begin{cases} \vec{A}_1 = \vec{A}_2 \\ \dfrac{1}{\mu_1}\dfrac{\partial \vec{A}_1}{\partial z} = \dfrac{1}{\mu_2}\dfrac{\partial \vec{A}_2}{\partial z} \end{cases} \tag{8}$$

In the case of harmonic excitation, one can compute the electric field by the following relation:

$$\vec{E} = -j\omega\vec{A} \tag{9}$$

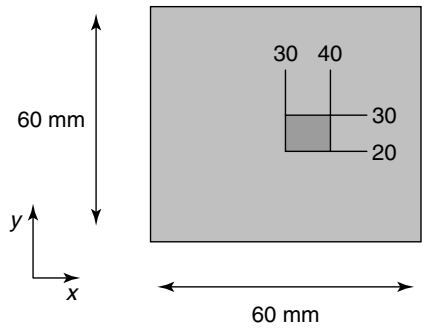


Figure 9. Location of the virtual damage.

3.3 Validation with experimental data

A computation of the in-plane component of the electric field E_y for three kinds of materials allows the results obtained for the same values measured to be compared by the HELP-Layer® system. Table 1 shows this comparison.

3.4 Modeling a damage inside the structure

Assuming that the defect is located at 1-mm depth inside the structure for $x \in (30, 40)$ mm and $y \in (20, 30)$ mm (Figure 9), generating a variation of one electric parameter of the structure (i.e., σ or ε). The first step is to construct the secondary sources *JAs* inside the structure. After that one performs the same process as presented previously, with new virtual sources *JA*2 and new secondary sources *JAs*2, medium 2 being the structure and medium 3 being the damaged zone of the structure (Figure 10). Note that we do not solve the global problem but a new elementary problem with new DPSM sources *JAs*2.

Two kinds of damages (damage 1 and damage 2) have been simulated, by variation of the local electrical conductivity σ and by variation of the local relative dielectric permittivity ε_r, respectively. Table 2

Table 1. Comparison between computed values and experimental values of the E_y component of the tangential electric field

Type of structures	E_y component ($\mathrm{V{\cdot}m^{-1}}$): computed values	E_y component ($\mathrm{V{\cdot}m^{-1}}$): experimental values	ΔE_y (%)
$[0_2, 45_2, 90_2, -45_2]_S$	1.59	1.65	+4
$[0_2, 45_2, 90_2, -45_2]_{2S}$	1.40	1.40	−3.5
$[0_2, 90_2]_{2S}$	1.60	1.65	+4

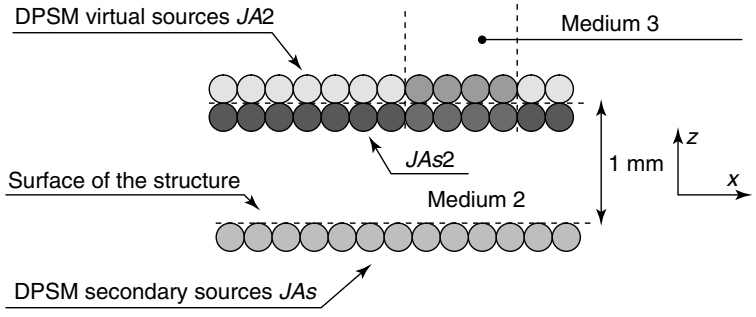


Figure 10. New DPSM sources.

Table 2. Numerical values of σ and ε_r for each kind of damage

Kind of damage	Sound area		Damaged area	
	σ (S·m^{-1})	ε_r	σ (S·m^{-1})	ε_r
1	10^4	4	5×10^3	4
2	10^4	4	10^4	2

shows the numerical values of these two parameters for each kind of damage.

Figure 11(a) and (b) present the modulus of the resulting electric field $|E_y|$ and $|E_x|$, respectively for the damaged structure 1. Figure 12 (a) and (b) present the same parameters for the damaged structure 2. These electric fields expressed in volts per millimeter represent for each one of the damaged structures (i.e., 1 or 2) the difference between the electric field obtained from a perfectly sound structure and the electric field obtained from the damaged structure.

One can see that in the case of a damage generating a variation of the electric conductivity σ, the y component of the electric field is dominating. In Figure 11(a) and (b), showing $|\vec{E}_y|$ and $|\vec{E}_x|$ respectively, the maximum value of the y component is 10^{-3} V·mm^{-1}, whereas the maximum value of the x component is only 1.3×10^{-7} V·mm^{-1}. On the contrary, in the case of a damage generating a variation of the dielectric permittivity ε_r, the major contribution on the electric field is given by the x component (Figure 12a and b): i.e.,

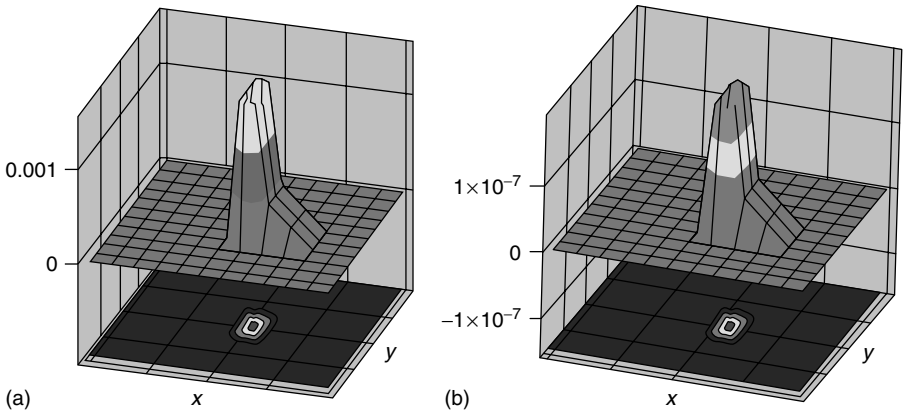


Figure 11. Electric fields resulting from a damaged structure 1. (a) Modulus of the E_y component and (b) modulus of the E_x component.

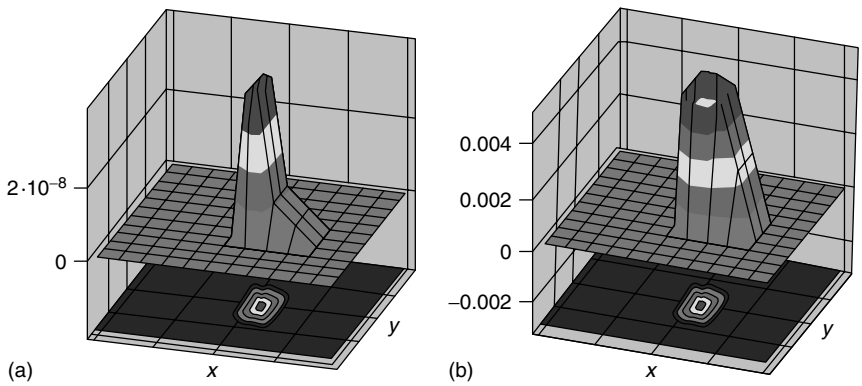


Figure 12. Electric fields resulting from a damaged structure 2. (a) Modulus of the E_y component and (b) modulus of the E_x component.

the maximum value of the y component is only $2.5 \times 10^{-8}\,\mathrm{V{\cdot}mm^{-1}}$, while the x component is $3 \times 10^{-3}\,\mathrm{V{\cdot}mm^{-1}}$. This is a very important result because it allows a kind of damage (i.e., the origin of the damage, *see* **Electric and Electromagnetic Properties Sensing**) to be determined.

3.5 Solving the inverse problem

To solve the inverse problem (i.e., computation of the new values of σ and ε_r, and determination of the kind of damage), the preceding remarks allow an algorithm to be developed (Figure 13). The first step helps to determine if the modulus of the x component of the electric field measured by the HELP-Layer® system E_{x_m} is significant (i.e., whether $>10^{-3}\,\mathrm{V{\cdot}m^{-1}}$). If the E_{x_m} component is not significant, the damage has a mechanical origin (delamination, fiber breaking, crack); so, one computes the modulus of the E_y component (resulting from the difference: sound structure—damaged structure) with variation of the σ value and compares it to the experimental E_{y_m} component. When the computed E_y value is equal to the experimental E_{y_m} value, the last value of σ is the local conductivity of the structure due to the damage. If the E_{x_m} component is significant, the damage has no mechanical origin, but may be due to burning, liquid ingress, etc. One compares the computed E_x value with the experimental E_{x_m} value and determines the local permittivity ε_r due to the damage, by the same process. To determine the possible local variation of σ one can then perform the same process with the E_y components.

The first evaluation concerns a quasi-isotropic plate of 2-mm thickness $[45_2, 0_2, -45_2, 90_2]_S$, including various delaminations generated by calibrated impacts (impact energies of 0.75, 2, 2.5, 3, and 4 J). The HELP-Layer® system measures the two components of the electric field (i.e., E_{x_m} and E_{y_m}), while the simulation program computes the values ε_r and σ on the damaged area corresponding to the same electric field component variations. For a sound structure, the electrical parameters are $\varepsilon_r = 4$ and $\sigma = 10^4\,\mathrm{S{\cdot}m^{-1}}$. Table 3 shows the results obtained. One can see that a delamination resulting from an impact of energy lower than 2.5 J does not cause a fiber breakage and consequently cannot induce a variation of σ that would detect the fiber breakage. The threshold of detection for this kind of defect is a delamination resulting from an impact of about 2.5 J for a 2-mm-thick plate of such a composite.

The second evaluation concerns a 2-mm-thick orthotropic plate $[0_2, 90_2]_S$, including various burning generated by electric sparks, for electric energies of 10, 40, 80, and 120 J. The field variations measured are given in Table 4. Assuming the same sound material properties as in the first case, the simulation

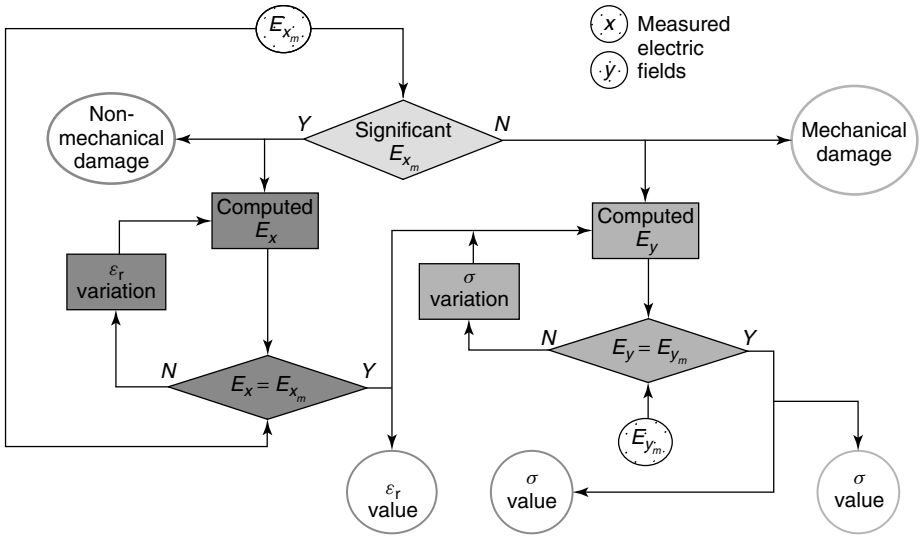


Figure 13. Algorithm allowing to solve the inverse problem.

Table 3. Equivalent σ and ε_r for impact delamination

Impact energies (J)	Measured fields		Deduced properties	
	ΔE_x (V·m^{-1})	ΔE_y (V·m^{-1})	σ (S·m^{-1})	ε_r
0.75	NS[a]	NS	X[b]	X
2	NS	NS	X	X
2.5	NS	1.1×10^{-2}	8×10^{3}	X
3	NS	1.6×10^{-2}	7×10^{3}	X
4	NS	5.0×10^{-2}	4.7×10^{3}	X

[a] Nonsignificant values.
[b] Same values as a sound structure.

Table 4. Equivalent σ and ε_r for electric burning

Electric energies (J)	Measured fields		Deduced properties	
	ΔE_x (V·m^{-1})	ΔE_y (V·m^{-1})	σ (S·m^{-1})	ε_r
10	0.07	1.8×10^{-3}	NV[a]	3.93
40	0.26	6×10^{-3}	NV	3.80
80	0.53	1.3×10^{-3}	NV	3.56
120	0.80	2×10^{-3}	NV	3.39

[a] Nonsignificant variation (see subsection 3.5).

leads to values of ε_r and σ as given in this table. Here an extra hypothesis was necessary. It was assumed that in the case of light burning, the low variation of the conductivity of the resin is masked by the relatively high conductivity of the carbon and that in the case of a fiber breakage phenomenon induced by hard burning, the increase of σ in the resin because of pyrolysis and the decrease of σ in the carbon fiber bundles would compensate each other. On the basis of these considerations, it has been assumed that only the permittivity ε_r was affected. In fact, it would have been more accurate if the two parameters, ε_r and σ, could have been varied. However, that would require the establishment of a new experimental procedure allowing the identification of both parameters. This improvement of the method is an objective for future development.

4 CONCLUSION

This article has been dedicated to an original system of structural health monitoring called the *HELP-Layer® system*, which is based on the interaction of an electromagnetic field and materials such as dielectric or conductive composite structures. The system shows that it is perfectly possible to detect, localize, and characterize various damages in a composite structure by using electromagnetic methods. Simulations allow validation of this electromagnetic method for use in a SHM system such as the HELP-Layer®, based on the analysis of electrical properties of materials. They also permit to distinguish between two types of damage (e.g., mechanical and thermal damage). The main interest of this method lies in its high sensitivity to "nonmechanical" damages such as resin pyrolysis and liquid ingress.

Work in progress consists in determining the effect of electrically qualified damage on the residual mechanical behavior of the composite material linking directly the HELP-Layer® system measurements to the degradation of structural performance, which, in fact, is the most interesting point.

REFERENCES

[1] Lemistre M. Low frequency electromagnetic techniques. *Structural Health Monitoring*. ISTE: London, UK, 2006; Chapter 6, pp. 412–461.

[2] Lemistre MB, Placko D. *HELP-Layer System*, French Patent FR0403310, 2004, US Patent US2005/0228 208A1, 2005.

[3] Lemistre MB, Balageas DL. A new concept for structural health monitoring applied to composite materials. In *Structural Health Monitoring*, Balageas DL (ed). DEStech Publications, 2002; pp. 493–507.

[4] Lemistre MB, Balageas DL. Structural health monitoring system based on diffracted Lamb waves analysis by multiresolution processing. *Smart Materials*. IOP Publishing, 2001; Vol. 10, No. 3, pp. 504–511.

[5] Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press: New York, 1998.

[6] Lemistre MB, Placko D, Liebeaux N. Simulation of an health monitoring concept for composite materials, comparison with experimental data. *Proceedings of SPIE* 2003 **5047**:130–139.

[7] Placko D, Liebeaux N, Kundu T. *Modélisation par Sources Réparties*, Patent in progress no. 0214108.

[8] Dufour I, Placko D. An original approach of eddy current problems through a complex electrical image concept. *IEEE Transaction on Magnetics* 1996 **32**(2):348–365.

[9] Lemistre M. In *DPSM for Modeling Engineering Problems*, Placko D, Kundu T (eds). John Wiley & Sons, 2007; Chapter 10, pp. 333–347.

Chapter 79
Hybrid PZT/FBG Sensor System

Zhanjun Wu[1], Xinlin P. Qing[2] and Fu-Kuo Chang[1]

[1] *Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA*
[2] *Acellent Technologies, Inc., Sunnyvale, CA, USA*

1 INTRODUCTION

Damage detection techniques have been studied intensively in almost every aspect of engineering. In recent years, tremendous progress has been made. This is apparently a response to the high demand for safer structures and lower cost. Technological progress on novel smart actuators and sensors, such as piezoelectric transducers (lead–zirconate–titanate—PZTs) and fiber-optic sensors (fiber Bragg grating—FBG), has paved the way for the development of structural health monitoring (SHM) technology [1, 2]. Owing to their advantages, piezoelectric ceramics have been widely employed in noise and vibration control and active and passive sensing [3–6]. However, when the piezoelectric sensors are used in the active-sensing mode, such as using Lamb wave in pitch–catch manner to detect structural damage with a highly integrated diagnostic system, crosstalk between PZT actuation signals and sensor signals is always a problem. Further, electronic signals from PZT sensors to the data storage or processing unit may attenuate significantly during long-distance transmission, for example, from a bridge deck to its monitoring and control center. FBG sensor systems have offered an attractive solution to the problem of making strain measurements, as they have the advantages of small size, high resolution, high multiplexibility, electromagnetic immunity, and potentially high-density quasi-distributed measurements. FBGs serving as strain sensors have recently been used

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

in numerous harsh field environments, such as civil structures [7–9], aerospace vehicles [10], and ships [11] (*see* **Fiber Bragg Grating Sensors**).

A hybrid piezoelectric/fiber-optic sensor system offers the best decoupling of actuator and sensor signals (minimum interference), because the two devices use different mechanisms for signal transmission: the piezoelectric actuators use electrical channels while the FBG sensors use optical means [12, 13]. However, building a hybrid system is not simply a matter of putting the devices together. One of the most important issues arises from their different characteristics. PZTs are often used to actuate structures or detect dynamic responses, while FBG sensors are mainly used for quasi-static measuring or relatively low-frequency responses. This does not mean that FBG sensors cannot measure dynamic responses. But there are some issues relating to such sensors needed to be investigated, such as the relationship between sensitivity and grating length, the part of the optical fiber on which gratings are engraved, strain resolution, frequency range, and also the signal-to-noise ratio. Udd [14] demonstrated the capability of FBG sensors to detect acoustic emission and ultrasonic stress waves. Ogisu *et al.* [15] reported damage detection with the FBG sensor/PZT actuator hybrid system. Minardoa *et al.* [16] gave the full characterization of the ultrasonic waves by wavelength shift detection, indicating that the measurement can be achieved only if the grating length is smaller than the ultrasonic wavelength. Pierce *et al.* [17] successfully demonstrated the technique of damage inspection for CFRP (carbon fiber–reinforced plastic) plates with both fiber-optic Michelson interferometer and fiber-optic Mach–Zehnder interferometer using ultrasonic Lamb waves. However, this technique is not capable of multiplexing. Betz and his colleagues [18–20] have done a thorough study on both theoretical analysis and experimental validation of using PZT/FBG hybrid system to sense acoustic stress wave for damage detection.

In this article, a general review of the hybrid piezoelectric/fiber-optic SHM systems is presented, which includes the ultrasonic stress wave detectability study based on both experimental demonstration and theoretical analysis, the integrated hybrid PZT/FBG active diagnostic system and the applications, as well as the demonstrations of damage detection in both metallic and composite structures.

2 FBG AS ULTRASONIC STRESS WAVE SENSORS

2.1 Experimental demonstration

FBG is a selective reflector that reflects optical signals of a certain wavelength called the *Bragg wavelength* λ_b, according to the physical state of the gratings. An FBG consists of a series of periodically located cross sections with a higher refractive index (RI) at the core of a length of optical fiber. At each of the RI steps, a small fraction of the optical signal is reflected. With variations of temperature and strain, the fiber's physical properties change. This results in a change of the RI (n) and the period of the Bragg grating (Δ_0) and causes a Bragg wavelength shift (λ_b). This shift in the Bragg wavelength can be measured by an optical spectrum analyser (*see* **Acoustic Emission**; **Piezoelectric Wafer Active Sensors**; **Stanford Multi-actuator–Receiver Transduction (SMART) Layer Technology and Its Applications**).

$$\lambda_b = 2n_{\text{eff}_0}\Delta_0 \tag{1}$$

There are several methods for interrogating the FBG signals to effect quasi-static or low-frequency multisensor measurements [21]. Micron Optics fiber Fabry–Perot tunable filter (FFP-TF), developed by Micron Optics, is commonly used. It utilizes a Fabry–Perot etalon that passes wavelengths that are equal to integer fractions of the cavity (etalon) length; all other wavelengths are attenuated according to the Airy function [22]. However, for high-frequency measurements up to $\sim 10^5$ Hz or higher, another interrogation scheme, developed by Udd [14], is well suited. As shown in Figure 1, a matched fiber grating filter is employed instead of FFP-TF.

Udd and his coworkers presented a systematic study on acoustic emission detection using FBG sensors. In the first test, they attached an FBG sensor onto a PZT actuator and designed an interrogation scheme shown in Figure 1, in which a tunable matching FBG filter was employed to interrogate the

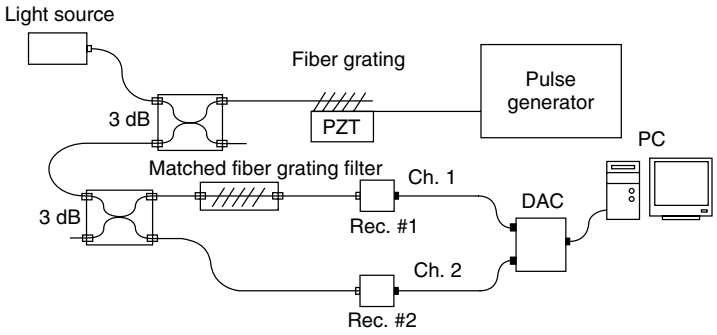


Figure 1. Schematic diagram of the FBG system with matched fiber grating filter. [Reproduced with permission from Ref. 14. © SPIE, 2001.]

optical response signal. Data were taken in the form of photoreceiver voltage response to the light signal. The ratio of signal from channel 1 to channel 2 can be correlated to the corresponding strain under measurement. From their experimental investigation, a lower detection limit has been established for acoustic emission (AE) detection using FBGs. They also carried out tests to study the capability of Lamb wave detection by FBGs on an aluminum plate. The detection results showed that the sensitivity of the FBG sensor varied according to the distance between the FBG sensor and the ultrasonic transducer. Furthermore, the directional dependency of the detection sensitivity was also studied.

2.2 Theoretical analysis

Minardoa *et al.* [16] investigated the response of an FBG to the dynamic strain induced by a symmetric mode Lamb wave theoretically and presented numerical simulation results. They studied the interaction between a uniform FBG of length L, written into the core of a standard single-mode fiber, and an ultrasonic wave. The FBG that is not subjected to any external stress is described by a modulation of the effective RI of the fundamental guided mode along the fiber axis z.

$$n_{\text{eff}}(z) = n_{\text{eff}\,0} - \Delta n \sin^2\left(\frac{\pi}{\Lambda_0} z\right) \tag{2}$$

where Δn is the maximum index change and Λ_0 is the grating pitch. The Bragg wavelength, as determined by applying the Bragg condition, is given by equation (1).

They assumed an ultrasonic plane wave, with the acoustic wavefront normal to the optical fiber. Strain field is modeled by a longitudinal strain wave propagating along the fiber axis. In addition, the time dependence is assumed to be sinusoidal and can be expressed as

$$\varepsilon(z, t) = \varepsilon_m \cos(k_s z - \omega_s t) \tag{3}$$

Here, ε_m denotes the strain wave amplitude, ω_s its angular frequency, and k_s its wave number related to its wavelength by $k_s = 2\pi/\lambda_s$.

The strain wave influences the Bragg grating response by modulating its geometrical and physical properties, which is the same as quasi-static measurement using FBG sensors. Then the new effective RI of the Bragg grating under the ultrasonic wave action can be evaluated as the sum of two contributions. The first one is a mechanical contribution, due to the modulation of the grating pitch under the strain wave, which can be determined by the deformation along the z axis of the grating; the second one is an optical contribution due to the change in RI via the elasto-optic effect. Through the study the sensitivity of FBG response has been shown to decrease when reducing the ratio between the ultrasound (US) wavelength and the grating length. Moreover, a significant spectrum shape distortion has been shown to occur for high-power ultrasonic waves. It also showed there are essentially three main operating regions that can be distinguished. The first region, when $\lambda_S/L \ll 1$, the wavelength shift sensitivity S_λ approaches zero, and so no apparent Bragg wavelength modulation occurs. In the second region, corresponding to $\lambda_S/L \approx 1$, S_λ increases with the ratio λ_S/L, and in the third region,

corresponding to $\lambda_S/L \gg 1$, S_λ approaches the static value $S_m \approx 0.78$.

3 DAMAGE DETECTION

3.1 Lamb wave source location

Betz and Lamb [18] suggested a fit function to predict the directivity dependence of FBG sensors.

$$\hat{A}(\alpha) = a_2 \sin^2\left(\pi \frac{\alpha - a_1}{180}\right) \tag{4}$$

where $\hat{A}$ is the normalized amplitude of the strain amplitude, α is the angle between the direction of FBG sensors and the longitudinal wave propagation, and a_1 and a_2 are fit parameters that can be determined by test. He designed a test scheme to obtain those parameters and then used the inverse of equation (4) to calculate the direction of the Lamb wave, and by two sets of FBG sensor rosettes the location of the Lamb wave source can be found.

3.2 Acoustic emission detection

Udd [14] and his coworkers studied the acoustic emission detection ability of FBG sensors. However, a further study was carried out by Baldwin and Vizzini [23]. Baldwin and Vizzini successfully demonstrated the ability of an FBG sensor to detect a pencil lead break event, metal to metal impact events, and AE events from composite specimens with loosely mounted and embedded FBG sensors using a similar interrogation technique as Udd adopted.

3.3 Composite delamination detection

Ogisu *et al.* [15] reported damage detection with hybrid FBG sensor/PZT actuators using small diameter FBG sensors. They proposed a novel FBG sensor signal interrogation system capable of detecting frequencies of up to 1 MHz. The system employs an arrayed wave guide grating (AWG)-type filter to obtain a high-sensitivity filter characteristic for detecting small displacements in the grating of the FBG sensors. Furthermore, the AWG-type systems are also capable of interrogating multiple sensors at the same time. The sensor length in their study was less than one-seventh of the wavelength of the Lamb wave to ensure sensitivity. They also carried out double-lap joint-type composite coupon tests with the proposed method to demonstrate the damage detection capability.

4 ACTIVE HYBRID PIEZOELECTRIC/FIBER-OPTIC SHM SYSTEM

To take the advantages of both PZT and fiber-optic sensor, an active hybrid piezoelectric/fiber-optic SHM system has been developed recently [13]. With this system, the emerging concept of SHM can become a commercially viable option in structural engineering, allowing a new generation of safer, more reliable, and lower maintenance structures. As shown in Figure 2, the developed structural diagnostic system can permit quantitative characterization and event determination pertaining to aerospace and civil structures in hostile service environments. More specifically, the hybrid system can potentially be used to perform:

- *in situ* material property characterization;
- detect material and structural defects;
- detect damage including delaminations and corrosion; and
- characterize load environments (fatigue, overload).

4.1 Principle of the PZT/FBG hybrid active structural health monitoring system

The hybrid diagnostic system uses piezoelectric actuators to input a controlled excitation to the structure and fiber-optic sensors to capture the corresponding structural response. The system consists of three major parts: a diagnostic layer with a network of piezoelectric elements and fiber gratings to offer a simple and efficient way to integrate a large network of transducers onto a structure; diagnostic hardware consisting of an arbitrary waveform generator and a high-speed fiber grating demodulation unit together with a high-speed data-acquisition card to provide actuation input, data collection, and

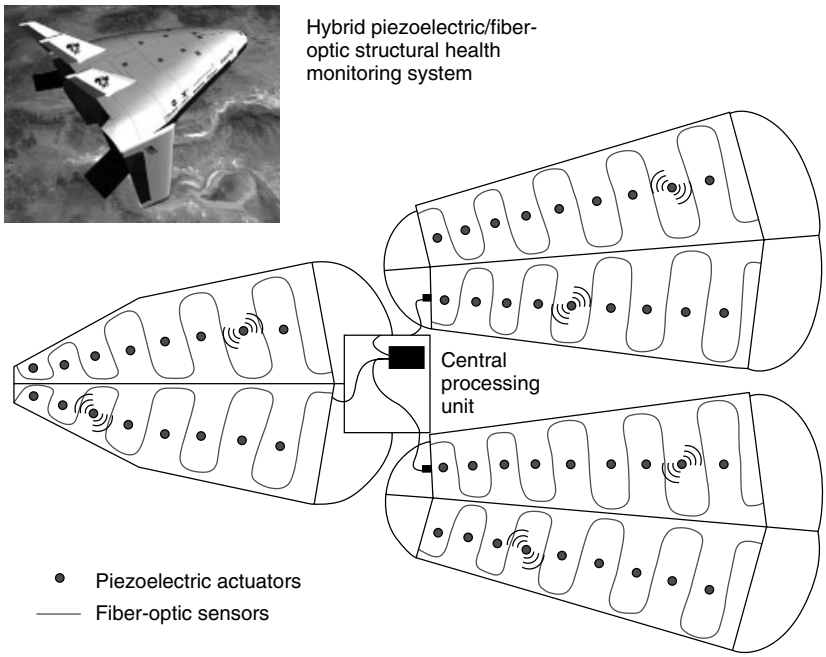


Figure 2. Schematic of a hybrid piezoelectric/fiber-optic structural health monitoring system for aerospace vehicles. [Reproduced from Ref. 13. © Institute of Physics Publishing, 2005.]

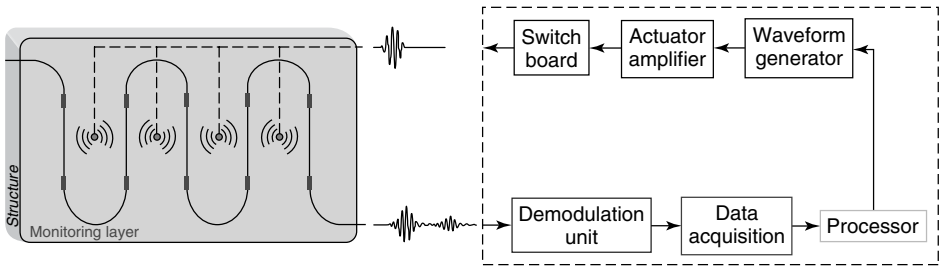


Figure 3. Piezoelectric/fiber-optic hybrid sensing system scheme. [Reproduced from Ref. 13. © Institute of Physics Publishing, 2005.]

information processing; and diagnostic software to determine the condition of the structure. Figure 3 shows a schematic diagram of this system.

One major advantage of the hybrid diagnostic system is that it offers the best actuator/sensor decoupling (minimum interference between actuation input signal and sensor output signal) because the transducers use different mechanisms for signal transmission: the piezoelectric actuators use electrical channels while the fiber-optic sensors use optical means. Since they use two separate mechanisms for transmitting signals, there is virtually no interference between them.

4.2 Hybrid piezoelectric/fiber-optic sensor sheets

Hybrid piezoelectric/fiber-optic (HyPFO) sensor sheets have been developed [13, 24]. The concept of a HyPFO sensor sheet is a generalization of the concept of a SMART layer [25], which is a device that comprises a thin dielectric film containing an embedded network of distributed piezoelectric actuator/sensors. Such a device can be mounted on the surface of a metallic structure or embedded inside a composite material structure during fabrication of the structure. Besides the piezoelectric

and fiber-optic sensors, other types of sensors, such as strain gauges, MEMS (microelectromechanical systems) and TRD (time-rate-of-decay) temperature sensors, can also be integrated in the sensor layer. The advantages of a hybrid sensor layer include the following:

- It is not necessary to install each sensor individually on a structure. Sensors are embedded in thin, flexible films that can easily be mounted on structures in minimal amounts of installation time.
- Multiple measurements can be performed. For example, fiber-optic sensors can be used to measure temperatures, PZTs can be used to measure concentrations of hydrogen, and sensors of both types can be used to monitor acoustic emissions.

4.3 Demonstration of damage detectability of the system

In order to demonstrate the damage detection capability of the active hybrid diagnostic system, tests were conducted on both aluminum and fiber-reinforced composite plates.

4.3.1 Damage detection on an aluminum plate

Damage detection tests were performed on an aluminum plate with a dimension of $500 \times 500 \times 1.5\,\text{mm}^3$. Four PZT actuators, each having a diameter of 6.35 mm and thickness of 0.25 mm, and a single grating sensor were mounted on the top surface of the aluminum plate. The single FBG is written at the center wavelength of 1550 nm and has a gauge length of 10 mm. The layout of the PZT actuators and the FBG sensors is shown in Figure 4. In the tests, five-peak burst waveforms were used to excite the structure [26].

To demonstrate the ability of the hybrid PZT/FBG system to detect damage, a $50 \times 8 \times 2\,\text{mm}^3$ small stick-on patch was attached to the path between actuators and fiber grating sensors to simulate damage. Using the same five-peak burst waveforms as above, a series of tests with different frequencies of actuation signal were carried out. The sensor output was recorded and compared with the baseline taken before a stick-on patch was put on the aluminum plate. The typical test results are shown in Figure 5.

There are many modes of Lamb waves that can be actuated in a plate with a PZT. In our test, according to the product of frequency and thickness, only two kinds of modes were activated, i.e., A_0 and S_0. When

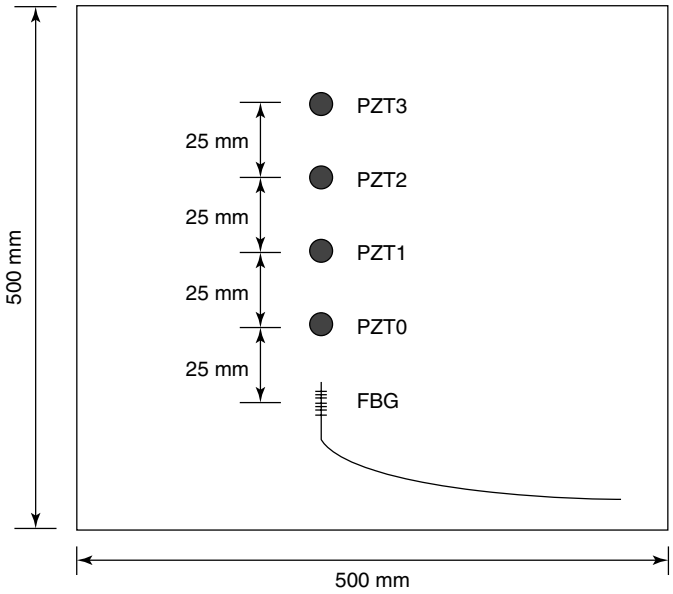


Figure 4. Layout of PZT actuators and the FBG sensor on the aluminum plate.

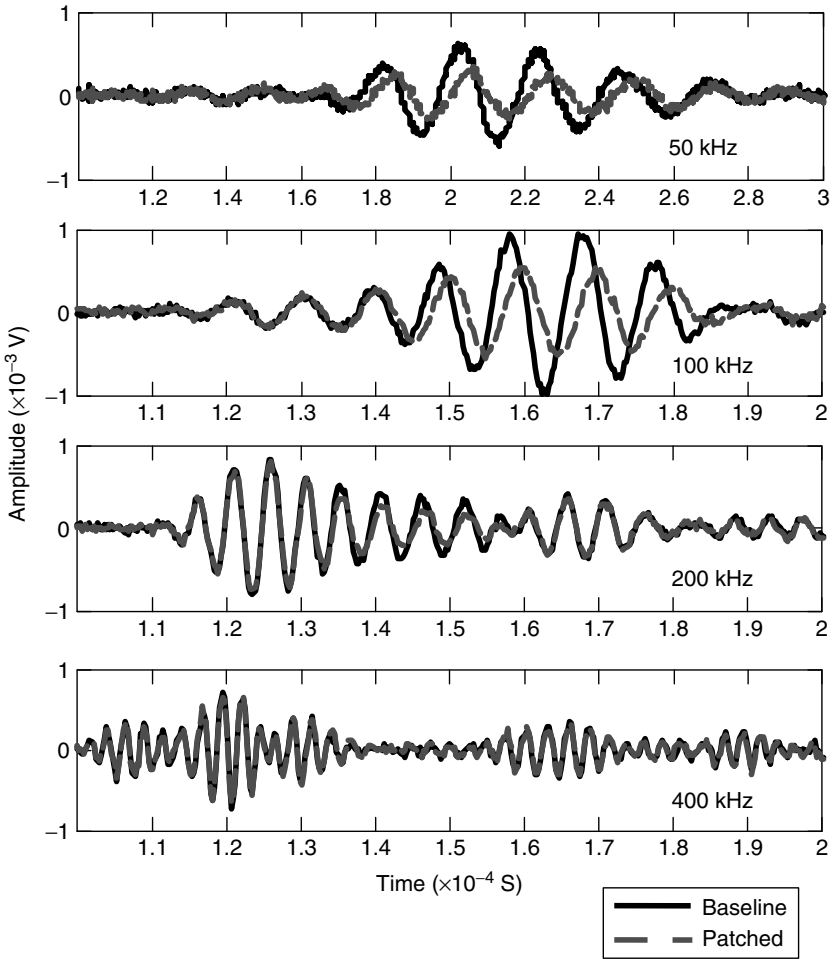


Figure 5. Damage detection test results from the aluminum plate.

the wavelengths are large in relation to the thickness of the plate, the fundamental symmetric mode and the fundamental antisymmetric mode are equivalent to the extensional and flexural waves, respectively.

The data presented in Figure 5 were signals actuated by PZT1 and collected by the FBG, the direction of which is parallel to the path of signal transmission. It was demonstrated that the signal amplitude is related to the distance and the cross angle between the signal transmission path and the FBG's direction [25]. From the signals we obtained, it can be seen that both A_0 and S_0 mode waves are detected. S_0 mode waves travel faster in this case, so they arrive first. When the center frequencies of the actuation signal are 50 and 100 kHz, respectively, S_0 mode waves carry less energy and are not significant. When the center frequency of the actuation signal is 200 kHz, S_0 mode waves and A_0 mode waves are both significant. We cannot see apparent changes in S_0 mode waves owing to the stick-on patch. However, we can see significant changes in both amplitude and phase in A_0 mode waves, which arrived later than S_0 mode waves. The results are consistent with established theories. The A_0 mode flexural waves show high sensitivity to weight added on their transmission path. The simulated damage was thus clearly identified. When the center frequency of actuation signal is 400 kHz, there are no A_0 mode waves showing up in the sensor signal, which is caused by the limitation of FBG sensors. When FBG sensors are utilized for stress

wave detection, they cannot pick up a clear signal when the wavelength of the stress wave in question is smaller than the gauge length of the FBG sensor. In the case of 400 kHz, the wavelength of the A_0 mode decreases to about 6 mm, which is less than the 10-mm gauge length of the FBG. As Minardo [16] has demonstrated, no A_0 mode stress wave could be detected. There are only S_0 mode signals detected in this case, which are extensional waves and not sensitive to weight added on its traveling path, and hence the signal after damage and baseline signal are almost identical.

4.3.2 A quasi-distributed damage-detection scheme for composite plates

A damage detection scheme for a composite plate was implemented using the hybrid PTZ/FBG active-sensing system. The configuration of the specimen tested is given in Figure 6. Two PZT actuators and three FBG sensors (engraved on the same optical fiber) were embedded in a Hybrid SMART layer [13], which was then bonded onto the composite plate.

A delamination damage was introduced by repeated impacts on the center of the zone covered by the sensor network. A pitch–catch test was then carried out with a five-peak wave actuation signal of 100 kHz. Typical signals are shown in Figure 7. It can be seen that there are significant changes in both the amplitude and phase of the detected signals due to the delamination damage. To locate and evaluate the extent of damage, a damage index can be employed. When multiple paths are affected by the damage area, which results in big damage indices for these paths, their effects add up to show a heightened intensity of colors. This display technique can be used as a fast imaging method to help visualize the approximate location and extent of damage [28], as shown in Figure 8.

The extent of delamination damage in the composite plate was examined by an ultrasonic scan. The results clearly showed that severe damage was inflicted on the center area of the zone. The damage detected with the hybrid system is consistent with the result of the ultrasonic scan and X-ray image of the impact damage. In summary, it has been demonstrated that debond in composite structures can be conveniently identified by the proposed hybrid PZT/FBG active-sensing scheme [26].

4.4 Potential application for monitoring a large area

As described in [29], there is a major challenge in the networking of a multitude of piezoelectric sensors applied to physically large structures because of a large number of connection wires and big signal noise from long-distance communication. The hybrid system could be a potential solution for the applications of SHM on large structures, such as the health monitoring of bridge and long pipeline structures [27].

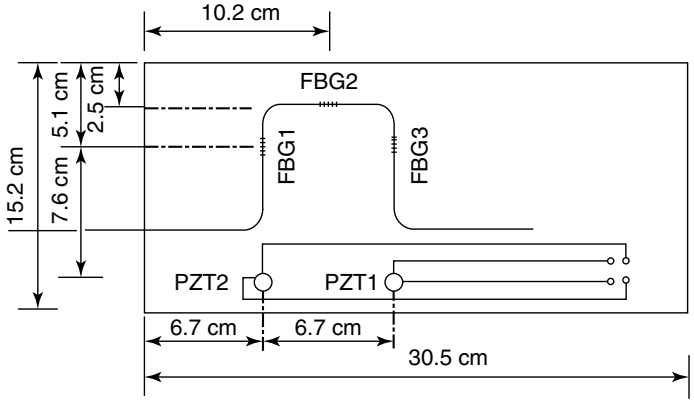


Figure 6. Configuration of the composite specimen. [Reproduced with permission from Ref. 27. © 2006, Qing *et al.*]

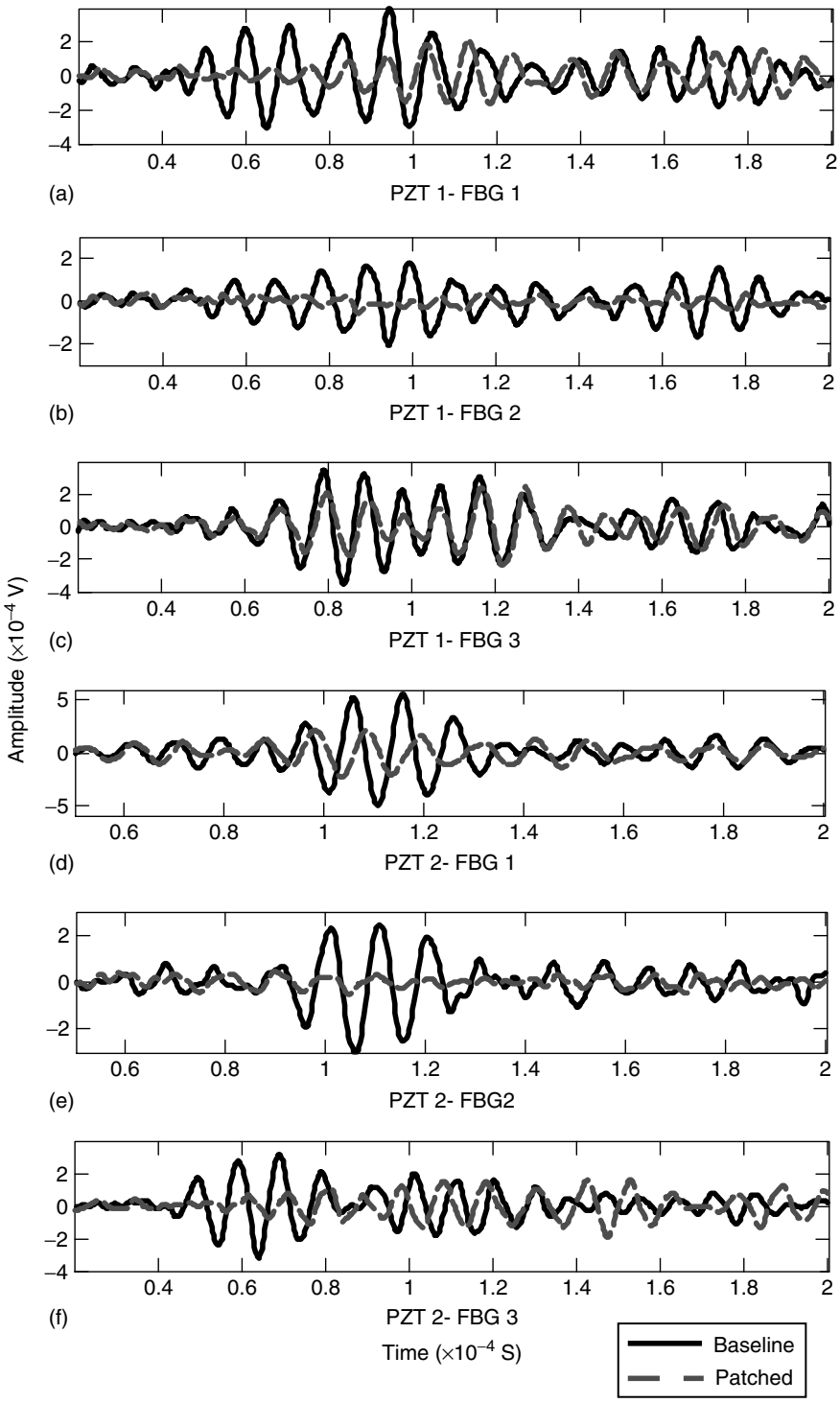


Figure 7. Damage detection results with the hybrid active-sensing system.

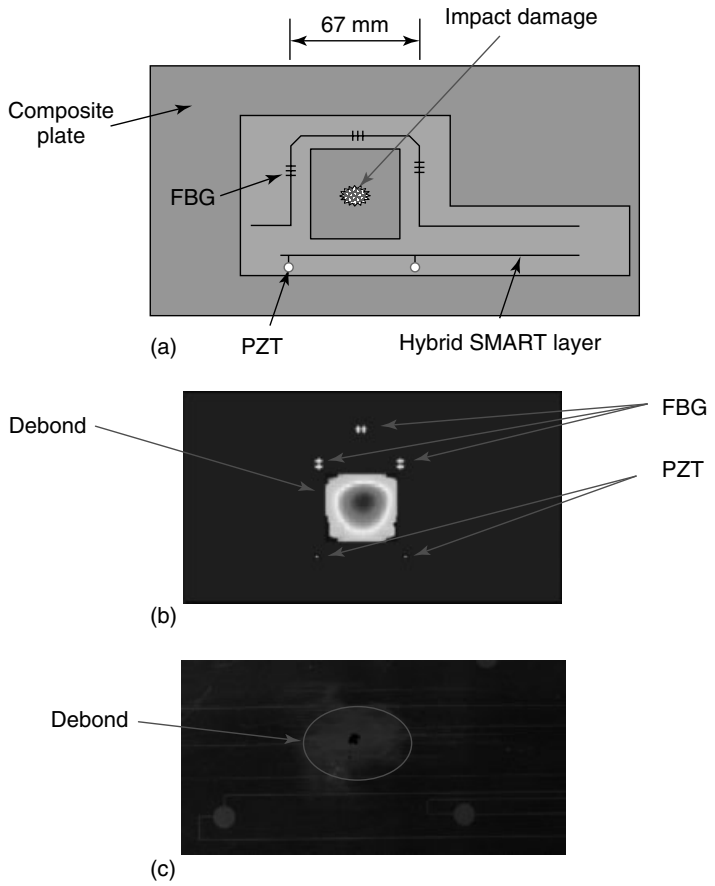


Figure 8. Hybrid sensor network used for damage detection. (a) Hybrid layer mounted on the surface of a composite plate, (b) diagnostic image of impact damage on the composite plate, and (c) X-ray image of the damage on the composite plate.

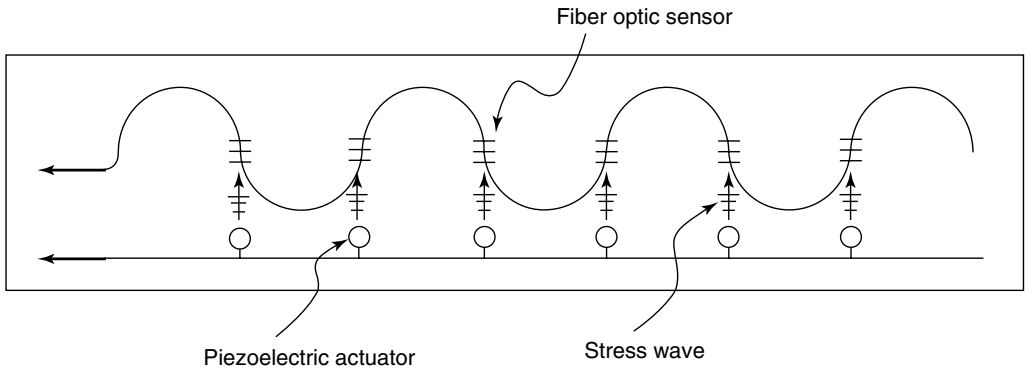


Figure 9. A typical design of hybrid piezoelectric/fiber-optic sensor network. [Reproduced with permission from Ref. 27. © 2006, Qing *et al.*]

Figure 9 shows a typical design of hybrid piezoelectric/fiber-optic sensor network used in the SHM system. In the network, all piezoelectric actuators are connected to a diagnostic instrument through a shielded cable. The distributed FBGs are used as sensors. The piezoelectric actuators can be simply connected together in series or in parallel, and then they generate stress waves in the structure simultaneously. By using a set of control wires, each piezoelectric actuator can also be used to generate stress wave in the structure individually. The smart cable with piezoelectric actuators and the optic fiber with distributed gratings can be placed on a thin carrier film, and then will be permanently bonded on the composite or metal structure to be monitored. The changes caused from the degradation of material properties (or corrosion) around all diagnostic paths will be identified. The design is particularly good for global damage monitoring of large composite or metal structures, such as fuel storage tanks for space ships, bridges with composite repairs, and metal pipelines. The advantages of this design include the following: (i) simpler wiring; (ii) long-distance sensor signal transmission; and (iii) capability to detect ultrasonic stress waves, quasi-static strain, and impacts. With all those features, an ideal health management solution based on hybrid FBG/PZT network can be offered by monitoring multiple physical parameters using the same system. At first, strain, temperature, and impact can be monitored continuously using FBG sensors, which can provide environment loading information and is critical to structural heath assessment. Then, when there is any abnormal sign, such as high stress or impact at certain locations, the active system can be triggered to send out diagnostic signals using the PZT actuators to detect whether any damage occurred.

5 CONCLUDING REMARKS

In order to resolve some issues in the piezoelectric sensor–based SHM system, such as coupling between sensors and actuators and long-distance signal transmission for large structure applications, extensive research on the development of hybrid piezoelectric/fiber-optic diagnostic systems has been conducted. A general review of the efforts made is presented in the article. An integrated hybrid active system and its application approach are also introduced, and damage detection capability with the system is demonstrated. One of the key issues for the hybrid PZT/FBG systems is the interrogation technique for FBG sensors, which dominates the sensitivity and dynamic range of the system. Besides, the fabrication technology of FBG sensors also needs to be improved to achieve high sensitivity for ultrasonic stress wave, which requires smaller grating length compared to quasi-static strain measurement.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from NASA to Acellent and the National Science Foundation (Grant No. CMS-0200399) to Stanford University for this development.

REFERENCES

[1] Choi K, Chang FK. Identification of impact force and location using distributed sensors. *Journal of American Institute of Aeronautics and Astronautics* 1996 **34**(1):136–142.

[2] Kersey AD, Davis MA, Patrick HJ, LeBlanc M, Koo KP, Atkins CG, Putnam MA, Friebele EJ. Fiber grating sensors. *Journal of Lightwave Technology* 1997 **15**(8):1442–1463.

[3] Han JH, Rew KH, Lee I. An experimental study of active vibration control of composite structures with a piezo-ceramic actuator and a piezo-film sensor. *Smart Materials and Structures* 1997 **6**(5):549–558.

[4] Qing X, Beard S, Kumar A, Chan H, Ikegami R. Advances in the development of built-in diagnostic system for filament wound composite structures. *Composite Science and Technology* 2006 **66**:1694–1702.

[5] Rose JL, Rajana K, Hansch MKT. Ultrasonic guided waves for NDE of adhesively bonded structures. *Journal of Adhesion* 1995 **50**:71–82.

[6] Rose JL, Ditri J. Pulse-echo and through transmission lamb wave techniques for adhesive bond inspection. *British Journal Of Non-Destructive Testing* 1992 **34**(12):591–594.

[7] Merzbachery CI, Kersey AD, Friebele EJ. Fiber optic sensors in concrete structures: a review. *Smart Materials and Structures* 1996 **5**(2):196–208.

[8] Idrissy RL, Kodindoumay MB, Kersey AD, Davis MA. Multiplexed Bragg grating optical fiber sensors for damage evaluation in highway bridges. *Smart Materials and Structures* 1998 **7**(2):209–216.

[9] Todd M, Johnson GA, Vohra S, Chen-Chang C, Danver B, Malsawma L. Civil infrastructure monitoring with fiber Bragg gratings sensor arrays. In *Proceedings of Structural Health Monitoring 2000*, Chang FK (ed). Technomic: Lancaster, PA, 1999, pp. 359–368.

[10] Fox JJ, Glass BJ. Impact of integrated vehicle health management (IVHM) technologies on ground operations for reusable launch vehicles (RLVs) and spacecraft. *IEEE Aerospace Conference Proceedings*, 2000; Vol. 2, pp. 179–186.

[11] Wang G, Pran K. Ship hull structure monitoring using fiber optic sensors. *Proceedings of European COST F3 Conference on System Identification and Structure Health Monitoring*, Universidad Politécnica de Madrid: Spain, 2000; Vol. 1, pp. 15–17.

[12] Lin M, Powers WT, Qing X, Kumar A, Beard SJ. Hybrid piezoelectric/fiber optic SMART layers for structural health monitoring. *Proceeding of 1st European Workshop on Structural Health Monitoring*, France, 2002; pp. 641–648.

[13] Qing X, Kumar A, Zhang C, Gonzalez IF, Guo G, Chang FK. A hybrid piezoelectric/fiber optic diagnostic system for structural health monitoring. *Smart Materials and Structures* 2005 **14**(5): 98–103.

[14] Perez I, Cui HL, Udd E. Acoustic emission detection using fiber Bragg gratings. *Proceeding of SPIE* 2001 **4328**:209–215.

[15] Ogisu T, Shimanuki M, Kiyoshima S, Okabe Y, Takeda N. Damage growth detection of composite laminate structure using embedded FBG sensor/PZT actuator hybrid system. *Proceedings of SPIE* 2005 **5758**:93–104.

[16] Minardo A, Cusano A, Bemini R, Zeni L, Giordano M. Fiber Bragg gratings as ultrasonic waves sensors. *Proceedings of SPIE* 2004 **5502**:84–87.

[17] Pierce SG, Philp WR, Culshaw B, Gachagan A, McNab A, Hayward G, Lecuyer F. Surface-bonded optical fiber sensors for the inspection of CFRP plates using ultrasonic lamb waves. *Smart Materials and Structures* 1996 **5**(6):776–787.

[18] Betz DC. Lamb wave detection and source location using fiber Bragg grating rosettes. *Proceedings of SPIE* 2003 **5050**:117–127.

[19] Betz DC, Thursby G, Culshaw B, Staszewski WJ. Identification of structural damage using multifunctional Bragg grating sensors: I. Theory and implementation. *Smart Materials and Structures* 2006 **15**(5):1305–1312.

[20] Betz DC, Staszewski WJ, Thursby G, Culshaw B. Structural damage identification using multifunctional Bragg grating sensors: II. Damage detection results and analysis. *Smart Materials and Structures* 2006 **15**(5):1313–1322.

[21] Kersey AD, Davis MA, Patrick HJ, LeBlanc M, Koo KP, Askins CG, Putnam MA, Friebele EJ. Fiber grating sensors. *Journal of Lightwave Technology* 1997 **15**(8):1442–1463.

[22] Kersey AD, Berkoff TA, Morey WW. Multiplexed fiber Bragg grating strain-sensor system with a fiber fabry–perot wavelength filter. *Optics Letters* 1993 **18**(16):1370–1372.

[23] Baldwin CS, Vizzini AJ. Acoustic emission crack detection with FBG. *Proceedings of SPIE* 2003 **5050**:133–143.

[24] Lin Mark, Qing X, *Hybrid Piezoelectric/Fiber Optic Sensor Sheets—Multiple Sensors of Different Types Could be Installed On Or In Structures*, MFS-31846-1, *NASA Tech Briefs*, July, 2004.

[25] Lin M, Qing X, Kumar A, Beard S. SMART Layer and SMART suitcase for structural health monitoring applications. *Proceedings of SPIE on Smart Structures and Material Systems* 2001 **3329**: 98–106.

[26] Wu Z, Qing X, Chang FK. Debond detection for composite laminate plates with a distributed hybrid PZT/FBG sensor network. *Journal of Intelligent Material Systems and Structures (Submitted)*.

[27] Qing X, Wu Z, Chang FK, Ghosh, K, Karbhari V, Sikorsk C. Monitoring the disbond of externally bonded CFRP composite strips for rehabilitation of bridges. *Proceeding of the Third European Workshop on Structural Health Monitoring*, Granada, Spain, 2006; pp. 463–470.

[28] Beard S, Qing PX, Hamilton M, Zhang DC. Multifunctional software suite for structural health monitoring using smart technology. *Proceedings of the 2nd European Workshop on Structural Health Monitoring*, Munich, 2004; pp. 101–108.

[29] Qing X, Beard S, Kumar A, Yu P, Chan HL, Zhang D, Ooi T, Marotta SA. Practical requirements for implementation and usage of SHM systems on aerospace structures. *Proceedings of the 5th International Workshop on Structural Health Monitoring*, Stanford University, 2005; pp. 1502–1509.

FURTHER READING

González IF, Wu ZJ, Chang FK. Health monitoring by Means of hybrid diagnostic System. In *Proceeding of Structural Health Monitoring*, Chang FK (ed). DEStech Publications: Stanford, CA, 2005, pp. 732–740.

Chapter 77

Energy Harvesting using Thermoelectric Materials

Daniel J. Inman[1] and Henry A. Sodano[2]

[1] *Center for Intelligent Material Systems and Structures, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*
[2] *Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ, USA*

1 INTRODUCTION

The concept of developing completely self-power electronics has received significant interest over the past decade partially fueled by the recent advances in wireless and structural health monitoring (SHM) technology [1]. As described in earlier articles, SHM is an integrated process of data acquisition, signal processing, and statistical inference used to track and assure the safety and performance of a structure. SHM has immediate benefits and market potential. However, because of the need for a widely dispersed sensor network to effectively monitor large civil structures and the need to dispense with increasing wiring harnesses in vehicle applications, it has been realized that wireless SHM sensors should be used to reduce the system complexity (*see* **Wireless Sensor Network Platforms**). However, when using wireless sensors it is necessary that a portable power supply be used, such as batteries. Because batteries must be periodically replaced, the use of batteries to power wireless sensors greatly restricts sensor placement and location. For instance, it is often necessary to embed the sensor in the structure to achieve the desired response or to be close to areas of high damage probability. By implementing methods of obtaining ambient energy from the sensors surroundings, a wireless device could be designed to function indefinitely. Additionally, one major use of SHM systems would be to quickly inspect civil structures or military vehicles after catastrophic events, which typically cause widespread power outages thus making the use of wired power sources impossible.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

The recent advances in low power electronics and wireless technology have made the hardware necessary to perform the required tasks available, but an effective power source has yet to be identified. Therefore, the development of wireless SHM systems revolves around the ability to capture ambient energy surrounding the device and convert it into usable electrical power. In many studies, piezoelectric materials have been utilized to capture the ambient vibrations around a system and convert them into electrical power [2] (*see* **On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System**). However, the energy generated by piezoelectric materials is typically far too small to directly power most electronic devices. The issue of too little energy can be compensated for by using energy storage methods to accumulate sufficient energy to power the electronics in short bursts. A second and more developed method of obtaining energy from ambient sources is through the use of thermoelectric generators (TEGs), which capitalize on thermal gradients. TEGs use the Seebeck effect [3], which describes the current generated when the junction of two dissimilar metals/semiconductors experiences a temperature difference. Using this idea, numerous *p*-type and *n*-type junctions are arranged electrically in series and thermally in parallel to construct the TEG. Thus, if a thermal gradient is applied to the device, it will generate an electrical current that can be utilized to power other electronics. By implementing power harvesting devices, autonomous portable systems can be developed that do not depend on traditional methods for providing power, such as the battery, which has a limited operating life.

TEGs are an established technology and have been used for capturing ambient energy in various applications. Lawrence and Snyder [4] suggest a potential method of retrieving electric energy from the temperature difference that exists between the soil and the air. To test their concept, a prototype was built without the TEG and the heat flow was measured to estimate the amount of power that could be obtained. The results showed that a maximum instantaneous power of approximately 0.4 mW could be generated by the thermoelectric device. Rowe *et al.* [5] investigate the ability to construct a large TEG capable of supplying 100 W of power from hot waste water. The system tested used numerous thermoelectric devices placed between two cambers, one with flowing hot water and the other with cold water flowing in the opposite direction thus maximizing the heat exchange. With a total of 36 modules, each with 31 thermocouples, 95 W of power could be generated. Fleming *et al.* [6] investigated the use of TEG to provide electrical power in micro air vehicles. A TEG was mounted on the exhaust system of an OS max 61 internal combustion engine and was shown to generate 380 mW of power.

Several authors have studied the use of TEGs for obtaining waste energy from the exhaust of automobiles. Birckolz *et al.* [7] worked with Porsche to develop a TEG unit that would fit around the exhaust pipe of the 944 engine. The unit was experimentally tested and found to generate an open circuit voltage of 22 V and a total power of 58 W. Similarly, Matsubara [8] constructed an exhaust system using ten TEG modules and a liquid heat exchanger to maximize the thermal gradient. The system was tested on a 2000-cc class automobile and shown to produce 266 W of power. Bass *et al.* [9] investigated the placement of a TEG in the vertical muffler of a class 8 diesel truck. The system generated 1 kW of power, thus allowing it to be employed as a substitute for the truck's alternator. By removing the alternator from the engine, the power delivered to the driveshaft was increased by 3–5 hp, providing an increase in fuel efficiency and a reduction in emissions. For more information on TEG applications in automobiles, see Vázquez *et al.* [10]. A review of previous research in TEGs shows that there are several key parameters that dictate the amount of energy generated; the most important of these are the surface area in contact with the thermal source, temperature gradient across the device, and the thermal conductivity between the TEG and the source.

The idea to use thermoelectric devices to capture ambient energy from a system is not new. However, TEGs have typically been used simply to determine the extent of power capable of being generated rather than investigating applications and uses of the energy. Furthermore, most previous research efforts have utilized liquid heat exchangers or forced convection to significantly improve heat flow and power generation, but require complex cooling loops and systems. These previous studies also commonly do not consider the amount of energy applied to the cooling system and therefore only report gross levels of power. In this article, the use of TEGs

as power harvesting devices that do not have an active heat exchanger, but function as a completely passive power scavenging system with SHM applications, is reviewed. The motivation for investigating a passive power generation device stems from the need to identify effective power sources for the development of self-powered wireless SHM systems. These systems could be placed in a desired location without regular replacement of batteries or maintenance, as most wireless devices currently do. Because of the remote placement of the TEG, it becomes impossible to incorporate active heat exchangers into the system.

When deploying a wireless SHM network, the sensors are placed in a variety of locations, some of which have ambient vibration and others that have ambient thermal gradients. Because the ambient source of energy may vary around the structure, it is important to have a variety of methods to capture the energy. In the case of vibrating structures such as a bridge or tall building, piezoelectric materials may form the ideal energy harvesting device, but if ambient vibration is not present, then alternative power generation devices must be studied. Here two potential locations for a wireless SHM system that does not have significant vibration are examined: first, those that are exposed to the sun but do not vibrate, such as a dam; and second, structures that are subjected to high temperatures, such as a boiler, rocket, or combustion engine. When performing SHM, it is typically not necessary to continuously check its integrity, but rather to perform an analysis once a day or perhaps a week. Therefore, the energy harvesting device should be able to store the energy until it is needed by the electronics. The ability to store the generated energy from piezoelectric materials in a rechargeable battery has been shown by Sodano *et al.* [11, 12]. Here two TEG systems are discussed in the context of using them to recharge a completely discharged nickel–metal hydride battery and the relative charge time will be compared with that of piezoelectric-based power harvesting systems.

2 BASIC THEORY

TEGs use the Seebeck effect, which describes the current generated when the junction of two dissimilar metals experiences a temperature difference. This effect is also responsible for the current generated by a thermocouple. Using this idea, numerous p-type and n-type semiconductor junctions are arranged electrically in series and thermally in parallel to construct the TEG. Thus, when an electrical current is applied to the TEG a thermal gradient is generated, allowing the device to function as a small solid state heat pump. Inversely, if a thermal gradient is applied to the device, it will generate an electrical current that can be utilized to power other electronics. The simplest model of TEG is

$$V = \alpha \Delta T \tag{1}$$

Here V is the voltage generated, ΔT is the temperature difference (gradient) across the TEG and the constant of proportionality α is called the *Seebeck coefficient*. Thus the basic TEG model produces a dc voltage proportional to the temperature difference across the device. If the TEG is connected to a load, then one simple model is to consider the TEG electrically as a voltage source with some internal resistance applied to some resistive load. This is illustrated in Figure 1. In the figure V_{TEG} is the voltage generated according to equation (1), R_{TEG} is the internal electrical resistance provided by the TEG material and R_{load} is the electrical resistance provided by an external load to the TEG. The load resistance is that of the electrical resistance associated with a direct use device or the resistance offered by a storage device (either a capacitor or battery). For the case when R_{load} is a battery or a capacitor being charged, a diode must be included in this circuit between the generator and the load to prevent a backflow of energy from the battery to the TEG when ambient energy is not being harvested. As a general rule, the power delivered is maximized when the load resistance or impedance is equal to the internal resistance of the TEG.

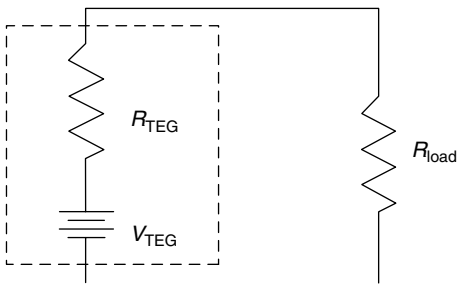


Figure 1. Simple electric circuit model of a TEG.

Practical TEG devices consist of many *pn* junctions connected electrically in series and thermally in parallel (so the same thermal gradient appears across each *pn* pair) as illustrated in Figure 2. If n is the number of such junctions, then the voltage can be modeled simply as

$$V = n\alpha\Delta T \tag{2}$$

Therefore, the voltage or power output is increased by increasing the number of junctions. Traditional thermoelectrics are built by physically assembling discrete blocks of n-doped or p-doped thermoelectric elements onto electrical circuits as illustrated in Figure 2. The circuits are typically on ceramic plates that have metal lines to route the electrical current. The thermoelectric elements are placed onto the ceramic plates using pick-and-place assembly equipment and attached to the metal lines on the ceramic plates using solder. On the basis of this method of manufacturing, the ability to scale the size of the thermoelectrics to smaller scales has been limited.

Scaling thermoelectric devices to a smaller size is desirable because the power density (watts per square centimeter that it can generate) is inversely proportional to the length (height in Figure 2) of the thermoelectric element. By scaling to smaller geometries, higher power densities can be achieved, leading to a more efficient device. This advantage has spawned a rapid growth of research into microelectromechanical system (MEMS) thermoelectric devices constructed using thin-film technology. Thin-film thermoelectric materials can be grown by conventional semiconductor deposition methods, and devices can be fabricated using conventional semiconductor microfabrication techniques. The resulting TEG devices can be produced in large quantities with small dimensions, allowing them to be easily integrated into the SHM sensing platform.

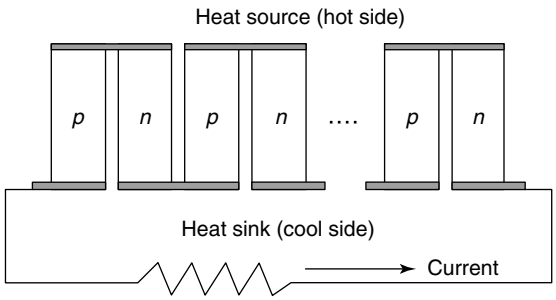


Figure 2. Arrangement of a thermal electric generator device.

3 EXAMPLES

Two thermoelectric systems will be presented here to determine their ability to convert solar energy and waste heat into electrical energy and its storage in a conventional rechargeable battery. The results will be compared with energy harvesting using photovoltaic materials (solar cells). The first system developed used solar radiation to heat one side of the TEG, while the cold side was bonded to a metallic structure that functioned as a heat sink with a large thermal mass or capacity. Because direct solar energy was not sufficient to heat the hot side of the TEG, the idea of a "greenhouse" was combined with a solar concentrator to elevate the temperature. A greenhouse functions by allowing visible light emitted by the sun to pass through the transparent surface, thus heating the objects inside and converting the visible light into thermal infrared waves that cannot penetrate the surrounding medium to exit, causing the heat to build up. Because dark objects do not reflect much visible light, but rather convert almost all of it into thermal energy, a blackbody heat sink was placed inside the container to increase the thermal energy stored. To further increase the amount of thermal energy applied to the hot side of the heat sink a solar concentrator was used. A solar concentrator focuses a large area of sunlight onto a smaller area, thus increasing the thermal energy stored in the greenhouse. A schematic describing the layout of the system is shown in Figure 3. This system does not preclude the use of a solar cell, but could be applied with one to capitalize on both the incident light as well as the thermal energy released.

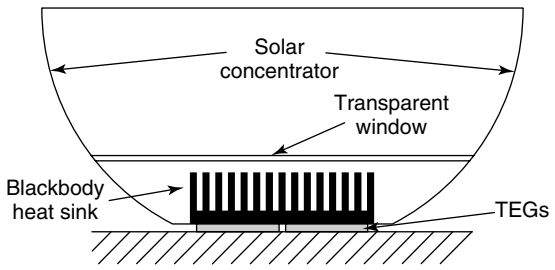


Figure 3. Schematic of solar harvesting device.

The second method of capturing ambient energy was from waste heat that would be available from combustion engines, boilers, or furnaces. To simulate the energy available in these locations, a hot plate was used. A heat sink was attached to the hot plate to allow more energy to be removed from the cold side of the TEG and facilitate a larger power output. To investigate completely passive power harvesting methods, the systems did not include a means of providing forced convection, which would greatly increase the energy produced by the TEG, but would require additional energy for the active cooling systems.

4 EXPERIMENTAL TESTING

A prototype of each power harvesting system was constructed and experimentally tested to identify their ability to capture thermal energy, convert it to electrical power and then use the electrical output to charge a discharged nickel–metal hydride battery. Both systems used eight Melcor thermoelectric coolers (model HT-4-30) to generate the electrical signal. The physical properties of this device are shown in Table 1. The first system used solar energy to develop a thermal gradient over the TEG. To increase the amount of solar energy applied to the power harvesting device, an aluminum parabolic solar concentrator was used to reflect the visible light onto a black body aluminum heat sink ($7.62 \times 7.62 \times 3.175$ cm). The heat sink was painted flat black to allow more of the visible light to be converted into thermal energy. Above the heat sink was a Plexiglas window that trapped the thermal energy, thus allowing it to accumulate and increase the thermal gradient over the TEG. The solar concentrator, heat sink, and TEGs were bonded to a steel plate, which acted as the host structure; this setup is shown in Figure 4.

Table 1. Dimensions and electrical properties of the TEG

Property	Value
Dimensions ($w \times l \times h$)	30 mm × 34 mm × 3.2 mm
Maximum temperature difference	77 °C
Number of thermocouple junctions	127
Device resistance	3.78 Ω
Resistivity	1.37 Ω cm^{-1}

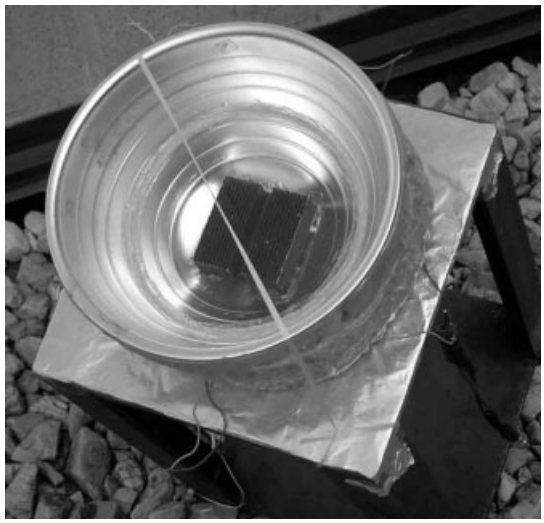


Figure 4. Experimental setup of the solar harvesting system.

The second source of ambient thermal energy tested here is simply a structure with an elevated temperature available from aircraft engines, Heating Ventilation and Air Conditioning (HVAC) systems, boilers, or furnaces. To simulate the energy available in these locations, a hot plate was used. Because the purpose of this study is to investigate completely passive power harvesting methods, the systems did not include a means of providing forced convection, which would greatly increase the energy produced by the TEG, but would require additional energy for the active cooling systems. A prototype of the power harvesting system was constructed and experimentally tested to identify its ability to capture thermal energy, convert it to electrical power, and then use the electrical output to charge a discharged nickel–metal hydride battery. The setup used eight Melcor thermoelectric coolers (model HT-4-30) to generate the electrical signal. To simulate this environment, the TEGs were fixed between a heat sink and a thin aluminum plate, which was then attached to a hot plate using thermal grease. The experimental setup is shown in Figure 5. To monitor the temperature of the hot and cold sides of the TEG, Omega CO-1 thermocouples were used. Because the thermocouple was only 0.13-mm thick, it could be bonded to the hot and cold sides of the TEG.

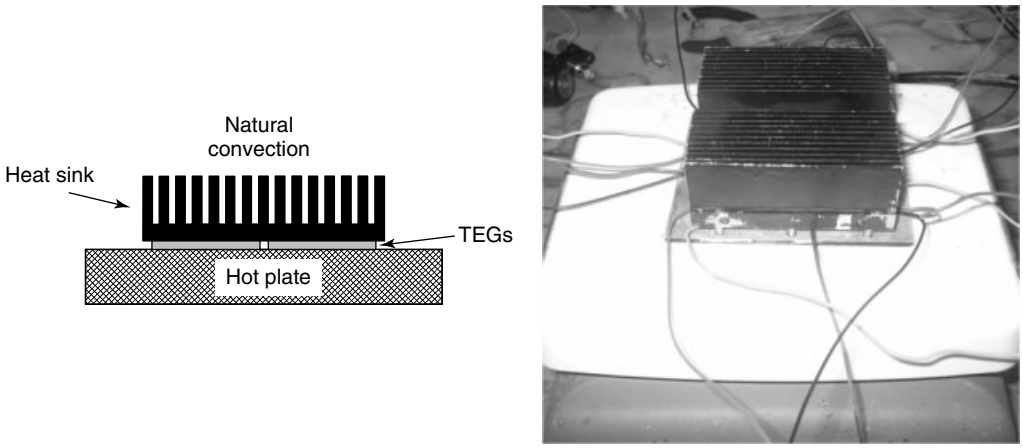


Figure 5. Experimental setup of the energy harvesting system.

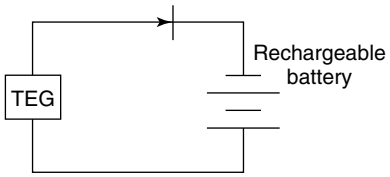


Figure 6. Diagram of circuit used to recharge batteries.

When an SHM system is applied, it is often only necessary that the integrity of that structure is not monitored continuously, but rather only periodically—every day, week, or even month. Continuous testing is not necessary because in many cases, damage will occur over an extended period of time from aging or fatigue. If the SHM system operates only for a small percentage of the time, the energy generated by the power harvesting system must be stored until needed by the electronics. The storage of the electricity is also necessary to ensure that when the electronics are switched on, sufficient power is available. Typically, if energy is to be stored for an extended period of time, it is more effective to use a rechargeable battery. Therefore, a simple circuit can be been constructed to take the electrical energy generated by the two power harvesting devices and store it in a nickel–metal hydride battery. The circuit used in this study is shown in Figure 6. The diode is a necessary piece of this circuit because it forces current to only flow in one direction. If the diode is not present, the TEG would draw power form the battery during times when the voltage generated was less than the voltage of the battery or if the thermal gradient is reversed. Because the output of the TEG is a dc signal, it does not require a means of rectifying, which is a source of energy loss in piezoelectric power harvesting.

5 RESULTS

To experimentally illustrate the power generated by the TEG, it is placed on a hot plate, thus allowing the applied temperature to be accurately monitored. For the system of eight TEGs electrically in series used in this power harvesting device, the Seebeck coefficient is not defined and therefore was fit using experimental data and equation (2). To determine this coefficient, the voltage output of the TEG was measured as the temperature of the hot plate was varied for two different configurations of the TEG modules. A schematic of the different TEG configurations is provided in Figure 7. The resulting voltage output for each configuration and the linear fits for each case are shown in Figure 8. From this figure, it can be see that the Seebeck coefficient for the two different configurations varies. The change in the Seebeck coefficient occurs because the stacking of the TEG modules causes variation in the thermal gradient over each module and in the heat sink's ability to remove energy from the cold side due to a change in surface area in contact with the thermal source. The effect of the configuration on the heat sink's performance can be

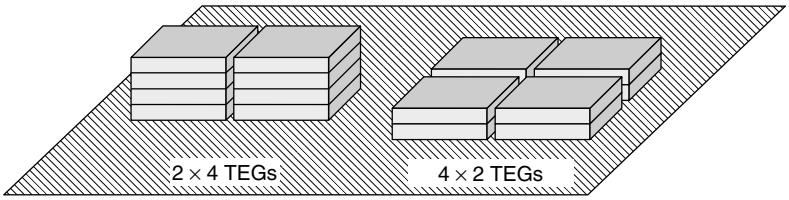


Figure 7. Three different configurations of the TEGs with the hot plate, the footprint is 20.4 cm^2 with two TEGs and 40.8 cm^2 with four TEGs.

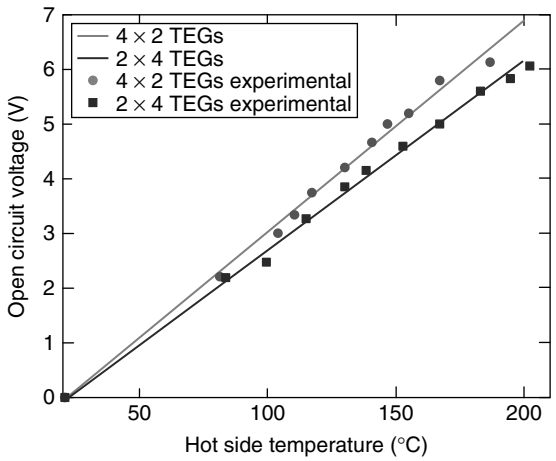


Figure 8. Output voltage of the thermoelectric generator as a function of the hot side temperature.

illustrated by the air temperature around the heat sink. When the heat sink is close to the hot plate, the temperature of the air surrounding it is higher and thus reduces the total thermal difference between the hot and cold sides. However, if too many TEGs are stacked, the gradient between the hot plate and heat sink is divided over each TEG and a large portion of energy is allowed to be transmitted from the sides of the stack, thus reducing the overall power output. The resulting Seebeck coefficient of the configuration using two layers of four TEGs and four layers of two TEGs are 2.96×10^{-4} and 2.68×10^{-4} V K^{-1} respectively.

The current output for each of these systems can now be found using the following relationship [12]:

$$I = \frac{G\alpha D_T}{2\rho} \tag{3}$$

where I is the current output, G is the area per unit length of the thermoelectric element, and ρ is the resistivity. Now the power generated by each system can easily be determined by equating the power output of the energy harvesting system to the product of the output voltage and current. The results show that the power harvesting module does indeed vary linearly and the thermal gradient or number of modules necessary to generate a particular amount of power for an electronic device can be easily determined. Furthermore, the amount of power that can be generated from TEGs when placed on a 200 °C surface can be as high as 40 mW without any form of convective heat transfer, as shown in Figure 9. As a comparison, when a piezoelectric device subjected to the level of vibration typically found in an automobile engine, it can only generate a maximum of 2 mW [10]. This point illustrates the substantial difference in the power available from TEGs and the idea that more powerful electronics can be powered when using them.

After the amount of power generated by the TEG for various hot side temperatures had been quantified, the ability of each device to recharge a nickel–metal

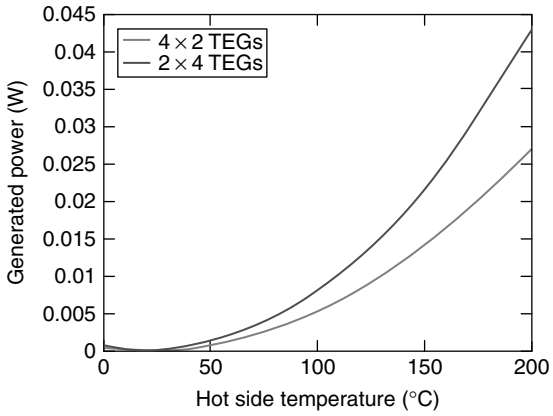


Figure 9. Power generated by the thermal harvesting system as a function of the hot side temperature.

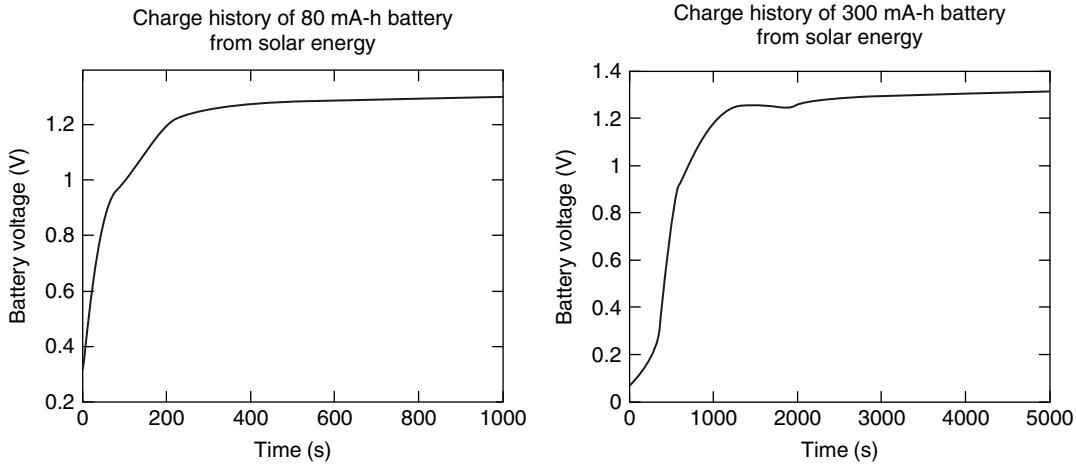


Figure 10. Charge histories of rechargeable batteries from solar harvesting system for an 80 mA·h battery and a 300 mA·h battery.

hydride battery was determined. The circuit layout shown in Figure 6 was used to collect the electrical output of the TEG and store it in the rechargeable battery. For this study, an 80 and a 300 mA·h battery were tested with the solar harvesting system and only a 300 mA·h battery was tested for the waste heat device. (The unit milliampere-hour "mA·h" indicates the capacity of a battery: an 80 mA·h capacity implies the battery will last for 1 h if subjected to 80-mA discharge current.) An 80 mA·h battery is roughly the size of a watch battery and a 300 mA·h battery is slightly less than half the size of a typical AAA battery.

First, the solar harvesting device was investigated by placing the system in direct sunlight with an ambient temperature of 29.4 °C. Both batteries were charged on the same day with relatively short time between each test, therefore limiting the variation between tests. The hot side temperature of the solar harvesting device was measured to be approximately 52.8 °C. The resulting charge times for both batteries are shown in Figure 10. From these figures it is apparent that the TEG system developed in this study can effectively harvest solar radiation and use that energy to quickly recharge a discharged battery. Next, the waste heat system was tested and the resulting charge time is shown in Figure 11. Once again, it is clear that the TEG is very capable of storing converted energy in a battery.

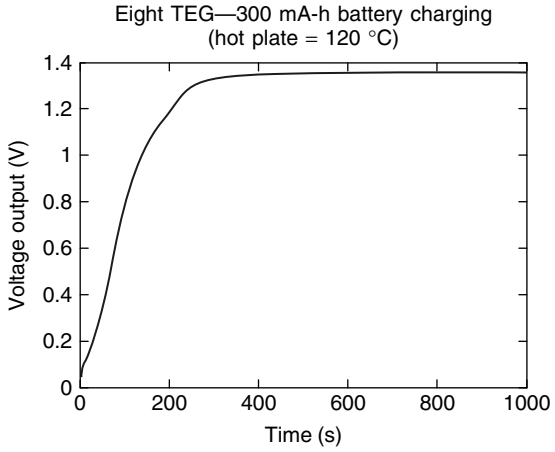


Figure 11. Charge history of a 300 mA·h nickel–metal hydride battery from waste heat harvesting system.

6 COMPARISON OF HARVESTING TECHNOLOGIES

After identifying the ability to use a TEG for the purpose of recharging small batteries, the results were compared with those found using piezoelectric materials by Sodano *et al.* [11]. The time required for each system to charge the battery to a cell voltage of 1.2 V was the measure in each case and the results are provided in Table 2. The charge time listed for the TEG devices are from Figures 10 and 11,

Table 2. Time needed to achieve the battery's cell voltage by a TEG with solar energy and waste heat, and for a piezoelectric material experiencing random vibration [11]

Capacity of the battery (mA·h)	Charge time from solar energy (min)	Charge time from waste heat (min)	Charge time from vibration using piezoelectric materials (h)
80	3.3	NA	2
300	17.3	3.5	5.8

NA, Not Available

while the charge time for the piezoelectric system are for a PZT patch excited at an amplitude and random frequency measured from a typical automobile [11]. The results from charging a battery using piezoelectric materials are only provided to give an idea of the results from previous studies performed under realistic conditions and do not represent that a TEG will always outperform piezoelectric materials. (Note: this comparison is not intended to imply that piezoelectric materials are less effective or efficient than a TEG and these tests were performed on two very different systems.) The charge times in Table 2 may seem to be lower than possible; the time listed is simply to achieve the battery's cell voltage and not a full charge. The time needed to take the battery up to capacity would be longer, however, the times listed for both the piezoelectric and the TEGs both represent the time required to reach the cell voltage from a fully discharged state. To determine the time needed to provide a complete charge to the battery, a charge controller would be needed. The TEG is capable of quick recharge times because of its large current output, whereas the piezoelectric material supplies a very high voltage at a low current. To give an idea of the difference in these two devices, the impedance of one TEG is approximately 3 Ω while that of the piezoelectric device is approximately 10 000 Ω. Owing to the lower impedance of the TEG, eight modules had to be connected electrically in series to boost the output voltage to the required 1.2 V of the battery; however, this lower impedance also makes the TEG far more suited for use with rechargeable batteries, which charge faster with larger currents.

7 CONCLUSIONS

The usefulness of wireless SHM systems is established in many articles in this encyclopedia. Energy harvesting is an enabling technology for SHM. The long term goal of many SHM systems is to develop a completely self-powered electronic module that may be placed in a remote location without the concern of replacing the power supply. Generating energy from ambient sources using piezoelectric materials to harvest the vibration energy surrounding a system is one approach. For those cases where the structure in question does not experience sufficient vibration to generate the power levels needed to sustain the operation of the electronics, thermoelectric harvesting may be appropriate. This article has focused on using TEGs to generate an electrical signal from ambient thermal gradients. Two potential examples of this technology are presented: harvesting of solar radiation and harvesting of waste heat. For each application, an experimental prototype was constructed and tested to determine the effectiveness in recharging a discharged nickel–metal hydride battery.

Each application of the thermal generator was implemented such that only conductive heat transfer was present. Most of the previous studies have utilized an active heat exchanger to increase the energy output. However, the use of active convection causes the net power output to be reduced due to the energy applied to the heat exchanger. Therefore, this article has focused on demonstrating that with fairly small thermal gradients and only conductive heat transfer the thermal electric generator can form an effective power harvesting source. The results indicate that TEG does produce significantly more power than a piezoelectric device and that the time needed to recharge a battery is significantly lower. Thus in applications where thermal gradients are present, TEGs make a suitable choice for powering SHM applications [12].

REFERENCES

[1] Spencer BF, Ruiz-Sandoval ME, Kurata N. Smart sensing technology: opportunities and challenges.

Journal of Structural Control and Health Monitoring 2004 **11**(4):349–368.

[2] Sodano HA, Park G, Inman DJ. A review of power harvesting using piezoelectric materials. *The Shock and Vibration Digest* 2003 **36**(3):197–206.

[3] Goldsmith HE. *Applications of Thermo-Electricity*, *Methuen's Monographs on Physical Science*. John Wiley & Sons: 1960.

[4] Lawrence EE, Snyder GJ. A study of heat sink performance in air and soil for use in a thermoelectric energy harvesting device. *Proceedings of the 21st International Conference on Thermoelectronics*. Portland, OR, 25–29 August 2002; pp. 446–449.

[5] Rowe MD, Min G, Williams SG, Aoune A, Matsuura K, Kuznetsov VL, Fu LW. Thermoelectric recovery of waste heat—case studies. *Proceedings of the 32nd Intersociety Energy Conversion Engineering Conference*. Honolulu, HI, 27 July–1 August 1997; pp. 1075–1079.

[6] Fleming J, Ng W, Ghamaty S. Thermoelectric-based power system for unmanned-air-vehicle/microair-vehicle applications. *Journal of Aircraft* 2004 **41**(3):674–676.

[7] Birckolz U, Grob E, Stohrer U, Voss K. Conversion of waste exhaust heat in automobile using FeSi2 thermoelements. *Proceeding of the 7th International Conference on Thermoelectric Energy Conversion*. Arlington, VA, 1988; pp. 124–128.

[8] Matsubara K. Development of a high efficient thermoelectric stack for a waste exhaust heat recovery of vehicles. *Proceedings of the 21st International Conference on Thermoelectronics*. Portland, OR, 25–29 August 2002; pp. 418–423.

[9] Bass JC, Elsner, NB, Leavitt FA. Performance 1 kW thermoelectric generator for diesel engines. *Proceedings of AIP Conference* 1994 **316**(1): 295–298.

[10] Vázquez J, Sanz-Bobi MA, Palacios R, Arenas A. State of the art of thermoelectric generators based on heat recovered from the exhaust gases of automobiles. *Proceedings of the 7th European Workshop on Thermoelectrics*. Pamplona, Spain, 2002, Paper No. 17.

[11] Sodano HA, Park G, Inman DJ. Generation and storage of electricity from power harvesting devices. *Journal of Intelligent Material Systems and Structures* 2005 **16**(1):67–75.

[12] Sodano HA, Park G, Inman DJ. Comparison of piezoelectric energy harvesting devices for recharging batteries. *Journal of Intelligent Material Systems and Structures* 2005 **16**(10): 799–807.

FURTHER READING

Grisso BL, Kim J, Farmer JR, Ha DS, Inman DJ. Autonomous impedance-based SHM utilizing harvested energy. In *Structural Health Monitoring 2007*, Chang F-K (ed). DEStech Publications, September 11–13, 2007; pp. 1373–1380.

Chapter 85
Loads Monitoring in Aerospace Structures

Steve Reed
QinetiQ, Farnborough, UK

1 INTRODUCTION

The aim of this article is to describe the philosophy behind operational loads monitoring (OLM) in aerospace structures, within the context of in-service structural health monitoring (SHM). In-service OLM or measurement can be defined as *a structural usage substantiation activity*. This activity involves the capture, analysis, and reporting of representative, directly measured strain data or derived loads and associated flight parameters from a sample of suitably instrumented in-service air vehicles within a fleet. Within this article, the principle aims of loads monitoring programs are described. Thereafter, the basic approaches taken and the use of data captured within these programs are outlined. Readers are referred to **Operational Loads Sensors**; **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**; **Eddy-current *in situ* Sensors for SHM**; **Fiber-optic Sensor Principles**; and **Directed Energy Sensors/Actuators** for further information on sensor technologies. Furthermore, although military usage papers dominate the loads monitoring literature (*see* **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Flight Demonstration of a SHM System on a USAF Fighter Airplane**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**; **Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance**; and **Aerospace Applications of SMART Layer Technology**), there is a growing realization of the importance of such programs in the civil arena and this is reflected in **Commercial Fixed-wing Aircraft**; **History of SHM for Commercial Transport Aircraft**; **Video Landing Parameter Surveys**; and **Landing Gear**.

OLM is related to flight loads measurement (FLM) or flight loads survey (FLS) activity; however, these are distinct activities with differing aims. FLM or FLS programs are primarily associated with the substantiation of loads models where instrumented aircraft

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

are flown to particular points in the flight envelope, whereas OLM is concerned with substantiating the usage of the aircraft in service.

2 AIMS OF LOADS MONITORING PROGRAMS

The aims of the loads monitoring program dictate the measurands, conduct, and output from the program. An illustration of the most common aims is provided in Figure 1.

2.1 Substantiation of design and qualification fatigue usage spectra

During the design of an aircraft a great many assumptions about the usage of the aircraft. These assumptions are drawn from a variety of sources including data from previous aircraft types in similar roles with adjustments made for changes in capability and data from operators specifying intended usage, role, and stores carriage for the new type. Nowadays, a large proportion of major military aircraft programs are multinational projects for which design and qualification spectra are generally developed from a combination of intended usages. Furthermore, in response to rapidly changing military threats, it is highly unlikely that an aircraft will be used solely within its originally intended role throughout its life.

The fatigue spectra derived from these assumptions and used in the design and qualification of the aircraft cover a wide range of loading actions including maneuver-driven symmetric and asymmetric aerodynamic and inertia loads, gust loading, buffet loads, landing impact loads, ground handling loads, thermal loading, and engine thrust or propeller torque loading.

The increased use by the military of modified civil aircraft adds further issues to consider. Usage in a military environment may be considerably different to that assumed in civil design. Also, the modifications made to the aircraft (e.g., conversion from a passenger aircraft to an air-to-air tanker or refueling aircraft; *see* **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**) may introduce significant changes in local loads within the structure and an associated change in the usage of the aircraft in service.

Therefore, one of the essential initial elements of the loads monitoring program is to capture the

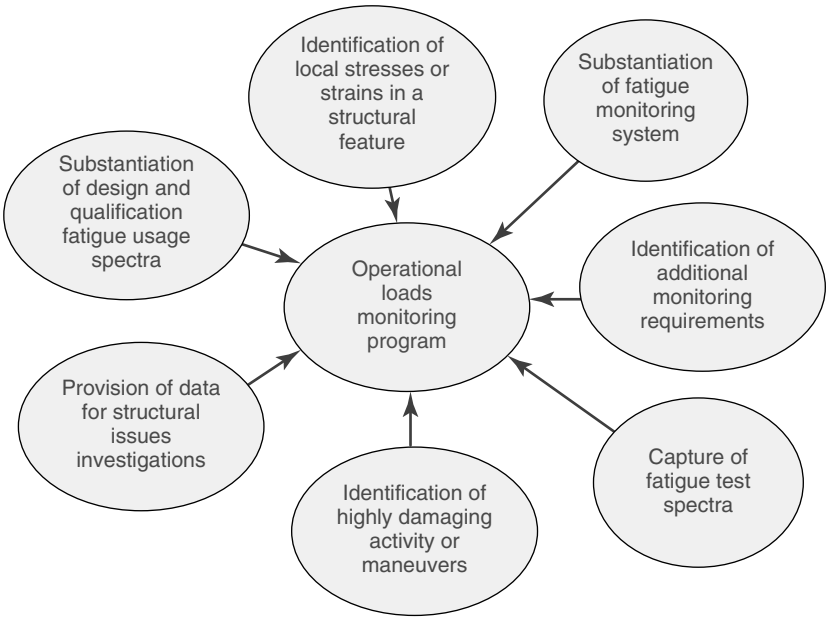


Figure 1. Aims of loads monitoring program.

fatigue design and qualification spectra and identify appropriate methods of assessing these spectra against in-service usage [1]. Thereafter, it is necessary to identify which of these spectra can be substantiated by the loads monitoring program and how this will be done. Direct measurement of loads or strains using electrical resistance strain gauges [2, 3] has generally been the preferred method of substantiating fatigue spectra. However, there are circumstances where measuring strains may not be practicable and substantiation of spectra using parametric methods or a combination of strain and parametric data may be required.

From a safety perspective, it is essential that any usage spectra more severe than assumed in design are identified and remedial action put in place. Secondly, usage more benign than assumed has the potential for life extension or to accommodate increases in usage severity in the future.

2.2 Identification of local stresses or strains in a structural feature

OLM may also be required to determine the local stresses or strains in a structural feature from overall forces and moments, often in association with finite element or boundary element modeling. This scenario might occur when a feature cannot be exercised adequately on a fatigue test and hence comparison with test loading would not be appropriate.

2.3 Substantiation of fatigue monitoring system

The usage of all aircraft is tracked in some way (*see* **Usage Management of Military Aircraft Structures**). This ranges from counting flight cycles or flying hours to full aircraft monitoring using strain-gauge fits installed on every aircraft in a fleet. All of these tracking methods contain usage assumptions and even the most sophisticated SHM systems cannot measure usage in all structural components of the aircraft directly and hence assumptions have to be made. Although OLM substantiation of a structural monitor can expose lack of understanding of the loading actions, generally it is the usage assumptions that prove to be in greater error.

2.4 Identification of additional monitoring requirements

A prime function of the loads monitoring program is often to identify the need for any additional monitoring requirements where there are insufficient margins of safety or unacceptable levels of risk, using the initial tracking measurands. Where the tracking of these unmonitored structural components, using these simple measurands, such as flight cycles, does not correlate sufficiently with fatigue damage accrual or where there is insufficient margin of safety, then additional monitoring options may need to be considered.

Novel aircraft types where there are no historical data (such as unmanned air vehicles (UAV)) are likely to have a significant level of uncertainty in their design usage data. Also, projects that use civil-designed aircraft in military roles are most likely to face uncertainty in usage or deviation in usage from that assumed in design. Challenging fatigue-life requirements or targets, limited fatigue testing, and an aim to reduce inspection burdens may drive the designer to consider additional monitoring to invoke the reduced safety factors associated with monitored structure. In these cases, OLM data can be used to support monitor development.

2.5 Capture of fatigue test spectra

One of the aims of an OLM program can be the generation of fatigue test spectra and additional data to validate these spectra (e.g., detailed stresses, accelerations, and pressurization cycles). Where this is a requirement, this will become a significant design driver for the OLM program and will necessitate a Skopinski-type [4] calibration in a loads rig (discussed later). Such calibration allows the OLM data to be described in global loading terms of shear, bending moment, and torque for fatigue test load spectra generation.

2.6 Identification of highly damaging activity or maneuvers

Identification of highly damaging activities or maneuvers is a key aim. Comparisons between maximum and minimum load/stress or strain conditions should

be made with limit-load and maximum fatigue allowable conditions. If such risks are identified, then either additional evidence should be required to support an increase in the loads envelope or appropriate limitations may need to be introduced into the operational release for the aircraft.

Secondly, the aim is to identify the structural cost, in fatigue damage terms, of activities or maneuvers. This information can then be used by operators to make informed decisions as to the need to undertake these maneuvers. These data can also be used to assist in the education of aircrews in methods of reducing fatigue consumption without significantly affecting mission capability.

2.7 Provision of data for structural issues investigations

OLM programs have frequently been used to provide data to support structural investigations and to exploit potential fleet life extension. In reality, structural issues or exploitation of life extension potential have often been the catalyst needed to initiate an OLM program for a fleet.

3 OLM MEASUREMENT SYSTEM DESIGN

3.1 Review of flight loads measurement and previous OLM installations

The OLM measurement requirements should be driven by the program aims. As previously discussed, OLM is, to a certain extent, a natural progression from FLM or FLS. Generally, FLM instrumentation will be far more extensive than can be justified for an OLM program and FLM installations are often not sufficiently robust to be used in-service. However, FLM experience should be invaluable in establishing the most suitable elements of the instrumentation, strain range, and sample rate data for inclusion in the OLM fit and the FLM data should be reviewed with this aim in mind.

3.2 Strain-gauge installations

Although electrical resistance strain gauges have been in use for over 50 years [2, 3], they are still the mainstay of instrumentation used in OLM data capture. Fiber-optic strain sensors (*see* **Fiber-optic Sensor Principles**; **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**; **Fiber Bragg Grating Sensors**; **Novel Fiber-optic Sensors**; and **Reliable Use of Fiber-optic Sensors**) have made inroads into civil and marine application and aerospace technology demonstrators (*see* **The Character of SHM in Civil Engineering**; **Monitoring Marine Structures**), and it is highly likely that these systems will be used for in-service aerospace applications in the near future [5]. Electrical resistance strain gauges rely upon the Wheatstone bridge principle, whereby a strain change in the structure to which the measuring arms of the bridge are attached causes an imbalance of resistance across the bridge and generate an output voltage. This voltage is measured, conditioned, and converted to a measure of the strain/stress/load in the structure, using appropriate calibration equations (discussed later). Strain-gauge bridges can be arranged in several different configurations, such as a 1/4 bridge, a 1/2 bridge, a full bridge or rosette depending on the measurement requirement [6].

Where the aims of the program include fatigue spectra substantiation, strain gauges will have to be located and configured to meet this aim. For example, for direct comparison with a full-scale fatigue, the OLM installation would be a subset of the fatigue test strain-gauge fit based upon knowledge from analysis of the loads paths, strain responses, and constraint effects of the test article. If the test does not exercise a feature thoroughly, the probability of that feature failing at the end of test is extremely small. Hence, for unfailed features, when in-service data are captured and if the OLM spectra are more severe than the test spectra, low-fatigue lives will be identified and supplementary evidence will be required to support a fatigue-life clearance.

Also, although assessment in terms of fatigue damage (or crack growth rate) is the ultimate fatigue-life currency, this can often be quite a blunt tool. Damage comparison is only really valid when the stress or strain spectra are similar shapes. Therefore, if the strain range or spectrum shape seen on the fatigue test is highly dissimilar to that seen on the in-service aircraft, a direct damage comparison may not be appropriate and the feature may not be substantiated by the test. Therefore, once data are available,

test and in-service spectra comparison, in addition to fatigue damage comparisons, should be undertaken.

Alternatively, if the aims of the OLM program were to include provision of data for assembly of fatigue test load spectra, then OLM strain gauging would have to be located and configured so that the outputs from the strain-gauge bridges can be combined to produce bending moment, shear force, and torque values in the structural components required to be tested and suitable loads calibration methods would be required (discussed later).

Where program aims include substantiation of fatigue monitoring systems, strain-gauge installation would have to be located to capture the prime loading in the critical features protected by the monitor. For wing fatigue monitoring, this has generally been wing-to-fuselage or wing center-line joint features, but the critical features will be aircraft specific.

In practice, most OLM programs will have several of the aims described above and a combination of configurations will be required. Where there is some leeway or options to be considered for locating strain-gauge installations, they should be located in regions of low strain gradient. Locations should be chosen to avoid vulnerability to in-service damage from maintenance traffic and general use. Historically, the reliability of strain-gauge installations across OLM and direct strain-based fatigue monitoring systems has been mixed. Poor access for installing gauges does increase the risk of a poor-quality bonding and the associated risk of repair being required. Surface preparation is essential to obtain a high-quality bond [6]. For most applications, the in-service temperature range will require elevated temperature cure for gauge epoxy-based (or similar) adhesives. A temperature survey will be required to identify heat sinks in the structure to ensure accurate curing temperatures for the strain-gauge adhesive. Twisted-pair wiring lengths from the strain gauge to the signal conditioning unit should be kept to a minimum to reduce the effects of cable resistance and interference. The use of distributed master/slave data acquisition unit (DAU) connected by digital signals is often recommended to combat excessive cable lengths. Additionally, the path of strain gauge and instrumentation cables should be considered from an interference perspective to avoid sources such as high-power electric motors. Finally, where required, the method of calibrating the strain-gauge arrangement should be considered. For example, for easily removable components, loads calibration undertaken on static test machines represents an attractive and cost-effective option, although consideration has to be given to end constraints and stiffness. It is efficient practice to install secondary or backup gauges during installation, where a similar response from primary and secondary systems is expected.

3.3 Parametric and discrete measurands

The instrumentation most commonly used in OLM programs to obtain parametric and discrete data include accelerometers, synchros, potentiometers, tachogenerators, gyroscopes, flow meters, switches, thermometers, and pressure transducers. The requirements for parametric data captured within an OLM program should again be driven by program aims and applicable aircraft environment standards [7, 8]. Traditionally, the prime role of the parametric data has been to provide a basis for understanding the strain/stress/loads. A simple example of this would be comparing an inner wing bending moment plot with the normal acceleration parameter. Combinations of parameters, such as airspeed, altitude, accelerations, roll, and pitch rates, can be used to identify a flight condition or particular maneuvers. This may be essential in cases where airborne calibration or check calibration of strain gauges is required. Furthermore, where direct strain monitoring of a feature is impractical, it may be necessary to use parametric data or a combination of strain and parametric data to substantiate the fatigue spectra. Substantiation of the performance of an aspect of the parametric monitoring systems, such as normal acceleration-based systems, is also generally undertaken using flight test instrumentation standard accelerometers.

3.4 Capturing parametric data from data bus systems

The majority of modern aircraft used in military roles are fitted with data bus systems. Military aircraft generally use Mil Std 1553B or similar data bus systems and civil-derived aircraft tend to use ARINC 429 or similar systems. Many of the parameters needed to support an OLM program may well already

be available on these data bus systems. Although capturing the data stream from a data bus is an attractive option, care should be taken to ensure that the parameter provided on the bus is fit for the purposes of OLM in terms of sample rate, refresh rate, processor prioritization effects, time synchronization, accuracy, and resolution of the source sensor.

3.5 Capturing parametric data from aircraft instrumentation

Where an aircraft is not fitted with a data bus or where the required parameters are not available or suitable on the bus, tapping into existing aircraft instrumentation or the installation of OLM-specific instrumentation will be required. From a design, qualification, and through-life support perspective, use of existing aircraft instrumentation represents a far more cost-effective approach. However, tapping into primary flight instrumentation systems, such as pilot's altimeter, should generally be avoided owing to the increased risk to these systems. Additionally, as with data bus systems, an understanding of the specification and output signal format of the instrumentation is essential. Also, consideration has to be made as to whether the tapping loads up the signal. Where aircraft are fitted with accident data recorders (ADRs) or crash survivable memory units (CSMUs) tapping into input or output ports for these systems can also be considered for OLM data capture. Interfacing with ADR systems has generally been discouraged in the past owing to the risk of interfering with these systems; however, modern systems, in particular, are often fitted with output data ports and these options should be considered.

3.6 Introducing OLM-specific instrumentation

It is most likely that OLM-specific instrumentation will be required (such as accelerometers). Consideration should be given to equipment reliability, vulnerability in the operational environment, long-term support, additional data-acquisition signal conditioning requirements, as well as fitness for the purpose. Furthermore, experience has shown that even short-term OLM programs last considerably longer than ever envisaged at the outset and long-term support of the installation should be considered.

3.7 Numbers of instrumented aircraft

Historically, instrumentation of around 10% of a large fleet, ideally with a minimum OLM fit of two aircraft for small fleets, has proven a reasonable proportion for planning OLM installations. However, each program should be judged on a case-by-case basis and the factors considered should include program aims, fleet size and disposition, roles within the fleet, fleet-within-fleet issues (including structural build standard), OLM installation capability and reliability, cost, likely attrition, and volume of data required against collection time.

3.8 Data acquisition

The purpose of the data-acquisition system is to take measurements of the parameters (e.g., acceleration, velocity, or displacement) and provide data, which are used for analysis. The basic block diagram, as illustrated in Figure 2, illustrates the essential elements of an instrumentation system. The parameter is sensed by a transducer, which in turn produces data, often in the form of an analog voltage. These raw data are passed to the signal conditioning unit to be processed into a suitable form having been scaled, filtered, and usually converted into digital data. Thereafter, the data are transmitted to a data recorder unit.

3.9 Data sampling rate and antialiasing filters

Experience has shown that data should be sampled at a minimum of 10 times the highest frequency of the structurally significant modes, with appropriate antialiasing filters, to have confidence in capturing a close approximation to the strain peaks and troughs in the time history. Where there is insufficient data to identify the highest significant structural mode with confidence, an initial data sampling plan should be introduced with significantly higher rates than are likely to be needed and only reduced when analysis of the time histories shows that the sample rate can be reduced safely. Methods such as subsampling

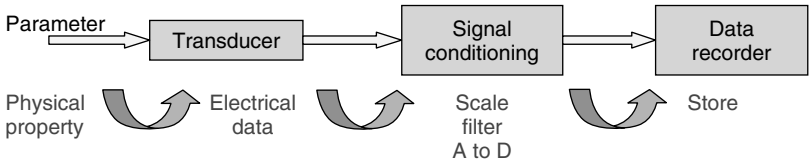


Figure 2. Data-acquisition cycle.

and repeat damage sums as well as identification of maximum and minimum values of subsampled data will identify when the sample rate reduction is significantly affecting the data capture. This is illustrated in Figure 3. Some may argue that to capture the frequency content in a signal it only needs to be sampled at twice that frequency. While it is correct that the frequency content will be identified, the magnitude of the peak and trough of a cycle cannot be reliably captured by only two points in the cycle. Regions particularly vulnerable to higher frequency dynamic loading, particularly for combat and associated trainer aircraft, may include the following:

- tailplane/foreplane/canard
- fin
- rear fuselage
- outer wing
- flying controls and lift devices (e.g., rudder, airbrake, slats, flaps, ailerons leading edges devices, and leading edge root extensions)
- external stores and pylons
- undercarriage and support structure.

4 CALIBRATION

Calibration refers to the process of determining the relation between the output (or response) of a measuring instrument and the value of the input quantity or attribute. Calibration can be undertaken using various methods with different aims and it is imperative that the calibration tests performed are relevant to the program aims and produce data consistent with these aims. Some of the more commonly used calibration methods, primarily for loads to fatigue test

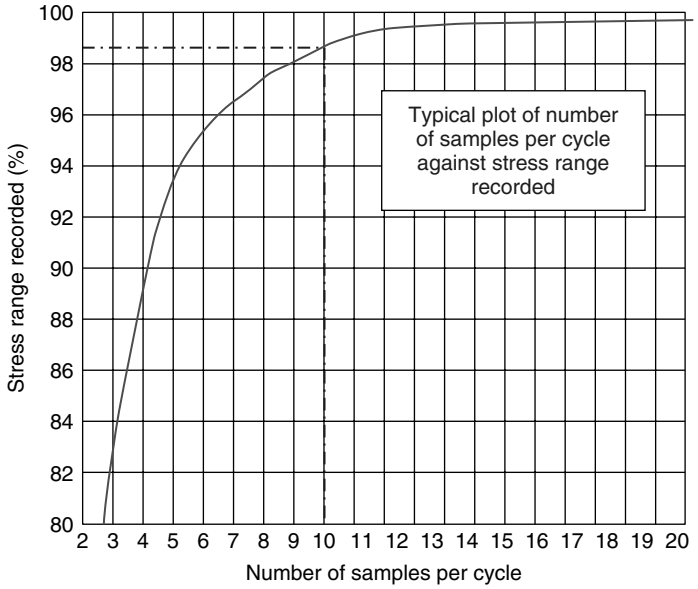


Figure 3. Identification of sample rate effects.

or loads input to a stress model, are discussed in the following sections.

4.1 Setting of physical datum values

Irrespective of the method of calibration used, there will be a need to establish the appropriate physical loading condition that corresponds to the zero output from the strain-gauge installation. Strain-gauge bridges are frequently balanced to give zero output in the calibration state with no external load applied. Bridge balancing can sometimes be achieved with the instrumented component in a zero load state but, in general, the "zero position" of a gauge installation output has to be set to a physical known output or datum before analysis can proceed. Frequently, ground conditions are used in this exercise since mass information is likely to be readily obtainable and an expected output can be derived from suitable models.

4.2 On-aircraft strain-gauge load calibration

Where the aims of the OLM program require the identification of loads, such as in the determination of fatigue test spectra, a more complex approach to calibration is required. As previously mentioned, the methods generally used for loads calibration are developed from the work undertaken at NACA (forerunner of NASA) in the early 1950s and published by Skopinski *et al.* [4]. Load calibration is a method that permits the measurement in flight of the shear, bending moment, and the torque on the principal lifting or control surfaces of an aircraft. Although it is accepted that the stress in a structural member may not be a simple function of the three loads of interest, processes have been developed for numerically combining the outputs of several strain-gauge bridges in a way that the loads may be obtained. A simple typical installation for a two-spar structure, using four-active-arm strain-gauge bridges is presented in Figure 4.

Ideally, gauges would be placed so that a bending moment bridge would respond only to bending and a shear bridge would respond only to shear and, so on. However, this is only true for an elementary truss-type beam arrangement and, in practical structures, more complex interactions between loading mechanisms exist. With care, gauge locations can be chosen and configured so that the bending bridge output contains predominantly bending moment effects and, so on. The loads on the surface of a wing, for example, can be specified by three orthogonal forces, normal forces, longitudinal forces, and lateral forces (Figure 5) and by three orthogonal moments, bending moment, torque, and in-plane bending. Thereafter, loads models are used to transfer these overall loads to local loads within the structure.

The strain in a structural member or component can therefore be expected to be some function of these six quantities and this strain response must be taken into account in any method of relating strain-gauge bridge output to load. In real structures, these relationships can be complicated further as the strain in a wing root, for example, may be affected not only by the loads outboard of the measurement station but also by loads on the opposite wing or inboard of the measurement station. This *carryover effect*, as it is

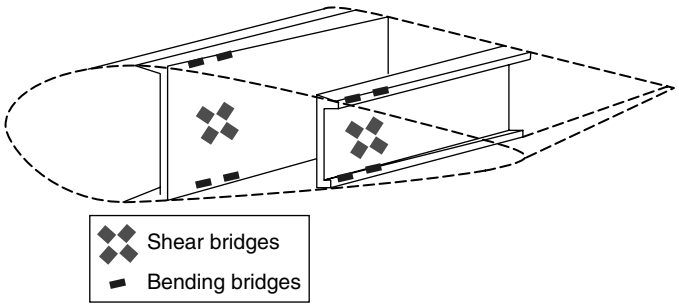


Figure 4. Example of strain-gauge bridges for loads calibration.

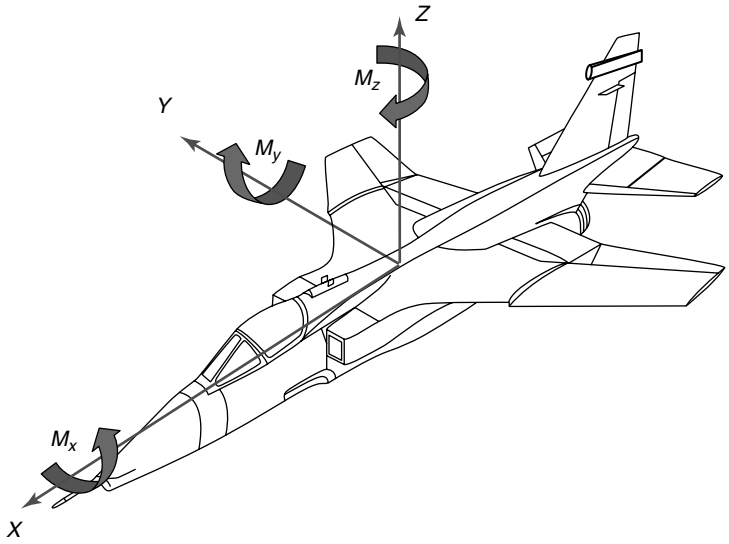


Figure 5. Forces and moments acting on an aircraft.

termed, can be particularly significant for unsymmetrical loading actions [4]. However, certain simplifications to these six quantities can be made. For a wing structure, the stress in the structural member and hence the output from the strain gauges mounted upon that structure may be taken to be a function of the three principal terms of aerodynamic load investigation, shear, bending moment, and torque. Additional loading actions may need to be considered for other structural features; for example, drag loads are generally highly significant for external stores and hence may need to be considered.

For major lifting surfaces, the method of obtaining the relationship between the output of the strain-gauge bridges and the shear, bending moment, and torque loads is by applying a series of point loads and employing the principle of superposition. This assumes that the strain at a particular location due to loads applied simultaneously at several points on the structure is the algebraic sum of the strains due to the same loads applied individually. This is the fundamental basis of Skopinski-type calibrations. Developments from Skopinski's methods have been used widely and include the application of distributed loads to produce check cases [9].

In theory, load calibration to limit-load conditions is desirable. However, in practice, application of a point load anywhere near limit-load condition is likely to cause significant local damage to the structure. Furthermore, the extent of loading achievable is often determined by how an aircraft can be constrained within the loads rig and how these constraint loads can be reacted within the structure. Therefore, detailed design and analysis are required to ascertain achievable load ranges. Loads should generally be applied at regions with local reinforcement, such as spar and rib intersections, or at external load application points, such as aileron or rudder hinges, to prevent local damage. Major attachment locations, such as wing, fin, and tailplane mounts, engine mounts, slinging points, and undercarriage attachments, are generally used as reaction points.

Methods such as influence coefficients can be used to identify the response of individual bridges to changes in the applied load location. If the normalized bridge output (bridge output/applied load) remains relatively constant with change in location of applied load, then the bridge is more sensitive to shear. If the normalized bridge output varies linearly with variation in the span-wise location of the applied load, then the bridge is sensitive to bending moment. However, if the output varies with chord-wise location of the load, then the bridge is responsive to torque.

Once the bridge-to-load equation has been optimized, where practicable, this may be cross-checked

by predicting the distributed load cases and determining the percentage accuracy. R values of 0.995 or R^2 values of 0.990 are realistic aims for shear and bending moment load calibrations. Such high levels of correlation are often difficult to achieve in practice for torque values. This is partly because the torque value can often be a relatively small number but driven by the difference between two large numbers (i.e., front and rear spar shears).

Loads calibration is a complex and potentially costly exercise. There are many potential pitfalls and experience has shown that, the more checks that can be introduced into the process, the more likelihood there is of retaining a high level of confidence in the results of the exercise.

4.3 Off-aircraft loads calibration

It may not be necessary to calibrate the component or structure on the aircraft in a full rig facility; wing attachment links are a good example of such a component. Great care has to be taken with off-aircraft calibration methods to ensure that the loading applied to the component during the calibration exercise can be related directly to the loading seen by the component in service.

4.4 Strain-gauge correlation to fatigue test damage

Where program aims are to compare in-service usage directly with fatigue test spectra, a method of correlating the strain outputs from the OLM aircraft to the damage introduced into the fatigue test can be used. In this method, strain-gauge bridge arrangements on the fatigue test are replicated on the OLM aircraft and the fatigue analysis is fitted to either test failure points or the end of test, with appropriate safety factors applied. The fitting factor (stress factor applied to the test spectra to predict a failure criterion at the end of test) is then applied to strain-gauge bridge output from the OLM aircraft. Thereby, in-service flying can be compared directly with that replicated on the fatigue test.

4.5 Strain-gauge airborne and on-ground calibration

Airborne and on-ground calibration methods have been used in several programs as either prime calibration or as backup or confidence checks in support of loads calibration or calibration to test damage methods. Airborne calibration is most likely to be chosen as an alternative method where Skopinski-type loads calibration is deemed prohibitively costly. The principles of this method are relatively simple, but the application in practice can be highly problematic and the method is considered unlikely to be successful for all applications.

The method requires the aircraft to be flown in several predetermined flight conditions or maneuvers or data corresponding to these conditions be extracted from OLM data. For example, for an airborne calibration of a wing root bending bridge on a combat aircraft, a range of symmetric pull-up and push-over maneuvers over a range of normal acceleration values and a series of wind-up turns[a] could be used to develop slope and offset correction (i.e., $y = mx + c$) for the bridge in question.

With sufficient data and a good knowledge of the global and local loads or stress distributions within the aircraft, airborne calibration is feasible for wing locations and possibly fuselage locations driven by bending loads, for example. However, having confidence in the application of this method to a tailplane and, in particular, a fin, where relationships between flight conditions and stresses in the component are more complex, is considered highly challenging and a high-risk strategy. Parametric and discrete OLM instrumentation should be subject to manufacturers' recommended calibration procedures.

5 DATA ANALYSIS

5.1 Data capture and analysis

As with so many aspects of an OLM program, capture and processing of the OLM data is a technically complex function. The analysis aims will be program specific, but a typical top-level process is illustrated in Figure 6.

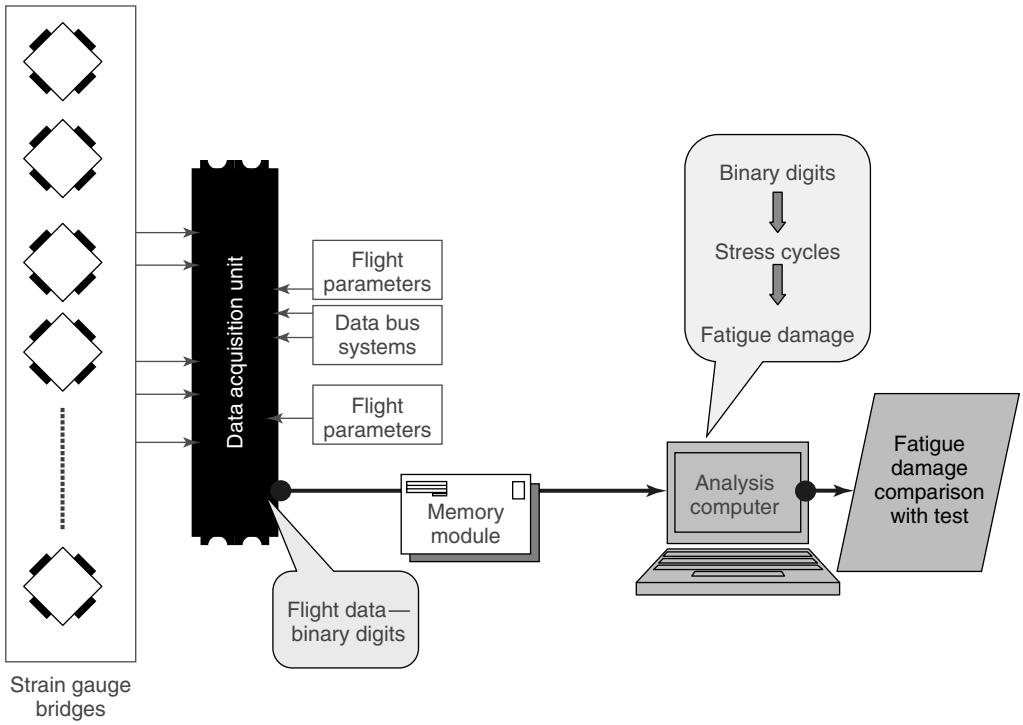


Figure 6. Example of top-level capture and analysis process.

5.2 Analysis process

An example of data analysis process is illustrated in Figure 7. The fatigue analysis process and postprocessing activity, such as comparison of OLM data with fatigue test spectra, will again be program-aim specific and there is insufficient space within this article to describe the various methods used. However, there are key functions required irrespective of program-specific aims and these are described briefly within this section.

5.3 Data anomaly and confidence checks

Retaining the integrity of the OLM data is essential to the success of any OLM program and anomaly and confidence checks generally include the following:

- identification of any regions of lost data;
- identification of bit failures;
- identification of error flags;
- collation and reporting of DAU-produced data health reports;
- identification of provisional data status against established criteria (e.g., data loss $<2\%$);
- check to ensure data loss does not coincide with highly damaging events (e.g., high-g);
- data replacement and identification process for minor losses (e.g., last known good value);
- exceedances of channel maximum and minimum expected strain/stress or load values;
- excessive rate of change values (where practicable) compared with performance data;
- failure of correlation between measurands expected to have a high degree of correlation;
- mismatch between flight record and OLM data;
- data trending checks with similar sortie descriptors;
- outlier analysis of known condition values (Figure 8);
- comparison with expected theoretical values for flight conditions (Figure 9).

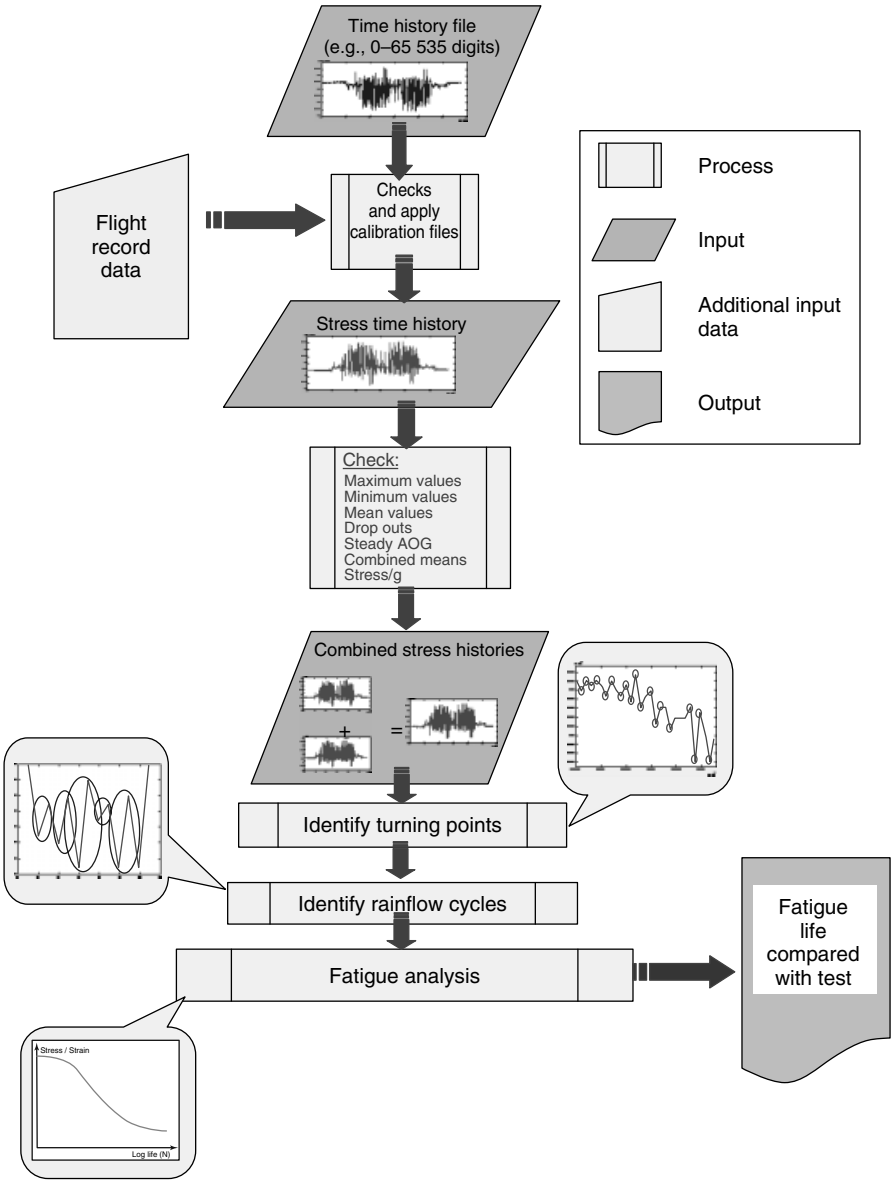


Figure 7. Example of data analysis process schematic.

5.4 Data representation

In-service time history data are analyzed in various ways to undertake a comparison with design and qualification spectra. The method used should be consistent with the fatigue design process. Traditionally, cycle data have often been presented in a cycle frequency of occurrence matrix or FOOM, usually by cycle range and mean or amplitude and mean. An example of FOOM data representation is presented in Figure 10.

Exceedance diagrams are a frequently used method of data reduction and permit comparison between data sets, e.g., test and usage spectra. An example of exceedance diagram is reproduced in Figure 11.

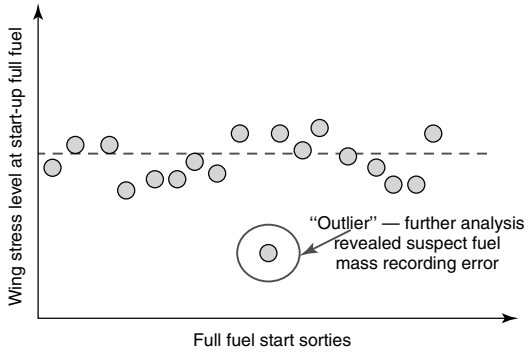


Figure 8. Illustration of trending of wing stress during start-up.

Care should be taken when using exceedance diagrams and the method chosen to reduce each data set should be consistent with the method used to produce the comparison data initially. Users also need to be aware of the effects of the loss of cycle mean stress information.

5.5 Fatigue data presentation

Assessments of in-service usage in fatigue damage terms are usually undertaken using data expressed in standard forms (i.e., per 1000 flying hours or per 1000 sorties). Additionally, fatigue damage or crack

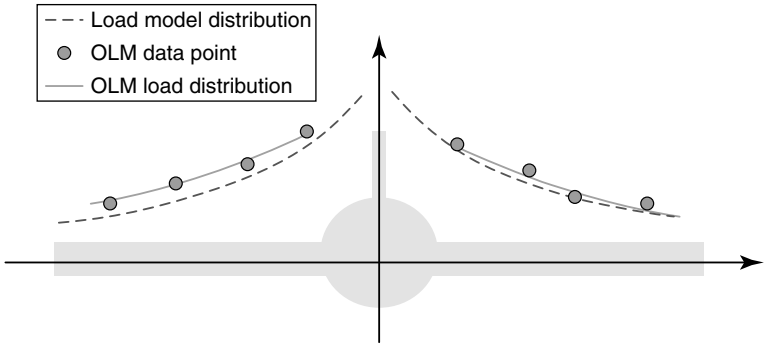


Figure 9. Example of span-wise load distribution plot.

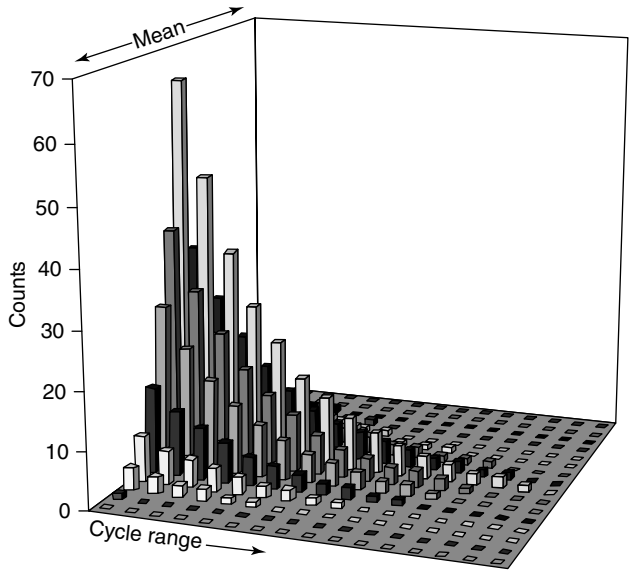


Figure 10. Example of FOOM Representation (16 × 16 FOOM illustrated—generally larger).

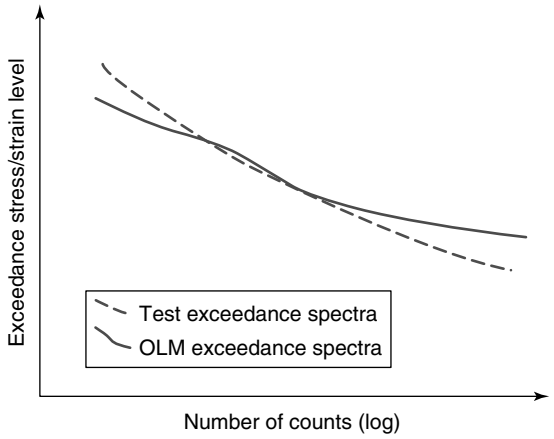


Figure 11. Example of exceedance diagram comparing test and OLM spectra.

growth accumulated through sorties on a flight-by-flight basis provides a valuable indication of highly damaging maneuvers or flight regimes. Examples of typical wing bending stress/strain time histories and corresponding damage accumulations are presented in Figure 12 for a combat/trainer aircraft and in Figure 13 for a transport aircraft.

6 DATA CAPTURE PROGRAM

6.1 Criteria affecting data capture

Even for fleets well provided with OLM aircraft, OLM data will rarely account for more than 2% of flying. For many fleets with few OLM aircraft and only periodic recording, OLM will represent significantly lower percentages of overall flight data. Therefore, the aim of detailing a data capture program is to ensure that the program aims are met with representative data.

6.2 Identification of data capture requirements

Data capture requirements should be specified for the initial OLM program. A review of available sortie

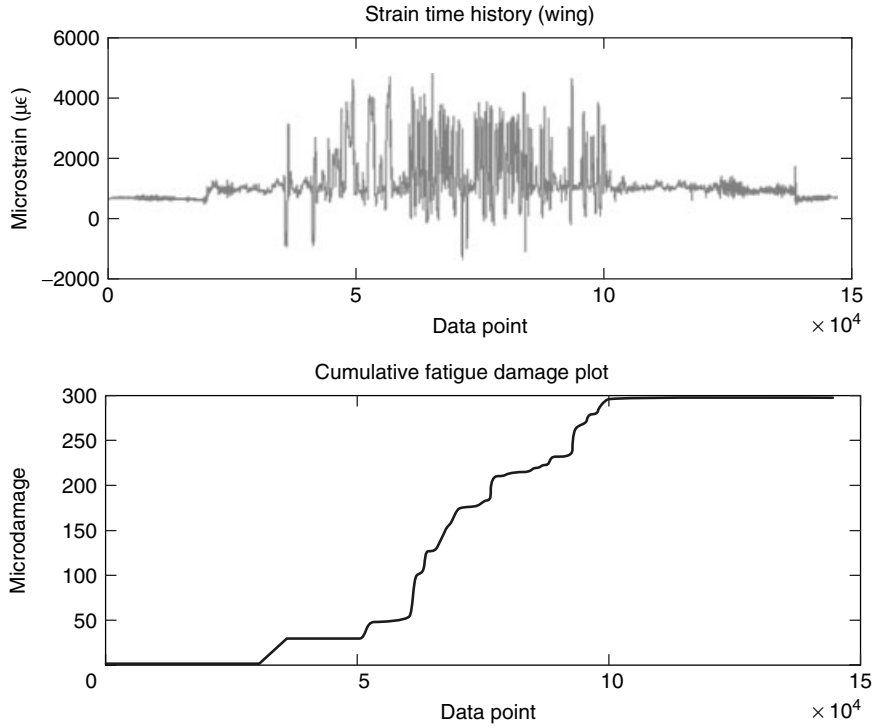


Figure 12. Example of combat/trainer aircraft wing strain/stress time history and fatigue damage accumulation plots.

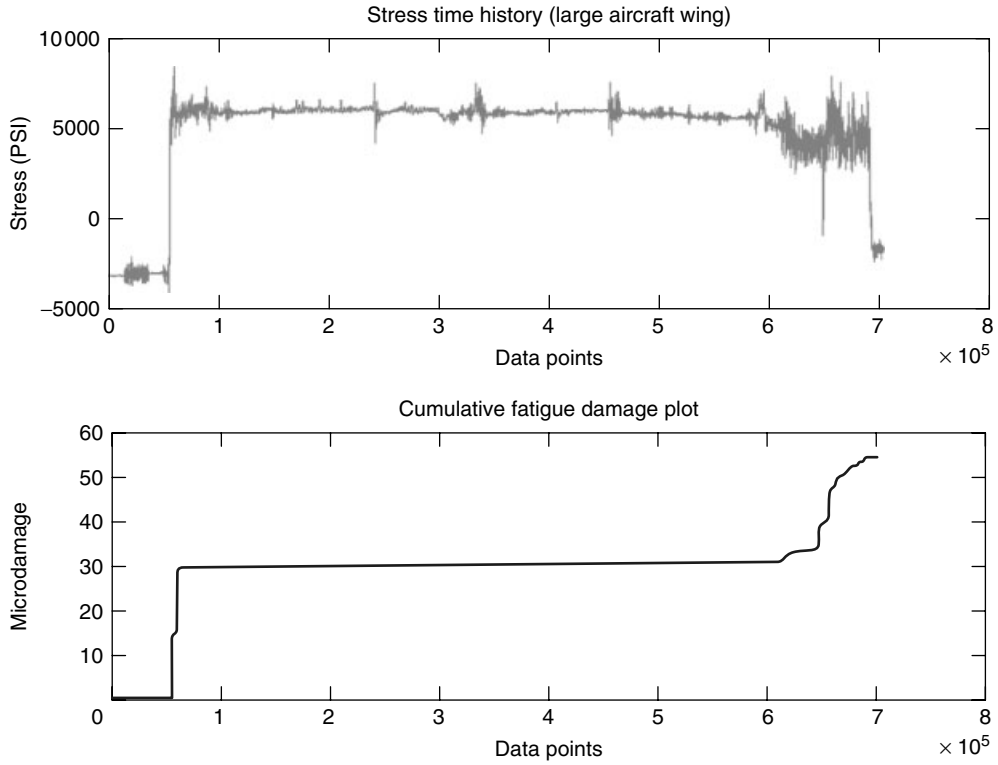


Figure 13. Example of transport aircraft wing strain/stress time history and fatigue damage accumulation plots.

usage data (or data from previous aircraft in the role) should be undertaken as the basis for the data capture requirements and the factors that should be considered in determining the program should include the following:

- program aims (i.e., need for test spectra, for example, may be first priority);
- sortie-type distribution flown by the fleet;
- likely variance within sortie type (e.g., air test—low variance, air combat—high variance), based upon cumulative average data from monitoring system if available;
- *a priori* fatigue damage distribution across sortie type (e.g., from monitoring system);
- fleets-within-fleets and matching OLM capability;
- fleet distribution, deployments, maintenance, and modification program;
- initial simplification or stores and role equipment variations;
- seasonal and syllabus variations (at least one year's data should be captured);
- impact of scheduled maintenance on data capture.

Where usage data are available for a fleet and the fleet is fitted with even a rudimentary monitoring system, these data can be used to provide an estimate of the number of sorties required within each sortie type. A simple time ordered cumulative average fatigue damage plot (Figure 14); can provide an indication of the minimum number of sorties needed in each sortie type to gain a reasonable representation of that sortie type *a priori*. (It is accepted that this heuristic method depends upon the monitor being reasonably accurate or at least relatively consistent.) Where no usage data exist, first-cut estimates can be developed from previous aircraft types in similar roles.

Once data capture commences and damage or crack growth[b] data are available, these initial estimates of the minimum data capture requirements could be refined using OLM-derived damage or crack growth rates by sortie profile code and by major structural component. In addition, where scatter is found to

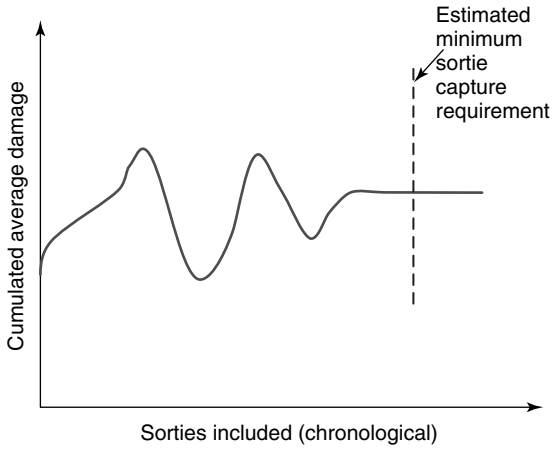


Figure 14. Example of cumulative average fatigue damage plot.

be larger in a particular sortie type than expected, particularly where that sortie type contributes highly to fatigue damage accrual, a greater number of sorties will be needed to provide representative data.

7 REPORTING

7.1 Post-air-test/initial analysis and report

Following the air test or initial flying after the installation of the OLM system on each OLM aircraft, postflight analysis should be undertaken promptly. This analysis should include the following:

- identify the serviceability status of all data channels;
- identify any data losses or anomalies;
- compare channel values with expected values;
- compare character of data with that expected for channel;
- check channel ranging values in engineering units;
- check frequency content for channels susceptible to high-frequency loading;
- review channel sample rates;
- recommend remedial action as required;
- check analysis software.

Having confidence in the data provided from an OLM program is essential in ultimately achieving the aims of the program. Where confidence in the data is lost, remedial action to recover the situation can be extremely costly and wasteful. Therefore, rapid initial analysis and reporting following initial sorties is required.

7.2 Interim and final reporting

At a point determined by the program aims, interim and final reports should be raised. The final report, in particular, should form a natural progression from the interim reports but should be considered as the definitive historical record for the program and hence should be a stand-alone document but with references to lower-level detailed reports generated during the course of the program. The final report should include details of the following:

- program aims
- OLM installation
- calibration process
- data analysis process
- data capture program
- data capture achievement
- data quality
- data analysis
- conclusions by program aims
- recommendations for remedial airworthiness action or further work
- OLM report references
- lessons identified
- future OLM recommendations, including data capture, system enhancements, or care and maintenance requirements.

ACKNOWLEDGMENTS

The author is pleased to acknowledge the contribution of OLM practitioner colleagues throughout the UK aerospace industry, who provided valuable input to the recently produced UK OLM guidance paper [10], on which some of the content of this article is based. In particular, the author wishes to acknowledge the contribution of his coauthor on that project, Mrs Dorothy Holford.

END NOTES

[a.] Wind-up turn: In a wind-up turn, the aircraft is rolled into a turn at each test condition and, keeping the speed and power constant through the maneuver by allowing the height to vary around the nominal test altitude, the normal acceleration is increased progressively until a defined limiting condition is reached.

[b.] Linear elastic fracture mechanics methods assume that the crack elongation is a function of its original length as well as the incremental applied stress/load spectrum, whereas the stress–life approach considers the incremental damage contribution from the spectrum to be independent of the *a priori* damage accumulation. Meaningful comparison of crack growth increments on a flight-by-flight basis will require calculation from a standard initial crack length.

REFERENCES

[1] UK Ministry of Defence, *Design and Airworthiness Requirements for Service Aircraft*, Defence Standard 00–970, Part1/2, Issue 4, 2006.

[2] de Forest AV. The rate of growth of fatigue cracks. *Transactions of the ASME* 1936 **58**:141.

[3] Simmons Jr EE. *Material Testing Apparatus*, US Patent No: 2,292,549, February 1940.

[4] Skopinski TH, Aiken Jr WS, Huston WB. *Calibration of Strain Gauge Installations in Aircraft Structures for Measurement of Flight Loads*, NACA Report 1178, 1954.

[5] Lloyd PA. Structural health monitoring evaluation tests. In *Health Monitoring of Aerospace Structures—Smart Sensor Technologies and Signal Processing*, ISBN 0-470-84340-3, Staszewski WJ, Boller C, Tomlinson GR (eds). John Wiley & Sons, 2004, pp. 207–259.

[6] British Society for Strain Measurement (BSSM), *British Society for Strain Measurement Code of Practice for the Installation of Electrical Resistance Strain Gauges CPI*, 1992.

[7] Military Standard (US), *Department of Defense Test Method Standards for Environmental Engineering Considerations and Laboratory Testing*, Mil Std 810F, 2000.

[8] British Standard BS 3G 100, *Specification for General Requirements for Equipment for Use in Aircraft*, 1991.

[9] Jenkins JM, DeAngelis VM. *A Summary of Numerous Strain-Gage Load Calibrations on Aircraft Wings and Tails in a Technology Format*, NASA Technical Memorandum 4804, 1997.

[10] Reed SC, Holford DM. *Guidance for Aircraft Operational Loads Measurement Programmes*, UK Military Aircraft Structures Airworthiness Advisory Group (MASAAG) Paper 109, May 2007.

Chapter 140
Loads and Temperature Effects on a Bridge

Ming L. Wang
Department of Civil and Materials Engineering, University of Illinois, Chicago, IL, USA

1 INTRODUCTION

Bridges are an essential part of a highway network. They are open to traffic, resistant to natural disaster, and subjected to millions of loading cycles per year. However, they are quite expensive to maintain and do occasionally fail [1]. The fact that many bridges are carrying greater average loads than predicted during their design has significantly increased the need to monitor bridge performance over the past few years. To effectively manage bridges today, there is a great need to monitor the real-time conditions of bridges, and the deterioration rates of their components, so that efficient and proactive measures can be taken [2–5]. By using the latest state-of-the-art technologies, it is possible to utilize health monitoring systems on highway bridges to determine their behavior and condition, and assess maintenance and inspection needs [6, 7]. Bridge health monitoring systems have historically been implemented for the purpose of understanding bridge behavior under various loads and environmental effects [8–13].

The Kishwaukee Bridge is a five-span precast, posttensioned segmental concrete box girder bridge across the Kishwaukee River in Rockford, Illinois. The structure has been under increasingly stringent inspection since extensive cracking adjacent to the piers in the webs was observed [14–18]. To this end, the author has developed instrumentation that continuously monitors the bridge on the basis of local strain, displacement, temperature measurements, and global frequency measurements [17, 19].

The ultimate focus of the instrumentation is the measurement of permanent deformations in the web at key locations that may indicate yielding of the steel shear reinforcement and consequent changes to the load-carrying capacity. This is accomplished through local and global monitoring, for maximum redundancy. The local monitoring relies on strain gauge pairs installed on the web faces in the middle of each main span to estimate the loads on the bridge, and by two linear variable displacement transducers (LVDTs) displacement-strain rosettes that were installed at the location identified during prior static load tests as having the most severe reinforcement stresses. The global measurements are derived from ambient vibration accelerometer records.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Frequency measurements are relatively insensitive to deterioration of shear capacity. However, they are analyzed for two reasons. First, measurements were obtained during 1999–2000; despite visual inspections, these data are directly relevant to assessing long-term changes in the structure [20], although recent experience clearly indicates the need for a more thorough understanding of thermomechanical effects [21]. Secondly, both the benchmark frequency measurements and the current data are available at high resolution and accuracy, and the requisite sensitivity of frequency shifts can be explored through finite element method (FEM) analysis [22]. Moreover—and relevant to this article—these two factors enable the use of bootstrap methods to develop quantitative change-point criteria for automated alarm monitoring. This is an important step, given the time and expense for manual analysis of data. These frequency data thus complement the interpretation of long-term strain and displacement measurements from moving ambient loads, data which can be beset with possible drifts and large uncertainties. Temperature correlation was performed after noting that the average lag between internal and external free air temperatures was roughly 5 h.

This article provides a supplement to the current research on bridge health monitoring and improves the understanding of bridge behavior. Specifically, load and temperature effects are addressed in detail.

2 KISHWAUKEE BRIDGE MONITORING

2.1 Background

Kishwaukee River Bridge (Rockford, IL) is a posttensioned, precast segmental concrete box girder bridge opened to traffic in 1980. The bridge has five spans with lengths of $51.8\,\text{m} + 3 \times 76.2\,\text{m} + 51.8\text{m}$. As the first generation of segmental structures, the Kishwaukee Bridge engineers chose the design of a single shear key joint usually located close to the center of gravity of the cross section. These joints are quite vulnerable especially during polymerization of the glue.

Problem arose during and after the completion of the bridge. The epoxy applied between segments was not hardened properly in some joints. The epoxy was unable to carry the shear stress fully and was instead acting as a lubricant that caused reduction of shear resistance capacity. Therefore, a substantial part of the shear force was concentrated at the shear keys [16]. As shown in Figure 1, the inclined cracks went through only for the length of one segment but not continuously to the next segment. Many steel pins were inserted to the webs between segments to stop the propagation of shear cracks. It has proved to be effective in successfully slowing down the progress of cracking. However, there is

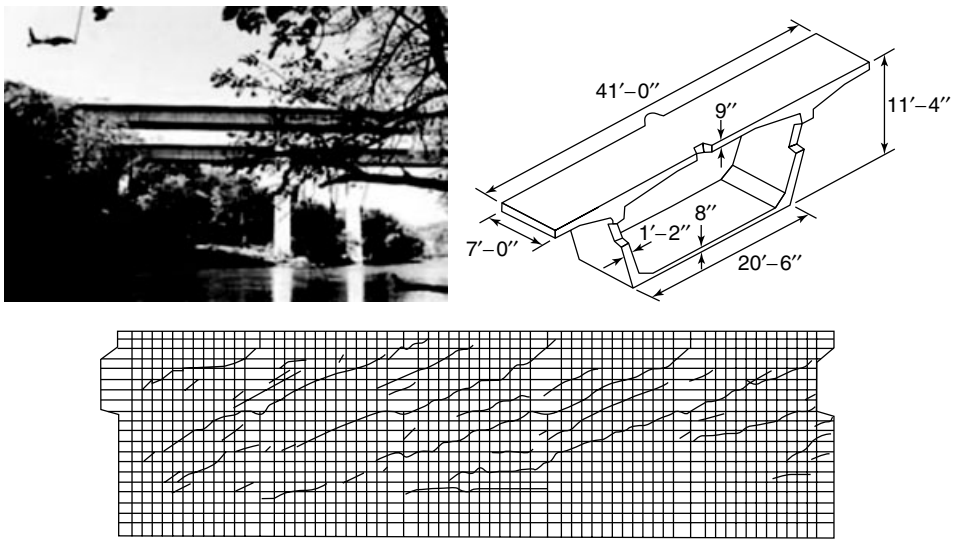


Figure 1. Shear cracks on the webs of Southbound Kishwaukee Bridge.

a need to know current shear carrying capacity as well as the extent of propagation of cracks. A long-term monitoring program is therefore imperative for structural surveillance of Kishwaukee Bridge to provide continuous health records for the bridge management department [23].

Since the major problem of Kishwaukee Bridge is shear cracks on the webs, a specifically designed structural health monitoring system was installed on the Kishwaukee Southbound Bridge in 2003 [24]. This system monitors and records the strain, crack opening displacement (COD), and acceleration of the bridge, as well as the temperature outside and inside the bridge girder. Using shear strain measured through a rosette of three CODs across cracks, its corresponding shear resistance capacity can be predicted. Thus, the operation and management for this bridge is transformed into a more objective and quantitative process. This process provides for optimal integration of experimental, analytical, and informational system components [25]. In addition, the outcome of the process provides valuable information for current evaluation of structural integrity, durability, and reliability [19]. Using this information, composed from the sensory system, data-acquisition system, and health assessment system, the bridge owner and maintenance authorities can make rational decisions in assigning the budgets for both maintenance and repair.

2.2 Static load test

In order to determine the actual health status of Kishwaukee Southbound Bridge and set up the baseline for the following long-term health monitoring, we did two static load tests in 2000. The bridge was tested for service loading conditions. The weight of the trucks was comparable with a weight of the design truck defined in AASHTO LRFD Bridge Design Specifications. The design weight of the truck (72 kips) was amplified by a dynamic factor of $\delta = 1.320$. Four different positions of the trucks were proposed on the bridge, as shown in Figure 2.

Visual inspection of the shear cracks determined the most damaged webs in the bridge. Web SB2-N4-E was chosen for *in situ* measurement of deformation due to shear forces. Three LVDTs were installed at the interior surface of the web, close to the neutral axis, as shown in Figure 3. Measured displacements were used for estimation of the web's shear stiffness. The load stage using trucks located in position 2 (second load) was selected for assessment, because neither transverse bending moments nor vertical axial stresses accompanied the imposed load in the web of segment SB2-N4. Shear forces generated by trucks located in position 3 (second load) were used for calculation of the steel stress increment in reinforcement.

Average shear strain and shear stiffness of the cracked web can be assessed by equations (1) and (2):

$$\Delta\gamma_{Lt} = \frac{2[\Delta\varepsilon_{\alpha} - \Delta\varepsilon_{L}\sin^2(\alpha) - \Delta\varepsilon_{t}\cos^2(\alpha)]}{2\cos(\alpha)\ \sin(\alpha)} \quad (1)$$

$$(GA)_{ir} = \frac{\Delta V}{\Delta\gamma_{Lt}} \quad (2)$$

To determine the global flexural stiffness of the bridge, strain gauges were installed at the webs of segments located next to the closures. Measured concrete strains were used for the calculation of curvature in equation (3) and evaluation of the modulus by equation (4):

$$\Delta\chi_{\text{meas}} = (\Delta\varepsilon_{\text{upper}} - \Delta\varepsilon_{\text{lower}})/d_{gs} \quad (3)$$

$$E_c = \Delta M/(\Delta\chi_{\text{meas}}\ I_i) \quad (4)$$

The assessed value of modulus was still in the range of 35 000–40 000 MPa. This value is very similar to the design value, while the tangent shear modulus has been reduced about 50–55% by shear cracks. Therefore, the change of bridge shear stiffness has little influence on its flexural stiffness. Dynamic tests and FEM simulation also corroborate the negligible effect of shear stiffness on flexural stiffness.

2.3 Half-scale experiment of concrete girder

In order to determine the shear carrying capacity of Kishwaukee Bridge, a half-scale I-beam model was cast in the laboratory. The following parameters are measured in this experiment: deflection of the end of cantilever, flange strains, web strains, reinforcement

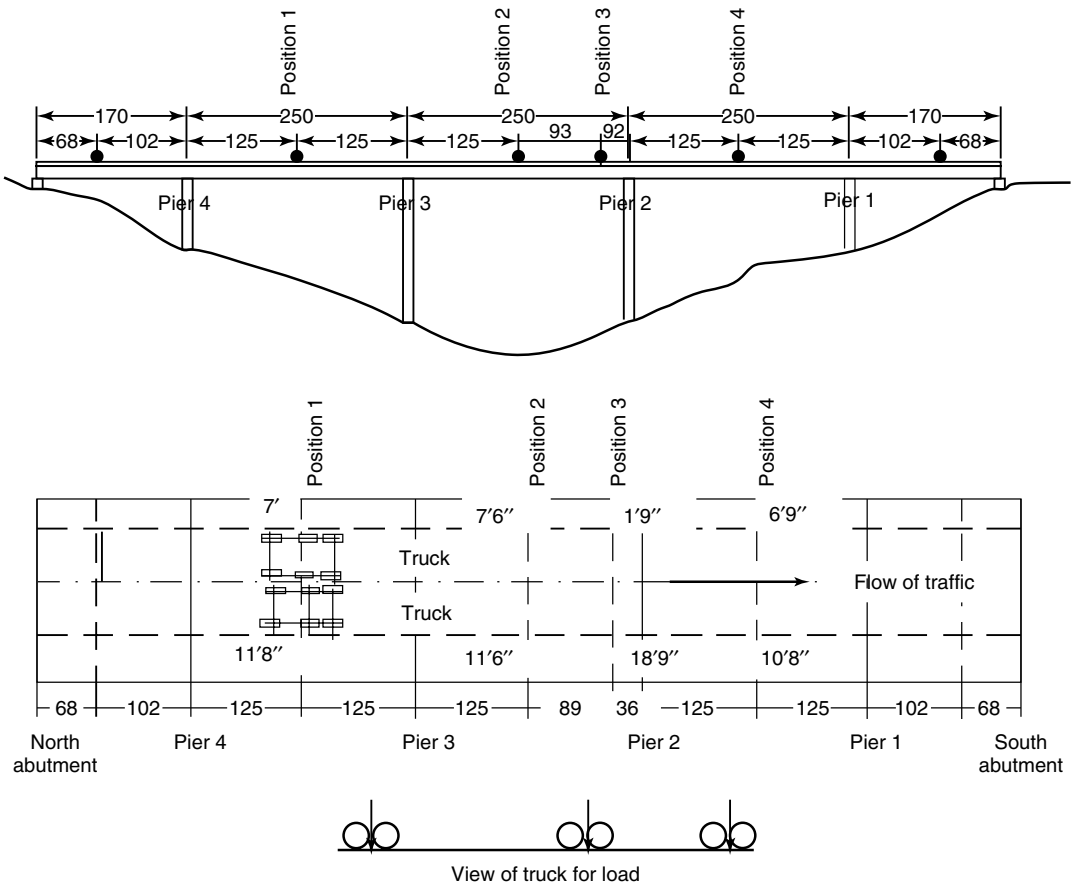


Figure 2. Static diagnosis load test on the southbound Kishwaukee Bridge.

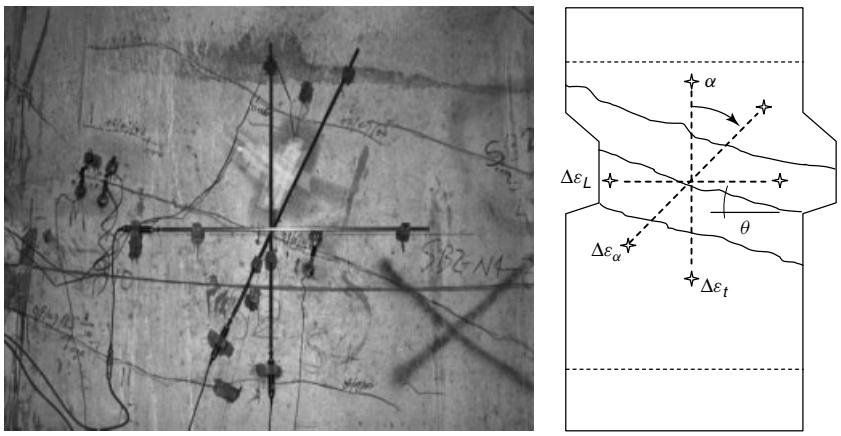


Figure 3. LVDT sensors installed on the inner surface of the web.

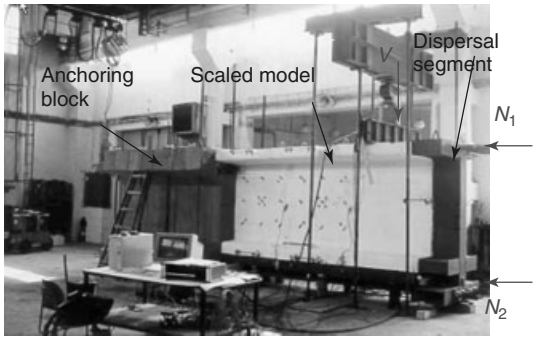


Figure 4. Arrangement of half-scale experiment.

strains, prestressing force, and elongation of the strands.

Figure 4 presents the arrangement of the half-scale experiment. The three half-scale I-beam segments were put between the anchoring block and the dispersal segment. The friction between the joints of three segments was reduced to a certain degree to simulate the unhardened epoxy problem on the southbound Kishwaukee Bridge. The prestressing reinforcement was low relaxation strands with a high yielding point of 1800 MPa. The upper prestressing force N_1 was applied with eight 3-strand tendons on the top flange while four 4-strand tendons were used to apply the lower prestressing force N_2 on the bottom flange. The shear force V was applied with a hydraulic jack.

The propagation of shear cracks on the I-beam webs is shown in Figure 5. According to the visual inspection records, the type and inclination of these shear cracks are similar to the actual cracks on the Kishwaukee Bridge as shown in Figure 1. It indicated that the experiment successfully simulated the actual damage condition of Kishwaukee Bridge. The relationship between shear stress and shear strain is shown in Figure 6. The original tangent shear modulus G_c was 15 000 MPa. After the linear elastic phase, the concrete cracked and the steel reinforcement carried much of the shear force itself, which is evident from the graph. About the yielding point of shear reinforcement, the reduced tangent shear modulus G_{ct} was roughly 5300 MPa. According to the analysis of the static load test, the current tangent shear modulus is approximately 6740 MPa at the

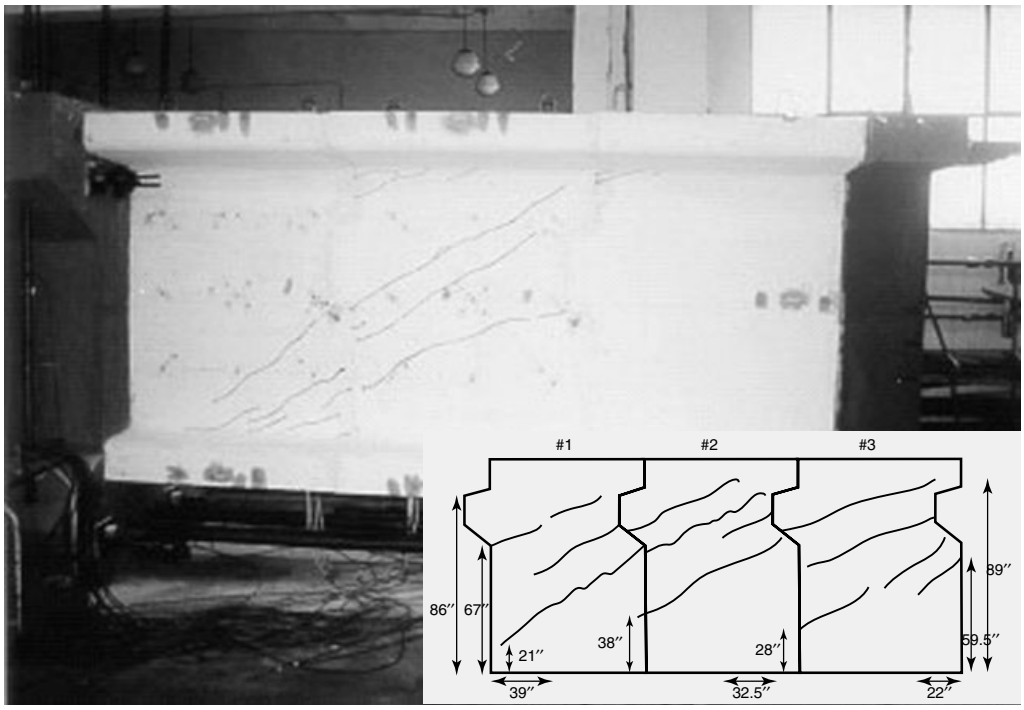


Figure 5. Propagation of shear cracks on the specimen.

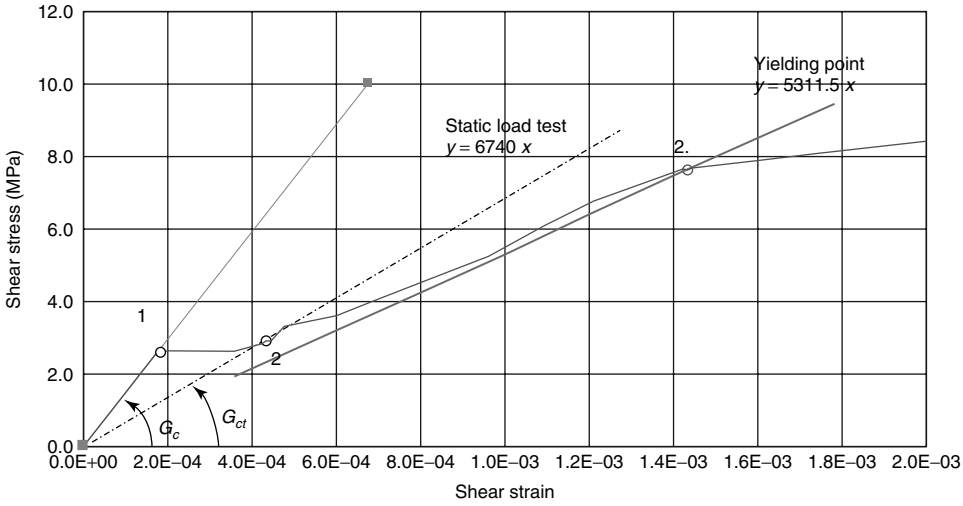


Figure 6. Relationship between shear stress and shear strain measured from half-scale experiment.

worst location (SB2-N4) of the southbound Kishwaukee Bridge. This means the shear steels are still in the safe region of stress–strain curve. However, more strict regulations should be encouraged so that no overweight truck passes over the bridge, which might induce further damage.

2.4 Real-time global and local health evaluation

Kishwaukee Bridge monitoring system has been collecting and processing data and generating evaluation and health reports for five years. The system can analyze the frequency distribution, COD, shear strain in the web, and traffic information in real time. Automated warning/alarm system is in effect to warn against any further local structural damage on the bridge, system problems, sensor dysfunction, and data errors.

Subsequent to temperature compensation and local-in-time change-point, detect the second step in the analysis strategy is to compare the current population with off-line reference data sets, which have been preserved for this purpose. Our strategy using the bootstrap has been discussed previously [23].

After establishing the baseline of bridge health, a long-term monitoring system can be used to provide the continuous health information of a bridge. For concrete bridges, cracks, especially the shear cracks, act as the main role of local damage and the global health information of a bridge can be represented with the bridge stiffness. Both global and local conditions of bridges need be evaluated in order to determine their in-service behavior and justify rehabilitation and repair plans. The global health information can be provided by dynamic measurement, while the local damage can be captured by sensors such as strain gauges and LVDT. The raw acceleration data are collected and preprocessed to obtain the natural frequencies in a sensor substation. Then the acceleration and frequency data are transferred into the database server via the Internet in real time, as shown in Figure 7(a). After that, the application server will analyze the data to get the hourly bootstrap mean and its confidence intervals, as shown in Figure 7(b).

On the basis of dynamic tests from 1999 to 2000, the global dynamic characteristics of the bridge are obtained from the acceleration data with the related temperature values. These structural parameters are set up as the baseline of global health assessment. Temperature has a significant influence on the natural frequencies of a structure. Hence, it is important to derive the relationship between temperature and natural frequencies, i.e., how much the frequency will change owing to the variation of 1 $^{\circ}$C. To a certain degree, this relationship can represent the change in the bridge bending stiffness due to the temperature variation.

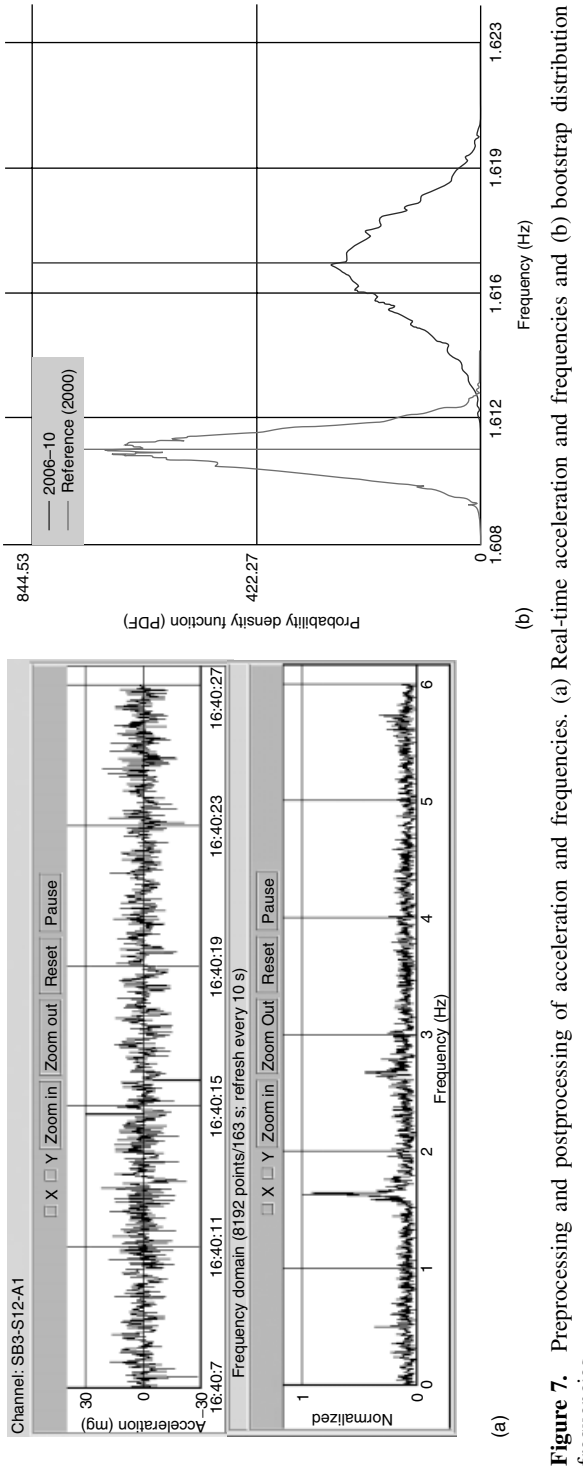


Figure 7. Preprocessing and postprocessing of acceleration and frequencies. (a) Real-time acceleration and frequencies and (b) bootstrap distribution of frequencies.

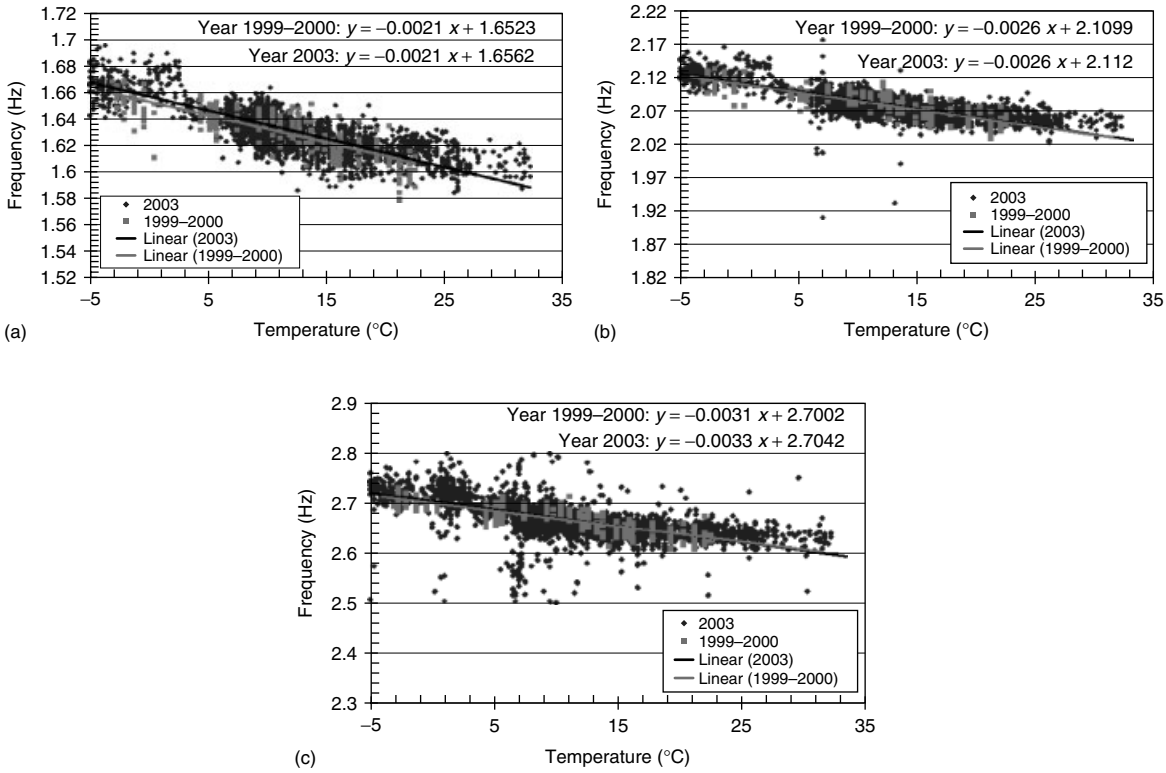


Figure 8. Relationship between temperature and first three modes; (a), (b), and (c) corresponding to mode 1, mode 2, and mode 3, respectively.

Figure 8 provides the temperature-related regression curves and parameters of the first three modes from 1999 to 2003. As shown in the figure, the changes in the first two modes due to the variation of 1 °C during 2003 are almost the same as the baseline (1999–2000). However, the analysis about mode 3 during year 2003 shows a little increase in the frequency change due to the unit temperature variation.

According to the theory of dynamic analysis, the changes in higher modes usually reflect the development of local damage. In order to verify the result of global health assessment and inspect the state of local damage, it is necessary to carry out the specific local health evaluation.

The raw data of COD are preprocessed in sensor substation and transferred into the database server via the Internet, as shown in Figure 9(a). Then, based on the hourly record of CODs, the application server will analyze these data to obtain their bootstrap distribution and confidence intervals. The result is shown in Figure 9(b).

In order to find out whether the shear cracks propagated during 2004, it is necessary to analyze the relationship between temperature and CODs, i.e., to find how much the COD will change due to a variation of 1 °C.

On the basis of the data of five years, the analysis result of CODs (A) is given in Figure 10 according to their locations. In five years, the CODs (B) excluding the temperature effect show some differences between the west web and the east web. As shown in the graph, on the west web, the total accumulated COD is about 120 μm. However, on the east web of SB2-N4, the corresponding COD is about 75 μm. Both values indicate accumulation of damage in terms of increases in COD. The difference between the west web and the east web is possibly due to the heavier traffic on the west side of the southbound bridge.

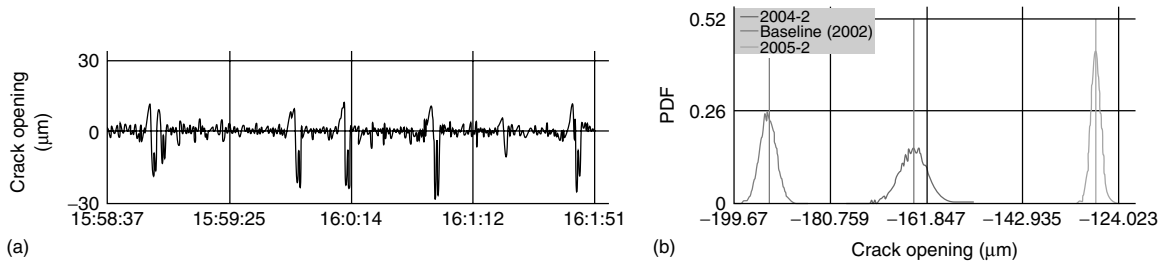


Figure 9. Preprocessing and postprocessing of crack opening displacement. (a) Real-time crack opening displacement and (b) bootstrap distribution of COD.

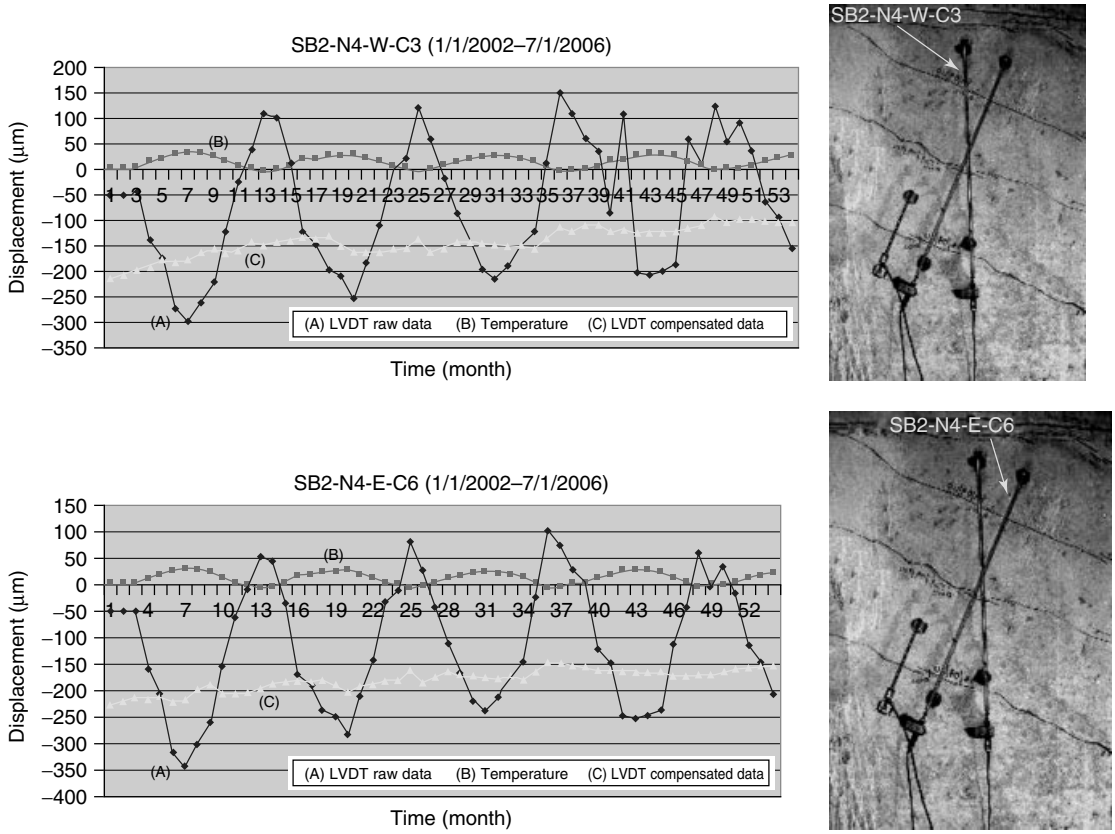


Figure 10. Relationship between crack opening displacement and temperature.

This result conforms to the visual inspection of Illinois Department of Transportation. According to the analysis on COD, we can evaluate the average shear strain of the cracking webs.

On the basis of the global and local measurements, the rule-based expert system gives the shear carrying capacity of the bridge with ductile mode of failure. Figure 11 presents the shear stress–shear strain curve of the cracked web at segment SB2-N4. As shown in the graph, the monthly maximum shear strain is over the baseline of the static load test in 2000. However, this value is still far below the yielding point. It indicates that the most damaged segment (SB2-N4) is still working in the nonlinear elastic zone.

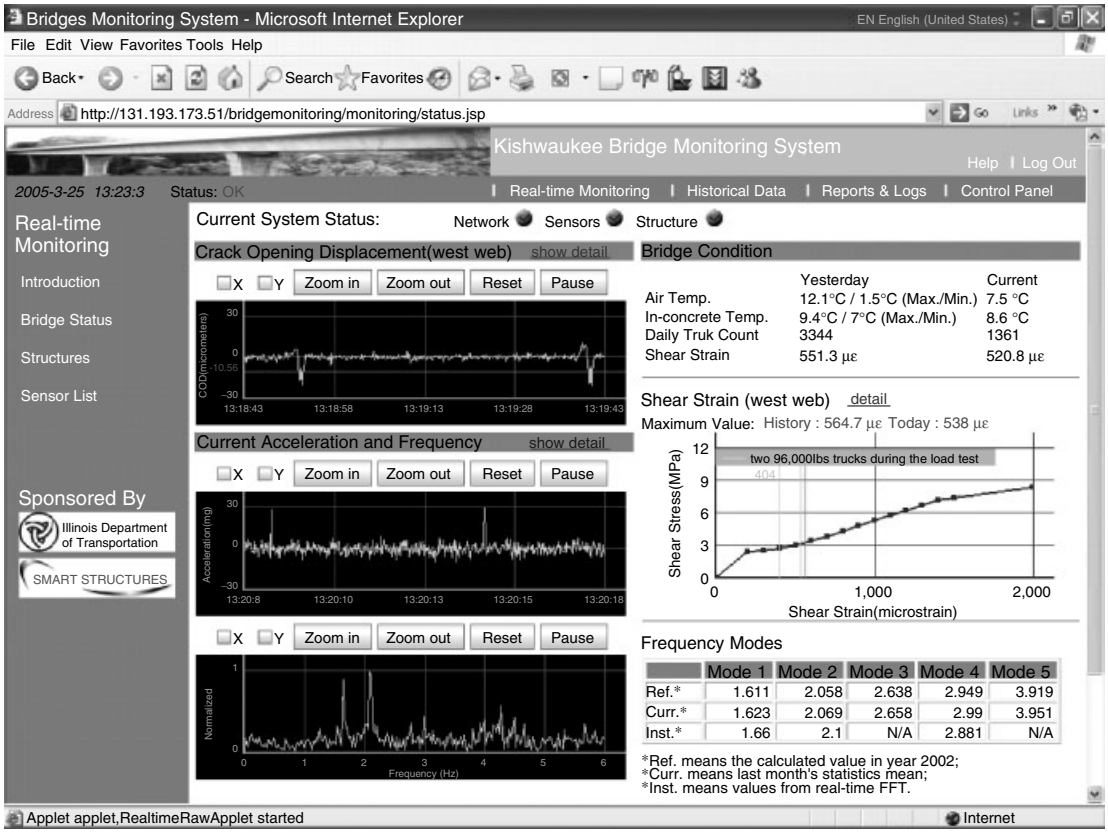


Figure 11. Main interface page.

2.5 Temperature effect

Measurements obtained from the monitoring system are available continuously and with greater precision; data studied in this article are given in Figure 12. The new data allow comparison with theory to better estimate characteristic thermal times and the effect of heat transfer on modal estimates. A reasonable 1-D model for transient heat conduction is given by equation (5),

$$\overline{T}(x,t) = \sum_{n=1}^{\infty} c_n(\beta_n) X(\beta_m, x) \mathrm{e}^{-Fo_n} \qquad (5)$$

where $\overline{T}$ is normalized temperature, $\beta_n = f(h,k)$ are the eigenvalues (with units of m^{-2}), Fo_n are Fourier numbers, c_n depends on a uniform initial condition of unity, and $X(\beta_m, x)$ are the eigenfunctions. Heat transfer at interior surfaces is neglected.

For long times (several hours), the sum is well approximated by the first term, $\overline{T} = c_1 X_1 \mathrm{e}^{-Fo_1}$ (with $c_1 X_1 \sim 1$–1.2 here). Transport properties of concrete were based on suggested values [26, 27]; the largest unknown is the heat transfer coefficient h, which was estimated from data when the free stream temperature, T_5, was nominally constant, as shown in Figure 13(c). Frequency variations for mode 1 and mode 3 are shown in Figures 13(a and b). A characteristic thermal time for a $1/e$ change to propagate through the thickness of the webs is then obtained from $Fo_n \sim \alpha\beta_1^2\tau \equiv 1 - e^{-1}$. For the case $h = 10$, $\tau = 14.2$ (h); for $h = 25$, $\tau = 8.4$ (h), corresponding times for average temperature, $\overline{T}_{\mathrm{avg}} = L^{-1}\int_0^L \overline{T}(x,t)\,\mathrm{d}x$, are 13.5 and 7.5 (h), respectively. The corresponding values of the Biot number $Bi = hL/k \sim 1$–5 together with τ indicate the presence of

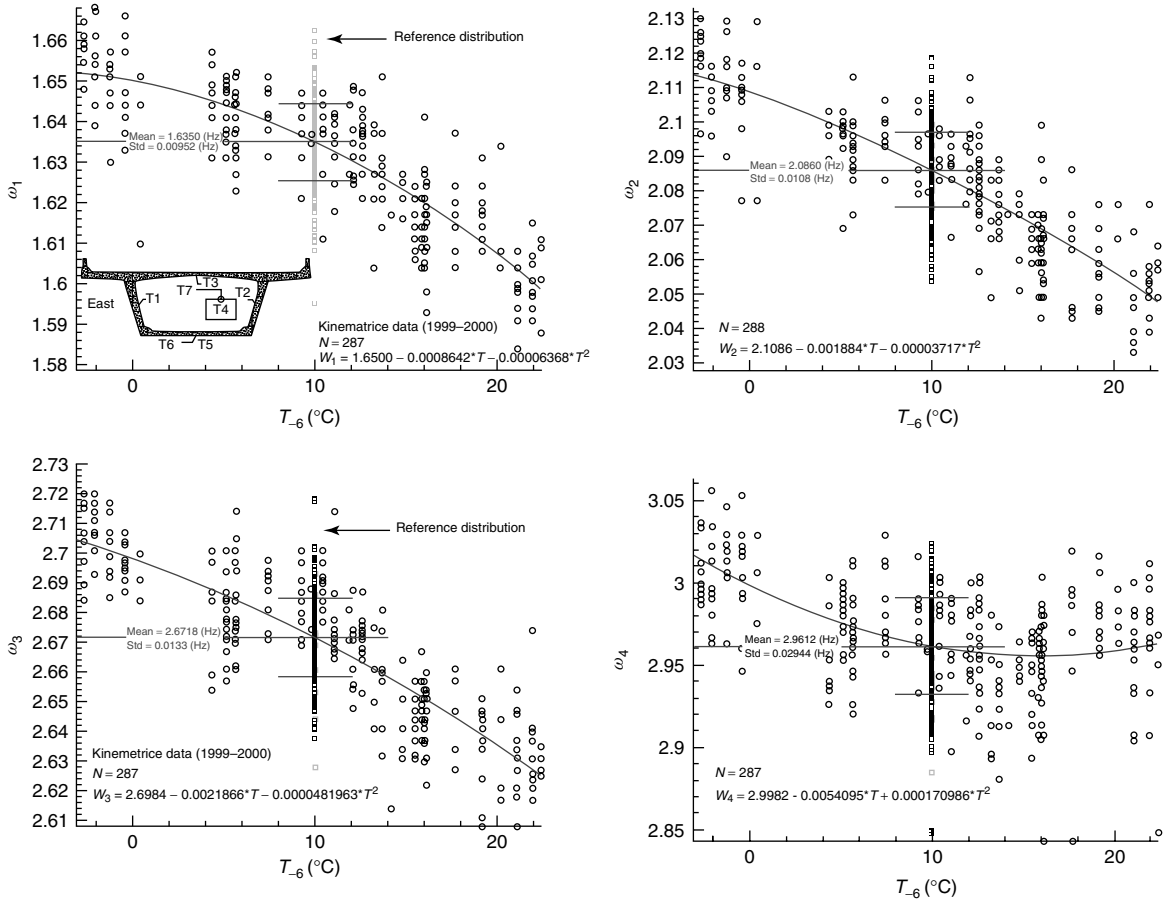


Figure 12. Open circles denote reference frequency data obtained from 1999 to 2000 for the Southbound Kishwaukee Bridge (primarily between 1600 and 2200 h). Deterministic temperature bias inferred from T_5 (delayed 5 h) was identified using least squares regression, and all frequency estimates were adjusted to a common reference temperature, $T_{\text{ref}} = 10(^{\circ}\text{C})$, according to the equation, $\omega_{i,\text{corr}} = \hat{\omega}_i(T_i) - [\hat{\omega}_i(T_i) - \omega(T_{\text{ref}})]$. (Open squares denote data obtained at 1600 h from 4 December 2003 through February 2004).

large thermal gradients and stresses [28]. The analysis and new data indicate that thermal equilibrium is rare for this structure.

2.6 Traffic effect

In this study, we have measured COD of shear crack in real time. The average COD subjected to a truck loading was about 25 μm as shown in Figure 9(a) of a real-time plot. The average COD due to temperature in a day was about 40 μm which is only about two to three times of the displacement due to traffic. In general, temperature effect could be 10 times more than the traffic effect if the member is free of damage [7]. However, it is difficult to separate the accumulation of damage that is due to traffic or temperature cycling. It is estimated that the average number of trucks passing daily is about 3000 using real-time data as shown in Figure 14. Figure 15(a, b) show the yearly COD cycles due to temperature.

Figure 15(c, d) are the estimated effects due to traffic. These are obtained by subtracting the temperature effect and permanent accumulation of damage due to fatigue as shown in Figure 16 from the original data as shown in Figure 17. We knew overall

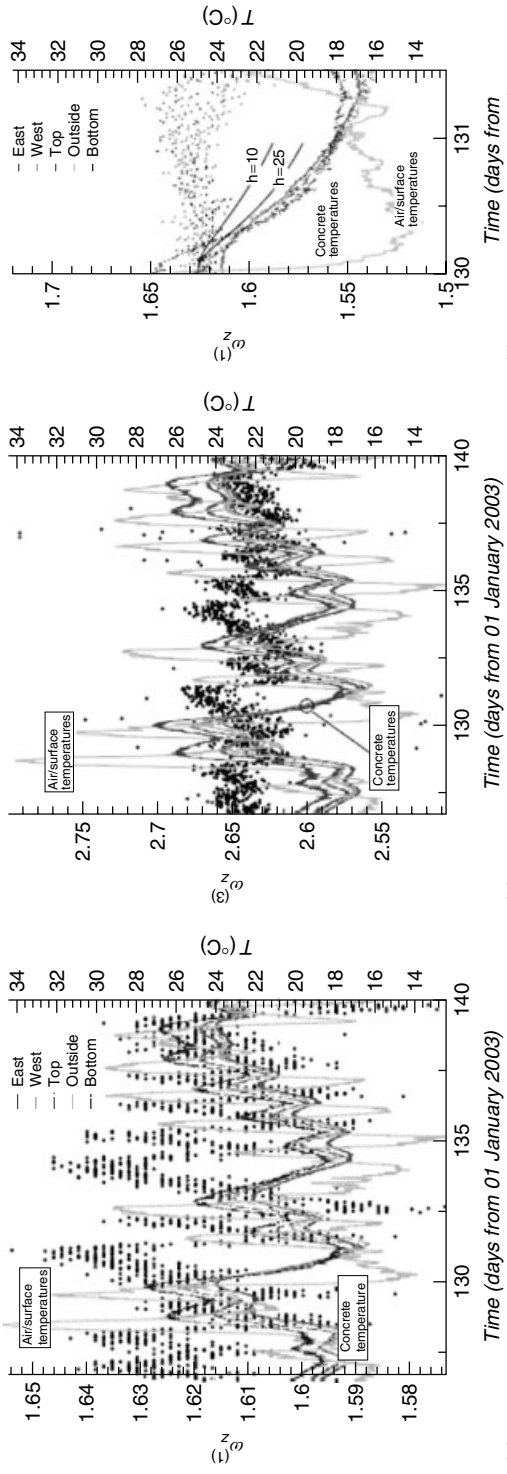


Figure 13. FFT estimates from May 2003: (a) time variation of $\omega_z^{(1)}$ and temperatures; (b) time variation of $\omega_z^{(3)}$; and (c) estimates of heat transfer coefficient based on concrete thermal properties $k = 3\,\mathrm{W\,m^{-1}\,K^{-1}}$, $\rho = 2242\,\mathrm{kg\,m^{-3}}$, and $c = 1000\,\mathrm{J\,kg^{-1}\,K^{-1}}$ [23].

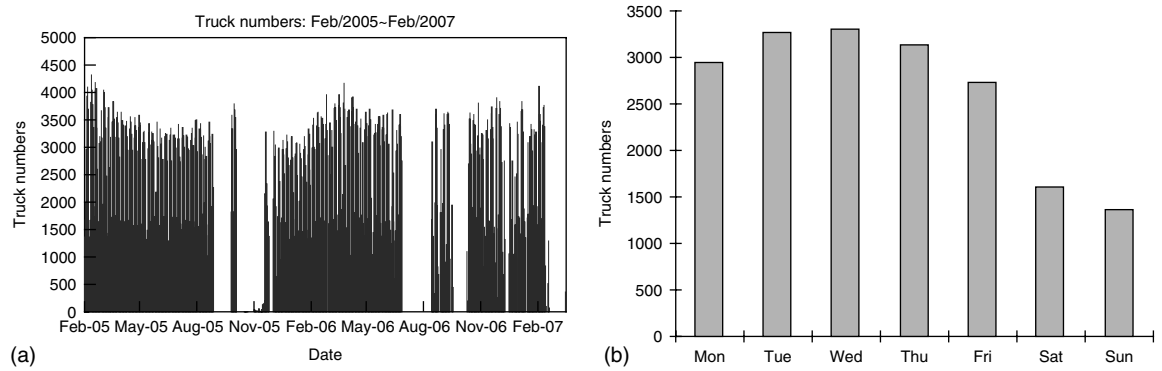


Figure 14. Average number of trucks in two years. (a) Passing daily and (b) average for each day of the week.

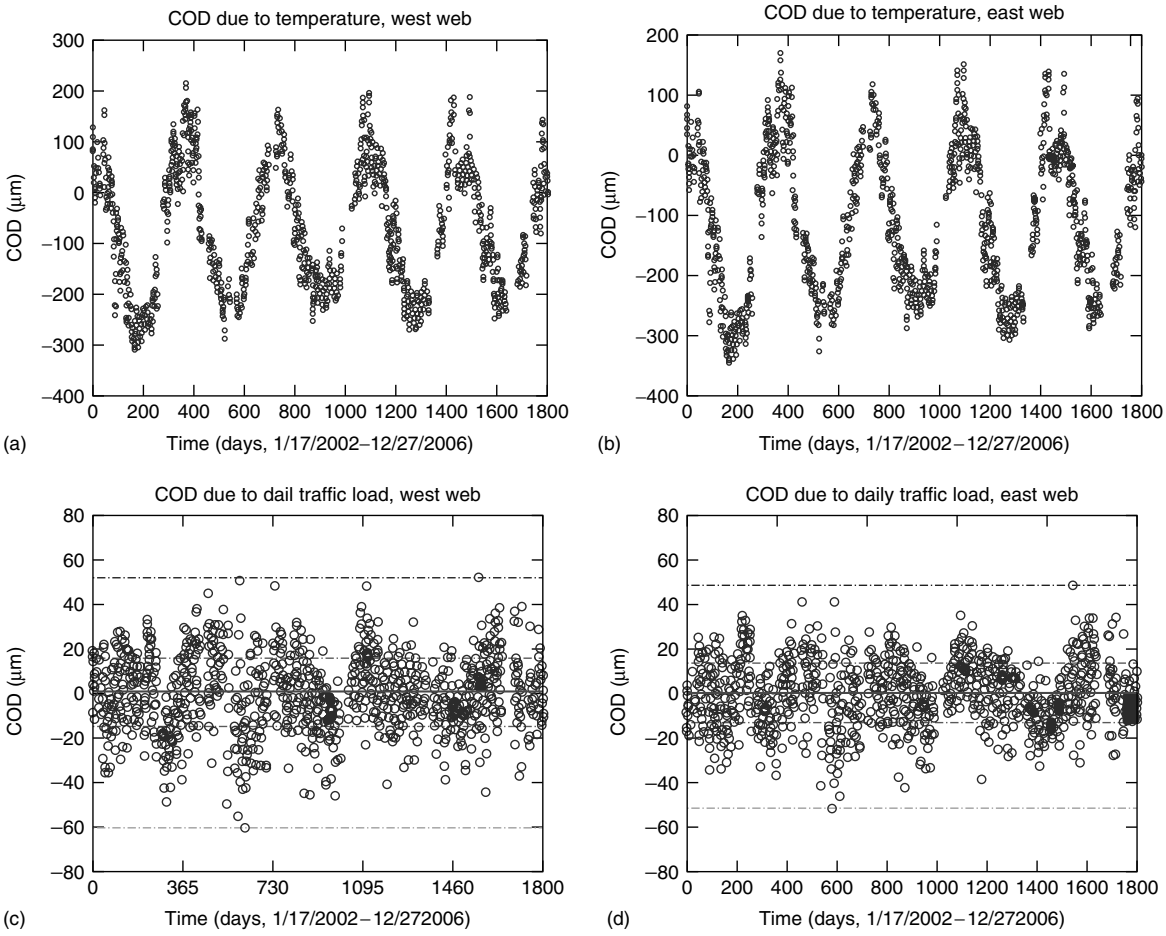


Figure 15. Crack opening displacement: due to temperature for west web (a) and for east web (b); and traffic load for west web (c) and for east web (d).

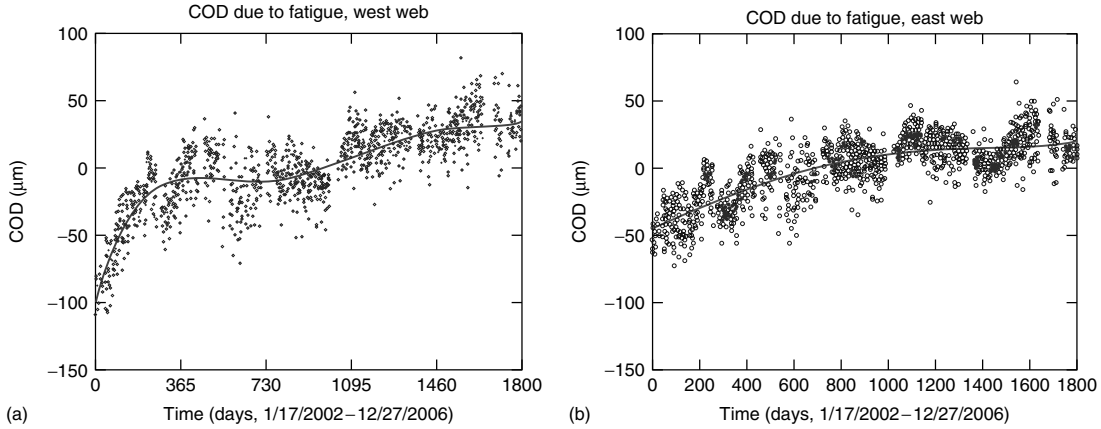


Figure 16. Crack opening displacement due to fatigue for west web (a) and east web (b).

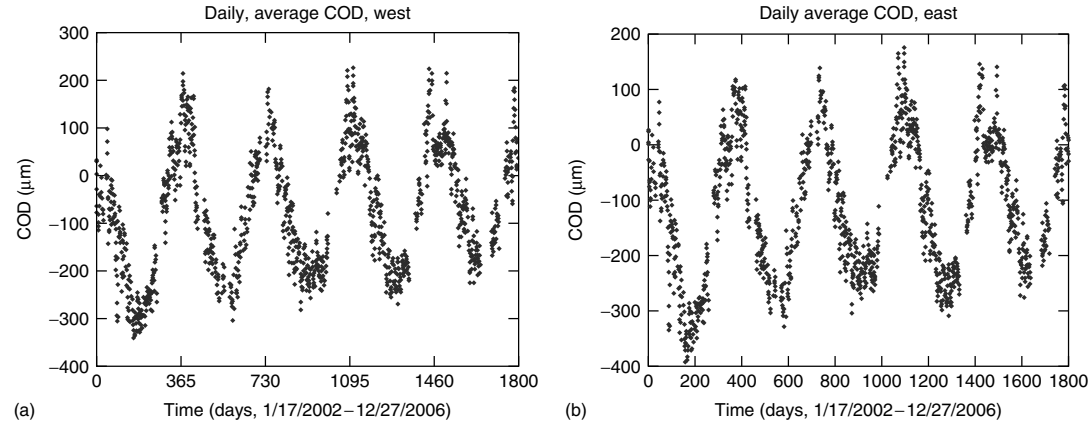


Figure 17. Crack opening displacement (raw data) for west web (a) and east web (b).

accumulation of damage in terms of COD is about 120 μm in five years from 2002 to 2006. Therefore, one can conclude, although temperature effect is about two to three times of traffic effects in terms of COD values, the accumulation of damage may be attributed mainly to truck loadings because of their sheer number of cycles over the years. Overweight trucks are usually cleared by State Highway Department. Unexpected loadings that produced very large COD are frequent. Upon inspection of all data gathered, it is seen that the cracks are active and progressing in a slow pace. Overall, the shear stiffness has reduced by 50%. However, bending stiffness and frequency have not shown the same trend and proportion. On the contrary, frequency has shown negligible effect from continuing local shear cracking.

3 CONCLUSION

A real-time bridge monitoring system includes a real-time data acquisition, a real-time data analysis, and a health reporting system. It should provide the current health status of the bridge in real time. It should be able to determine the current strength and resistant capacity of a structure. The key point is to use the minimum number of sensors to collect, process, and analyze the real-time dynamic and static data from the most critical positions of the bridge.

A large-scale real-time monitoring system can generate huge amount of data every day. The excessive information will overwhelm and decrease the productivity of the bridge engineers if there is no integrated program of data preprocessing and health

diagnosis in the monitoring system. All the real-time raw data shall be preprocessed in the bridge to increase the speed of data transmission, save the capacity of database, and improve the efficiency of health assessment.

The effect of the structure in terms of COD due to temperature is about two times the effect due to traffic loadings in terms of their daily magnitude. However, the accumulation of damage due to fatigue may be mainly attributed to traffic loadings because of the sheer numbers of cycle.

Automatic measurement should not be considered to be the be-all and end-all of bridge health monitoring. Its place is firmly entrenched in assisting engineers to conveniently carry out the damage detection, analysis, and evaluation of bridges. A retrofit mechanism was recommended for Kishwaukee Bridge. Retrofit is underway.

ACKNOWLEDGMENTS

The author gratefully acknowledges the initiation and continuous support in funding from the Illinois Department of Transportation, US National Science Foundation and Smart Structure Inc, and support from Slovak University of Technology in Bratislava is greatly appreciated.

REFERENCES

[1] DeWolf J, Descoteaux T, Kou J, Lauzon R, Mazurek D, Paproski R. Expert systems for bridge monitoring. *Computing in Civil Engineering: Proceedings of the Sixth Conference*. American Society of Civil Engineers: Atlanta, GA, 1989; pp. 203–210.

[2] Carder DS. Observed vibrations of bridges. *Bulletin, Seismological Society of America* 1937 **27**:267–303.

[3] Catbas FN, Grimmelsman KA, Aktan AE. Structural identification of the Commodore Barry bridge. *Proceedings of SPIE*, Vol. 3995, 2000; pp. 84–97.

[4] Cheung MS, Tadros GS, Brown J, Dilger WH, Ghali A, Lau DT. Field monitoring and research on performance of the Confederation Bridge. *Canadian Journal of Civil Engineering* 1997 **24**:951–962.

[5] Ashkenazi V, Roberts GW. Experimental monitoring of the Humber bridge using GPS. *Proceedings of the Institution of Civil Engineers* 1997 **120**:177–182.

[6] DeWolf J, Lauzon RG, Fu Y, Lengyel TF. Long-term monitoring of bridges in Connecticut for performance evaluation of structures. *Performance of Structures: From Research to Design, 2002 Structures Congress*. American Society of Civil Engineers: Denver Colorado, 2002; pp. 195–196.

[7] Taljsten B, Hejll A, James G. Carbon fiber-reinforced polymer strengthening and monitoring of the Grondals bridge in Sweden. *Journal of Composites for Construction* 2007 **11**(2):227–235.

[8] Bampton MCC, Ramsdell JV, Graves RE, Strope LA. *Deer Isle-Sedgwick Suspension Bridge. Wind and Motion Analysis*, Report FHWA/RD-86/183, 1986.

[9] Barr IG, Waldron P, Evans HR. Instrumentation of glued segmental box girder bridges. *Monitoring of Large Structures and Assessment of their Safety*. IABSE: Colloquium Bergamo, 1987.

[10] Brownjohn JMW, Bocciolone M, Curami A, Falco M, Zasso A. Humber Bridge full-scale measurement campaigns 1990–1991. *Journal of Wind Engineering and Industrial Aerodynamics* 1994 **52**:185–218.

[11] Lau CK, Wong KY. Design, construction and monitoring of three key cable-supported bridges in Hong Kong. *Proceedings of the 4th International Kerensky Conference on Structures in the new millennium*. Hong Kong, 1997; pp. 105–115.

[12] Leitch J, Long AE, Thompson A, Sloan TD. Monitoring the behaviour of a major box-girder bridge. *Structural Assessment Based on Full and Large-Scale Testing*. Butterworths, BRE: Garston, 1987; pp. 212–219.

[13] Macdonald JHG, Dagless EL, Thomas BT, Taylor CA. Dynamic measurements of the second severn crossing. *Proceedings of the Institution of Civil Engineers: Transport* 1997 **123**(4):241–248.

[14] Nair RS, Iverson JK. Design and construction of the Kishwaukee river bridge. *PCI Journal* 1982 **27**(6):22–47.

[15] Shiu KN, Russell HG. Knowledge gained from instrumentation of the Kishwaukee river bridge. *PCI Journal* 1983 **28**(5):32–53.

[16] Wang ML and Lloyd GM. etc., *Health Assessment of the Kishwaukee River Bridge*, Technical Report to IDOT. University of Illinois, April 2001.

[17] Lloyd G, Wang ML, Wang X, Halvonik J, Bootstrap analysis of long-term global and local deformation measurements of the Kishwaukee bridge. In *The 4th International Workshop on Structure Health*

Monitoring, Chang F.-K (ed). Stanford University: Stanford, 2003b; pp. 163–171.

[18] Halvonik J. *Stress State Analysis of South-Bound Kishwaukee Bridge*. Habilitation, Department of Concrete Structures and Bridges, Slovak University of Technology, 2002.

[19] Lloyd G, Wang ML, Wang X, Love J. Recommendations for intelligent bridge monitoring systems: architecture and temperature-compensated bootstrap analysis. In *Smart Structures and Materials 2003: Smart Systems and Nondestructive Evaluation for Civil Infrastructures*, Liu S.-C (ed). SPIE: San Diego, 2003a; Vol. 5057, pp. 247–258.

[20] Lloyd G, Wang ML, Satpathi D. The role of eigenparameter gradients in the detection of perturbations in discrete linear systems. *Journal of Sound and Vibration* 2000a **235**(2):299–319.

[21] Peeters B, Maeck J, De Roeck G. Vibration-based damage detection in civil engineering: excitation sources and temperature effects. *Smart Materials and Structures* 2001 **10**:518–527.

[22] Lloyd GM, Wang ML, Singh V. Observed variations of mode frequencies of a prestressed concrete bridge with temperature. In *Condition, Monitoring of Materials and Structures*, Ansari F (ed). ASCE Press: Reston, VA, 2000b; pp. 179–189.

[23] Lloyd G, Wang ML, Wang X. Thermo-mechanical analysis of long-term global and local deformation measurements of the Kishwaukee bridge using the bootstrap. *Earthquake Engineering and Engineering Vibration* 2004a **3**(1):107–115.

[24] Lloyd G, Wang ML, Wang X. Thermo-mechanical analysis of the Kishwaukee bridge from global and local deformation measurements. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Liu S-C (ed). SPIE: San Diego, CA, 2004b; Vol. 5391, pp. 618–623.

[25] Lloyd G, Wang ML, Wang X. *Components of a Real-time Monitoring System for a Segmental Precast Concrete Box Girder Bridge*. Structural Materials Technology (SMT): NDE/NDT for Highways and Bridges 2004: Buffalo, NY, 2004c.

[26] Baant ZP, Kaplan MF. *Concrete at High Temperatures: Material Properties and Mathematical Models*. Longman Group Limited, 1996.

[27] Bentz DP, Clifton JR, Ferrais CF, Garboczi EJ. *Transport Properties and Durability of Concrete: Literature Review and Research Plan*, NISTIR 6395. U.S. Department of Commerce, September 1999.

[28] Basole M. *Thermal Modeling and Field Temperature Measurement of Segmental Box Girder Bridges in Florida*, M.S. Thesis. Florida Atlantic University, 1992.

FURTHER READING

Efron B. Bootstrap methods: another look at the Jackknife. *The Annals of Statistics* 1979 **7**(1):1–26.

Friswell MI, Penny JET, Garvey SD. A combined genetic and eigensensitivity algorithm for the location of damage in structures. *Computer and Structures* 1998 **69**:547–556.

Hunter NF, Paez TL. Applications of the bootstrap to mechanical systems analysis. *Experimental Techniques* 1998 **22**(4):34–37.

Kay SM. Fundamentals of statistical signal processing, detection theory. *Prentice Hall PTR*, 1998 **2**:60–89.

Lloyd GM, Wang ML, Singh V, Dixit PA. A probabilistic analysis of the temperature dependent mode frequencies of a prestressed concrete bridge using a bootstrap statistic. *Proceedings: 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability*. University of Notre Dame, July, 2000c.

Miyata T, Yamada H, Katsuchi H, Kitagawi H. Fullscale measurement of Akashi-Kaikyo bridge during typhoon. *Journal of Wind Engineering and Industrial Aerodynamics* 2002 **90**:1517–1527.

Porter PS, Rao ST, Ku J, Poirot RL, Dakins M. Small sample properties of nonparametric bootstrap t confidence intervals. *Journal of Air and Waste Management Association* 1997 **47**:1197–1203.

Vincent GS. Golden gate bridge vibration study. *ASCE Journal of the Structural Division* 1958 **84**(ST6): 1817–1840.

Wang X, Wang ML. Smart health monitoring system of a prestressed box girder bridge. *HK Proceedings of ICANCEER 2002*. Hong Kong Polytechnic University: Hong Kong, 2002.

Wang X, Wang ML, Zhao Y, Chen H, Zhou LL, Smart health monitoring system for a prestressed concrete bridge. In *Smart Structures and Materials 2004: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Liu S-C (ed). SPIE: San Diego, CA, 2004; Vol. 5391, pp. 597–608.

Chapter 82

Principles of Structural Degradation Monitoring

Charles R. Farrar[1], Keith Worden[2] and Janice Dulieu-Barton[3]

[1] *Engineering Institute, Los Alamos National Laboratory, Los Alamos, NM, USA*
[2] *Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*
[3] *School of Engineering Sciences, University of Southampton, Southampton, UK*

1 INTRODUCTION

It is a fundamental requirement of engineering structures and systems that they operate within limits specified by the environment in which they will be used. At the design stage, the loading of the structure is defined and appropriate material choices are made on the basis of their properties. Prototypes may be built and stringent experimental testing may be carried out. Because of limitations of the experiments and approximations made in any numerical models developed, a true picture of the behavior of a structure is not available until it is in service.

In the past, engineers designed ideal structures, i.e., making the assumption that at all points through the life cycle the structure was damage free. The design philosophy here is the *safe-life* philosophy. The assumption is that the structures or components in question have *no* margin for failure. The safe life of the structure is estimated and safety factors are applied to support the fatigue analysis. The safe operating life is "guaranteed" by selecting a large margin of safety. This overdesigned approach usually leads to high costs and/or poor performance, and structural failures are statistically still possible because of, among other things, unpredicted loads. With the pressure to introduce new lightweight structures, this concept is no longer valid and new design philosophies have emerged. One might consider a *fail-safe* approach, which allows for crack initiation in

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

individual components on the condition that the entire structure operates safely until the cracks are detected. This philosophy will lead to cost reductions and better performance, but will entail increased maintenance. Arguably, the most desirable approach to structural design is the *damage tolerant* approach [1]. This assumes that structural damage is inevitable. The damage size is estimated from load sequences and fracture mechanical models. The main assumption is that the damage location is known, damage initiation can be reliably detected, and the damage severity can be monitored. It is clear that this approach requires reliable monitoring methods. In response to this requirement, the discipline of structural health monitoring (SHM) has emerged.

In the field of condition monitoring (CM) of machinery, there has been a transition in the maintenance philosophies from *run-to-breakdown maintenance*, through *time-based maintenance* to *condition-based maintenance* over the years, and each of the steps has entailed more comprehensive monitoring capabilities.

All these points indicate the desirability of implementing a reliable monitoring strategy. The basis of this article is to establish the requirements for such a monitoring system and it is argued that this necessarily has to be an *intelligent* monitoring system. If intelligent monitoring systems are incorporated, then structures can be designed to operate at the margin of safety without extended periods of inspection, and in the case of components that are overdesigned, material costs would be significantly reduced.

Where the type of health monitoring described in the paragraphs above is in a sense "safety critical", i.e., life by monitoring with no margins, such systems are currently prohibited from incorporating "intelligence" by regulatory bodies (such as the UK Civil Aviation Authority (CAA) and the US Federal Aviation Authority (FAA)). It is not a simple matter to overcome such objections, but the potential benefits of doing so justify the effort. Having said this, there are substantial benefits to be obtained from monitoring that addresses noncritical failures, where monitoring is used to convert unplanned to planned maintenance. This falls outside the scope of the safety-critical regulations, which could greatly accelerate the acceptance of intelligent systems.

In devising an intelligent fault detection system, the primary consideration is an unambiguous definition of damage as a prerequisite to providing a unified approach to damage evaluation across all the engineering disciplines. This requires a taxonomy for the relevant concepts, i.e., a precise definition of what constitutes a fault, damage, and defect. A specification for operational evaluation is an essential feature; the proposed approach makes use of a hierarchical damage identification scheme. An approach to sensor prescription and optimization must also be defined; in the present article issues such as placement and validation are covered. Finally, a data-processing methodology is required—in the current approach, this is based on a data fusion model. Recently, it appeared possible to summarize much of the received wisdom from the field of SHM in a series of axioms [2]. These axioms encode much information, which may provide guidance in designing an effective monitoring system.

The purpose of this article is to set down all the issues involved in progressing to intelligent fault detection and to suggest an overall approach. The authors believe that a holistic approach is required for intelligent fault detection. The structure and the sensors must be treated as one, i.e., at the design stage, provision should be made for fault detection. The holistic approach can be implemented on existing structures, so the issues discussed in the article should not be regarded as only relevant to new structures. The basis of the holistic approach is that not only should the structure be monitored but so should the sensors; hence dealing with sensor failure is a primary concern.

2 TAXONOMY

To fully explore what is meant by intelligence in the context of SHM, first of all, it is imperative to define in an unambiguous manner exactly what constitutes damage and when a system is considered to no longer be operational. The following discussion establishes what the authors consider to be coherent and workable definitions of faults, damage, and defects:

- A *fault* is when the structure can no longer operate satisfactorily. If one defines the *quality* of a structure or system as its fitness for purpose or

its ability to meet customer or user requirements, it suffices to define a fault as a change in the system that produces an unacceptable reduction in quality.

- A *damage* is when the structure is no longer operating in its ideal condition, but can still function satisfactorily, i.e., in a suboptimal manner.
- A *defect* is inherent in the material, and, statistically, all materials contain some unknown amount of defects; this means that the structure can operate at its design condition if the constituent materials contain defects.

The above definitions allow a hierarchical relationship to be developed; i.e., defects lead to damage, and damage leads to faults. In the proposed approach, it is necessary to introduce monitoring systems to obtain a damage tolerant structure, so that it can be decided when the structure is no longer operating in a satisfactory manner. This means that a fault has to have a strict definition, e.g., the stiffness of the structure has deteriorated beyond a certain level. In some cases, a simple definition based on one parameter may not be sufficient. A good example of this is when a crack is propagating in a stable manner, there is an increase in strain in the component and hence a reduction in component stiffness. Once the strain has increased above a certain level, a decision may be made to take the component out of service. On inspection it may be observed that the crack is growing in such a direction that the component will fail safe, so service could have been prolonged. So along with a strain monitor, it would be useful to monitor the direction of crack growth. With this in mind, it is clear that the choice of the fault monitoring system needs to take into account the material type and the operating environment. This approach ensures that the quality of the measurement is optimized and hence an overarching issue is the limitations of the sensor. From the above, it can be seen that the question—*what is a fault*—is based on not only the structure's operating environment but also the type of monitoring system that is used.

Another consideration is the level of damage tolerance required. Some systems may not require any damage tolerance. An example of this is the crumple zones in a car; these are designed to fail on impact in order that the energy is absorbed by the structure and not the driver and passengers. There would be little point in monitoring these areas in a car and the crumple zones could be defined as *areas with infinite fault tolerance*.

Finally, a means of identifying procedural faults must be included in any intelligent fault monitoring system. If a structure is being incorrectly used, this may cause a fault; however, it is important to distinguish between this type of fault and the more general *systemic* fault. An example of a procedural fault would be mode of operation that increases the load beyond the design load for a limited period of time. The increase in load does not cause immediate failure, but significantly reduces the product life. A type of alarm system that triggers when the design loads are exceeded could be used; this would alert the operator that there is a problem and hence prevent a repeat of the incident and thus increase the product life. An alternative monitoring philosophy is in current use in the aircraft industry. Health and usage monitoring systems (HUMS) do not just consider overloads, but record or examine all loads and modify the remaining life in the light of usage.

3 INTELLIGENT FAULT DETECTION

The discussion can now proceed to matters of detection and how it can be achieved with intelligence. The first observation one might make is that fault detection is, in a sense, trivial, as a fault is defined as a change in the condition of the structure that produces an unacceptable reduction in quality. By implication, such a change will be evident. Thus, *intelligent* fault detection actually entails detecting the damage that will, if not corrected, lead to a fault.

Detection of damage is a facet of the broader problem of *damage awareness* or *damage identification.* The objective of a monitoring system must be to accumulate *sufficient* information about the damage for appropriate remedial action to be taken to restore the structure or system to high-quality operation or at least to ensure safety. Also, efficiency demands that only the *necessary* information should be returned by the monitor. With this in mind, it is helpful to think of the identification problem as a hierarchical structure, in the same way as one can think of the evolution of the fault as a hierarchical structure. This

train of thought began with Rytter in his PhD thesis [3]. The original specification cited four levels, but an additional one is given here—that of classification. The modified structure is

1. detection: the method gives a qualitative indication that damage might be present in the structure;
2. localization: the method gives information about the probable position of the damage;
3. classification: the method gives information about the type of damage;
4. assessment: the method gives an estimate of the extent of the damage;
5. prediction: the method offers information about the safety of the structure, e.g., estimates a residual life.

While the location of the classification step is arguable—some might put it before localisation—few would argue that the structure above summarizes the main issues in damage identification. The vertical structure is clear; each level largely requires that all lower-level information is available (with the proviso just mentioned). Note that the damage identification scheme should, if possible, be implemented on-line, i.e., during operation of the structure; in this case, prediction must also be understood as an estimate of the residual safe life of the structure obtained during operation. For an aircraft in flight, for example, this is critical. If the diagnostic system signals serious damage, but fails to indicate that there is time to land, the aircraft may be lost needlessly and at great expense, when the crew bail out.

Classification is added to the original scheme because it is important, if not vital, for effective identification at level 5 and possibly at level 4. Level 5 is distinguished from the others in that it cannot be accomplished without an understanding of the physics of the damage, i.e., characterization. Level 1 is also distinguished in the following sense—it can be accomplished with no prior knowledge of how the system will behave when damaged. To explain this, a slight digression on pattern recognition (PR) or machine learning is needed.

Many modern approaches to damage identification are based on the idea of *pattern recognition*. In the broadest sense, a PR algorithm is simply one that assigns to a sample of measured data a class label, usually from a finite set. In the case of damage identification, the measured data could be vibration mode shapes, full-field thermoelastic data, scattered wave profiles, etc. The appropriate class labels would encode damage type, location, etc. To carry out the higher levels of identification using PR, it is almost certainly necessary to construct examples of data corresponding to each class. That is, to establish that a given set of measurements from a composite panel shows the presence of a delamination, the algorithm must have prior knowledge of what data from a delaminated panel looks like as opposed to one with, say, a resin-rich area. Each possible fault class should usually have a *training set* of measurement vectors that are associated uniquely with it. Many PR algorithms work by *training* a diagnostic, for example, a neural network can learn by example; it is shown the measurement data and asked to produce the correct class label; if the result differs from the desired label, the network is corrected. Typically, many presentations of data are required. This type of learning algorithm in which the diagnostic is trained by showing it the desired label for each data set is called *supervised learning*.

If supervised learning is required, there will be serious demands associated with it; data from every conceivable damage situation should be available. The two possible sources of such data are computation or modeling, and experiment. Modeling presents problems if the structure or system of interest is geometrically or materially complex, for example, finite element (FE) analysis of structures requiring a fine mesh can be extremely time consuming even if the material is well understood. Structures with composite or viscoelastic elements may not even have accurate constitutive models. The damage itself may be difficult to model; it may also make the structure *dynamically* nonlinear, i.e., an opening–closing fatigue crack, which also presents a formidable problem. Unfortunately, the situation is no better for experiment. To accumulate enough training data, it would be necessary to make copies of the system of interest and damage it in all the ways that might occur naturally; for high-value structures like aircraft, this is simply not possible. This is arguably the main problem associated with data-driven approaches to SHM.

Fortunately, there is an alternative to supervised learning—*unsupervised learning*. However, this

mode of learning only applies to level 1 diagnostics, i.e., it can only be used for detection.

The techniques are often referred to as *novelty detection* or *anomaly detection* methods [4–6]. The idea of novelty detection is that only training data from the normal operating condition of the structure or system is used to establish the diagnostic. A model of normal condition is created; later, during monitoring, newly acquired data is compared with the model, and if there are any significant deviations, the algorithm indicates novelty. The implication is that the system has departed from normal condition, i.e., acquired damage. The advantage of such an approach is clear. If the training data is generated from a model, only the unfaulted condition is required, and this simplifies matters considerably. From an experimental point of view, there is no need to damage the structure of interest. Although novelty detection is only a level 1 approach, there are many situations where this suffices, i.e., safety-critical systems where any fault on the system would require it to be taken out of service.

It is an important qualifier that the novelty detectors should flag only *significant* deviations from normal operating condition. All real systems are subject to measurement noise and usually operate in a changing environment; the monitor must be able to distinguish between a statistical fluctuation in the data and a real deviation from normality. This means that of the various flavors of PR existing [7], the most appropriate one is *statistical pattern recognition* (SPR). Another important observation is that there may be variations in the normal condition that are not statistical, i.e., the characteristics of the structure may vary with changing environmental conditions, and this must be addressed. In general, it is important that the algorithms used for damage identification should account properly for sources of uncertainty and variation in the data. The algorithms should also, as far as possible, return a confidence interval with their diagnosis.

The term *normal operating condition* requires some discussion. As stated in the previous section, defects are always present in a structure to some extent. The normal operating condition therefore means a state of the system when there is some assurance, statistical or otherwise, that the system is fit for purpose. In some cases, there may be macroscopic damage, i.e., a fatigue crack; however, if it is known that the crack will not grow under, for example, the standard loadings on the system, the state qualifies as in a normal operating condition. Novelty detection will then look for new cracks or unexpected growth of the old crack.

The discussion above is intended to show that there is often a trade-off between the level of a diagnostic system and the expense of training it adequately. Given this fact, the main requirement of an *intelligent* fault detection system then is that it should return information at the apposite level for the context. It should measure the appropriate data and process this with the appropriate algorithm. It should take proper account of uncertainty in the data and return a confidence level in its diagnosis.

4 SENSOR ISSUES

4.1 Operational evaluation

The first demand of an intelligent fault detection system is that it should measure the appropriate data. This simple statement hides a multitude of problems. The holistic approach to fault identification requires that the diagnostic system should be carefully designed with the objectives in mind. This means that, at the very least, decisions must be made about the type of sensors to be used and the placement of those sensors. Before this, a preplanning or evaluation stage is desirable, like the *operational evaluation* stage defined by Farrar and Doebling [8]. This requires the architect of the monitoring system to first

- provide economic and/or life-safety justifications for performing the monitoring;
- define system-specific damage including types of damage and expected locations; e.g., failure modes effects analysis (FMEA) and failure modes effects and criticality analysis (FMECA) [9];
- define the operational and environmental conditions under which the system functions;
- define the limitations on data acquisition in the operational environment.

Each of these requirements can substantially support the design process. The second of the four raises two important issues. First, as observed in the previous section, a supervised learning scheme

for the diagnostic requires training data; to build a model or specify an experimental program to generate training data, one must specify the expected damage classes (type, severity, location, etc.). Secondly, given *a priori* information about likely damage locations, one can make an informed choice of whether to design for *local* or *global* monitoring, or a hybrid of the two. For example, if certain "hot spots" on a structure are identified as critical regions, they can be inspected using a high-resolution local method like active ultrasound, leaving the rest of the structure to be monitored using a less sensitive, but global, vibration-based method.

Environmental conditions must be considered when designing a monitoring system. Consider the novelty detection methods described in the last section; if the data used to characterize the normal operating condition does not span the whole range of operational and environmental conditions observed in practice, it is likely to signal novelty when a previously unseen condition occurs. The system then erroneously diagnoses a fault. Examples of systems where this occurs abound: an offshore structure changing its mass due to oil storage or marine fouling, an aircraft before and after dropping a store, and a bridge in a desert area undergoing substantial temperature changes as day changes to night.

There are likely to be obstructions to implementing the optimum monitoring system. This places constraints on the optimization problem for sensor placement. In situations where the monitoring system is needed for a preexisting structure, certain useful, if not critical, locations may be inaccessible. For a truly holistic approach to damage monitoring, the monitoring system should be designed as an integral part of the structure or system. In the case of control, a minimum requirement for the sensor system is to measure the current state of the plant during normal condition. This may be insufficient for fault detection as the deviations from normal condition may be orthogonal to the measured dimensions. It is clearly necessary, therefore, to anticipate this during the design process; FMEA may give sufficient guidance for expected faults.

If one is taking a unified approach to damage identification, where the problem may be to monitor a civil engineering structure, a machine, or a chemical process, one should add a further requirement to the operational evaluation stage:

- identify context-specific features and decide the appropriate level for monitoring.

This should be the first or second consideration.

4.2 Sensor placement

The most critical issue in the specification of a monitoring system is the type and location of the sensors; there can be no monitoring without the appropriate sensors. Specification of the type of sensors is arguably a matter of operational evaluation; knowledge of the expected damage types is valuable if not crucial there. Also the question of local versus global monitoring must be settled at this stage. The optimal positioning of the sensors is critical for true intelligence in monitoring. In principle, this problem can be solved using an optimization scheme; various approaches are discussed in **Sensor Placement Optimization** in this volume.

4.3 Sensor validation and failure safety

The holistic approach to health monitoring demands that the sensor network be an integral part of the structure or system. It immediately follows from this that damage to the system may manifest itself as damage to the sensor network. A fault in the sensor network—which must be regarded as a systemic fault—will have undesirable consequences, whether or not the integrity of the overall structure is compromised. A sensor failure that causes an unnecessary alarm may cause the system to be needlessly taken out of service. A sensor failure that causes a fault to go unnoticed may have severe cost or safety implications—the monitoring system might as well not be present. The implication of this argument is that the sensor network itself should be monitored. Redundancy or diversity should be built in, if necessary, to avoid the monitoring system itself making a significant contribution to fault reporting.

There are essentially two approaches to monitoring the sensors. First, the individual sensors can be self-monitoring. This is the approach taken by Henry and Clarke in the SEVA program [10]. This is currently only an option with larger sensors—the Coriolis mass flow meter described in the SEVA work can

accommodate a communications bus. The sensor is capable of self-diagnostics and also returns the on-line uncertainty of the measurements. The alternative for sensors like accelerometers, which cannot accommodate substantial electronics, is to allow the sensors to monitor each other as described, for example, by Kramers [11, 12]. The crucial feature of the latter approach is that the sensor outputs are correlated, i.e., there is *redundancy* in the sensor network. Kramers' approach not only allows the identification of errant sensors but also generates an approximate version of what the deviant sensor *should* read. The development of smart sensors like piezoceramics and piezopolymers also opens up possibilities. The ability of piezoelectrics to act as both sensors and actuators allows the possibility of active validation, i.e., each sensor, in turn, can act as a signal generator, and the consequent readings on the remaining sensors can be compared with a template to give a health report.

The question of redundancy has been raised. This should also be a requirement of an intelligent monitoring system. If a sensor is diagnosed as defective, the information that would have been delivered by the sensor should be available elsewhere. This issue should be considered during the design stage and made part of the sensor optimization. In the future, it may be possible to perform on-line reallocation of the sensors; at present, the only option is to include backup sensors. The simplest approach is to collocate sensors at critical points, but to keep only one active at a given time; redundant sensors are switched in when damage occurs to an active sensor. This approach is clearly suboptimal. Also certain causes of damage, e.g., impact, are likely to affect the collocated pairs. Another approach is to overdesign the sensor network in such a way that damage to a single sensor still leaves a network that is optimized for monitoring in some sense. This approach to fail-safe sensor optimization is discussed in [13, 14].

Finally, it must be stated that sensors alone do not suffice for any real level of damage identification. The one possible exception is for level 1, in situations where novelty detection amounts to exceedance of a sensor reading above a given threshold. In all other cases, some form of data processing is required; intelligent use of signal-processing algorithms is essential, and a principled strategy for this is discussed later.

5 DATA PROCESSING FOR DAMAGE IDENTIFICATION

Once the operational evaluation stage has passed and the sensor network has been designed, the health monitoring system can begin to deliver data. The choice and implementation of algorithms to process the data and carry out the identification is arguably the most crucial ingredient of an intelligent fault detection strategy. Before even choosing the algorithm, it is necessary to choose between two complementary approaches to the problem:

- damage identification is an inverse problem;
- damage identification is a PR problem.

The first approach usually adopts a model of the structure and tries to relate changes in measured data from the structure to changes in the model; sometimes, locally linearized models are used to simplify the analysis. The algorithms used are mainly based on linear algebra or optimization theory, and an excellent survey of the dominant methods can be found in [15] (*see* **Modal–Vibration-based Damage Identification**).

The second approach is based on the idea described in Section 3, whereby measured data from the system of interest are assigned a damage class by a PR algorithm. This is the approach that is chosen here for detailed discussion. There is no implied criticism of the inverse problem approach; the authors are simply concentrating on the framework with which they are most familiar. For a critical appraisal of inverse problem approaches to damage identification, the reader can refer to [16].

The data-processing element of a monitoring system consists of all actions on the data upstream from the point of acquisition by the sensors. The ultimate product of the analysis is a decision as to the health of the system. The analysis has been neatly summed up by Lowe [17] as the D2D (data to decision) process. The principled approach—the intelligent approach—to the D2D process is based on ideas of *data fusion* (*see* **Data Fusion of Multiple Signals from the Sensor Network**).

Sensor and data fusion as a research discipline in its own right emerged largely because of various defense

organizations attempting to formalize procedures for integrating information from disparate sources. There are many definitions of data fusion, but one of the first—which has endured—came from the North American Joint Directors of Laboratories, who were charged by the US military to establish a standard data fusion terminology [18]. Their definition of data fusion was as follows:

> A multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data from single and multiple sources.

The object of the exercise was to determine a battlefield situation and assess threat on the basis of data from various sources. The philosophy of data fusion was quickly recognized to have broader applications: initially in the fields of meteorology and traffic management, but later in the medical field and in nondestructive evaluation (NDE). The first models of the D2D process were couched exclusively in military terms, but as time wore on, a more general terminology was adopted. One of the first general models for data fusion was the *waterfall model* developed in DERA, the UK Defence Evaluation Research Agency [19]. It is no longer widely used; however, it suffices to illustrate the main stages of the D2D process. The model is illustrated in Figure 1.

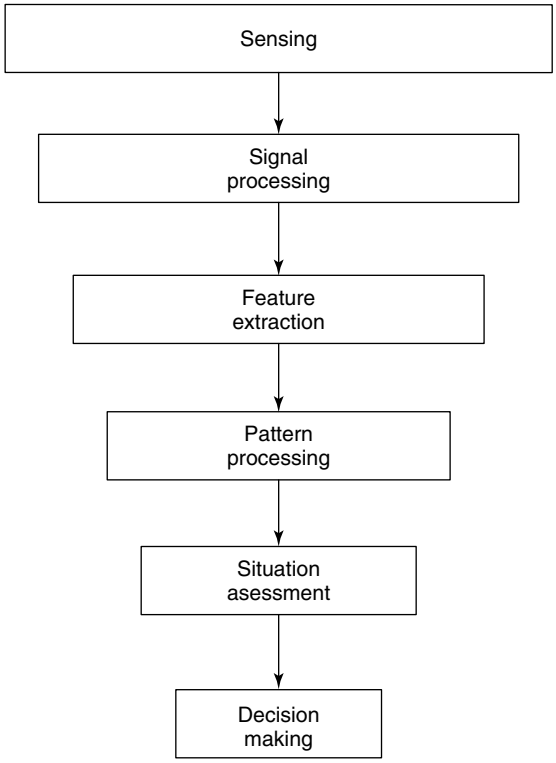


Figure 1. The waterfall model.

Apart from the situation assessment stage, which is an artifact of the old military terminology, all the important processing stages for SHM are present in the model.

Beyond the sensor level, which generates the raw data, the first stage is *signal processing*. This should more properly be called *preprocessing*. The purpose is to prepare the data for feature extraction, but this is dealt with in detail later. The preprocessing stage can encompass two tasks. The first of these is *data cleansing*. Examples of cleansing processes are filtering to remove noise, spike removal by median filtering, removal of outliers (care is needed here as the presence of outliers is one indication that the data is not from normal condition), and treatment of missing data values. The second (optional) preprocessing stage is a preliminary attempt to reduce the dimensions of the data vectors and further denoise the signal. For example, given a random time series with many points, it is often useful to convert the data to a spectrum by Fourier transformation. If the signal is divided into contiguous blocks before transformation and the resulting spectra are averaged, the number of points in the spectrum can be much lower than in the original time history and noise is averaged away. Another advantage of treating the time signal this way is that the data vector obtained should be independent of time. If the original time series is random, it makes little sense to compare measurements at different starting times. The preprocessing is usually carried out on the basis of engineering judgment and experience. At this stage, the aim would be to reduce the dimension of the data set from possibly many thousands to perhaps a hundred.

The second stage is *feature extraction*. The term *feature* comes from the PR literature and is short for *distinguishing feature*. Recall that the fundamental problem of PR is to assign a class label to a vector of measurements. This task is made simple if the data contains dominant features that distinguish it from data from other classes. In general, the components of the signal that distinguish the various damage classes

are hidden by features that characterize the normal operating condition of the structure, particularly when the damage is not yet severe. The task of feature extraction is to magnify the characteristics of the various damage classes and suppress the normal background. Suppose the raw data from the sensors is a time series of accelerations from the outside of a gearbox casing. Further, suppose that the time data has been preprocessed and converted into an averaged spectrum. Feature extraction, in this situation, could be extracting only the spectral lines at the meshing frequency and its harmonics, as these lines are known to be sensitive to damage. So feature extraction can be carried out on the basis of engineering judgment also. Alternatively, statistical or information theoretic algorithms like principal component analysis (PCA) can be used to reduce the dimension. The resulting low-dimensional data set is the *feature vector* or *pattern vector* the PR algorithm will use to assign a class. The aim of this stage would be to generate a feature vector of dimension less than 10. A low-dimensional feature vector is a critical element in any PR problem as the number of data examples needed for training grows explosively with the dimension of the problem. Care must be taken at this stage that the information discarded in the dimension reduction is not relevant for diagnosing the damage. Feature extraction should only discard components of the data that do not distinguish the different system states and thus concentrate the information about damage.

The next stage is *pattern processing*. This is the application of an algorithm that can decide the damage state on the basis of the given feature vector. An example would be a neural network that has been trained to return the damage type and severity when presented with say, condensed spectral information from a gearbox. Three types of algorithms can be distinguished depending on the desired diagnosis.

1. Novelty detection
As discussed earlier, the algorithm must simply indicate whether the data comes from normal operating condition. This is a two-class problem, which has the advantage that unsupervised learning can be used. Methods for novelty detection include outlier analysis [20], kernel density methods [5], autoassociative neural networks [21], Kohonen networks [22], growing radial basis function networks [23], and methods based on SPC control charts [24].

2. Classification
In this case, the output of the algorithm is a discrete class label. To apply such an algorithm, the damage states must be quantized, i.e., for location, the structure should be divided into labeled substructures. In this case, the algorithm could only locate to within a substructure, so resolution of what is essentially a continuous parameter may not be good unless many labels are used. However, this type of algorithm is useful in the sense that the algorithms can be trained to give the probability of class membership; this gives an inbuilt confidence factor in the diagnosis. In the case, where the desired diagnosis is from a discrete set, e.g., for diagnosing damage type, this class of algorithms are singled out. Examples of algorithms include neural network classifiers trained with the *1 of M* rule, linear and quadratic discriminant analysis, kernel discriminant analysis, and nearest neighbor classifiers. A comparison of some of these approaches on a damage classification problem is given in [25].

3. Regression
In this case, the output of the algorithm is one or more continuous variables. For location purposes, the diagnosis might be the Cartesian coordinates of the fault, and for severity assessment, it could be the length of a fatigue crack. The regression problem is often nonlinear and is particularly suited to neural networks. As in the classification case, it is often possible to recover a confidence interval for a neural network prediction [26].

In all cases, the pattern processing is subject to an important limitation. There is a trade-off between the resolution of the diagnosis and the noise rejection capabilities of the algorithm. Put simply, if the data is always noise-free, there will be very little fluctuation in the measurement from normal operating condition; in this case, small damages will cause detectable deviations. If there is much noise on the training data, it will be difficult to distinguish fluctuations due to noise and deviations due to damage, unless the damage is severe. One of the tasks of feature extraction is to eliminate, as far as possible, fluctuations on the normal condition data. This optimization for

performance is a requisite feature of intelligent fault detection.

The concepts of PR and machine learning have been applied in the framework of fault detection on many occasions. Detailed discussions can be found in this volume (*see* **Statistical Pattern Recognition**; **Machine Learning Techniques**; **Artificial Neural Networks**; **Novelty Detection**).

The final stage in the D2D chain for the waterfall model is the *decision*. This is a matter of considering the outputs of the PR algorithm and deciding whether action needs to be taken, and what that action should be.

The waterfall model is a fusion model in the sense that it allows for a sensor network with different types of sensors, all of which generate relevant information that can be fused together at the preprocessing or feature extraction stages. In general, fusion models also allow information to be combined at the pattern level or the decision level. The objective, at all times, is to reach a decision with higher confidence than can be reached using any of the information sources alone.

A better fusion paradigm than the waterfall model is the *omnibus* model. This is illustrated in Figure 2.

This has several advantages over the waterfall model. First of all, it includes the possibility of *action*. In the context of SHM, this could be repair. In the event of a sensor failure, this could entail the reallocation of sensors or the switching in of redundant sensors. The model has a loop structure

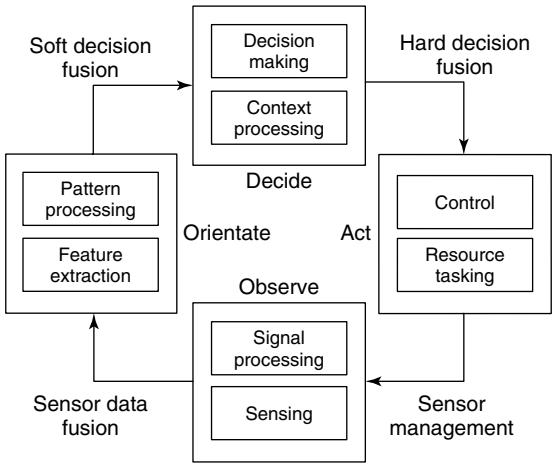


Figure 2. The Omnibus model for data fusion.

that makes clear the fact that action need not interrupt the monitoring process, but may enhance it.

6 THE AXIOMS OF SHM

On the basis of information published in the extensive amount of literature on SHM over the last 20 years [15, 27], the authors feel that the field has matured to the point where several fundamental axioms, or accepted general principles, have emerged. They are presented here in the spirit that they should guide the design and implementation of an SHM system. At the risk of repeating a little of the material discussed earlier, the axioms are discussed in some detail below. The material follows the discussion in [2] very closely. The said paper can be consulted for detailed case-study material relating to each axiom.

The word "axiom" is being used here in a slightly different way to that it is understood in the literature of mathematics and philosophy. The word "axiom" here is used to represent a fundamental truth at the root of any SHM methodology. However, the axioms are not sufficient to generate a given methodology. First of all, the authors are not suggesting that the axioms proposed here form a complete set; it is possible that there are other fundamental truths that have been omitted here. Secondly, the axioms here do not specify the "operators" for SHM; to generate a methodology from these axioms, it is necessary to add a group of algorithms that will carry the SHM practitioner from the data to a decision. It is the belief of the authors that these algorithms should be drawn from the discipline of SPR and algorithms of this type are used in the illustrations throughout this article.

The structure of the following discussion is simple: each of the following sections discusses the proposed axioms in turn.

Axiom 1. *All materials have inherent flaws or defects.*

Metals are never perfect single crystals with a perfect periodic lattice. Broberg [28] provides an excellent account of the inception and growth of microcracks and voids in the *process region* of a metal. In fiber-reinforced plastics (FRPs), defects can also occur at the macrostructural level because of voids produced in manufacturing. These defects compromise the

strength of the material, as coalescence of defects in extreme loading regimes will lead to macroscopic failure at the component level and subsequently at the system level. However, engineers have learned to overcome the design problems imposed by the inevitability of material imperfections by adopting design philosophies like those discussed in the introduction here.

In many engineering materials, the effects of nano-/microstructural level defects can be subsumed into the average material properties such as yield stress or fatigue limit and are not typically considered as "damage". However, in other circumstances, like in composite materials, this may not be the case, and the void content of the material should be regarded as initial "damage". There is no way (under dynamic load) of preventing damage evolution from voids and the associated degradation of the material properties.

As all materials contain imperfections or defects, the difficulty is to decide when a structure is "damaged". This is where the taxonomy of Section 2 comes into play. The strict definition of failure coupled with a monitoring system allows one to consider the concept of prognosis. Here prognosis means the prediction of the structure's future operational life given some assessment of its current condition and some prediction of its anticipated future operational environment.

Axiom 2. *The assessment of damage requires a comparison between two system states.*

This is possibly the most basic of the proposed axioms. It is necessary to state it explicitly as it is sometimes stated that some approach or other "does not require a baseline". It is argued here that this statement is simply never true, and any misunderstanding lies with the assumed meaning of "baseline". In the usual PR approaches to SHM, a training set is required. In the case of damage detection where novelty detection approaches can be applied, the training set is composed of samples of features that are representative of the normal condition of the system or structure of interest. For higher levels of diagnosis requiring estimates of damage location or severity, the training data must also contain samples of normal condition data, but also must be augmented with data samples corresponding to the various damage conditions. In this case, there is no argument that the normal condition data constitutes the "baseline". As discussed earlier, the basic level of damage identification is detection and this can be accomplished using novelty detection algorithms. All novelty detectors in current use rely on the acquisition of a normal condition training set; this includes autoassociative neural networks [21], probability density estimators [5], and, more recently, negative selection algorithms [29].

The necessity of Axiom 2 is not confined to damage detection methods based on PR, but it is also a requirement of the large class of algorithms based on linear algebraic methods. It is perhaps most obviously needed in the case of FE updating methods [30], a class of algorithms that have proved successful at the higher levels of damage identification, i.e., in the location and quantification of damage [31, 32]. The FE updating methodology is based on the construction of a high-fidelity FE model of the structure of interest in its normal condition. To assure that the model provides an accurate description of the virgin-state system, it is usually updated on the basis of experimental data, i.e., the parameters of the FE model (e.g., stiffness indices) are adjusted to bring it into closer correspondence with the experimental observations. The process of damage identification then proceeds by further updating the model on the basis of monitored data. Clearly, any further need for parameter adjustment will arise because the system has changed, and this change is assumed to be caused by damage. The particular elements adjusted will pinpoint the location of the damage and the size of the adjustment provides an estimate of the severity of the damage.

Several approaches claim to operate without a baseline data, which might be interpreted as implying that they do not to require a comparison of systems states. It is argued here that this is just a matter of terminology. One such method is the strain energy method of Stubbs and Kim [33]. This approach appears only to operate on data from the damaged structure. Roughly speaking, an estimate of the modal curvature is used to locate and size the damage. In fact, one might argue that there is an implicit baseline or model for this data on the assumption that the undamaged structure behaves as an Euler–Bernoulli beam. Also, the feature computed—the curvature—cannot be used without a threshold of significance, which is computed on the understanding that most of

the estimated curvature data comes from the rest of the structure that is undamaged.

The fact that damage detection algorithms require a comparison of system states is at the root of one of the main problems in SHM. If the normal condition or baseline state changes because of environmental or operational variations, the application of a novelty detection algorithm may yield a false-positive indication of damage. This issue is discussed in more detail when Axiom 4 is considered. The choice of baseline is one of the issues that may be addressed during the operational evaluation phase.

Axiom 3. *Identifying the existence and location of damage can be done in an unsupervised learning mode, but identifying the type of damage present and the damage severity can generally only be done in a supervised learning mode.*

The statistical model development aspect of SHM is concerned with the implementation of the algorithms that operate on the extracted damage-sensitive features to quantify the damage state of the structure. The algorithms used in statistical model development usually fall into three categories. When data is available from both the undamaged and damaged structure, the statistical PR algorithms fall into the general classification referred to as *supervised learning. Group classification* and *regression analysis* are categories of supervised learning algorithms and are generally associated with either discrete or continuous classification, respectively. *Unsupervised learning* refers to algorithms that are applied to data not containing examples from the damaged structure. *Outlier* or *novelty detection* is the primary class of algorithms applied in unsupervised learning applications. All of the algorithms analyze statistical distributions of the measured or derived features to enhance the damage detection process.

As previously discussed, the damage state of a system can be described in hierarchical terms as in Rytter's structure discussed in Section 3.

Answers to the questions posed in the damage identification hierarchy in the order presented represent increasing knowledge of the damage state. When applied in an unsupervised learning mode, statistical models can typically be used to answer questions regarding the existence and location of damage. As an example, if a damage-sensitive feature extracted from measured system response data exceeds some predetermined threshold, one can conclude that damage has occurred. This conclusion also must rely on the knowledge that the change in the feature has not been caused by operational or environmental variability. Many approaches to damage detection in rotating machinery, such as those that examine change in the kurtosis values of the acceleration amplitude response to identify bearing damage, are based on such outlier analysis [34]. Similarly, changes in features derived from relative information obtained from an array of sensors can be used to locate the damage as is done in many wave propagation approaches to SHM [35]. In general, these statistical procedures cannot distinguish between possible damage types or assess the severity of damage without additional information.

Axiom 4a. *Sensors cannot measure damage. Feature extraction through signal processing and statistical classification are necessary to convert sensor data into damage information.*

Sensors measure the response of a system to its operational and environmental input. Therefore, there is nothing surprising about the fact that sensors cannot *directly* measure damage. In a more basic context, it is similarly impossible to measure stress. The solution is to measure a quantity—the strain—from which one can infer the stress. (In fact, things are a little more indirect than this. The important point is that the sensor yields a value linearly proportional to the physical quantity of interest. Knowledge of the material properties then allows one to infer the stress from the strain measure.) In other words, the stress σ is a known function of the strain ε,

$$\sigma = f(\varepsilon) \tag{1}$$

In this case, the function f is known from observations of basic physics and is particularly simple. The situation is a little more complicated for damage. Suppose, for the sake of simplicity, that the damage state of a given system is captured by a scalar D. The first objective of damage identification is to measure some quantity, $\mathbf{x}$, usually vectorial (or alternatively—multivariate), which is a function of the damage state,

$$\mathbf{x} = \mathbf{f}(D) \tag{2}$$

The main difficulty for SHM is that the function **f** is generally not known from basic physics and must usually be *learned* from the data. The data in question may often be of high dimensionality such as a computed spectrum or a sampled wave profile.

The main problem associated with the machine learning or PR techniques used to learn the function in equation (2) is their difficulty in dealing with data vectors of high dimensionality. This limitation is sometimes termed the *curse of dimensionality*. If one considers methods depending on the availability of *training data*, The curse is simply that, to obtain accurate diagnostics, the amount of training data theoretically grows explosively with the dimension of the patterns [36].

From a pragmatic point of view, there are two solutions to the problem. The first is to obtain adequate training sets. Unfortunately, this will not be possible in many engineering situations, because of the limitations on the size and expense of testing programs. The second approach is to reduce the dimension of the data to a point where the available data is sufficient—this is the feature extraction discussed earlier. However, there is a vital caveat: the reduction or compression of the data must not remove the influence of the damage. (Also, one should be aware that if the high-dimensional features are insensitive to damage, no amount of dimension reduction will help.) This caveat means that the feature extraction must be tailored to the problem. This approach may be feature selection based on engineering judgment as in selecting the meshing harmonics from gearbox vibrations. More principled approaches to dimension reduction may be pursued, but care should be taken. For example, if one uses PCA one certainly obtains a reduced dimension feature; however, this vector is obtained using a criterion that may not preserve the information from damage.

Axiom 4b. *Without intelligent feature extraction, the more sensitive a measurement is to damage, the more sensitive it is to changing operational and environmental conditions.*

This section discusses what are called *intelligent feature extraction*. The concern being addressed here is that the features derived from measured data will not only depend on the damage state but may also depend on an environmental and/or operational variable. Temperature θ is used here for illustrative purposes. Equation (2) then becomes,

$$\mathbf{x} = \mathbf{g}(D, \theta) \tag{3}$$

The machine learning problem is complicated by the fact that one wants to learn the dependence on D, despite the fact that some of the variation in the measurand is likely to be caused by variation in θ. The problem of feature extraction is then to find a reduced dimension quantity that depends on the damage, but not the temperature.

An alternative approach to the problem posed in this section is to learn both the dependence on damage and the environment (D *and* θ). Note that as this learning problem is a mixed one, supervised learning is used to obtain the temperature dependency, but unsupervised learning is used to detect the damage. This approach is not considered here, but a discussion can be found in [37]. Finally, the simplest approach to the overall problem here would be to directly select features that are sensitive to damage, but not to environmental and operational variations. This is not easy, in general; an interesting example of this strategy in the context of wave propagation features can be found in [38].

Axiom 5. *The length and timescales associated with damage initiation and evolution dictate the required properties of the SHM sensing system.*

Axiom 1 introduced the concept of the length scales associated with damage and pointed out that defects are present in all materials beginning at the atomic length scale and spanning scales where component and system level faults are present. In terms of timescales, damage can accumulate incrementally over periods of time exceeding years, such as that associated with some types of fatigue or corrosion damage accumulation. Damage can also result on fraction-of-second timescales from scheduled discrete events such as aircraft landing impacts and from unscheduled discrete events such as blast loadings and natural phenomena hazards like earthquakes.

Axiom 4a states that a sensor cannot measure damage. Therefore, the goal of any SHM sensing system is to make the sensor reading as directly correlated with, and as sensitive to, damage as possible. At the same time, one also strives to make the

sensors as independent as possible from all other sources of environmental and operational variability. To best meet these goals for the SHM sensor and data-acquisition system, the following sensing system properties must be defined:

1. types of data to be acquired;
2. sensor types, number, and locations;
3. bandwidth, sensitivity, and dynamic range;
4. data acquisition/telemetry/storage system;
5. power requirements;
6. sampling intervals (continuous monitoring vs monitoring only after extreme events or at periodic intervals);
7. processor/memory requirements; and
8. excitation source (active sensing).

Fundamentally, there are five factors that control the selection of hardware to address these sensor system design parameters:

1. the length scales on which damage is to be detected;
2. the timescale on which damage evolves;
3. how varying and/or adverse operational and environmental conditions will affect the sensing system;
4. power availability; and
5. cost.

The development of a damage detection system for the composite wings of an unmanned aerial vehicle (UAV) can be used as an example. In one case, damage is assumed to be initiated by foreign object impact on the wing surface. Such damage is often very local in nature and may manifest itself on a length scale of the order of $10\,\mathrm{cm}^2$ or less. Accurate characterization of the impact phenomena occurs on a microsecond to millisecond timescale, which requires the data-acquisition system to have relatively high sampling rates (greater than 100 kHz). This damage may then grow to a fault after being subject to numerous fatigues cycles during many hours of subsequent flight. The timescales associated with the damage initiation and evolution influence sensing system properties 2–5 above.

This example clearly demonstrates that the length and timescales associated with damage initiation and evolution drive many of the SHM sensing system design parameters. *A priori* quantification of these length and timescales will allow the sensing system to be designed in an efficient manner.

Axiom 6. *There is a trade-off between the sensitivity to damage of an algorithm and its noise rejection capability.*

This axiom is self-evident in many ways. To detect small damage, a set of features combined with a PR algorithm must be very sensitive to the structure of the data. Under those circumstances, it is reasonable to assume that the presence of noise in the data would be likely to bias the results of the algorithm.

A pragmatic consequence of this axiom is that one should attempt to reduce the level of noise in the measured data or the subsequently extracted features as much as possible; this can be accomplished by many means including wavelet denoising, analog or digital filtering, or even simple averaging.

Axiom 7. *The size of damage that can be detected from changes in system dynamics is inversely proportional to the frequency range of excitation.*

In the field of ultrasonic nondestructive testing, the diffraction limit is often associated with the minimum size of flaw that can be detected as a function of ultrasonic wavelength. This limit may suggest that flaws of a size comparable with half a wavelength are detectable. The diffraction limit is actually a limit to the resolution of nearby scatterers, i.e., if two scatterers are separated by more than a half-wavelength of the incident wave, they will be separable. In fact, a flaw will scatter an incident wave for wavelengths below this limit and this amount of scattering decreases with increasing wavelength. This result means that, if instrumentation is available to detect arbitrarily small evidence of scattering, i.e., arbitrary small reflection coefficients, then arbitrarily small flaws can be detected. However, as described above, scattering is always substantial when the size of the flaw is comparable with the wavelength, and so it is advantageous to use small wavelengths to detect small flaws.

The wavelength λ is related to the wave phase velocity v and frequency f by

$$\lambda = \frac{v}{f} \tag{4}$$

From this simple relationship, it is clear that for constant velocity the wavelength will decrease as the frequency increases, which in turn implies that the damage sensitivity will increase.

Evidence for damage detection well below the diffraction limit for the size of the flaw can be found in numerous places in the literature, e.g., in [39] and [40].

The relationship between damage sensitivity and wavelength can be extended to more general types of vibration-based damage detection methods. In these applications, the wavelength of the elastic wave traveling through the material is replaced by the "wavelength" of the standing wave pattern set up in the structure, which is interpreted as a mode of vibration. The technical literature is replete with anecdotal evidence that such lower-frequency modes are not good indicators of local damage [15, 27]. The lower-frequency global modes of the structure that have long characteristic wavelengths tend to be insensitive to local damage. For the case of civil engineering infrastructure such as suspension bridges, these mode-shape wavelengths can be on the order of hundreds of meters [41], and flaws such as fatigue cracks that must be detected to assure safe operation of the structure are of the order of centimeters in length. Even if one allows for the fact that the cracks may have influence over a larger distance than the crack dimensions, this distance is still small compared to the mode-shape wavelengths.

The observations regarding the relationship between the characteristic wavelength and the flaw size have led research to explore other high-frequency active-sensing approaches to SHM. These methods are based on Lamb wave propagation [39], impedance measurements [42], high-frequency response functions, and time reversal acoustics [43]. In these applications, the excitation frequency can be as high as several hundred kilohertz and the corresponding wavelengths are on the order of millimeters. However, as the frequency increases and wavelength decreases, scattering effects (e.g., reflection of elastic waves off grain boundaries and other material interfaces) will eventually increase the noise in the measurements and place limits on the sensitivity of the damage detection process. Optimal frequency ranges for damage detection can be determined on the basis of the wavelength of the standing wave pattern and the condition that the damage is located in areas of high curvature associated with the deformation of the wave pattern.

Finally, there will generally be an increased energy requirement necessary to maintain a comparable amplitude excitation at the higher frequencies; this is because higher frequencies have higher attenuation. As a result, higher frequency excitation procedures are typically associated with more local damage detection procedures.

7 CHALLENGES AND BARRIERS TO SHM IMPLEMENTATION

The discussions of the previous sections have thrown light on various aspects of the design of SHM systems. However, one should recognize that there are still some significant barriers to the general implementation of damage identification systems. Some of these barriers are discussed in this section.

7.1 Structural monitoring versus structural health monitoring

Many sensor systems currently being deployed on real-world structures are actually structural monitoring systems as opposed to SHM systems. They are simply sparse arrays of sensors deployed with no definition of the damage to be detected and no definition of the methods for feature extraction and statistical classification that will be used to identify damage. Many bridge monitoring systems currently being deployed are examples of such systems that are often referred to as *SHM systems*. It is the authors' opinion that a true SHM system will have a quantified and predefined definition of the damage to be detected, as well as a predefined and validated approach to feature extraction and statistical classification of the features. Such a system would have performed false-positive studies where the chance of misclassification of damage has been quantified.

7.2 Local versus global damage detection

The need for quantitative global damage detection methods that can be applied to complex structures has led to the development of, and continued research

into, methods that examine changes in the vibration characteristics of the structure. The basic premise of vibration-based damage detection is that damage will alter the stiffness, mass, or energy dissipation properties of a system, which, in turn, will alter the measured global dynamic response properties of the system. Although the basis for vibration-based damage detection appears intuitive, its actual application poses many significant technical challenges [16]. The most fundamental challenge is the fact that damage is typically a local phenomenon and may not significantly influence the lower-frequency global response of a structure that is normally measured during vibration tests, particularly those where the response to ambient excitation is measured.

More recently, researchers have been studying hybrid multiscale sensing approaches to SHM [42]. Such approaches rely on active-sensing systems for local damage detection and then the same sensors are used in a passive mode to measure the influence of damage on the global system response. Here, the term *active* refers to systems where actuators are incorporated with the sensing system to provide a known input that is designed to enhance the damage detection process.

Fundamentally, there will always be a trade-off between the cost associated with deploying a local sensing system over a large area of the structure and the lack of fidelity associated with more global sensing systems. For most applications, some hybrid system will most likely be employed, which is based on *a priori* knowledge of specific areas on the structure that are most likely to experience damage, but such systems are still in the development and validation state.

7.3 Defining damage *a priori*

The success of any damage detection technique will be directly related to the ability to define the damage that is to be detected in as much detail as possible and in as quantifiable terms as possible. Here the definition of damage can include issues such as the type of damage to be detected (e.g., fatigue crack), the threshold level of damage that must be detected (e.g., 5-mm-long, through-thickness crack), the critical level of damage that produces failure or that will no longer allow for a planned safe shut down of the system (2-cm-long crack), locations where the particular type of damage accumulates in the structure (e.g., welded beam-to-column connections), and the tolerable or anticipated rate of damage growth.

In the examples cited above, the damage is quantified in terms directly related to the type of failure. There will be cases where the damage will be quantified in an indirect manner. As an example, damage may be defined as *delamination of a composite aircraft wing component*. The critical level may be defined as *an amount of damage that produces a certain change in the flutter characteristics of the aircraft*.

Because of the costs, system level testing to failure for the purpose of defining the damage to be detected is rarely done. Instead, these definitions are often based on prior observed behavior of damaged systems. Component level testing is a more cost-effective experimental approach that can also be used to develop such damage definitions. At other times, numerical simulations of the damaged system will be used to establish these definitions. Finally, there will be many cases where these definitions will be developed in an *ad hoc* manner based on the experience and intuition of people familiar with the equipment.

Large complex structural systems are usually made up of numerous components. Typically, there will be multiple damage mechanism that will be of concern. However, an SHM system that is optimal for detecting one type of damage may not be useful for detecting an alternate damage condition. Also, there will be cost limitations on the development and deployment of the SHM system. Therefore, it is imperative to define the various possible damage scenarios and to establish a priority for detecting damage associated with these scenarios. Such prioritization will be based on some study of the relative probability of occurrence associated with the various damage scenarios versus the consequences of the respective damage scenarios. With such a prioritization established, the SHM system can be designed to address the most critical damage concerns, which will make optimal use of a given SHM budget.

The definition of damage is the first step in identifying the required data-acquisition system capabilities, the candidate features to be extracted for damage detection, and the statistical models that are to be employed for the feature discrimination. To date,

many SHM studies suffer from taking the approach where the damage detection process is first developed and then the procedure is studied in an effort to define the damage that it can detect.

7.4 Defining the requisite sensing system properties

A significant challenge for SHM is to develop the capability to define the required sensing system properties before field deployment and, if possible, to demonstrate that the sensor system itself will not be damaged when deployed in the field. The first line of attack is clearly the establishment of an appropriate sensor network that can adequately observe the system dynamics for suitable signal processing and feature extraction. Sensor networks, generally speaking, contain three main components: the sensing unit itself, communications, and computation (hardware and, as appropriate, software control and processing algorithms).

As discussed earlier in relation to Axiom 5, the goal of any SHM sensor network system is to make the sensor reading as directly correlated with, and as sensitive to, damage as possible. At the same time, one also strives to make the sensors as independent as possible from all other sources of environmental and operational variability, and, in fact, independent from each other (in an information sense) to provide maximal data from minimal sensor array outlay. To best meet these requirements, various design parameters must be defined as in the first numbered list in discussion of Axiom 5 in Section 6. The five fundamental issues that control the selection of hardware to address these sensor system design parameters are also discussed in that section.

In addition, the feature extraction, data normalization, and statistical modeling portions of the process can greatly influence the definition of the sensing system properties. Before such decisions can be made, two important questions must be addressed. First, one must answer the question, "What is the damage to be detected?" (Section 7.3) Second, an answer must be provided to the question, "What are the environmental and operational sources of variability that must be taken into account?" To answer this question, one does not only have to have some ideas about the physical sources *per se*, but one also should have thought about how to accomplish data normalization. Typically, data normalization is accomplished through some combination of sensing system hardware and data interrogation software. However, these hardware and software approaches will not be optimal if they are not done in a coupled manner.

7.5 Accounting for operation and environmental variability

When deployed on a structure outside of a controlled laboratory setting, the damage detection process will have to deal with structures that experience changing operational and environmental conditions. These changing operational and environmental conditions will produce changes in the measured system response and it is imperative that these changes are not interpreted as indications of damage. Varying temperature and moisture levels are two common environmental conditions that must be accounted for during the damage detection process. Changing mass is a common operational variable that results from fuel usage and varying payloads. Running equipment or vehicles at varying speeds and through varying maneuvers are other operational parameters that can significantly influence kinematic quantities that are being measured as part of the damage detection process.

Often sensors will have to be added to monitor the changing operational and environmental conditions in an effort to develop a procedure that normalizes the data to remove trends caused by these effects. Also, the sensors themselves will have environmental and operational limitations under which they will function properly.

7.6 Need for long-term proof of concept studies

There are very few long-term SHM studies ongoing on real-world structures. These studies are difficult to perform because of costs and the rapid evolution of sensor technology. However, such studies are needed before structure owners and regulators will accept SHM as an acceptable means of condition-based maintenance. Research funding agencies will need

to feel that such studies are of merit. Without such studies, it will be difficult to develop the appropriate data normalization procedures that are needed to address the environmental and operational variability issues discussed in Section 7.5. To date, most SHM studies are performed in well-controlled laboratory settings.

7.7 Lack of data from damaged systems

As discussed in earlier sections, another fundamental challenge for SHM is that in many situations feature selection and damage detection must be performed in an unsupervised learning mode. That is, data from damaged systems is not available. Unsupervised approaches typically look for outliers from an established baseline condition and such changes can result from many things other than damage. Few system owners will allow engineers to damage their structure in an effort to validate a damage detection approach. Even if such studies were allowed, in almost all cases damage is introduced in an artificial manner and it is questionable if such "damage" is truly indicative of the actual damage that will be encountered in the field.

7.8 Time scales associated with damage evolution

Damage can accumulate over widely varying timescales, which poses significant challenges for the validation of SHM sensing systems in the field. In the case of civil engineering, infrastructure damage evolution can occur on timescales that are comparable to the career of the maintenance engineer. This challenge is supplemented by many practical issues associated with making accurate and repeatable measurements over long periods of time on complex structures often operating in adverse environments. There are few studies that have examined the long-term issues associated with the field deployment of such sensing systems.

7.9 Nontechnical issues

In addition to the challenges described above, there are other nontechnical issues that must be addressed before SHM technology can make the transition from a research topic to actual practice. These issues include convincing structural system owners that the SHM technology provides an economic benefit over their current maintenance approaches using quantified benefit–cost analyses and convincing regulatory agencies that this technology provides a significant life-safety benefit. In the case of intelligent damage detection systems, it will be necessary to overcome the current barriers to certification for safety-critical systems.

In general, the multidisciplinary nature of SHM system development makes it a more costly technology development to undertake. It requires people with diverse technical expertise and a significant amount of technology integration and validation. Such costs have to be quantified and addressed when performing initial benefit–cost studies for the SHM system development. These issues are coupled with short research time horizons in many industries, which are on the order of a 12–18-month time to market.

8 SUMMARY

In summary, assuming that this article has managed to convince the reader of the desirability of an intelligent monitoring system, the requirements for such a system are as follows:

- It should be designed from a holistic viewpoint, ideally at the design stage of the structure of interest. At the very least, an operational evaluation stage is necessary to establish the baseline requirements for and restrictions on the monitoring.
- The sensor distribution should be optimal. This means the smallest number giving the required resolution should be active at any given time. Note that the optimization may well be constrained. It should, for example, take account of inaccessible regions, and any upper limits on weight or power consumption. If failure safety is desired—and it should always be recommended—the distribution should include inactive units that can be switched on in the event of sensor failure. This implies that the sensors themselves should be monitored. This can be accomplished by using self-validating sensors or sensor

networks. Active and inactive units should be installed at the optimal locations.

- Data processing should be dictated by an optimal fusion model designed to give the most robust and confident decision process given the available sensor resolutions and confidences. Where possible, algorithms that are provably optimal should be employed.
- As far as the "axioms" discussed in this chapter are to be considered fundamental truths, they should be considered as valuable guidance in the design process, they give information on what is, and is not, feasible. The barriers to implementation discussed here are also to be considered in deciding what is feasible or, indeed, practical.
- If control is part of the system requirements, the actuator distribution should be optimal. The smallest number of actuators giving the required control actions should be employed and these should be located optimally.
- Feedback should be part of the process. If the decision process identifies an information gap and the system is capable of sensor/actuator redeployment, this should be implemented by whatever means available. This may be as simple as switching in inactive units, or could, in the future, entail physical movement of units.
- There should be a strategy for planning and implementing repair or replacement of the system or of subsystems. This may or may not go as far as specifying a backup system, depending on cost constraints.
- Monitoring should be as continuous as possible, with all levels of the D2D process updated as frequently as cost/technology constraints allow.

Some of these requirements are not currently within the reach of available technology. However, the purpose of this article has been, amongst other things, to specify a paradigm for intelligent fault detection. If the community considers that the pursuit of intelligent SHM systems is desirable, then many, if not all, of these requirements will be realized. However, the problem of global SHM is so significantly complex and diverse that it will not be solved in the immediate future. Like so many other technology fields, because of the issues discussed above and difficulties in surmounting them, advancements in SHM will most likely come in small increments over long periods of time.

Significant future developments of this technology will need to come by way of multidisciplinary research efforts encompassing fields such as structural dynamics, signal processing, motion and environmental sensing hardware, computational hardware, data telemetry, smart materials, and SPR, as well as other fields yet to be defined.

ACKNOWLEDGMENTS

The authors would like to acknowledge numerous useful discussions with their valued colleagues: Dr Graeme Manson and Professor Wieslaw Staszewski of the University of Sheffield, UK, Dr Gyuhae Park of the Engineering Institute, Los Alamos National Laboratory, USA, and Professor Hoon Sohn of KAIST, South Korea.

REFERENCES

[1] Reifsnider KL, Case SW. *Damage Tolerance and Durability of Material Systems*. Wiley Inter-Science, 2002.

[2] Worden K, Farrar CR, Manson G, Park G. The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society—Series A* 2007 **463**:1639–1664.

[3] Rytter A. *Vibration Based Inspection of Civil Engineering Structures*, Ph.D. Thesis. Department of Building Technology and Structural Engineering, University of Aalborg, Denmark, 1993.

[4] Bishop CM. Novelty detection and neural network validation. *IEEE Proceedings—Vision and Image Signal Processing* 1994 **141**:217–222.

[5] Tarassenko L, Hayton P, Cerneaz Z, Brady M. Novelty detection for the identification of masses in mammograms. *Proceedings of 4th International Conference on Neural Networks*. IEE Publication 409: Cambridge, 1995; pp. 442–447.

[6] Worden K. Structural fault detection using a novelty measure. *Journal of Sound and Vibration* 1997 **201**:85–101.

[7] Schalkoff R. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons: Singapore, 1992.

[8] Farrar CR, Doebling SW. *The State of the Art in Vibration-Based Structural Damage Identification*, CD-ROM, Los Alamos Dynamics, 2000.

[9] Leitch RD. *Reliability Analysis for Engineers—An Introduction*. Oxford Science Publications, 1995.

[10] Henry MP, Clarke DW. The self-validating sensor: rationale, definitions and examples. *Control Engineering Practice* 1993 **1**:585–610.

[11] Kramers MA 1991 Nonlinear principal component analysis using autoassociative neural networks. *American Institute of Chemical Engineers Journal* **37**:233–243.

[12] Kramers MA. Autoassociative neural networks. *Computers in Chemical Engineering* 1992 **16**: 313–328.

[13] Side S, Staszewski WJ, Wardle R, Worden K. Fail-safe sensor distributions for damage detection. *Proceedings of 2nd International Workshop on Damage Assessment using Advanced Signal Processing Procedures—DAMAS '97*. Sheffield, 2000; pp. 41–52.

[14] Staszewski WJ, Worden K, Wardle R, Tomlinson G. Fail-safe sensor distributions for impact detection in composite materials. *Smart Materials and Structures* 2000 **9**:298–303.

[15] Doebling SW, Farrar CR, Prime MB, Shevitz D. *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in their Vibration Characteristics*, Los Alamos National Laboratory Report LA-13070, 1996.

[16] Friswell MI, Penny JET. Is damage location using vibration measurements practical? *Proceedings of 2nd International Conference on Structural Damage Assessment Using Advanced Signal Processing Procedures (DAMAS 97)*. Sheffield, 1997; pp. 351–362.

[17] Lowe D. Feature extraction, data visualisation, classification and fusion for damage assessment. *Oral Presentation at EPSRC SIDAnet Meeting*. Derby, 2000.

[18] White FE Jr. Joint directors of laboratories data fusion subpanel report: SIGINT session. *Technical Proceedings of the Joint Service Data Fusion Symposium* DFS-90, 1990; pp. 469–484.

[19] Bedworth M. *Probability Moderation for Multilevel Information Processing*, DRA Technical Report, DRA/CIS(SE1)/651/8/M94.AS03BP032/1, 1994.

[20] Worden K, Manson G, Fieller NRJ. Damage detection using outlier analysis. *Journal of Sound and Vibration* 2000 **229**:647–667.

[21] Pomerleau DA. Input reconstruction reliability information. In *Advances in Neural Information Processing Systems 5*, Hanson SJ, Cowan JD, Giles CL (eds). Morgan Kaufman Publishers, 1993.

[22] Taylor O, MacIntyre J, Isbell C, Kirkham C & Long A. Adaptive fusion devices for condition monitoring: local fusion systems of the NEURAL-MAINE project. *Proceedings of 1st International Conference on Damage Assessment of Structures—DAMAS '99*. Dublin, Eire, 1999; pp. 205–216.

[23] Roberts S, Tarassenko L. A probabilistic resource allocating network for novelty detection. *Neural Computation* 1995 **6**:270–284.

[24] Sohn H, Farrar CR. Statistical process control and projection techniques for damage detection. *Proceedings of European COST F3 Conference on System Identification and Structural Health Monitoring*. Madrid, 2000; pp. 105–114.

[25] Worden K, Manson G. Damage identification using multivariate statistics: kernel discriminant analysis. *Inverse Problems in Engineering* 2000 **8**:25–46.

[26] Lowe D, Zapart C. Point-wise confidence interval estimation by neural networks: a comparative study based on automotive engine calibration. *Neural Computing and Applications* 1999 **8**:77–85.

[27] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature from 1996–2001*, Los Alamos National Laboratory Report LA-13976-MS, 2004.

[28] Broberg KB. *Cracks and Fracture*. Cambridge University Press, 1999.

[29] Dong Y, Sun Z, Jia H. A cosine similarity-based negative selection algorithm for time series novelty detection. *Mechanical Systems and Signal Processing* 2005 **20**:1461–1472.

[30] Friswell MI, Mottershead JE. *Finite Element Model Updating in Structural Dynamics*. Kluwer Academic Publishers, 1995.

[31] Goerl E, Link M. Damage identification using changes of eigenfrequencies and modeshapes. *Mechanical Systems and Signal Processing* 2003 **17**:103–110.

[32] Fritzen C-P, Bohle K. Global damage identification of the 'Steelquake' structure using modal data. *Mechanical Systems and Signal Processing* 2003 **17**:111–117.

[33] Stubbs N, Kim J-T. Damage localization without baseline modal parameters. *AIAA Journal* 1995 **34**:1–6.

[34] Rao, JS. *Vibratory Condition Monitoring of Machines*, Alpha Science International Ltd, 21 August 2000.

[35] Pierce SG, *et al.* Damage assessment in smart composite structures: the DAMASCOS programme. *Proceedings of SPIE Conference on Smart Materials and Structures*. Newport Beach, 2001; Vol. 4327, pp. 223–233.

[36] Silverman BW. *Density Estimation for Statistics and Data Analysis*, *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1986; Vol. 26.

[37] Worden K, Sohn H, Farrar CR. Novelty detection in a changing environment: regression and interpolation approaches. *Journal of Sound and Vibration* 2002 **258**:741–761.

[38] Manson G. Identifying damage sensitive environment insensitive features for damage detection. *Proceedings of International Conference on Identification in Engineering Systems*. Swansea, 2002; pp. 187–197.

[39] Alleyne DN, Cawley P. The interaction of Lamb waves with defects. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 1992 **39**:381–397.

[40] Valle C, Niethammer M, Qu J, Jacobs LJ. Crack characterisation using guided circumferential waves. *Journal of the Acoustical Society of America* 2001 **110**:1282–1290.

[41] Farrar CR, Cornwell PJ, Doebling SW, Prime MB. *Structural Health Monitoring Studies of the Alamosa Canyon and I-40 Bridges*, Los Alamos National Laboratory Report LA-13635-MS, 2000.

[42] Park G, Sohn H, Farrar CR, Inman DJ. Overview of piezoelectric impedance-based health monitoring and path forward. *Shock and Vibration Digest* 2003 **35**(6):451–463.

[43] Sohn H, Park G, Law KH, Farrar CR. Instantaneous online monitoring of unmanned aerial vehicles without baseline signals. *Proceedings of 23rd International Modal Analysis Conference*. Orlando, FL, 2005.

FURTHER READING

Bedworth M, O'Brien J. The omnibus model: a new model of data fusion. *DERA Malvern Preprint*, 1999.

Manson G, Worden K, Allman DJ. Experimental validation of a damage severity method. *Proceedings of 1st European Workshop on Structural Health Monitoring*. Paris, 2002; pp. 845–852.

Manson G, Worden K, Allman DJ. Experimental validation of structural health monitoring methodology II: novelty detection on an aircraft wing. *Journal of Sound and Vibration* 2003a **259**:345–363.

Manson G, Worden K, Allman DJ. Experimental validation of structural health monitoring methodology III: damage location on an aircraft wing. *Journal of Sound and Vibration* 2003b **259**:365–385.

Stubbs N, Kim J-T, Farrar CR. Field verification of a nondestructive damage localisation and severity estimation algorithm. *Proceedings of 13th International Modal Analysis Conference*. Nashville, TN, 1995; pp. 210–218.

Worden K. Cost action F3 on structural dynamics: benchmarks for working group 2—structural health monitoring. *Mechanical Systems and Signal Processing* 2003 **17**:73–75.

Worden K, Manson G, Allman DJ. Experimental validation of structural health monitoring methodology I: novelty detection on a laboratory structure. *Journal of Sound and Vibration* 2003 **259**:323–343.

Chapter 83
Design Principles for Aerospace Structures

Jens Telgkamp
Structure Design Principles, Airbus, Hamburg, Germany

1 INTRODUCTION

SHM is a promising technology for application to metallic and composite structures. Current and future structures will benefit from the use of SHM in several ways. While the SHM system developers bring several technologies closer to maturity, the business case for the application is becoming clearer; a general development is that for newly developed aircraft it is becoming more and more important to integrate the idea of SHM in the early design phase: SHM will provide its full benefit only when it is incorporated at an early development stage, while only a part of this benefit will be available if the idea is introduced after the structural design is already fixed.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 AIRCRAFT STRUCTURE DESIGN—A MULTIDISCIPLINARY CHALLENGE

Aircraft structures have, since the beginning of powered flight more than 100 years ago, undergone tremendous development. To give an example of the enormous performance increase over that period, Figure 1 gives an overview of the relative aircraft weight (related to range and payload) of aircraft structures from 1920 to today, where each dot marks an

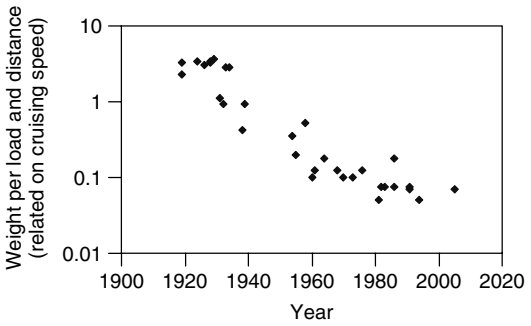


Figure 1. Relative weight of typical civil aircraft structure, from 1920 to today.

aircraft type typical of the relevant year's state of the art. Note that the scale is logarithmic, i.e., the structural performance has increased by a factor of roughly 100 over that time.

Up to current times, rapid development has continued wherein the market sets the development targets. Some general challenges are the requirements regarding

- range;
- size/capacity;
- cruising speed;
- cost (recurring/nonrecurring cost);
- need and responsibility to design environment-friendly aircraft.

However, these challenges partially contradict each other, so the ultimate challenge is to use the technologies and materials available to meet the best compromise to obtain the aircraft demanded by the market.

During the whole aircraft design process, the aircraft designers are influenced by a variety of factors. Some major factors are as follows:

- available materials and technologies;
- airworthiness regulations;
- environmental considerations;
- general aircraft requirements (mission profile, maintenance, operating cost, etc.);
- specific requirements for structural details;
- manufacturing capacities and capabilities;
- NDI (nondestructive inspection) and NDT (non destructive testing) capabilities, depending on the available mature technologies for NDI/NDT;
- design costs.

In aircraft design, there is an ongoing competition between metallic and nonmetallic materials, mainly composites. The composite materials have, during

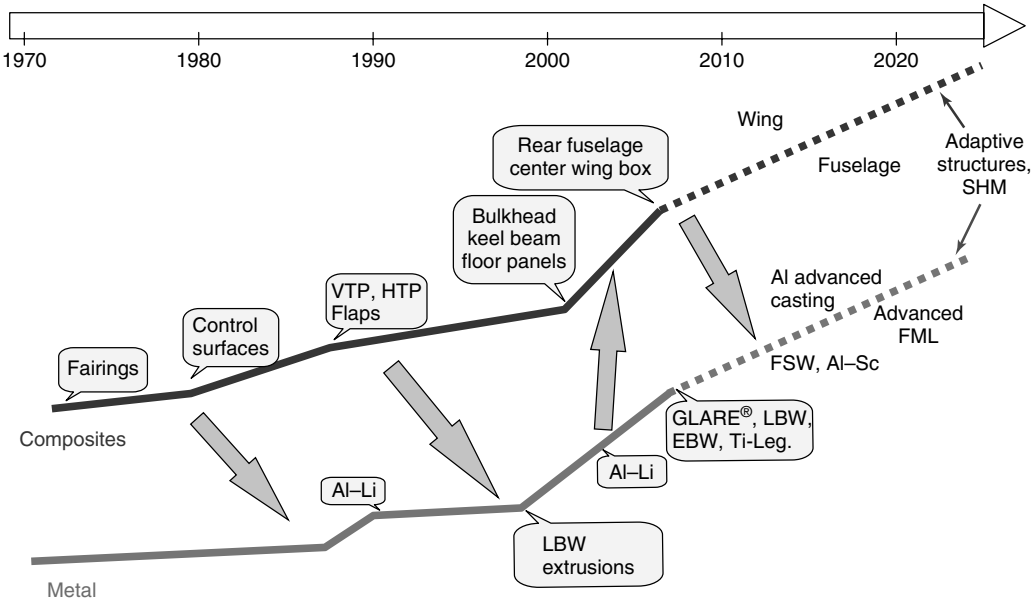


Figure 2. Application and technology introduction at Airbus for typical metallic and composite structures over time.

recent years, undergone a rapid development challenging the community developing metallic materials and new manufacturing processes for metals. Figure 2 displays some materials and technologies introduced by Airbus into their products from 1970 to today together with an anticipated technology extrapolation into the future. It is clear that the metallic and composite communities stimulate each other and the airframe manufacturer selects the best technology for the relevant aircraft under development.

In addition to the monolithic metallic materials and fiber–resin composite materials, fiber–metal laminates have also been used, for example, the GLARE® material on the Airbus A380, which combines some of the advantages of metallic designs (e.g., repairability, static behavior) with those of composite materials (e.g., tailoring of fiber layer orientations). The GLARE® fiber–metal laminate is a sandwich consisting of some thin aluminum sheets with fiber/prepreg layers in between. Figure 3 gives an overview of some of the main strengths and drawbacks of metallic materials, composites (here, carbon fiber reinforced plastics), and fiber–metal laminates.

Although aircraft designers continue to seek the technically best solution for each application, the general trend is in the direction of extended use of composites. Figure 4 displays the use of composites (relative to the overall structural weight of the aircraft) over the years in Airbus aircraft structures: from less than 10% in the Airbus A300 (first flight 1972) to over 50% in the A350 XWB, one of the designs currently under development at Airbus.

The choice of material for each structural application depends, of course, on the local loading and other design requirements at the location under consideration. To explain typical loading of the structure, an example is used: Figure 5 shows a stylized aircraft fuselage structure under in-flight and pressurization loading. The cylindrical barrel is a very strongly simplified model of an aircraft fuselage design (a typical design of real cylindrical fuselage structure is depicted in Figure 6). The pressurization causes tensile stresses in the fuselage skin in the circumferential as well as in the longitudinal directions. Additional mechanical load to the structure comes from the fuselage bending: The lifting force is introduced into the fuselage center at the location of the wing attachments—in the center of the fuselage. Gravity acts on the fuselage structure, causing down-bending in the forward as well as in the rear fuselage. On top of that, the resulting aerodynamic force acting on the horizontal tailplane (downward force in cruising condition) causes an additional down-bending of the rear fuselage.

Although this is only a simplified model of fuselage loading (and it takes into account only in-flight loading in cruising condition), it is clearly evident that different local loads put forth different challenges on the choice of materials and manufacturing technologies at the different locations: The upper shell is mainly characterized by biaxial tension load

	Strengths	Drawbacks
Metals (Al alloys)	• Standardization • Reparability • Static behavior • Improvement potential	• High density • Fatigue behavior • Corrosion behavior • High costs of new alloys
Composites (CFRP)	• Fatigue behavior • Low density • No corrosion • Best suited for smart structures	• Impact behavior • No "plasticity" • Reparability • Recycling
Fiber–metal laminates	• Improved fatigue • better tailoring • Higher fire resistance • Less corrosion (compared to Al alloys)	• Lower stiffness • Higher density (compared to CFRP) • Less industrialized process (compared to CFRP)

Figure 3. Main strengths and drawbacks of metallic, composite, and fiber–metal laminate structure.

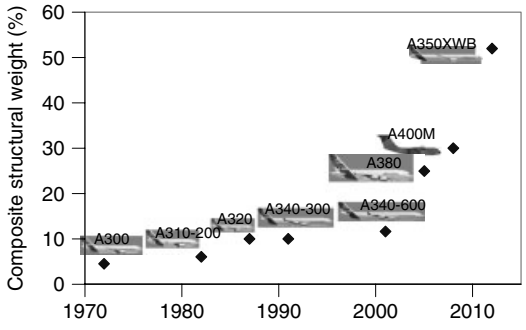


Figure 4. Composite structural weight fraction of Airbus aircraft programs over time.

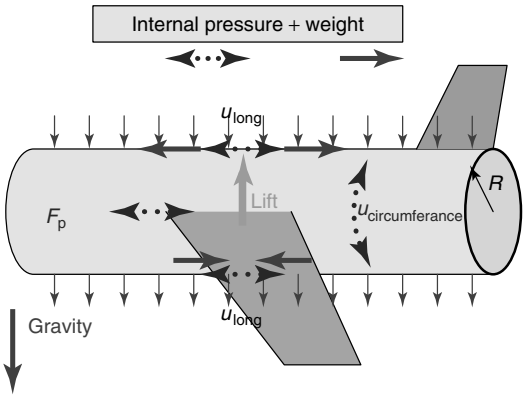


Figure 5. Simplified model of an aircraft fuselage structure in cruising condition with main loads coming from gravity and internal pressure.

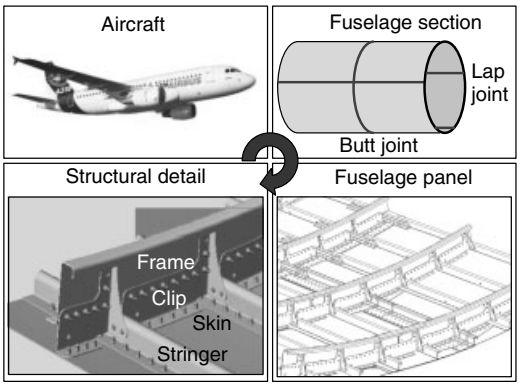


Figure 6. Aircraft fuselage structure in four different levels of detail.

(from pressurization and bending), while the lower shell areas are characterized by tensile loading in the circumferential direction (from pressurization) and compressive loading in the longitudinal direction (from fuselage bending). Therefore, the upper shell areas tend to be dimensioned by fatigue/crack growth, while the lower shell areas may be dimensioned, instead, by static stability.

It is once more emphasized that this example of fuselage bending and compression is an extremely simplified model, since it takes into account only one load case and a very simple structure modeling. However, it is obvious that in this simple model a number of areas with different local requirements of the materials and choice of tech can be identified. The final design and dimensioning has to take into account a variety of additional factors such as a long list of load cases, needs of system installation and passenger doors/windows, corrosion protection, repairability, inspectability, and many others.

Figure 7 displays a simplified Airbus A380 fuselage with colors/grayscales indicating dimensioning criteria. In the lower part of the same figure, the material choice for the fuselage is indicated, and shows some correlation to the dimensioning criteria: While the GLARE® and 2524 materials show good performance mainly in fatigue and crack-growth areas, the 6013 material has a high stiffness and therefore, exhibits good stability against buckling in the stiffened panels; the 7475 alloy shows strong static strength.

In summary, it is important to emphasize that aircraft structure design is a multidisciplinary challenge where aircraft top-level requirements, structure loading, and materials/technology selection influence each other [1, 2]. The final set of dimensioning criteria also dependent upon the material choice and only the right material/technology choice will finally deliver good structural performance. This article gives examples of how this performance can be further improved by integrating SHM technology into the structural design.

3 OVERVIEW OF POTENTIAL SHM APPLICATION TO AIRCRAFT STRUCTURE

This section gives an overview of possible ways to integrate SHM into the structure so that the operator of the aircraft will benefit from the technology. It

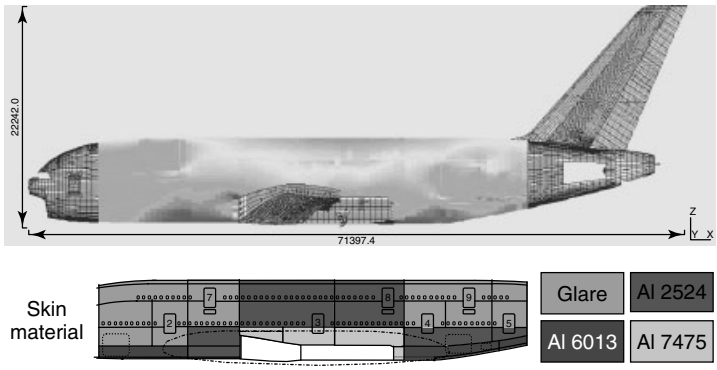


Figure 7. Schematic plot of dimensioning criteria for fuselage skin and material choice for the Airbus A380-800 aircraft.

is evident that for all these ideas the business case and the cost–performance trade-off have to be looked at in detail before taking any concrete decision for an individual aircraft application. However, some examples for potential applications are discussed.

3.1 Concepts for applying SHM to aircraft structure

Figure 8 gives an overview of some applications of SHM to aircraft structures.

According to this, the main benefits of SHM to aircraft structures can be subdivided as follows:

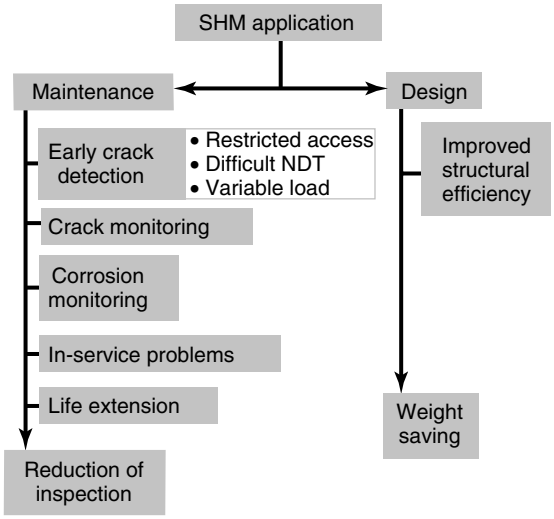


Figure 8. Different applications of SHM in civil aircraft structure.

- Applications related to maintenance
  - **Early crack detection**
    SHM can be used as an alternative to conventional NDI/NDT. This is especially beneficial either in areas where conventional inspection is very challenging and therefore, financially penalizing, or in areas of the structure where physical access to the area to be inspected is difficult.
  - **Damage monitoring**
    This means monitoring of known damage (cracks or delaminations) on individual aircraft. In this case, care must be taken in regard to the question of what it means, in practice, to operate a structure with a known flaw. In today's world (without SHM being used in series aircraft), it is the current policy not to operate aircraft with known flaws.
  - **Corrosion monitoring**
    Since all aircraft structures contain a certain percentage of metallic components, the possibility of corrosion development will always have to be taken into account in the aircraft design. In particular, the topic of galvanic corrosion needs to be looked at, since this type of corrosion is likely to develop in interfaces between metallic components and carbon fiber components. Therefore, Airbus is looking at technologies to monitor the metallic components in aircraft structure for corrosion. In the first generation, the application seems especially beneficial for "hot-spot"

areas, where possible corrosion issues have been identified in service.

– **In-service problems (mainly related to fatigue, damage tolerance)**
Sometimes, when a fleet of aircraft is already in operation, problems are detected either on in-service aircraft or on the full-scale fatigue test of the aircraft type (which, for civil transport aircraft, is normally not yet finished when the first aircraft enter into service). An example of this is given in Section 4.2.

– **Life extension**
Sometimes, in the case of aging aircraft fleet, the airframe manufacturer decides to certify and allow a longer service life than originally planned for the aircraft type. In some cases, SHM may help in future with this challenge. An example of this situation is discussed in Section 4.3.

- Applications related to design improvement
Generally, it is expected that including the use of SHM on a newly designed aircraft structure can be beneficial because it will help avoid design conservatism without compromising safety. Examples of this are given Section 5.

3.2 Some SHM technologies considered for potential application to aircraft structure

This section gives a compact overview of some of the technologies that are considered by Airbus. Airbus do not consider themselves as SHM technology developers, but they develop concepts for integrating SHM into their structure and therefore adapt commercially available technologies according to their needs.

Some technologies under investigation at Airbus are

- comparative vacuum method (CVM™);
- optical fiber sensors (mainly fiber Bragg gratings);
- Lamb waves (guided waves);
- foil piezoelectric (PVDF sensor);
- eddy-current foil sensors (ETFS);
- foil crack wire;
- acoustic emission.

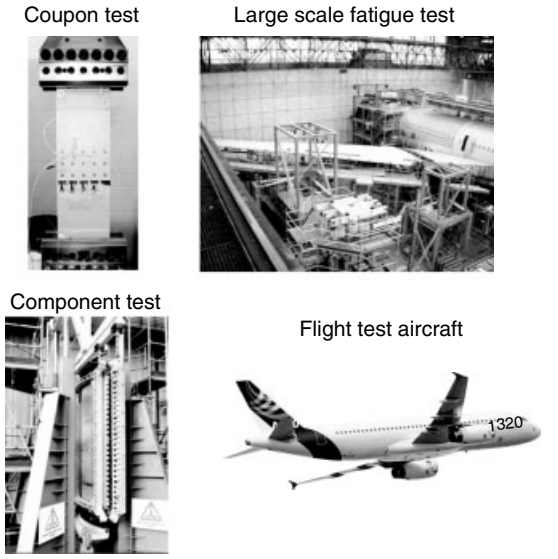


Figure 9. Four different levels of testing aircraft structure.

This list is not exhaustive, but only intended to give some examples. Details can be found in [3–5].

As already mentioned above, Airbus are not developing SHM technologies themselves, but they have to ensure there is enough practical experience with the individual SHM systems under consideration for use on aircraft. To do so, the technologies are being tested at various levels (Figure 9), such as

- coupon tests (small lab specimens);
- component tests (for example, shear/compression shells or curved panel tests);
- full-scale tests (full-scale static and fatigue tests as they are a part of a new aircraft structure certification);
- flight test aircraft.

4 IMPROVING AIRCRAFT STRUCTURE MAINTENANCE THROUGH USE OF SHM

4.1 Early crack detection in areas difficult to access or difficult to inspect

To ensure integrity and the safe operation of the aircraft structure, an inspection program is defined

for each individual aircraft type. Generally, these inspections are penalizing because of the following reasons:

- The inspections have to be performed by the operator (airline) of the structure, using their own resources and material.
- The inspection plan forces the operator to perform logistics to have each individual aircraft in the service routine at the appropriate time.
- During the inspection of an individual aircraft, this aircraft cannot be used for scheduled flights and does not give any benefit for the operators.

Analyses of individual maintenance examples have shown that the latter point is often the most penalizing one. The downtime due to inspection puts an economic burden on the operator, which is in many cases, is more penalizing to the operator than the cost of performing the maintenance action itself.

One example of an area that is difficult to access is the "frame feet" of a metallic fuselage structure. These are the areas where the frames are connected to the structure below (e.g., a center wing box fitting), as illustrated in Figure 10. If a scheduled inspection of this area turns out to be necessary, the inspection will be penalizing to the operator—even if the inspection itself is relatively simple. The main reason is that

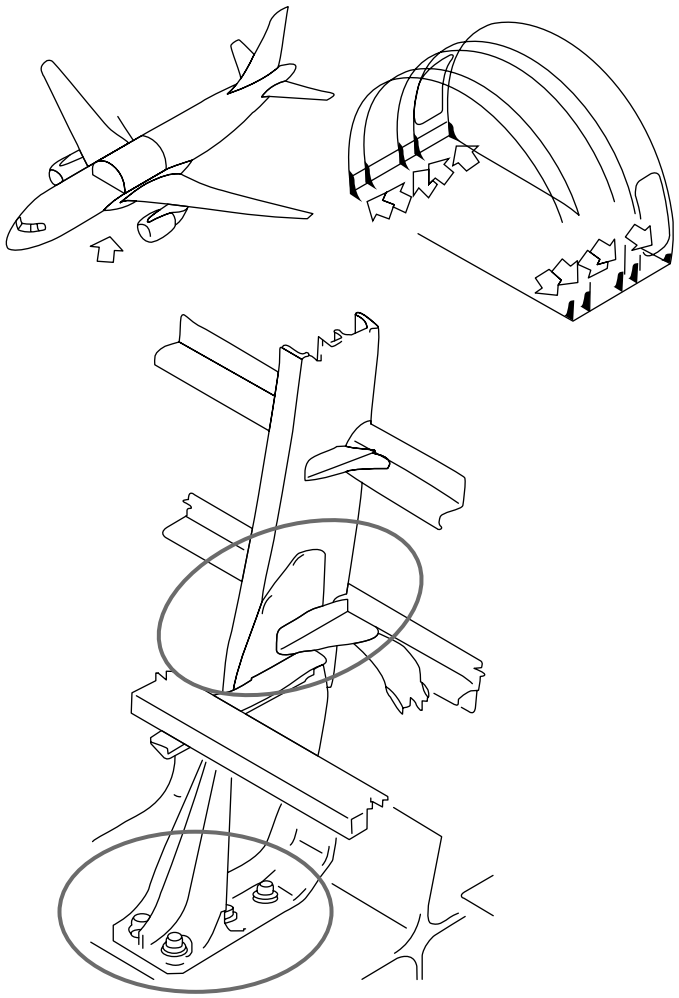


Figure 10. Connection of fuselage frames to center wing box.

the area under consideration is not easy to reach by an inspector to perform conventional NDI/NDT, and therefore significant effort and time are necessary to access the area. The result is additional work and, especially, additional downtime to perform the inspection. A suitable SHM-based solution would enable monitoring of the area without the need to access the area directly, because the sensors would be permanently installed on the structure.

An example of an area where the NDT inspection itself is difficult is illustrated in Figure 11: In a lap joint between a door panel and the surrounding fuselage structure panels, some signs of scratches were discovered. Since these scratches are possible sources of fatigue cracks, regular inspections of the area are necessary. However, the area is not easy to inspect with conventional inspection technologies because of the difficult local geometry. A suitable SHM solution would give the benefit of replacing the difficult (and therefore time-consuming and cost-intensive) inspection by an automated action.

The above-mentioned examples are only two possible benefits SHM may have when applied to structural areas, which are difficult to access or difficult to inspect. It is clear that a variety of different application cases can be defined and that each individual application will have to be investigated in regard to the benefit resulting from the SHM application, and also in regard to possible technologies to realize an SHM solution.

4.2 SHM application as a retrofit solution to in-service aircraft

This section is dedicated to the situation where a potential problem of an individual structural area is detected after a fleet of aircraft has already entered into service. This can happen in several ways. For example:

- A damage finding can occur on the full-scale fatigue test when some aircraft of the relevant type are already in service, since this test is, in most cases, not completely finished at the time when the scheduled service of the relevant aircraft type starts.
- A damage finding is identified on one or several aircraft of the in-service fleet.

In both cases, the question about the impact of the damage on the whole in-service fleet is immediately raised.

According to the state of the art, the damage is analyzed with regard to its impact and the result may be one of the following scenarios:

1. The operators have to modify the component on all in-service aircraft of the relevant type because a safe continued operation of the aircraft is no longer possible with the nonmodified structure.
2. The operators have to perform additional inspection actions on the relevant structural area for the whole fleet of that aircraft type, and for the whole remaining lifetime.

Sometimes, the operators are not forced to one of the two options above, but instead they are given the option to choose between them, i.e., each operator can choose between modifying all aircraft of the fleet or agreeing to an inspection plan for all aircraft.

Instead of modifying the structure immediately or agreeing to scheduled extra inspections, the operator may, in future, have the opportunity to decide on a third option with SHM: the maintenance workers have to get access to the structural area under consideration once to install the SHM technology. The inspections would then be carried out in an automated way by the SHM system, if necessary, for the whole remaining lifetime of the aircraft.

The main advantages of this option are as follows:

- The operator has to access the area only once: to install the SHM system. Further access for inspections will not be necessary since the system is locally installed and the inspection can be carried out without direct physical access.
- Also the downtime for the inspections is significantly reduced by using SHM inspection instead of conducting conventional inspections.

On the other hand, the business case is very dependent on the individual damage scenario examined. Normally, a probabilistic analysis needs to be carried out to calculate the probability that a significant part of the fleet will be affected by the damage, which was observed on one aircraft or on the full-scale test cell.

Figure 12 illustrates, in a simplified way, the comparison of cost between the classical options

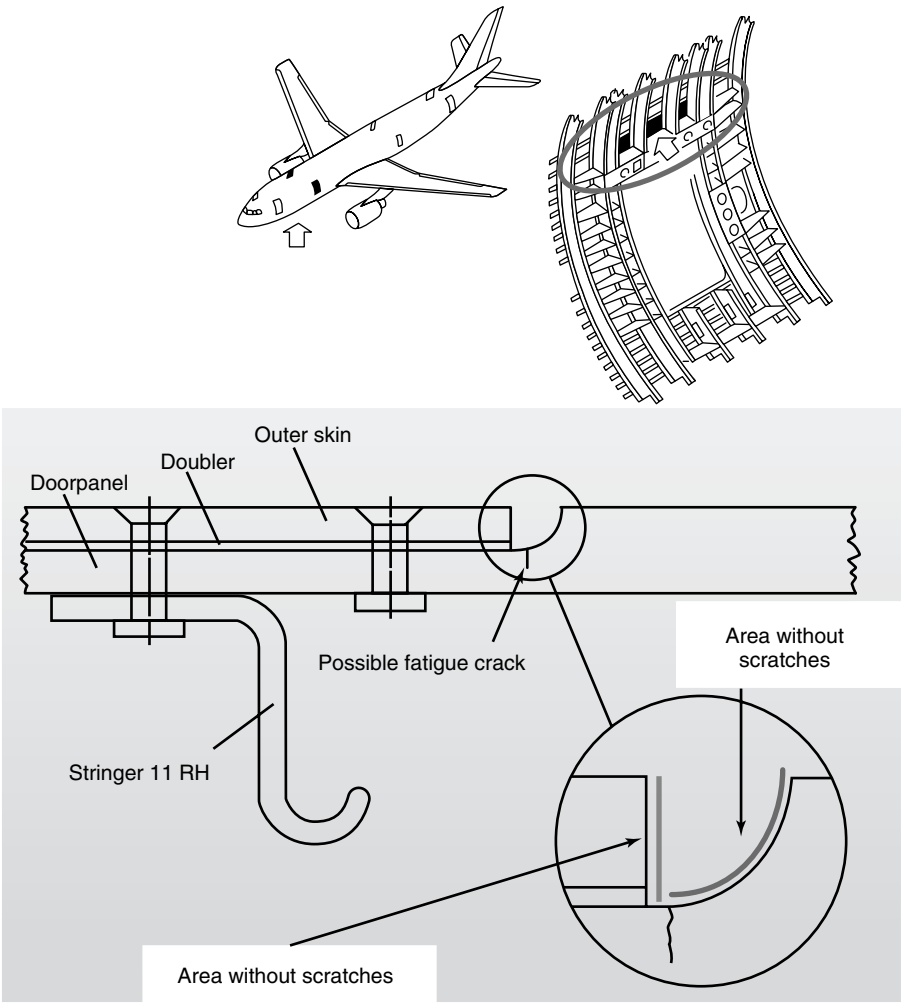


Figure 11. Chemically milled radius above a passenger door.

(structure modification or inspection requirements) on the one hand and the SHM alternative, on the other. In case of a conventional maintenance philosophy (without SHM), the aircraft manufacturer would either decide to have mandatory inspections for all fleet (case 1), or inspection requirements only (case 3), or a combination of both by giving the operator the freedom to chose (case 2).

- For mandatory modification, the cost for the operator is the cost for modification of the whole fleet, including additional cost for downtime, etc.
- In the case of inspection requirements, the costs for the operator are the cost for performing the scheduled extra inspections on the whole fleet. However, it may happen that some aircraft of the fleet will show a damage finding during the inspection. For these aircraft, the modification/repair has to be performed after the finding.

For the option using SHM in future, it is clear that the trade-off between the conventional solution and the SHM option depends strongly on the probability of the damage occurring on an aircraft during the service life. Therefore, probabilistic analysis is necessary before deciding on the business case for the application.

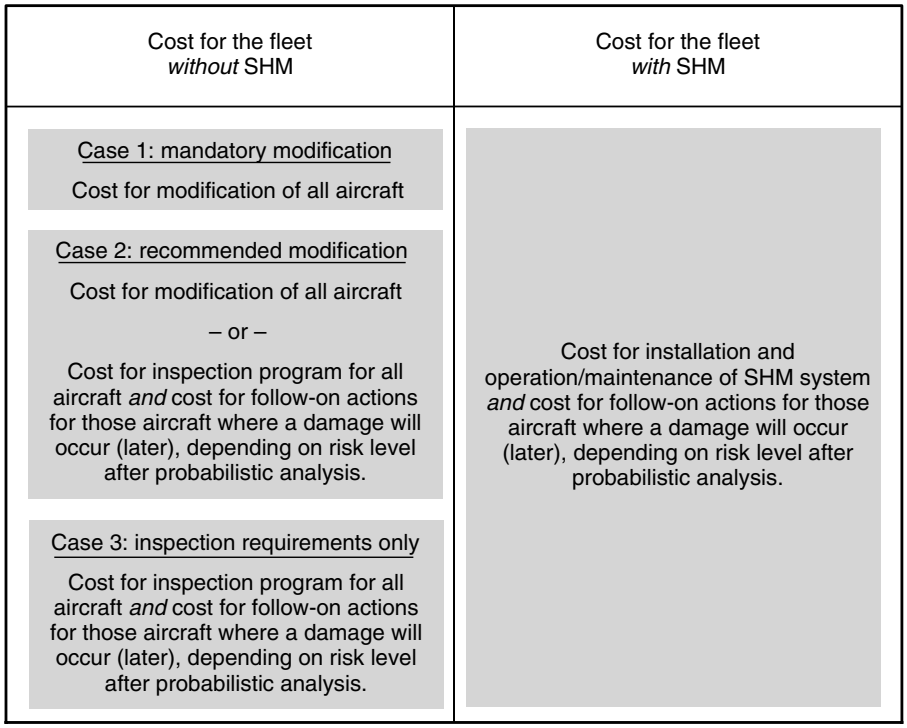


Figure 12. Cost comparison for in-service damages without and with SHM.

For example, if damage is found during a full-scale fatigue test, it is likely to happen after the design service goal of the aircraft, since the duration of a full-scale test is higher, e.g., 2.5 or 3.0 aircraft lifetimes. In that case, probabilistic analysis has to be carried out to identify how likely it is that the damage will occur on individual aircraft during the in-service life of one aircraft. This probability will indicate which fraction of an in-service fleet of aircraft will show the damage before reaching its design service life. If the damage is likely to occur on a significant fraction of the fleet, it may be more beneficial for the operator to decide for modification of the whole fleet at the beginning. If, however, the damage is expected to occur only on a small percent of the fleet, the operator may prefer to install the SHM system for monitoring the hot-spot area under consideration. The operator would then not be required to perform additional conventional inspections and the SHM (automated) inspections would not be penalizing to the operator. Only a small fraction of the fleet (if any) would show the damage later and only these aircraft would then need to be modified.

4.3 SHM application in aircraft life extension programs

It may happen that the aircraft manufacturer decides to operate an aircraft type that has proven to be reliable and robust for an extended service life i.e., for more flight cycles compared to what was planned when the aircraft was originally designed, certified, and entering into service. During the process of service life extension, the aircraft manufacturer has to demonstrate to the airworthiness authorities that the structure of the aircraft fleet can withstand the extended service goal. To do so, he has to demonstrate by calculations and/or tests that all components of the aircraft structure are sufficiently well designed to meet the extended requirements, for example, the "crack-growth" and "fatigue" requirements.

It may, of course, happen that some components do not have sufficient dimensioning reserves to withstand the new requirements, or at least cannot withstand them without additional inspections, modification, or component replacement. All of these options are clearly penalizing to the operator. As described in

the section above (application to in-service aircraft), it may be beneficial to include suitable SHM retrofit solutions to the structure. Again, the detailed business case for the SHM alternative depends strongly on the specific application under consideration and on the available SHM solution.

5 IMPROVING AIRCRAFT STRUCTURE DESIGN THROUGH USE OF SHM

This section is mainly dedicated to the possibility of improving aircraft design by using SHM. The idea is to exploit the aircraft structure in a more efficient way by having the benefit of knowing about the structure integrity throughout the service life rather than only during scheduled conventional inspections. In addition, structural information can be retrieved from areas that are not accessible with conventional NDT.

However, it must be clearly stated that this family of applications is complex: it requires suitable and very reliable SHM technologies and the certification of the optimized structure will depend on the SHM system. The certification of a structure, which has been weight optimized with the SHM benefit, would rely on the presence of permanent monitoring. In other words, the structure, in general, will not be allowed to operate without the SHM system.

5.1 General possibility of design improvements using SHM

The overall idea is to obtain a structural benefit from the use of SHM in the sense that some structural components may be sized in a less-conservative way compared to conventional design strategy but without compromising on safety. The design conservatisms mentioned before result from the fact that at present the operator performs scheduled inspections at rather large intervals, because very frequent inspections would be too penalizing. This has to be taken into account in the component's sizing in the sense that undetected damages between two inspections or growth of damages between two inspections have to be proved as not likely to turn into a dangerous scenario. This philosophy can, of course, be challenged if an automated inspection through SHM is performed permanently, or at least at small intervals compared to the conventional inspection. In any case, it is evident that the dimensioning of the component and the inspections performed are linked to each other.

The new philosophy would also fulfill the airworthiness requirements, but following a different strategy compared to that the conventional method. Therefore, the high level of safety can clearly be maintained.

An additional aspect comes from the fact that the dimensioning of the component will then only be valid as long as the SHM system is reliable and available throughout the whole structure service life. However, care must be taken not to penalize the aircraft operator by the new process of sizing and certifying the structure. For example, if the SHM system should fail during the aircraft operation, the operator should be able to continue operation for at least a fixed "grace period", so the operator can reach a hub or service center before he finally fixes the SHM system. Therefore, the justification has to show that no dangerous damages can grow even when operating the structure without SHM for a limited (and short) grace period.

These considerations put some strong requirements on the SHM systems considered for this application:

- The system has to find the relevant structural faults at extremely high mean time between failures (MTFB).
- For safety and certification reasons, it is clear that the system has to be equipped with self-diagnostic capabilities, i.e., the SHM system will give an indication if the system itself has a defect and cannot continue to ensure reliable monitoring.
- The system has to withstand a long lifetime, which has to be equal to or longer than the lifetime of the structure itself. Hence, the system has to be robust in the sense that environmental effects in service (including exposure to temperature differences, water, fuel, hydraulic fluids, etc.) should not compromise the reliability and durability of the system.
- False-positive indications have to be very rare, which means that only in very unlikely cases the SHM system will indicate a structural failure

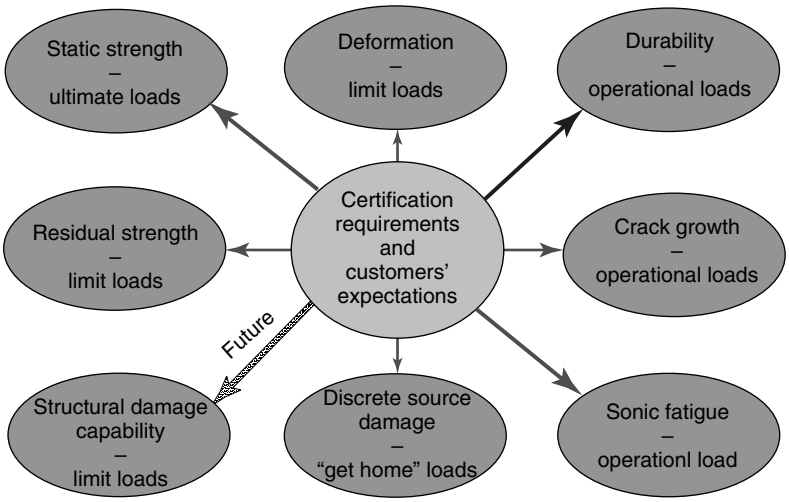


Figure 13. Main dimensioning criteria for aircraft metallic primary structure.

although the structure is intact. A false-positive indication would not compromise the safety of the structure, but it would force the operator to perform an additional conventional inspection to clarify the situation and therefore be a significant economic drawback on the operator.

Furthermore, it is clear that any application targeting a less-conservative dimensioning will have to be taken into account in the early design stages of the structure, in contrast to other applications, which may be also beneficial as retrofit solutions to existing aircraft [6, 7].

5.2 Example for metallic structure: monitoring stiffeners

During the design of aircraft structures, a wide range of aspects have to be considered to reach sufficient static strength, high durability, and excellent fatigue and damage-tolerance behavior. The end result of the iterative calculations is an optimized design regarding weight, costs, and aircraft performance. Figure 13 gives an overview of the major design criteria based on certification requirements and customers' expectations. These criteria and further aspects such as corrosion, repairability, and inspectability have to be applied to all structures and also to major modifications and repairs.

The fatigue and damage-tolerance regulations including the corresponding advisory circulars have to be met for certification. Fatigue and damage-tolerance analyses and complementary tests have to be performed for the whole primary structure to demonstrate high fatigue life, slow crack growth, and large damage capability for the structure. The major aspects of the fatigue and damage-tolerance regulation FAR 25.571 are summarized below:

- An evaluation of the structure has to show that a catastrophic failure due to fatigue, corrosion, or accidental damage, will be avoided throughout the operational life of the airplane.
- A structural inspection program has to be developed, considering probable damage locations, crack initiation mechanisms, crack-growth time histories, and crack detectability.

Figure 14 shows, in principle, the damage types to be considered during the fatigue and damage-tolerance analysis. Since the fuselage and wing structures for in-service aircraft are mainly externally inspected, the analysis assumes a skin crack above a broken internal stiffener, e.g., stringer or frame. Monitoring of the internal stiffener by SHM allows the application of a less-stringent crack scenario, i.e., the assumption of a skin crack above an intact stiffener. This leads to a reduced stress intensity delta K at the crack tips, which results in a slower crack growth and a longer critical crack length in the skin.

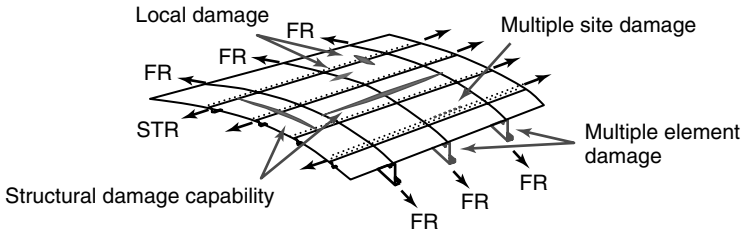


Figure 14. Damage types to be considered in fatigue and damage-tolerance justification.

The benefits due to SHM mentioned above exist in metallic areas dimensioned by damage tolerance, i.e., mainly by crack growth. The question of which areas of the structure are dimensioned by these criteria cannot be answered in a generalized way: the local dimensioning criteria for each structural area are dependent on the aircraft type, design criteria, mission profile, inspection program, and material. An example of aircraft structural area that is typically designed for damage tolerance/slow crack growth, are the skin panels in the upper metallic fuselage structure.

There are four different possibilities of SHM application at the fuselage structure:

- monitoring of stringers—to detect stringer cracks or failures;
- monitoring of frames—to detect frame cracks or failures;
- monitoring of skin—to detect circumferential skin cracks;
- monitoring of skin—to detect longitudinal skin cracks.

The idea discussed in this example is to challenge the "crack-growth" criterion for the aircraft fuselage skin by monitoring stiffeners (stringers and frames). Before going into detail, the principle of the crack-growth evaluation is briefly explained: The basic idea is that during the certification of a new airframe structure, the developers have to demonstrate to the certification authorities that any crack in the metallic structure growing during the operation of the aircraft will, in any case, be detected within the framework of the inspection program before it grows to a dangerous size. Figure 15 illustrates this principle: The calculated crack curve (crack length over number of flight cycles) is indicated, as well as the detectable and critical crack length. The aircraft manufacturer has to demonstrate that the crack-growth interval

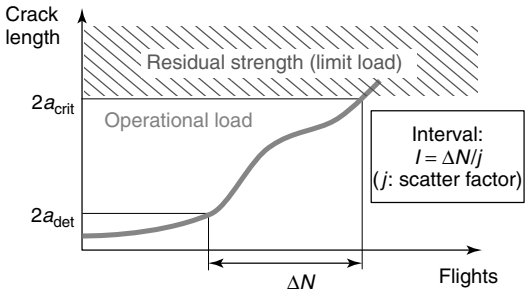


Figure 15. Schematic display of crack-growth calculation and determination of inspection interval.

from detectable to critical length is lower than the relevant inspection interval, divided by a scatter factor.

Figure 16 contains an example of an SHM application for the fuselage structure. It is assumed that the fuselage stringers are permanently monitored by SHM, whereby the skin is visually inspected from outside. The assumed crack scenario is the less-stringent damage scenario, i.e., a skin crack above an intact stringer. This crack scenario is compared with the traditional scenario, i.e., a skin crack above a broken stringer. The figure shows a significantly slower crack growth resulting from the application of SHM. The period over which the crack grows from $2a = 75\,\text{mm}$, which is detectable by general visual inspection, to the critical crack length is increased by a factor of roughly 2.5. This would either allow an increase in the intervals for general visual inspection of the skin by this factor or an increase in the allowable stress level by more than 15%. In analyses, it has turned out that for typical fuselage shells the calculated crack growth after a 15% stress increase with SHM still gives a better inspection interval than that reached without SHM and the original stress level in the skin. This significant improvement is not possible

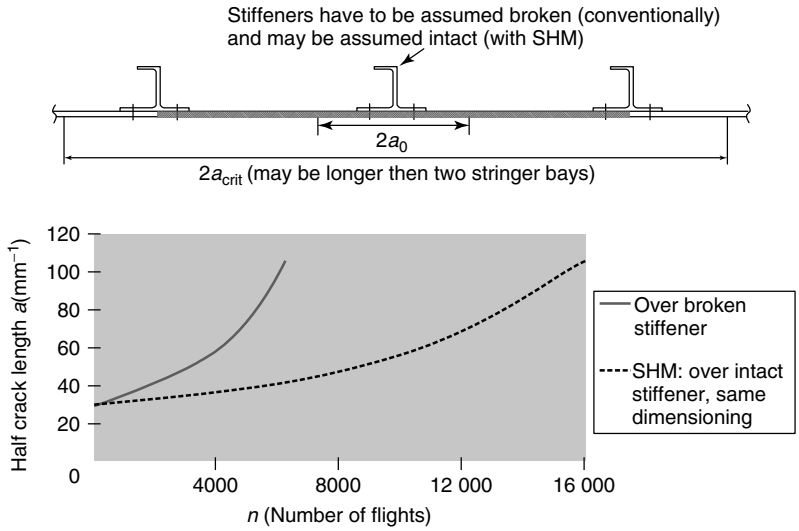


Figure 16. Calculated crack-growth curves for skin cracks, assuming broken or intact internal stiffeners (stringers).

for areas dimensioned by other design criteria, e.g., static strength. Every specific structural component has to be checked in detail for the extent to which a design/weight benefit would be possible by monitoring with an SHM system.

One example of a concept that integrates SHM technology into the structural design is presented here and is related to metallic components, e.g., stiffeners such as stringers or frames. A cavity is included into an extruded profile during the extrusion process. During operation of the component, this cavity is used to monitor the integrity of the component itself using the comparative vacuum monitoring (CVM™) method.

The CVM™ technology is based on the idea of maintaining a vacuum inside small cavities. The sensor consists of a foil or a film including the cavities and is applied to the structure in such a way that the structure surface forms a part of the cavity's boundary (Figure 17). A crack in the structure would now cause a leak in the vacuum, which can be detected by the relevant system attached to the sensor. Details can be found in **Comparative Vacuum Monitoring (CVM™)** and [4].

The Airbus VSI (Vacuum Sensor Integrated) concept takes advantage of the CVM™ principle. Figure 18 illustrates this idea. The cavities are included in the metallic component itself and are later connected to the relevant CVM™ technology (*see* **Comparative Vacuum Monitoring (CVM™)** and [4]). When a crack appears in the component (as indicated in the lower part of Figure 18) the leakage at the crossing points between cavities and crack will be detected.

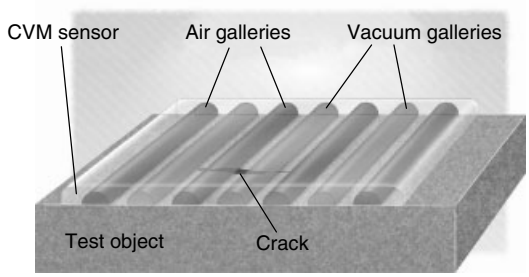


Figure 17. Basic principle of the CVM™ technology. [Reproduced with permission. © SMS Ltd.]

In the application described, the sensor and structural component are, in fact, one and the same part. This has two major advantages compared to conventional, surface-mounted sensors:

- The solution is very robust in the sense that the integration of the sensor (which is the cavity itself) into the profile makes it impossible to destroy the sensor without destroying the component itself. When using conventional surface-mounted sensors, care must be taken that the installation is robust to withstand roughly 30 years of service life (including humidity, temperature,

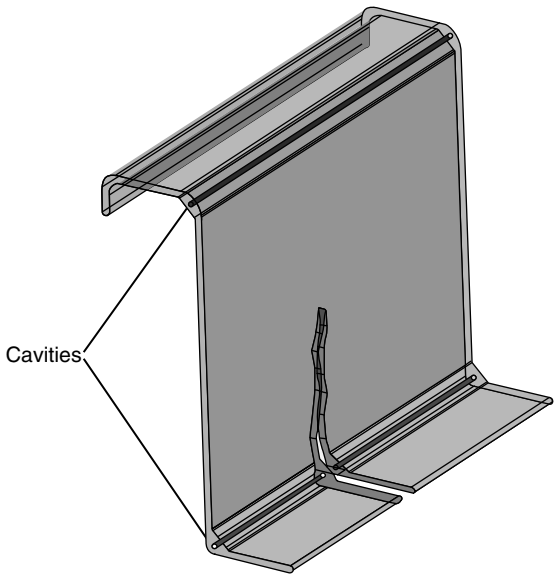


Figure 18. Schematic display of an extruded stiffening profile with integrated cavities for vacuum monitoring.

and hydraulic fluid exposure, workers stepping over the component, etc.).

- One of the main requirements for SHM systems is that the lifetime of the sensor must be at least as long as the lifetime of the structure that it is monitoring. Since the new component proposed here integrates the structural component and sensor, this requirement is always fulfilled by definition.

5.3 Example for composite structure: monitor for delaminations

In case of composite structures, some design conservatisms exist that may be challenged by using SHM technology without compromising on safety or reliability.

For example, composite structures are designed to withstand all required loads in a damaged status, as long as these damages cannot be easily found. If an internal delamination in a composite structure can exist in service, possibly because of an impact event on the component, the delamination will, often, not necessarily be found during operation. Consequently, the aircraft manufacturer has to design the whole structure to withstand the required loads in this damaged state. In general, this delamination damage could potentially exist anywhere on the structure.

If an SHM technology could be installed on the component guaranteeing to find delamination damage above a certain size, the structural dimensioning described could be challenged. The structure would be continuously monitored and could be designed to avoid the above-mentioned design conservatism, at least to a certain extent.

Figure 19 illustrates some of the damage types that can exist in composite structures, and can, above a certain size, be detected by conventional NDI methods such as ultrasonic inspection. When making the transition to SHM monitoring, instead of conventional inspections, the damage has to be detected by

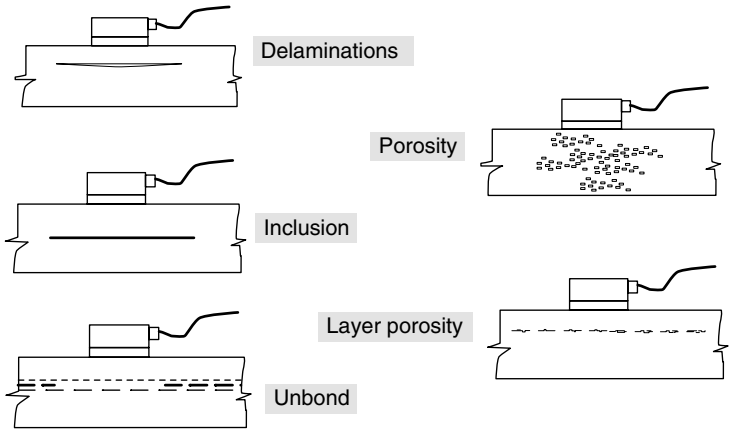


Figure 19. Different damages to be detected in a composite laminate.

suitable SHM technology. If the envisaged SHM solution is capable of finding even smaller cracks than what the conventional inspections can find, the benefit will be increased.

One method by which an SHM system can detect damage in composite laminates is to include sensing fibers (e.g., fiber Bragg gratings) into the laminate itself during manufacture (see [8, 9]). As in the example of metals above, the stringent requirements for the SHM system, such as reliability, lifetime, robustness, etc., also have to be fulfilled for this application. In addition, the repair aspect has to be considered as soon as the SHM system is embedded into the structural component itself instead of being applied on the surface. If the structural component is damaged and the damage is indicated by the SHM system, the structure will, of course, have to be repaired. For composite components, repair strategies are available today. However, if the operability of the structure depends on SHM, the repair of the SHM system itself is also necessary before the operation of the structure can be continued. The development of repair principles for this type of structure will therefore need to take into account both the repair of the structure and the SHM system, which would mean additional effort in the case of embedded sensor systems.

5.4 Improving global architecture through SHM

In this section, an example is given for the optimization of the global architecture by locally monitoring parts of the structure, using SHM. There are areas in the design where the installation of systems, tubings, pipes, etc. is determined by the inspectability of the structure, for example, when a highly loaded fitting is designed with consideration to special inspection requirements. The aircraft architects will then probably decide not to locate any system components in such a way that the structure inspection will be made more complicated. For the operator, it would be a significant economic burden to remove system components for each inspection of the structural component "hidden" behind it. Therefore, architects would have to consider a trade-off between "hiding" the structural area to be inspected behind the system and changing the installation of the system. A change of installation would then, most likely, not be the optimal from the system point of view, e.g., additional/longer tubings and pipes, and from global architecture point of view, e.g., additional space allocation through the system with a further impact on space-allocation planning.

An SHM solution for the structural component would solve the problem in the sense that with SHM permanently installed during the structure assembly, the inspections could be carried out in an automated way throughout the aircraft life without the need of getting direct physical access to the structural area. The system's architecture could then be optimized, leading the above-mentioned trade-off to an optimal solution, which would not be realistic without the SHM installation. As with previous examples, in this application, the detailed business case depends strongly on the details of structure and system installation needs of the particular area under consideration.

6 OUTLOOK AND CONCLUSIONS

With the growing maturity of SHM technologies and solutions, more and more applications are seen for civil aircraft. The "first generation" applications in Airbus are mainly related to replacing conventional inspections by SHM solutions and to retrofitting existing aircraft with SHM.

However, in future generations, more ambitious applications will be possible. One example for this is the idea of making the structural dimensioning less conservative, leading to a weight saving. Examples have been discussed for metallic as well as for composite structures, but it has to be kept in mind that the requirements for the SHM technology are extremely high and that the new technology has to first prove its reliability before use. This proof of reliability has to occur not only on a lab-scale, but also on flying aircraft before it can be considered to be used for "weight saving" type of applications.

The general development, which is foreseen for the future, is that multidisciplinary development of new structural solutions will become more important, integrating the disciplines of material development, stress/calculation, design, global architecture, technology development, and others to find the best solution for each relevant structural component. One future trend may be to look for more integration of structural components and SHM technologies.

The technology developers need to be triggered to work in the right direction to be able to deliver the right solutions for the structural concepts being developed. In addition, it is emphasized that only an early introduction of the new technologies into the aircraft planning and predesign process will finally deliver the optimum benefit for the new generation aircraft structure.

RELATED ARTICLES

Fiber-optic Sensors

Landing Gear

REFERENCES

[1] Assler H, Telgkamp J. Design of aircraft structures under special consideration of NDT. Keynote lecture Presented at the *9th European Conference on Non-Destructive Testing*. Berlin, 2006.

[2] Pacchione M, Telgkamp J. Challenges of the metallic fuselage. Presented at the *25th Conference of the International Council of the Aeronautic Sciences (ICAS)*. Hamburg, 2006.

[3] Beral B, Speckmann H. Structural health monitoring (SHM) for aircraft structures: a challenge for system developers and aircraft manufacturers. Presented at the *4th International Workshop on Structural Health Monitoring*. Stanford, CA, 2003.

[4] Stehmeier H, Speckmann H. Comparative vacuum monitoring (CVM™) of fatigue cracking in aircraft. Presented at the *2nd European Workshop on Structural Health Monitoring*. Munich, 7–9 July 2004.

[5] Paget C. Structure health monitoring for engineering asset management in aeronautics. *2nd World Congress on Engineering Asset Management And Fourth International Conference on Condition Monitoring*. Harrogate, 11–14 June 2007.

[6] Schmidt H-J, Telgkamp J, Schmidt-Brandecker B. Application of structural health monitoring to improve efficiency of aircraft structure. Presented at the *2nd European Workshop on Structural Health Monitoring*. Lancaster, PA, 7–9 July 2004.

[7] Telgkamp J, Schmidt H-J. Benefits by the application of structural health monitoring (SHM) systems on civil transport aircraft. *Proceedings of the 4 th International Workshop on Structural Health Monitoring*. *Stanford University,* DEStech Publications: Stanford, CA, 2003.

[8] Guemes A, Menendez JM. Embedded fibre Bragg gratings for design and manufacturing optimization of advanced composites. Presented at the *2005 International Workshop on Structural Health Monitoring*. Stanford, CA, 2005.

[9] Mrad N. Fiber optic based structural health monitoring. Presented at the *2nd European Workshop on Structural Health Monitoring*. Munich, 2004.

FURTHER READING

Price DC, Scott DA. An integrated health monitoring system for an ageless aerospace vehicle. *Presented at the 2003 International Workshop on Structural Health Monitoring*, 2003.

Scott M, Bannister M, Herszberg I, Li H, Thomson R. Structural health monitoring—the future of advanced composite structures. *Presented at the 5 th International Workshop on Structural Health Monitoring*, 7–9 July 2005.

Chapter 84
Design Principles for Civil Structures

Vistasp M. Karbhari
University of Alabama in Huntsville, Huntsville, AL, USA

1 INTRODUCTION

Structures associated with the built infrastructure are traditionally designed conservatively to ensure extremely small levels of risk associated with their sudden failure. The level of acceptable risk is orders of magnitude lower than that associated with structures more commonly associated with advanced technology applications such as aircraft and spacecraft. Melchers [1], for example, shows orders-of-magnitude difference in the risks associated with air and automobile travel than those associated with structural failure of the built infrastructure. The reason for this is threefold: (i) the built infrastructure interacts daily with man and hence forms the backbone of society and, consequently, is expected to be a stable anchor without probability of failure; (ii) unlike aircraft, automobiles, and machinery, which are extensively tested at the level of prototypes and are manufactured multiple times in exactly the same configuration, a majority of the built infrastructure is unique and therefore presents neither the opportunity for extensive prototype testing (except for ongoing tests on the structure during its service life) nor, to date, the means of accurately predicting future response, and is therefore designed without the same knowledge base; and (iii) the ambiguity and degree of uncertainty related to the actual life of the structure and the potential increases in load and usage, in the future, force the use of an extremely conservative design.

Classically, structures are designed by considering the relationship between the *capacity* of the components of the structure, and of the assembled system, to the *demand* anticipated for that structure. Since times immemorial, engineers have designed structures using a fairly simple method that requires that

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

the maximum stress due to anticipated loads on the structure are a fraction of the maximum stresses that the structure can withstand. The ratio of the stress at "failure" to the actual level determined for the structure under conditions of service is the factor of safety, which is generically a sufficiently large number, both due to the need for safety at the component level and due to the amassing of component-level factors of safety at the systems level.

This methodology has been used in various forms to develop codes and specifications based on the working stress, or allowable stress, design method. While the methodology has served fairly well for centuries, it does have some major disadvantages. First, the method is built on an approximation of deterministic values for load events, material characteristics, and a linearly equivalent model of structural response, which intrinsically removes all considerations of a statistical and probabilistic nature. Secondly, it leads to structures of the same class, but different configurations, being designed with nonuniform factors of safety. For example, the use of the same specifications for the design of short- and long-span bridges leads to long-span bridges having substantially higher factors of safety since the long-span bridges are governed by dead loads that can be determined with a high degree of certainty than the live loads that would govern the design of short-span bridges. Since the 1970s, this has led to the development of modern design codes and specifications that consider the concepts of structural reliability through limit-state and load and resistance factor design (LRFD) methods. In the LRFD method, which is still being developed in the United States, and is perhaps used more extensively on bridges than on other structures, the factor of safety is quantified by a reliability index, β, which relates the statistical characteristics of the combination of loads with those of the resistance of the components of the structure. In general, bridge codes are calibrated to a value of $\beta = 3.5$, which corresponds to the probability of failure of a component of the structure of 1 in about 2000 during the service life of the bridge (which has a nominal extent of 75–100 years).

Notwithstanding the use of a reliability basis for design, there is still significant conservatism since the value of β relates only to the notional failure of a component with the failure of the system itself having a much smaller probability of failure. In addition, the methodology does not provide a means for addressing the challenges associated with the uniqueness of each system and the lack of data regarding its anticipated service-life response. Additionally, there are challenges both in estimating the remaining life of a structure in service as well as in the real-time assessment of the optimum time to conduct repairs based on the currently used visual method of inspection.

2 SHM AND CIVIL ENGINEERING

Notwithstanding the design methodology initially used, all structures, including critical civil infrastructure facilities like bridges, pipelines, transmission systems, and highways deteriorate with time. This deterioration is due to various reasons including fatigue failure caused by repetitive traffic loads, effects of environmental elements and aging on the materials of construction, and extreme events such as an earthquake, hurricanes, and floods. In order to maintain the safety of the structures, various cognizant authorities prescribe methods of routine inspection. For example, in the case of "lifeline" structures such as highway bridges, the states are mandated by the National Bridge Inspection Program to periodically inventory and inspect all highway bridges on public roads. The National Bridge Inspection Standards [2], first implemented in 1971, prescribe minimum requirements for the inspection of highway bridges in the United States. A substantial amount of research has been conducted in this area to improve the speed and reliability of such inspections. According to a recent survey performed by the Federal Highway Administration [3], although visual inspection is slow and costly and has unproven reliability, it is still the primary tool used to perform these inspections. The implementation of these inspections consists of scheduled field trips to bridge sites at routine intervals, usually once every several years. If a significant increase in distress between inspections is noted, the period between inspections is decreased and the level of inspection is increased till such time that the distress has been corrected by replacement or repair. Not only is this method of time-based inspection inefficient in terms of resources, because all bridges are inspected at the same frequency, regardless of the condition of the bridge, but there is also a

potential danger that serious damage could happen to the bridge in between two inspections, thus posing a hazard to public safety. The adoption of structural health monitoring (SHM) principles assists in decreasing these effects.

With the increasing importance of lifelines, such as highways, to the national economy and the well-being of a nation, there is a need to maximize the degree of mobility of the system. During inspections, issues such as serviceability, reliability, and durability need to be answered in precise terms. More specifically, the owners (the Federal Highway Administration, State Departments of Transportation, etc.) need to be able to answer the following questions: (i) Has the load capacity or resistance of the structure (serviceability) changed? (ii) What is the probability of failure of the structure (reliability)? and (iii) How long will the structure continue to function as designed (durability)? [4]. This requires not just routine, or critical event (such as an earthquake)-based inspections, but rather a means of continuous monitoring of a structure to provide an assessment of changes as a function of time and an early warning of an unsafe condition using real-time data.

Although the term SHM has gained prominence recently, its basis and motivation can be traced to the very earliest endeavors of man to conceptualize, construct, worry about deterioration, and then attempt to repair (or otherwise prolong the life) of a structure. Thus, it represents an attempt at deriving knowledge about the actual condition of a structure, or system, with the aim of not just knowing that its performance may have deteriorated, but rather to be able to assess remaining performance levels and life. In its various forms, over the years, it has been represented as the process of conventional inspection, inspection through a combination of data acquisition and damage assessment, and more recently, as the embodiment of an approach enabling a combination of nondestructive inspection and structural characterization to detect changes in structural response. In recent years, it has often been considered as a complementary technology to systems identification and nondestructive damage-detection methods.

Most work, to date, on SHM systems for civil structures, has been useful, but resembles existing bridge management systems. These management systems focus on processing collected data, but are unable to measure or evaluate the rate of structural deterioration for a specific bridge. Housner *et al.* [5] provide an extensive summary of the state of the art in the control and monitoring of civil engineering structures, and the link between structural control and other forms of control theory. They also define SHM as "the use of *in situ* nondestructive sensing and analysis of structural characteristics, including the structural response, for detecting changes that may indicate damage or degradation". While most ongoing work related to SHM conforms to the definition given by Housner *et al.* [5], this definition also identifies the weakness associated with these methods. A health monitoring system, which detects only changes that may indicate damage or degradation in the civil structure without providing a measure of quantification, is of little use to the owner of that structure. For example, while data from the Strong Motion Instrumentation Program (SMIP) in California provides very important seismograph data related to the shaking of ground and structures during earthquakes through a sensor network through California [6] (which is used to both enhance the understanding of earthquakes and to improve design and construction methods), it does not assist engineers/inspectors in discerning whether structures should be kept open or closed after an earthquake—this has still to be done through manual inspections. While researchers have attempted to integrate quantitative nondestructive testing (NDT) with health monitoring, the focus, to date, has primarily been on data collection, and not on evaluation. It could be concluded that these works have produced better research tools than approaches to bridge management. It is only recently that the emphasis has been on the implementation of efficient approaches to not only collect data from a structure in service but to also process the data to evaluate key performance measures, needed by the owner, such as serviceability, reliability, and durability. For the current work, the definition by Housner *et al.* [5] is modified and SHM is defined as the use of *in situ*, nondestructive sensing and analysis of structural characteristics, including the structural response, for the purpose of estimating the severity of damage and evaluating the consequences of damage on the structure in terms of response, capacity, and service life. More simply, SHM represents the implementation of a Level IV [7] nondestructive damage-evaluation method, and for the purposes of use in civil infrastructure then it must have the ability to collect, validate,

and make accessible operational data on the basis of which decisions related to service-life management can be made [8].

On the basis of this, a more effective "condition-based" approach is the option considered for implementation, in which, the condition of the structure is constantly monitored using an appropriate nondestructive evaluation (NDE) technique without requiring the inspector to actually be on site, with the "in-person" visual inspection only being performed when necessary. The word "constantly" is used to indicate that the assessment of a structure's condition would be performed at much shorter intervals than would be possible using current visual inspection procedures. More importantly, the condition assessment would include more content than could be provided by a visual inspection. The following assessment, for example, would be typically desired by the bridge management authority: (i) damage to the structure and changes in the structural resistance, (ii) probability of failure or of the structure's performance falling below a certain threshold, and (iii) estimation of the severity of damage and the remaining service life. Condition-based monitoring can thus not only reduce resources needed for inspection but, when combined with advances in sensor, computational, and telecommunications technology, could also provide for continuous and autonomous assessment of the structural response.

With the increasing age of the world's built environment and the criticality of infrastructure systems and lifelines to the well-being and progress of society, there is an immense value in the efficient implementation of SHM—enabling, in an utopian world, the autonomous management of systems to prevent failure while simultaneously minimizing the cost of maintenance, reducing or eliminating downtime, and even upgrading structural systems as necessary. It must be noted that a critical feature of such systems is embedded in the concepts of durability and damage tolerance as developed in the aerospace world, and as shown schematically in Figure 1. It should be noted that while the concept is depicted in terms of residual static strength and damage size, it could just as easily be expressed in terms of another residual performance characteristic, or even remaining life and deterioration as measured through characteristics other than physical damage size.

3 CRITICAL NEEDS AND DESIGN PRINCIPLES

SHM intrinsically includes the four operations of acquisition, validation, analysis, and management. For the SHM system to operate as designed, key issues as identified in Figure 2 need to be addressed during the design of the SHM system itself prior to its implementation in the field. Also, the concept of monitoring prescribes that it be an ongoing process, albeit in some cases it could be used at preset intervals of time, or when activated by a threshold event (such as earthquake excitation, and overload). Thus, SHM is essentially the basis for condition-based rather than time-based monitoring, and the system should be

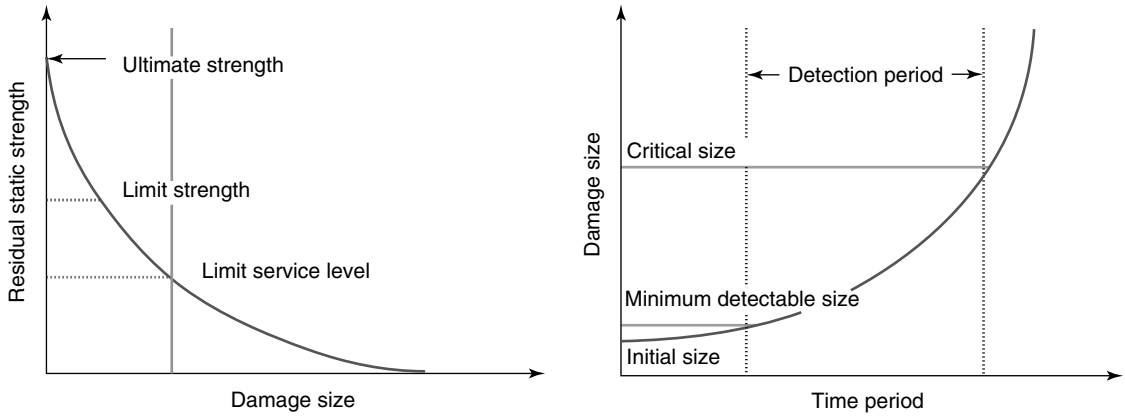


Figure 1. Schematic showing the application of concepts of durability and damage tolerance to design.

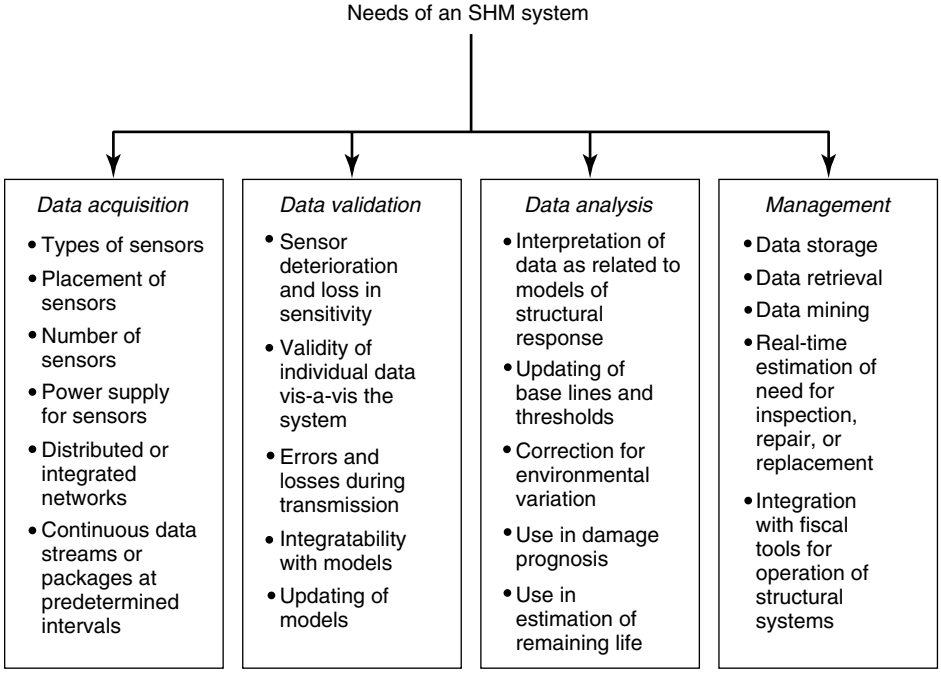


Figure 2. Critical components in the design of an SHM system.

capable of integrating real-time data on aging and degradation into the assessment of structural integrity and reliability. It should be noted that mere monitoring systems, i.e., those that just collect data, do not use an integrated approach to the design, implementation, and operation of the SHM system, resulting in the benefits of the system not being fully realized. Too often, a disproportionate emphasis is placed on the collection of data rather than on the management of this data and the use of decision-making tools that would support the ultimate aim of using the collected data to effect better management of the infrastructure system. Typically, even well-planned systems collect data on a continuous or periodic basis and transmit the data to a common point. The data is then compared with the results from a numerical model. The weakness lies in the fact that the systems may not attempt to update the model to reflect aging and deterioration, or changes made through routine maintenance or even rehabilitation. This can be done through the incorporation of a system's identification or nondestructive damage-evaluation algorithm to rapidly process the data with the ultimate goals of estimating capacity and service life.

Through a review of a large number of existing SHM projects on components of the built infrastructure, the following primary aspects can be identified as being crucial to the design of a viable SHM system:

- definition of clearly defined goals for the system;
- appropriate care in the selection of instrumentation and method of operation;
- predetermination of appropriate rates of usable (i.e., valid) data;
- development of approaches for real-time analysis of the validated data; and
- development and seamless implementation of tools for appropriate data interpretation and result implementation.

A brief discussion of each of the five aspects is given below.

SHM systems are often used on the basis of technology push rather than applications pull based on a clearly defined objective. While the enhancement of systems safety (through use of SHM as a mechanism of warning of failure) is desirable, SHM systems by themselves cannot ensure a higher level of safety, or even a better method of maintenance. SHM systems

by themselves also cannot decrease the level of maintenance, or even increase the periods between maintenance. If appropriately designed, however, they can reduce the amount of unnecessary inspections and ensure that deterioration/degradation is tracked such that the owner/operator has consistent and updated estimates of deterioration (quantity and general location), capacity, and remaining service life. While developing a design for an SHM system, a comprehensive examination of the structural characteristics and expected response of the structure to be monitored should first be considered. In addition, it is crucial that the requirements of the system be clearly identified, especially as related to the characteristics that are to be directly measured. In the case of a bridge, for example, the following aspects would need to be considered as the initial aspects in the design of the SHM system:

- type and class of bridge (the requirements and design principles will differ based on not only span but also the class of bridge—cable stayed, suspension, slab-on-girder, box girder, etc.);
- materials used in the fabrication of bridge (aspects to be monitored in steel bridges could be different from those in bridges fabricated of structural concrete);
- criticality of specific components of the bridge, their vulnerability to load and extreme events, and level of redundancy;
- forces that need to be considered in evaluating the bridge response (e.g., traffic load, wind load, temperature-induced loads, wave loads, seismic forces, etc.);
- determination of the characteristic parameters and response under static, normal load, load combinations, and extreme load events. This will provide the crucial basis for the selection of instrumentation;
- level of average daily traffic (ADT) and consideration of "permit" loads; and
- interaction of superstructure with substructure.

Thus, *in toto*, the design has to consider the three interacting aspects of working conditions/environment, structural configuration and state, and structural response. These must be considered in light of "hazards" that could cause changes in structural response, ranging from aging and deterioration of materials to the effect of events such as earthquakes and floods. Table 1 lists some of the aspects that should be considered during the design phase of SHM systems for different types of infrastructure.

While a large number of sensors are now available ranging in size, sensitivity, method of measurement,

Table 1. Examples of aspects for consideration in designing SHM systems

Class of infrastructure	Considerations
Bridges	Materials deterioration, time-dependent losses (creep, shrinkage, relaxation), corrosion, fatigue and cyclic load related changes, overstress, environmental extremes, unintentional damage, vandalism, soil–structure interaction, earthquake and wind forces, wave action
Buildings	Materials deterioration, corrosion, overload, use in unintended manner, wind and/or snow load, earthquakes, vandalism
Industrial (manufacturing-related) facilities	Materials deterioration, corrosion, overload, use in unintended manner, wind and/or snow load, damage due to earthquakes, vandalism, stress/damage due to manufacturing activities, vehicular impact
Energy generating facilities	Materials deterioration, corrosion, overload, wind/snow load, earthquakes
Offshore facilities	Materials deterioration, corrosion, overload, wave action, debris impact, ice-related damage, explosions (offshore oil platforms), machinery-related impact
Pipe lines	Materials deterioration, corrosion, overload/overstressing, blockage, temperature extremes, wind/wave/snow and earthquake loads
Dams, waterways	Materials deterioration, corrosion, overload, water pressure, soil–structure interaction, earthquakes
Port- and harbor-related facilities	Materials deterioration, corrosion, overload, vessel impact, wave action, debris impact, soil–fluid–structure interaction

and ability to work in a hard-wired and/or wireless mode, the intrinsic challenge still remains the same—appropriate selection (based on actual need, operational environment, and expected service life) and placement, and appropriate design of the system. A good review of different types of sensors crucial for use in civil infrastructure is given in [9]. On the basis of the characteristics of the structural system under consideration and the needs of the monitoring system, various devices/sensors can be used which measure either the absolute, or relative, value of characteristics such as deflections/deformations, strains, accelerations, temperature, moisture content, acoustic emission, electrical resistance or potential, load, etc. The sensors should clearly be selected on the basis of their ability to not only measure the specific characteristics needed but also do that at the appropriate level of sensitivity and with the reliability level needed in the exact operating environment. When used in a civil engineering environment, special care has to be taken to ensure compatibility with the changing environment and the vagaries of nature. This is of special importance when sensors are bonded or otherwise placed on surfaces, which may be subject to extremes of temperatures, large temperature variations (both daily and over seasons), moisture (ranging from humidity and precipitation to actual immersion such as when flash floods cause overtopping of bridges), and UV radiation. Compensation for temperature variation is well established for bonded resistance strain gauges and accelerometers. Yet, false alarms are often seen due to fast transient temperature changes especially in low-frequency accelerometers and this must be kept in mind during sensor selection and deployment. In cases where the sensors are bonded onto the concrete or steel substrates, care needs to be taken to ensure that the bond itself does not deteriorate with exposure and time, and further that deterioration and/or changes at the level of the substrate do not cause recording of erroneous measurements. While these aspects may appear to be trivial, they are of extreme importance in ensuring the long-term reliability of data. There are special concerns related to the use of SHM in cold regions beyond those associated with increasing brittleness of the materials used in the sensors, leads, and connections. Rime is a form of ice that forms when supercooled droplets, such as those in fog, freeze on contact with surfaces already below freezing point and this causes sensors to literally freeze. Also, this can cause brittle rupture of components themselves, or stretching, fracture, or even pullout of the connectors and leads. In addition, the overall design of the system must take into account the effect of accidental impact such as that from floating debris (Figure 3), which can effectively cause severe damage as a result of repeated impact. It is critical that sensor and system selection is based not just on considerations of sensitivity and durability but also on robustness and reliability, especially if the SHM system is expected to be in operation over an extended period of time. In this vein, it should be noted that although fiber-optic systems have good sensitivity and fairly good short-term durability, the drift can be significant if intended to be used continuously over long periods of time at very high levels of sensitivity. In addition, even though these systems have shown very good performance in extremely harsh environments (such as when mounted in automotive cylinders subject to instantaneous gas temperatures as high as 1500 $^{\circ}$C and continuous temperatures up to 300 $^{\circ}$C with pressures up to 300 bar), their reliability, to date, rarely exceeds 10 years. It must be emphasized that this is only a fraction of the expected service life of a bridge, thus pointing out the need for monitoring of the SHM system itself to ensure that parts are replaced prior to the end of their service lives.

All too often the small size, relative cost, or just the inability of the users to design a system results in the overuse of sensors. The mere capability of placing a million sensors on a structure does not automatically ensure a better SHM system. Rather, it almost always ensures failure since attention has not been paid to

Figure 3. Floating debris in the form of a tree trunk and roots.

design, nor of how to access and interpret data. Just as materials deteriorate under environmental exposure, through use, sensors can also degrade and this fact is often completely forgotten, resulting in expensive and highly complicated systems either delivering invalid, or no data, in very short periods of time. At the minimum, the expected service life of the SHM system must be in excess of that of the structural system being monitored. It is also important to keep in mind that unlike aircraft and spacecraft, which can be expected to be routinely brought in for thorough inspection, civil structural systems remain in the environment in which they were constructed for extremely long periods of time with significantly lower levels and extent of maintenance, and with long anticipated service lives (30–75 years or even more). Thus, even if the durability of the SHM system cannot be guaranteed for these long periods, there has to be a very high level of surety that a replacement system (or components) can be easily installed (at a reasonable cost) to continue the monitoring.

It must be emphasized that even in times when sensor costs can reasonably be expected to decrease substantially there is a fallacy in the argument that "more is better". In fact, it can easily be shown that the opposite is true—the increase of data channels not only complicates the issue of data transmission and synchronization but also increases the complexity of data validation. The mere possession of data of points in the structure that are unimportant and do not provide any characterization of response is a waste. Thus, there is a major challenge in minimizing (i.e., optimizing) the number of sensors based on the actual need and ability to characterize pertinent aspects of response.

Beyond the selection of the sensors themselves are the critical design principles of placement and arrangement. Although it is conceivable that a very large number of sensors could be placed on a structure such that the determination of the exact location was unimportant in reality, it is critical that sensors are placed at locations where one anticipates the signal representing the characteristic being measured shows an important value. Thus, it is essential that a comprehensive study of structural response is conducted prior to positioning the sensors so as to enable the identification of crucial locations. These could be locations of the highest stress or strain, nodes at which inflexion of deformation took place, structural connections or locations at which failures were anticipated to occur, etc. In addition, the layout needs to be designed so as to optimize the number of locations used and the accuracy of information available both through direct measurement and through interpolation. In the case of long-span bridges, for example, the following design principles could be used:

- Sensor arrangements should be such that the deformation of the superstructure is clearly determined. In the case of box-girder-based bridges, this would necessitate the determination of deflections of both the top and bottom surfaces.
- Sensor arrangement should be mapped as a function of traffic lanes and prior determination of locations of maximum load under different load combinations. Placement under the wheel-line or at locations such that determination of wheel-line effects is possible is important.
- When wind load is an important consideration, the sensor positioning must take this into account by considering the measurement at midspan of decks where deformations are likely to be the highest and at the top of towers, if the bridge design includes towers. In the case of long-span bridges where spans are long enough to show significant wind force response, positioning of sensors at quarter or eighth span is advantageous.
- In areas of seismic activity, it is important to not only have accelerometers positioned to accurately measure dynamic response but there also needs to be placement in the vicinity of the bridge at locations where ground motion can be measured isolated from the effects of structural response. It should be noted that the positioning to capture structural response of the bridge also needs to be able to capture the dynamic response of the bridge under excitation from ambient vibration such as traffic.
- Temperature and humidity effects are best measured at midspans and extreme locations. In addition, there is the need to be able to determine gradients between the top and bottom surfaces of decks and between the top and base of towers.
- While corrosion can take place at any location, areas where there is a change of section, joints and connections, high stresses, etc. are important points for location.
- The overall configuration (geometry) of the structure should also be mappable.

The challenge of assessing the validity of data is similar to the challenge of selection of the sensor and its placement. While large data streams may seem impressive to the routine observer (and unfortunately all too often even seem to be the ultimate goal of experienced engineers and scientists, without concern regarding the usefulness of the data), what is important is the ability to validate the data in real time, and thereby separate signals due to nonresponsive sensors from those providing the actual measurement of response. All too often, a structure or system has been declared severely damaged or conversely totally undamaged after an extreme event, because the validity of data streams was not checked. There are two generic approaches to validation—analytical redundancy and hardware redundancy. In the first approach, a mathematical model that allows for comparison of static and dynamic responses of the sensor to determine the anticipated value has to be implemented. Unfortunately, this requires addition of sensors, and increases time and complexity. In the case of hardware redundancy, validation is done through selection of data that is common to a majority of sensors. Again, for reliability, the number of sensors has to be increased, but depending on the structure even this may not ensure validation of the data stream. Further, this results in overall loss in sensitivity since the result is necessarily a statistical approximation. The obvious solution, albeit with significant implementation challenges, is the development of a knowledge-based system that applies reason through genetic algorithms, fuzzy logic, or other such tools to infer the right solution. This would enable validation in real time (unlike the two previously described approaches) through the use of reasoning under uncertainty. However, again, the major challenge is to find the right compromise between performance and precision [10, 11].

While it is useful to be able to both see data streams in real time and to archive them for future use, the real purpose of the data is to allow interpretation and analysis. If the system is set up such that there are large streams of data, but none of them can be accessed in real time, the system must be considered a failure. Obviously, data pertaining to damage done due to seismic excitation is largely useless if it is only accessible in a time period equivalent to that over which visual inspection shows that the bridge has collapsed. If one assumes that a large percentage of available data is valid, then the benefits are still substantially curtailed if a preliminary assessment and characterization of response cannot be completed almost in real time.

Unfortunately, the greatest challenge is that most SHM systems are designed merely to collect data, rather than to provide a means for its efficient management and interpretation. It is critical that the system provide a means not just of recording (and displaying) response but also (and more importantly) of characterizing the response and comparing it to an appropriately updated model to enable assessment of the critical aspects of capacity and service life. Thus, overall, the major principles for design initiate with the development of an appropriate plan for SHM, and continue through the selection and placement of the sensors, collection and transmission of data, checking of its validity, and finally to the actual use of the data in a manner required by the owner/operator of the structure. In reality, the primary aspect of design of such systems is in moving past the mere attractiveness of a sensor network to the actual development and implementation of a system capable of serving as a true tool for health monitoring—i.e., not just being able to state that the "patient" is sick, but rather of being able to pinpoint the location and reason, as well as the effect of the incapacity.

4 SUMMARY

Although the field of SHM as applied to civil structures is still in its infancy, it is already demonstrating significant advantages not only in the assurance of integrity of built infrastructure and its predicted serviceability but also in paving a new path for both a better understanding of structural response and for the development of design codes. It is expected that the further developments of SHM systems will result in the establishment of a comprehensive methodology for autonomous health monitoring of structural systems to the point where true condition-based physical inspection and monitoring would become a reality. The integration of sensing networks with the development of robust and effective tools for real-time data interpretation and use in developing measures of performance (such as capacity) and remaining service life provide useful tools for inspection and assessment of the health of our infrastructure. The integration of damage identification

and finite-element-based tools, as being currently developed, can further provide assistance to the engineer in assessing health immediately rather than having to resort to expensive closures while assessments are made off-line. While attention needs to be paid to the relevant principles associated with the design of such systems, it is also useful to consider additional advantages that would conceivably alter the design in minor ways. While the field of structural response is a mature one there are still significant ambiguities in our understanding of long-term response, especially as related to dynamics of assemblies and systems under combined natural forces. The use of an appropriately designed SHM system would enable further understanding of response through data analysis and interrogation, which would lead to better and more refined methods of structural design. In addition, the use of real-time data enables immediate updating of risk and resource allocation, thereby linking design, construction, and service-life maintenance together and enabling the development of a new paradigm for future development of design codes and specifications—one not based on the use of inordinately high factors of safety due to uncertainty, but one based on a continuous assessment and monitoring of risk and health. This would reenvision civil structural design and maintenance, leading to the development of a modern field of civil infrastructure systems.

REFERENCES

[1] Melchers RE. *Structural Reliability and Analysis Prediction*. Ellis Horwood, 1987.

[2] National Bridge Inspection Standards. *Code of Regulations*, U.S. Government Printing Office, 1998.

[3] Moore M, Rolander, D, Graybeal, B, Phares, B and Washer, G. *Highway Bridge Inspection: State-of-the-Practice Study*, Federal Highway Administration, Washington, Dc, FHWA-RD-01–033, 2001.

[4] Sikorsky, C. Development of a health monitoring system for civil structures using a Level IV nondestructive damage evaluation method. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. Stanford University, Palo Alto, 1999 pp. 68–81.

[5] Housner GW, *et al.* Structural control: past, present, and future. *ASCE Journal of Engineering Mechanics* 1997 **123**(9):897–971.

[6] State of California Strong Motion Instrumentation Program, http://www.consrv.ca.gov/cgs/smip/Pages/index.aspx, 2007.

[7] Rytter, A. *Vibrational based Inspection of Civil Engineering Structures*, Ph.D. Thesis. Department of Building Technology and Structural Engineering, Aalborg University: Aalborg, 1993.

[8] Guan H, Karbhari VM, Sikorsky C. Web-based structural health monitoring of a FRP composite bridge. *Computer-Aided Civil and Infrastructure Engineering* 2006 **21**:39–56.

[9] Mufti, A., *Guidelines for Structural Health Monitoring*, ISIS Canada Research Network Design Manual No. 2, ISIS Canada, Winnipeg, Manitoba, 2001.

[10] Ibarguengoytia PH, Sucar LE, Vadera S. Real time intelligent sensor validation. *IEEE Transactions of Power Systems* 2001 **16**(4):770–775.

[11] Andrew JF. Integrity-based self-validation test scheduling. *IEEE Transactions on Reliability* 2003 **52**(2):162–167.

Chapter 89
Maintenance Principles for Civil Structures

Dan M. Frangopol[1] and Thomas B. Messervey[2]

[1] *Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA*
[2] *Department of Mathematical Sciences, United States Military Academy, West Point, NY, USA*

1 INTRODUCTION

1.1 Design and management considerations for the incorporation of monitoring

This chapter explores the rapidly evolving field related to incorporating information obtained through structural health monitoring (SHM) into the design and management of civil structures. Herein, the employment of monitoring technologies is investigated with respect to the following topics:

- development of a strategic, top-down monitoring approach;
- collecting the most critical information, at the right time, with the right instrument;
- probabilistic formulation of the analysis;
- conducting structural assessment;
- updating with monitoring information;
- life-cycle costing; and
- inclusion of risk.

In particular, this area of study is somewhat complex and challenging, but is timely, needed, and offers great potential to improve current practice. Emphasizing its youth, references on maintenance principles themselves (in the absence of SHM) are relatively sparse and graduate courses on service-life design or life-cycle maintenance management (LCM) are rarely part of engineering curriculua. For the interested reader, references [1, 2] serve as an excellent resource to get started. Although new, the field is one of significant activity across research centers, professional publications, associations, and conferences worldwide. Several examples of research centers at major universities include ATLSS (Advanced Technology for Large Structural

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Systems at Lehigh University), LIST (Laboratory for Intelligent Structure Technology at the University of Michigan), CIMSS (Center for Intelligent Material Systems at Virginia Tech), SSTL (Smart Structures Technology Laboratory at the University of Illinois at Urbana-Champaign), CIBrE (Center for Innovative Bridge Engineering at the University of Delaware), SFB477 (Collaborative Research Center on Life-Cycle Assessment of Structures via Innovative Monitoring at the Technical University of Braunschweig, Braunschweig, Germany), SFB398 (Cooperative Research Center for Lifetime Oriented Design Concepts at the Ruhr-University Bochum, Germany), ISIS (Intelligent Sensing for Innovative Structures centered at the University of Manitoba, Canada), SISTeC (Smart Infra-Structure Technology Center at Kaist, Korea), EMPA (Swiss Institute of Materials Science and Technology, Dubendorf, Switzerland), and several others. Some examples of professional journals include *Structural Control and Health Monitoring* [3] edited by Lucia Faravelli, *Structure and Infrastructure Engineering* [4] edited by Dan M. Frangopol, and *Structural Health Monitoring* [5] edited by Fu-Kuo Chang. Several professional associations dedicated to this topic include IABMAS (International Association for Bridge Maintenance and Safety), ISHMII (International Society for Structural Health Monitoring), IALCCE (International Association for Life-cycle Civil Engineering), and SAMCO (European network of Structural Assessment Monitoring and Control).

Traditional structural design, or purchasing in general, usually focuses on obtaining the least cost solution that fulfills specified requirements. Over the past several decades, efficiencies have been gained through reductions in structural weight as material properties, construction methods, and design software technologies have improved. Although it is implied that a concrete structure will require repair, roofs will need to be replaced, and paints will need to be reapplied, the intended service life of a structure is often left unspecified. In such cases, project bids consider only the initial costs of design and construction and upon completion the structure is turned over to the owner with the absence of a maintenance plan. Maintenance and repair activities are then likely to become an *ad hoc* reaction whenever a defect manifests itself, at which time a maintenance program is developed [1]. Unfortunately, in terms of expense, research in the field of life-cycle management (LCM) has shown that the costs of inspecting, maintaining, and repairing a structure over its useful life span often dwarf those associated with the initial design. This is compounded by the frequent desire to extend the service life of a structure beyond that originally intended. Against this backdrop, monitoring technologies have the potential to improve the design and management of civil infrastructure in several ways: (i) performance-based design can be conducted by recording site-specific conditions such as wind, load demands, or temperature; (ii) inspections can be scheduled on an "as needed" basis driven by structure-specific data when indicated by monitoring data; (iii) the accuracy of structural assessments can be improved by analyzing recorded structural response data; (iv) as a result of more accurate information provided as input to analytical models, maintenance, repair, and replacement activities can be optimally scheduled, which results in cost savings; and (v) performance thresholds can be established to provide warning when prescribed limits are violated. However, these benefits also come with an associated life-cycle cost as monitoring systems must be purchased, installed, and maintained and their information processed and assessed. As a result, a truly optimal and efficient design needs to consider and evaluate the costs and benefits of different strategies and approaches. An oversimplified example highlights some of these considerations and issues.

Example 1 A tenant in a second floor apartment complex requests permission to construct a large wood balcony and agrees to share costs evenly with the building owner. Two estimates that meet local design requirements are provided by construction companies qualified to do the work. Company A bids \$13 000 and states that the balcony will need replacement every 20 years. Company B bids \$15 000 and states that the balcony will need replacement every 30 years by using higher quality wood. A safety inspection is required at 10-year intervals (excluding the years at which the balcony is replaced) at a cost of \$500 for each inspection. The apartment has an intended service life of 60 years at which time a major rehabilitation is planned. Assuming good inspection results, which company should get the work?

Answer As a temporary user, the tenant can immediately select the lower initial cost solution, Company A. As the long-term owner, the apartment manager must do more work and consider the life-cycle costs of each solution. Undiscounted, hiring Company A appears to cost \$40 500 with two rebuilds and three inspections and hiring Company B appears to cost \$32 000 with one rebuild and four inspections. However, the timing of the payments is important. In order to make a valid comparison, all costs must be converted to the net present value of money using

$$NPV = \frac{FV}{(1+r)^n} \tag{1}$$

where NPV is the net present value, FV is the future value, r is the discount rate, and n is the year in which payment occurs. Assuming a discount rate of 4% results in:

Company A

$$NPV = 13\,000 + \frac{13\,000}{(1+0.04)^{20}} + \frac{13\,000}{(1+0.04)^{40}} + \frac{500}{(1+0.04)^{10}} + \frac{500}{(1+0.04)^{30}} + \frac{500}{(1+0.04)^{50}} = 22\,203 \tag{2}$$

where the NPV is in dollars.

Company B

$$NPV = 15\,000 + \frac{15\,000}{(1+0.04)^{30}} + \frac{500}{(1+0.04)^{10}} + \frac{500}{(1+0.04)^{20}} + \frac{500}{(1+0.04)^{40}} + \frac{500}{(1+0.04)^{50}} = 20\,365 \tag{3}$$

where the NPV is in dollars.

In this analysis, it appears that Company B is the best choice. However, it must be noted that the result is extremely sensitive to the discount rate. The owner may wish to consider the impact of different discount rates or the disruption caused by multiple rebuilds in his/her decision.

Example 2 The apartment manager in the above example selects Company B and learns of some wood preservation strategies. A sealer is recommended with a cost of material and labor of \$750. The company suggesting this sealer claims this product will extend the service life of the wood well beyond 60 years as long as the sealer is kept in good condition. The sealer is expected to last from 5 years to 10 years at which point sanding and reapplication is recommended. The purpose of the sealer is primarily to keep water out of the wood. If the moisture content of the wood is kept below 22%, wood-destroying fungi cannot grow. Separately, the manager learns of some monitoring strategies for the deck. Small disks that turn color if the moisture content of the wood rises above 20% are available for a cost of \$50 each. These disks would be affixed to the wood underneath the sealer coating. Strain alarms are also available that indicate if a member surpasses a critical strain threshold. These alarms cost \$100 each and can be used to determine if the deck has been overloaded or if members are no longer distributing loads as originally intended. These monitoring actions could potentially replace the 10-year safety inspection requirement. What purchases should the manager make?

Answer Clearly, this becomes a more complicated analysis and the answer depends on how the structure and the different products perform. Although there is not enough information to completely answer this question yet, some significant insights can be obtained. For the sealer, a net present value calculation (as previously indicated) yields a present value cost of \$19 622 if the sealer is replaced every 5 years and a present value cost of \$17 974 if replaced every 10 years. The preventive maintenance strategy is cost effective and the sealer should be purchased. The moisture content disks have the potential to indicate when to reapply the sealer and thus optimize its usage. However, the number of disks utilized and the future performance of the sealer will determine if they are cost effective. To determine the quantity of disks required, spatial effects must be considered (i.e., where to place the disks on the structure). False positive readings are possible as are false negative readings. Further investigation requires a statistical analysis. The use of strain alarms also provides unique opportunities and challenges. Use of such gauges would require the selection of critical structural members and a determination of how many of these members to monitor. Although these

gauges could alter or replace the 10-year inspection requirement, their most significant contribution comes from an increased level of safety (and corresponding decrease in the probability of failure) through the knowledge that the structure has operated within specified limits. The owner may be interested in such an approach specifically to reduce the likelihood of a lawsuit associated with a structural collapse. In order to account for the utility (cost–benefit) of the strain alarms, risk-based decision making must be employed or the cost of failure incorporated into a life-cycle cost analysis.

These two examples highlight some of the differences between a traditional design approach, which only accounts for the initial cost of a structure, and a life-cycle approach which accounts for all costs over the useful life of the structure. At some point, for this second-story balcony, the apartment manager would likely decide that this particular type of structure is not critical enough to warrant a full-scale monitoring solution. Although increased information enables better decision making, increased information also requires that more decisions be made, as well as the creation of systems dedicated to the collection and processing of the information itself. In contrast, for other types of structures such as a large bridge, dam, or building, this type of approach may be fully justified and necessary.

1.2 Importance and timeliness

Treatment of the inclusion of SHM into LCM must address the question of "why now?" This question can be best answered in two parts: (i) because we can and (ii) because we must. We can, because technological development of sensors, power generation, more efficient batteries, more powerful computational platforms, and wireless capabilities are making it feasible to obtain site-specific response data cost effectively. Although many of these technologies and ideas have existed for some time, especially in the aircraft industry, they have typically required a controlled environment, hard wired cables, and immense effort to obtain data. Although reasonable for the design and testing of new aircraft, application for the assessment of civil structures in the field has been infeasible. We must address this challenge now because of the current and impending condition of existing civil infrastructure, which warrants further explanation.

Sustainable economic growth, productivity, and the well being of a nation are intimately linked to the reliability and durability of civil structures such as buildings, bridges, dams, and transportation networks [6]. To this end, comparisons across countries in varying stages of development can be used to show that Gross Domestic Product (GDP), life expectancy, and infrastructure development are highly correlated. As a result, society relies on its engineers and government to design, maintain, and regulate structures that are safe and perform as intended over their service lives. In terms of magnitude, new civil engineering construction is the largest industry in the world, representing approximately 10% of annual GDP. Of this 10% of GDP spending, an estimated 5–10% is the result of the failure (not necessarily collapse) of existing structures [1]. For most countries, existing structures and civil infrastructure are their most valuable assets and their upkeep represents one of their most significant investments. Unfortunately, these assets are deteriorating at an alarming rate due to overuse, overloading, aging, or damage [7].

Highway bridges provide an excellent example of the problem. Many of the bridges constructed in the United States as part of the Eisenhower Interstate expansion in the 1950s, 1960s, and 1970s are approaching the end of their planned service lives. Bridges, in particular, are vulnerable to and are constantly subjected to aggressive environments, which include chemical attack from de-icing salts, environmental stressors such as wind, temperature, and water, as well as continuously increasing traffic volumes and heavier truck loads [8]. It is undesirable, if not impossible, to replace these bridges simultaneously without severely impacting commerce and productivity. It is estimated that hundreds of billion ton-miles of goods and materials along with several trillion passage-miles are transported on highway networks in the United States every year [9] and that this connectivity has made significant contribution to the Nation's economy and quality of life [10].

The deterioration of highway bridges in North America, Europe, and Japan is well documented and publicized. In the United States, 25.8% of the 596 808 existing bridges were structurally deficient or functionally obsolete as of the end of 2006 [11].

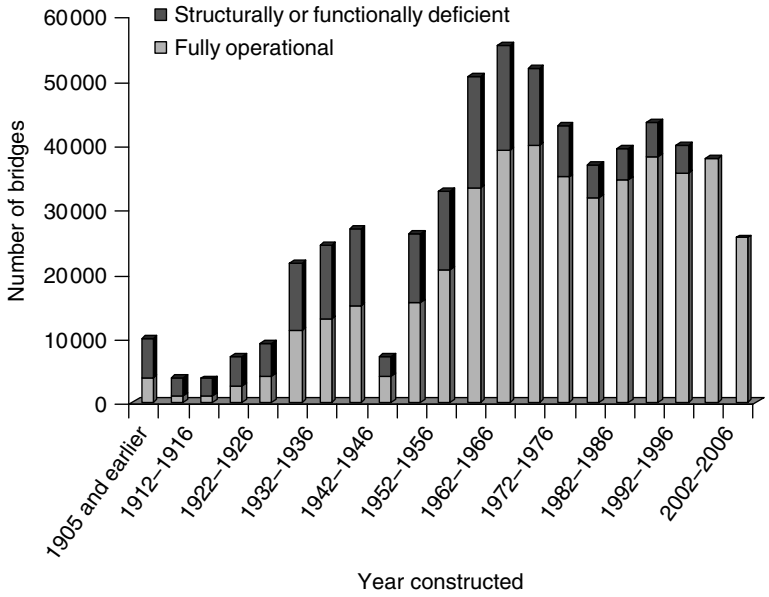


Figure 1. Quantity and classification of US bridges by year constructed. [Adapted from National Bridge Inventory Statistics.]

Figure 1 shows the quantity and classification of US bridges by the year constructed. Aside from the current number of deficient bridges, this figure shows the trend that, as bridges become older, a larger percentage become deficient. Because the majority of US bridges were constructed during the 1950–1980s, we are approaching what some describe as the age of "mass maintenance." Similar trends and results have been reported in Europe and in Asia [12, 13]. To resolve this problem in the United States, an estimated annual investment of $9.4 billion for the next 20 years is required [14]. Because so many structures need maintenance, repair, or replacement simultaneously, there is a great need for methods and technologies that accurately assess these structures, prioritize repairs, and enable the efficient allocation of funds.

2 WHAT TO MONITOR: DEVELOPING MONITORING STRATEGIES

The intent of this section is to offer some ideas for the formulation of monitoring strategies. Here, a top-down approach is developed, which is in sharp contrast to how SHM is typically utilized today. Currently, monitoring is most often used as a bottom-up diagnostic tool in response to an existing problem or defect. Equipment is brought to the site, measurements are recorded, the equipment is removed, and the data is studied. In time, as technologies, metrics, and methods are developed that are convincingly cost effective, the use of permanent (or systematic) monitoring systems will become more common. The formulation of a monitoring strategy should consider (i) historical failures and current assessment of the type of structure of interest; (ii) how the structure fits within a larger network; (iii) the type of measurement desired (global vs. local) and what sensing mechanisms are most appropriate; and (iv) what type of analysis is being conducted and what monitoring data will best improve the analysis. Across these considerations, cost effectiveness is imperative. Because maintenance demands will likely outpace available resources for the foreseeable future, infrastructure managers will likely not invest in monitoring unless it either becomes code driven or there is a return on their investment. Otherwise, money spent on monitoring will simply reduce the funds available for maintenance and repair.

2.1 Past failures and condition of existing structures

Historically, albeit unfortunately, structural failures and collapses have acted as the catalysts that have shaped design codes, construction methods, and management practices. Several notable studies have been conducted in this area and serve as an excellent resource. Matousek and Schneider [15] studied 800 reported failures and errors in the field of structural engineering across several classes of structures. Stewart and Melchers [16] summarized parts of a number of studies involving structural failures. Blind [17] analyzed initiating events and causes for dam failures. Bertrand and Escoffier [18] studied the failures of offshore structures; Anderson and Misund [19] studied the initiating events for the failures of pipelines; and Scott and Gallaher [20] studied the failure of components and systems in nuclear power plants. Perhaps, the most recent and applicable study to one of the most pressing needs today is a 2004 analysis of the reasons for reconstruction across 1691 bridges in Japan [21]. The results of this study are shown in Figure 2.

From this study, it might be concluded that monitoring strategies for concrete may be of particular interest for bridge managers as slab failure and concrete spalling/cracking accounted for most of the superstructure failures.

Although past failures certainly provide insight, the current condition and classification of existing structures must also be considered when developing a monitoring strategy. Maintained by the Federal Highway Administration (FHWA), the National Bridge Inventory (NBI) is a database of statistics for bridges in the United States [11]. The database has been created from construction records and over 40 years of bridge inspections, and provides bridge type, classification, location, age, and current condition. Statistics are also available that detail replacement, rehabilitation, and new construction projects as part of the Highway Bridge Replacement and Rehabilitation Program (HBRRP). Combining this data better enables the design of monitoring approaches for assessing existing structures, as well as those newly constructed. Using information from the NBI as of 2006, Figure 3 details (a) the makeup of the NBI by bridge type, (b) the makeup of deficient bridges by type, (c) the makeup of newly constructed bridges (2003 and 2004), and (d) the makeup of bridge replacement and rehabilitation projects (2003 and 2004).

From these statistics one can conclude that (i) most of the existing bridges in the United States are concrete; (ii) steel bridges represent the largest proportion of deficient bridges; (iii) most new construction is concrete, and (iv) most of the rehabilitation projects are of steel. Because steel bridges are generally older, it is not surprising that these bridges are disproportionately deficient. Additionally, since steel bridges make up the bulk of rehabilitation projects, it is reasonable to assume that many of these older bridges are located in urban areas where new construction is difficult without significantly disrupting traffic flow. From these statistics, if monitoring is viewed from the perspective of forming national priorities, effort should focus on concrete SHM for new construction and upon steel SHM for structures being assessed.

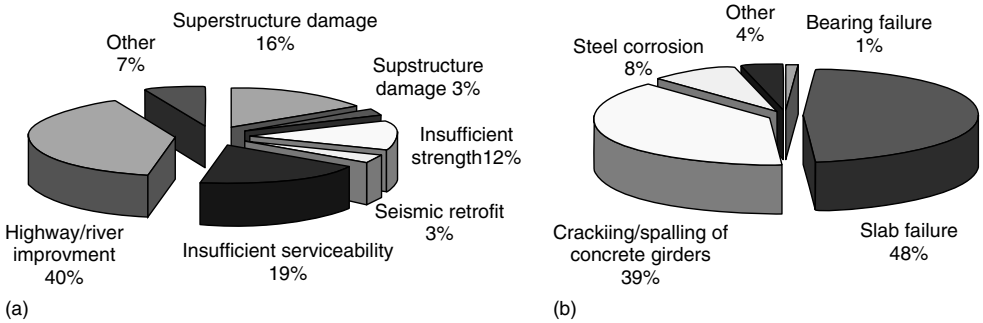


Figure 2. Study of 1691 recent bridge reconstruction projects in Japan: (a) reasons for reconstruction for all bridges and (b) for the bridges reconstructed due to superstructure damage, the primary cause of that damage. [Adapted from Ref. 21.]

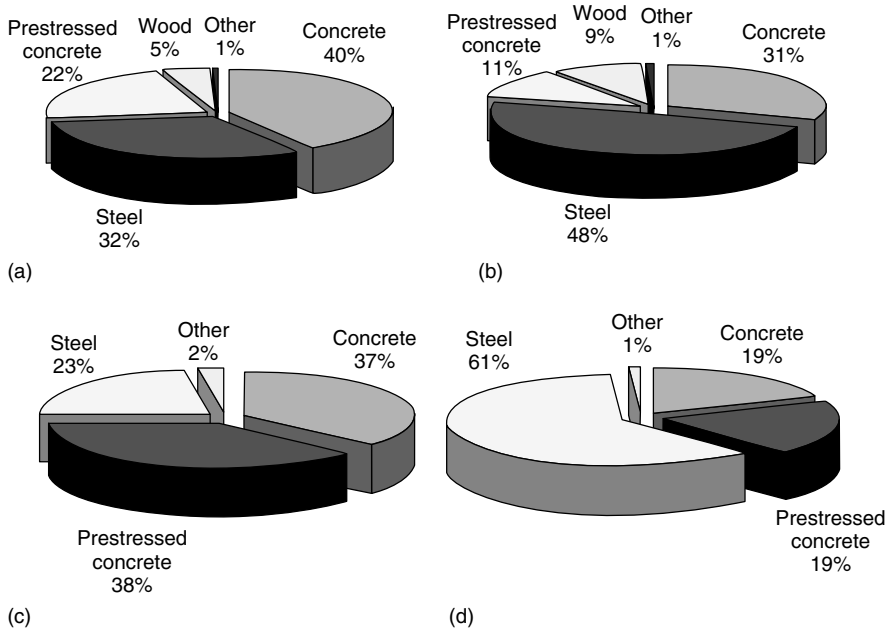


Figure 3. National Bridge Inventory (NBI) statistics of interest for developing national level monitoring priorities. (a) NBI bridge makeup, (b) deficient bridges by type, (c) newly constructed bridges (2003–2004), and (d) replaced and rehabilitated bridges (2003–2004). [Adapted from Ref. 11.]

2.2 Consideration of structures within a network

Rarely is the management of a structure considered in isolation. Whenever possible, inspections, assessments, and maintenance actions should be taken in context of where the allocated resources will provide the most benefit.

Example 3 A bridge manager of a network of 14 bridges (see Figure 4 and Figure 1 in Akgul and Frangopol [22]) located near Denver, Colorado, has $100 000 available for SHM technologies and desires to make the best use of these funds. Where should monitoring effort be focused and on what should money be spent?

Answer Clearly, not enough information is provided, but immediately, several questions and competing interests become apparent. Bridges within the network are of different types, ages, spans, and traffic characteristics. Most of the bridges are concrete. Since chloride ingress into concrete is best detected by embedded sensors installed during construction, the analysis is most likely limited to cracking of concrete, steel fatigue, or an assessment of load versus capacity through strain measurements, unless concrete decks are being resurfaced, in which case, embedded sensors might be a good investment. Noticeably absent from the provided information is any information concerning the current condition and/or safety classification of each bridge. Age directly correlates to the amount of deterioration and the average daily truck volume is the primary cause of fatigue through the number of accumulated strain cycles imparted upon the structure. Here, the bridge with the highest average daily truck traffic (ADTT) is also one of the newest bridges showing the necessity to consider age and ADTT in context. Even if condition and safety assessments were provided, the structure with the worst levels may not warrant monitoring priority if other bridges are more important to the overall performance of the network. Metrics to make such an assessment could range from the extra travel distance required if a particular bridge became non-operational, additional travel time associated with a loss of functionality (lane reduction), the highest cost of failure or consequence in the event of collapse, political importance, or historical/cultural value.

Bridge characteristics in Colorado highway network

Bridge name	Number of spans	Bridge length (m)	Year built	Average daily truck traffic
Prestressed concrete				
E-16-MU	1	34.1	1994	810
E-16-LA	2	77.9	1983	450
E-16-DM	2	44.5	1990	390
E-16-QI	2	74.1	1995	1335
E-16-LY	3	74.3	1985	1610
E-16-NM	2	64.6	1991	2955
E-16-MW	2	72.7	1987	230
Steel I-beam bridges				
E-16-FK	4	69.2	1951	1370
E-16-FL	4	54.0	1951	765
E-16-QI	5	82.3	1953	890
Steel plate girder bridges				
E-17-LE	4	68.6	1972	992
E-17-HS	4	64.5	1963	5
E-17-HR	4	64.0	1962	306
E-17-HE	4	67.7	1962	1290

* Adapted from [22]

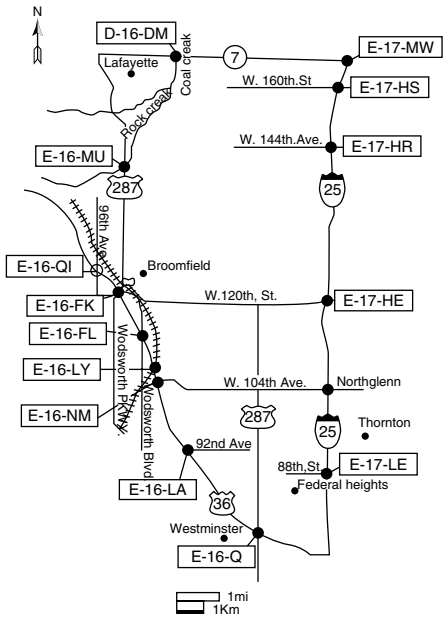


Figure 4. An existing bridge network near Denver, Colorado. [Reproduced with permission from Ref. 22. © American Society of Civil Engineers, 2003.]

Fortunately, some of these issues have already been addressed by researchers developing techniques for the reliability-based maintenance management of bridge networks based upon network theory. Akgul and Frangopol [22] established the network configuration and reliability analysis for the Denver area network depicted in Figure 4. Liu and Frangopol [23] used this network to introduce the bridge reliability importance factor (RIF) as

$$RIF_i = \frac{\partial \beta_{\text{net}}}{\partial \beta_{\text{sys},i}} \tag{4}$$

which relates the change in the reliability of the bridge network to the change in reliability of any particular bridge within that network. The bridge for which a change in reliability has the largest impact upon network performance will have the highest RIF. Liu and Frangopol [24] expand this work to include network connectivity and user satisfaction. Then, in Liu and Frangopol [25], 73 possible network maintenance actions are optimized with respect to life-cycle cost and network reliability. With respect to these approaches, the inclusion of SHM is mutually beneficial. SHM benefits the existing models by reducing uncertainties thus making them more accurate. Conversely, by providing metrics that prioritize bridges by importance, these models indicate where monitoring efforts would have the greatest impact.

2.3 Matching monitoring and analysis strategies: right tool for the right task

Monitoring strategies are broadly categorized into two groups, global and local. Both provide different types of information and, in general, support different analysis types. Figure 5 depicts global and local monitoring strategies, the type of information collected, and the associated measurement types.

Selecting an appropriate strategy might be dictated by the structure, type of analysis, or both. (Several common analysis modeling choices can be found in **Free and Forced Vibration Models**; **Civil Infrastructure Load Models for Structural Health Monitoring**; **Modal–Vibration-based Damage Identification**.) For example, one may be limited to a global monitoring approach when accessibility to specific parts of the structure is impossible. Conversely, one may desire global monitoring methods when working

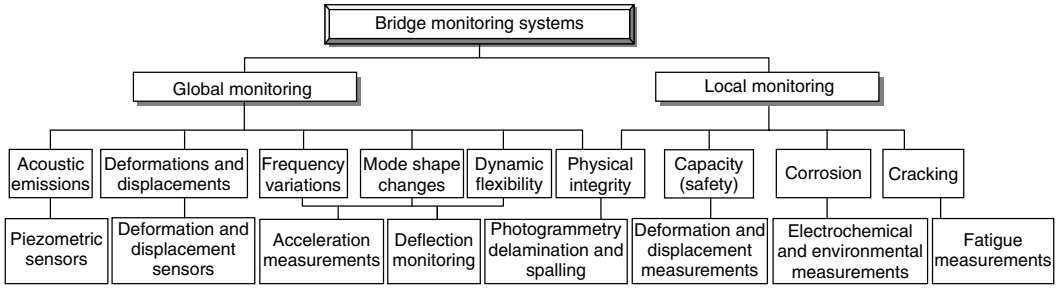


Figure 5. Infrastructure monitoring strategies [Adapted from Ref. 26.]

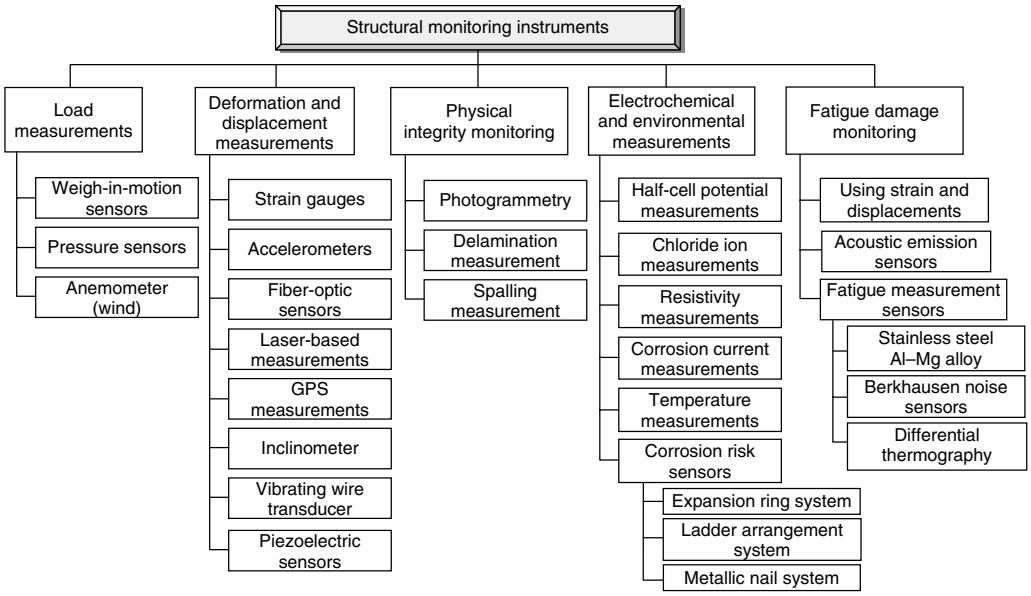


Figure 6. Common structural health monitoring instruments and associated uses. [Adapted from Ref. 26.]

with an analysis that creates an equivalent structure. A common example would be the characterization of the stiffness or modal characteristics in a finite element model. For this, accelerometers would be an appropriate instrument. This would not necessarily be the case when analyzing a specific structural failure mechanism such as flexure, shear, fatigue, or corrosion. In this case, information would be desired about member geometries, material properties, loads being imparted on the structure, and environmental effects. Figure 6 shows common SHM instruments for these types of measurements. A more detailed discussion of several of the most common sensor types can be found in **Fiber Bragg Grating Sensors**; **Microelectromechanical Systems (MEMS)**; **Reliable Use of Fiber-optic Sensors**; **Integrated Sensor Durability and Reliability**.

Example 4 Three options are being considered for the detection of corrosion during the design phase of a new reinforced concrete parking area on the roof of a university lecture hall in a harsh weather environment (Figure 7). The roof replacement is required to remedy existing cracks, leaks, and potential future safety concerns. The approaches under consideration include (i) periodic half-cell potential tests by an inspection team, (ii) embedded corrosion risk sensors 5 cm above rebar level, and (iii) corrosion sensors located on the rebar itself. What advantages and disadvantages should the design firm discuss with

Figure 7. Rooftop parking area being considered for replacement and redesign. [Photo by T.M. Messervey.]

the building owner when the project is still being developed?

Answer Each option implies a different maintenance strategy and different future actions required to be followed by the facility manager. Hence, any such discussion among designer, manager, and owner is positive. Periodic half-cell potential tests, option (i), require scheduling and user disruption during their execution, but are likely the least expensive option and no immediate funding is required. In scheduling such tests, there is the risk to inspect too often, thus allocating more funds than necessary to inspections, and there is also the risk of testing too infrequently and thus not detecting the onset of corrosion as desired. Embedding corrosion sensors above the rebar, option (ii), is intended to detect the deterioration mechanism (presence of chloride ions) before deterioration initiates on the rebar itself and as such is proactive in nature as shown in Figure 8(a). Such an approach lends itself to multiple repairs (resurfacing) of the top concrete layers to preserve the more demanding and expensive replacement of the entire deck. Detecting the presence of corrosion directly on the rebar and corresponding section loss, option (iii), lends itself to the complete replacement of the deck when necessary. Once corrosion has initiated, several models are available to predict the remaining safe life span of the rebar. The intended life span of the parking lot is a critical detail not provided in this scenario. Also left for further investigation are

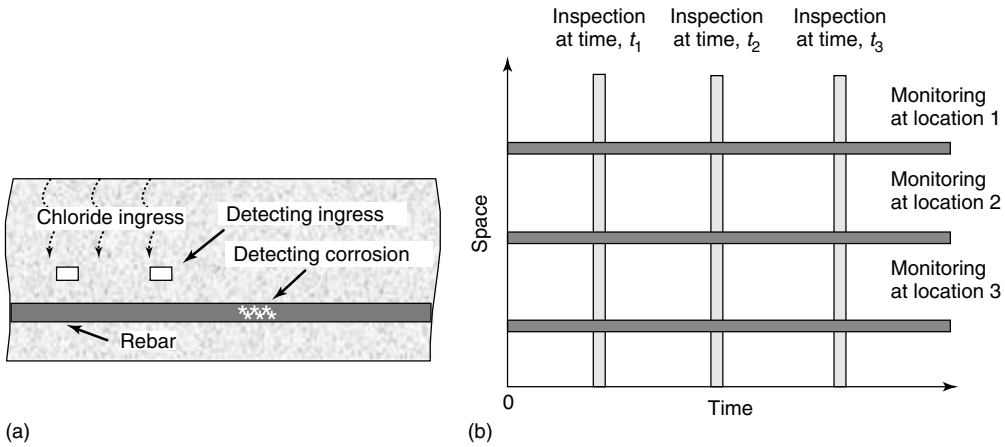


Figure 8. For a reinforced concrete design: (a) detecting the corrosion initiation mechanism versus detecting corrosion itself and (b) spatial and temporal differences between inspections and monitoring (adapted from [26]).

spatial and temporal effects. Generally, inspections or tests provide information about an entire structure at a point in time, whereas monitoring provides information continuously at a specific location over time as depicted in Figure 8(b). To balance these differences, inspections must occur at appropriate intervals of time, and sensors must be placed at the appropriate intervals in space.

2.4 Determining what and when to monitor at the structural level

2.4.1 Aleatory and epistemic uncertainty

Consideration of the uncertainty associated with critical loading and structural parameters is one of the most critical issues in assessing the condition of existing civil infrastructures [27]. The contributors to uncertainty in civil structural systems are discussed by Moon and Aktan [28]. Uncertainty can be considered in two broad categories, aleatory and epistemic. Aleatory uncertainty describes the inherent randomness of phenomenon being observed and cannot be reduced. Natural variations in temperature are an example of this type of uncertainty. Epistemic uncertainty describes the error associated with imperfect models of reality due to insufficient or inaccurate knowledge [29]. Error associated with predicting stresses in a structural member through use of an analytical model is an example of this type of uncertainty, as material properties, geometry, and loads are never deterministic. Both types of uncertainty play an important role in the monitoring of civil infrastructure. System identification, i.e., validating structural parameters through experimental testing, proof loading, or measurements from the structure of interest, all act to reduce epistemic uncertainty by improving the accuracy of model input parameters. Efforts should naturally be focused on the parameters for which better information has the greatest impact on model improvement. In many cases, the recording and inclusion of aleatory uncertainty greatly improves assessment and prediction of structural performance. Including temperature effects is again a good example as temperature significantly contributes to variations in strain or modal characteristics. An example of this approach, using SHM technologies to better estimate condition assessment, is provided in [27]. In this example, temperature effects recorded via SHM are utilized to enhance system identification and an ensuing reliability analysis. In this case, inclusion of this data creates a significant decrease in the component and system reliability indexes.

2.4.2 Structural systems and performance over time

Structures are composed of individual members that form structural subsystems, which interact on a global level to fulfill the desired outcome. How a component functions in a system provides insight on where to focus monitoring priorities [30]. For members in series as shown in Figure 9(a), the critical or weakest member is the one with the highest probability of failure, Member #1. For members in series, the *weakest* member should receive monitoring priority. In contrast, if these same three members are arranged in parallel, as shown in Figure 9(b), then the critical member becomes the member with the lowest probability of failure p_f, i.e., the *strongest* member, Member #3. As such, for members arranged in parallel, the *strongest* member (i.e., the member with the lowest probability of failure) should receive monitoring priority.

Varying rates of deterioration may also affect monitoring priorities or when monitoring is appropriate. Figure 10(a) depicts the reliability profiles of two members arranged in series. Member #1 deteriorates more rapidly than Member #2. As such, monitoring priority would first be given to Member #2 until the reliability indexes intersect at point X after which priority would shift to Member #1. In contrast, if these same two members are arranged in parallel, as shown in Figure 10(b), monitoring priority would first be given to Member #1 and then to Member #2 after the intersection of the reliability profiles. In both cases, the concept of monitoring the weakest member

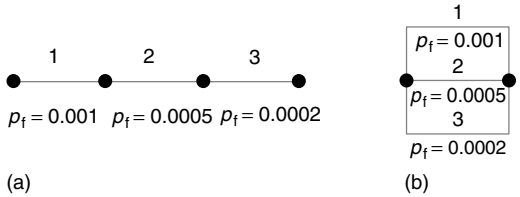


Figure 9. System analysis of members in (a) series and in (b) parallel where p_f = probability of failure.

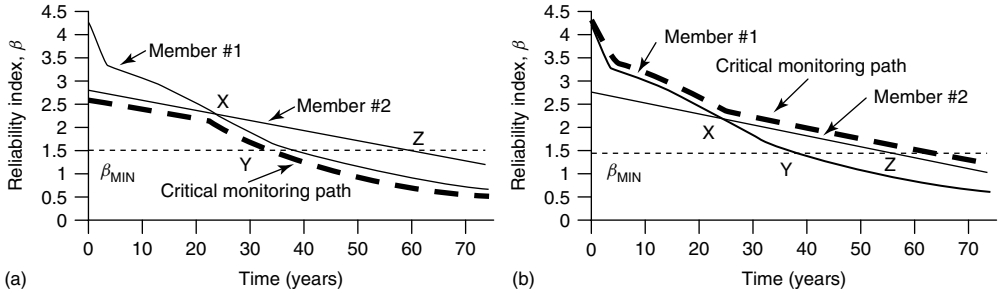


Figure 10. Time-variant system analysis of members in (a) series and (b) parallel.

in series and the strongest member in parallel remain the same, but the critical member changes over time because of varying deterioration rates. Although fairly intuitive for different structural components or components made of different materials, this also finds application among like members with different exposure levels. Examples include exterior steel girders being subjected to a higher corrosion rate than interior girders on a bridge because of a greater exposure to de-icing salts, or exterior versus interior beams and columns in a building.

Such an analysis could also be used to answer the question of when to monitor. Again using the graphs in Figure 10, if a minimum reliability threshold is established, one could conclude that monitoring would be appropriate on Member #1 at the time corresponding to point Y in series and on Member #2 at point Z in parallel given perfect information. Monte Carlo simulation of the model parameters can be used to estimate the earliest possible crossing of the minimum reliability threshold. This would be appropriate for a monitoring system with high operational costs that can be turned on or off, or for a nonpermanent monitoring solution that must be scheduled and brought to the site.

3 INCORPORATING MONITORING DATA INTO DESIGN AND MANAGEMENT

The role of design is to create structures that fulfill their intended purpose over a specified time horizon with the appropriate level of safety in the face of uncertainty. Unfortunately, no structure is perfectly safe, due to the possibility of unforeseen or catastrophic events such as earthquake, fire, flood, or terror. Even if a perfectly safe structure were possible, it would be cost prohibitive. Hence, engineers have the responsibility to reach an optimal balance between safety and cost. Design codes and design methodologies are intended objectively to quantify this balance. As design methods have improved and evolved over time, the role and treatment of uncertainty has changed as well. Allowable stress design (deterministic) accounts for uncertainty through the use of a Factor of Safety. This factor has traditionally come from expert opinion gained over time. In Load and Resistance Factor Design (semi-probabilistic), reliability methods are utilized to calibrate factors that increase loads and reduce resistances. Once these factors are established, the treatment of an individual problem (i.e., the calculations) becomes essentially deterministic in nature. In performance-based design, the engineer is responsible for the appropriate determination and treatment of the uncertainties associated with both loads and resistances in a probabilistic manner. This requires treating loads and resistances as random variables and the characterization of their probability distributions. Performance-based design provides the most flexibility, especially to incorporate newly developed materials, but also places the most responsibility on the designer [31]. Although structural monitoring could be incorporated into any of the design approaches, it is best-suited for a performance-based approach. Because monitoring provides statistical data about a particular structure or the environmental conditions in which it exists, it naturally lends itself to a probabilistic approach for both the design and management of structures. Design considerations for the employment of monitoring can be cross references in **Design Principles for Civil Structures**.

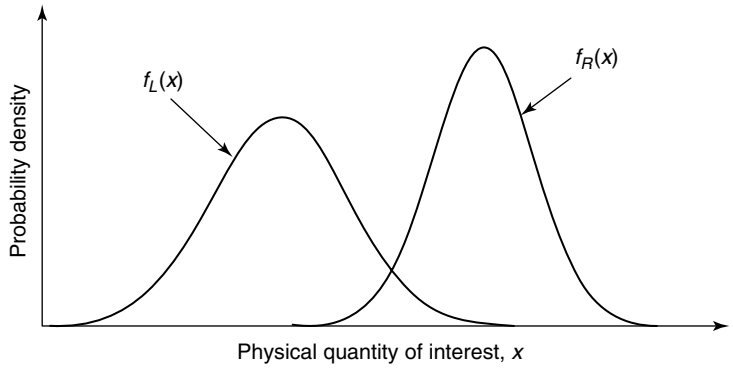


Figure 11. Probability density functions of resistance and load effect.

3.1 Reliability-based assessment of an existing structure using structural health monitoring

Structural reliability is utilized to conduct a probabilistic assessment of the performance function $g = R - L$, where R and L are the resistance and load effect, respectively. Provided that the capacity, R, and load effect, L, are random and can be quantified, the probability of safe performance, p_s, can be expressed as

$$p_s = P(R - L > 0) = \iint_{R>L} f_{R,L}(r, l)\, \mathrm{d}r\, \mathrm{d}l \quad (5)$$

where $f_R(r)$ and $f_L(l)$ are the probability density functions (PDFs) of R and L are defined in Figure 11, and $f_{R,L}(r, l)$ is their joint PDF. Most often, the capacity R and demand L are themselves functions of many other random variables. In such cases, a limit state function $g(\boldsymbol{X}) = 0$, describes the performance of the system in terms of the vector of basic random variables, $\boldsymbol{X}$, and defines the failure boundary, which separates a survival region from a failure region. Formulation of the probability of safe performance p_s then becomes

$$p_s = \int_{g(\boldsymbol{X})>0} f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}x \quad (6)$$

which represents the volume integral of the PDF of $\boldsymbol{X}$, $f_{\boldsymbol{X}}(\boldsymbol{x})$, over the safe region $g(\boldsymbol{X}) > 0$. The solution of this integral can quickly become too complex to solve in closed form or using numerical methods, and is typically solved using approximation methods (FORM/SORM) or by using Monte Carlo simulation [32–34]. Commercially available software for such calculations include STRUREL [35], CALREL [36], PROBAN [37], and RELSYS [38].

Typically, the probability of safe performance is a number close to 1, such as $p_s = 0.9772$ which can also be reported in terms of the probability of failure, $p_f = 0.0228$. To make this easier to express, the probability of failure is usually reported in terms of the reliability index, β. In the special case where R and L are independent and normally distributed, the reliability index can be related to the probability of failure as $p_f = \Phi(-\beta)$, where

$$\beta = \frac{\mu_R - \mu_L}{\sqrt{\sigma_R^2 + \sigma_L^2}} \quad (7)$$

where μ_R and μ_L are the mean values of the resistance and load effect respectively, and σ_R and σ_L are the standard deviations. For a probability of failure, $p_f = 0.0228$, the corresponding reliability index is $\beta = 2$, which is sometimes accepted as the lowest acceptable safety level when assessing structures.

Example 5 The Lehigh River Bridge SR-33 was constructed in 2001. During construction, strain gauges were affixed at several key locations as shown in Figure 12 to monitor the strain (i) after erection of the steel truss, (ii) after the emplacement of a concrete deck, and (iii) during the load test of a 326-kN (74 100 lb) design truck. The bridge is constructed of M270 grade 50 W steel with a nominal yield strength of 345 MPa (50 ksi). Strain readings for the key events are transformed to stress based upon

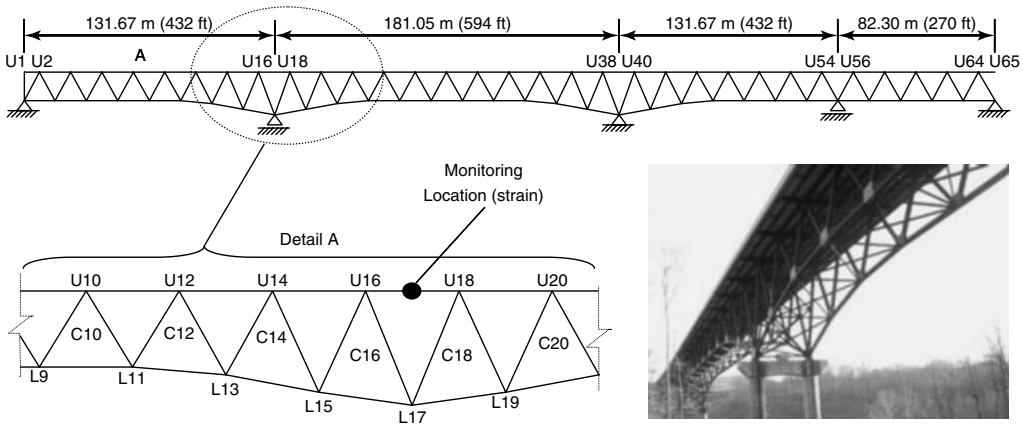


Figure 12. Monitoring of an upper chord on Lehigh River Bridge SR-33. [Adapted from Ref. 39.]

the cross-sectional area of the members as shown in Table 1. Is this member safe?

Answer A deterministic approach to the problem is fairly simple where the Factor of Safety (FS) is calculated as

$$FS = \frac{345 \text{ MPa}}{131.75 \text{ MPa}} = 2.62 \tag{8}$$

Although this approach does quantify a level of safety, incorporating the variability of any of the parameters is not possible. An LRFD approach would incorporate safety through the use of load and resistance factors, but the end result would simply be that the structure was adequate for this particular load with the successful checking of an inequality equation. The best solution is a performance-based approach using structural reliability. For such an analysis, reasonable assumptions must be made to make use of the provided information in a probabilistic manner. Dead loads can vary because of differences in material densities, thicknesses, cover of asphalt, spacing of rebar, etc. Traffic live loads on bridges are typically the most uncertain random variables in a reliability analysis. Quantifying these loads is difficult because they are a function of truck weights, axle spacing, vehicle speeds, and the relative positioning of trucks on the bridge. Ghosn and Moses [40] and Nowak [41, 42] have developed commonly utilized live load models for traffic live loads. Weigh-in-motion (WIM) sensors have the possibility to more accurately capture live load effects through direct measurements. A recent study of over 10 years of WIM data across multiple highways can be found in [43]. Regarding structural capacity, manufacturers typically produce, on average, higher yield strengths than that reported by specification to allow for error in production. A reasonable estimate of the coefficient of variation for the dead load, live load, and yield strength can be taken from Nowak and Yamani's study [42]. Using these coefficients, Table 2 summarizes the descriptors for each random variable. These random variables are also assumed to be normally distributed and independent.

Table 1. Maximum recorded stresses on an upper chord of Lehigh River Bridge SR-33 [39]

	Reason for strain recording	Load type	Total measured strain transformed to stress (MPa)	Incremental stress (MPa)
August 2001	Truss constructed	Dead	61	61
October 2001	Concrete deck emplaced	Dead	109	48
January 2002	Design truck crawl test	Live	131.75	22.75

Table 2. Probabilistic modeling of the provided monitoring data

Variable	Distribution	Descriptors (mean, SD) (MPa)	Coefficient of variation (COV)	Source
Dead load truss	Normal	(61, 3)	0.05	Nowak and Yamani [42]
Dead load concrete	Normal	(48, 2.4)	0.05	Nowak and Yamani [42]
Crawl test live load	Normal	(22.75, 5.91)	0.26	Nowak and Yamani [42]
Capacity: yield strength	Normal	(386, 42.5)	0.11	Nowak and Yamani [42]

Combining the mean values and standard deviations for the load terms yields

$$\mu_L = 61 + 48 + 22.75 = 131.75 \text{ MPa} \quad (9)$$

and

$$\sigma_L = \sqrt{3^2 + 2.4^2 + 5.91^2} = 7.05 \text{ MPa} \quad (10)$$

and solving for the reliability index, β

$$\beta = \frac{\mu_R - \mu_L}{e_s\sqrt{\sigma_R^2 + \sigma_L^2}} = \frac{386 - 131.75}{1.04\sqrt{42.5^2 + 7.05^2}} = 5.67 \quad (11)$$

which corresponds to a probability of failure $p_f = 0.000000007$. In this calculation, a sensor error, e_s, of 4% is included.

It is important to note that this example illustrated only one metric of safety, which best answers the question of *how safe was the structure for the load that acted upon it?* This type of analysis of a crawl test (or proof load) is extremely useful to quantify the benefit or retrofit of repairs through comparison with a previous study [44–46]. Similarly, changes in load distribution or structural performance can be observed. However, this method does not completely address the question of structural safety because time is not taken into account. Using probabilistic modeling, the longer a structure is in service, the more likely that a load from the upper tail of the distribution will occur. Typically, this is taken into account by designing or assessing for the 75-year live load using the statistics of extremes [47].

3.2 Using the statistics of extremes to determine live loading

The design and analysis of civil infrastructure is often concerned with the largest or smallest (extreme values) of a number of random variables. Buildings must withstand maximum wind loads, dams maximum flood levels, and bridges maximum live loads for a given time period [29]. With respect to monitoring, this concept is very useful because it allows the selection of what data to keep. In permanent monitoring approaches, the magnitude of a continuous data stream across hundreds of sensors becomes a management problem in itself. Selecting, logging, and maintaining peak values is one way to efficiently manage data [48].

Extreme values are the largest (or smallest) values from a set of n samples from a known distribution. For example, let X be daily recordings of peak wind velocities. Y_n is then defined as the maximum value from a specified period of time, or number of observations, n. Peak annual wind velocities would come from a set of $n = 365$ samples.

$$Y_n = \max(X_1, X_2, X_3, \ldots, X_n) \quad (12)$$

As the process repeats itself each year, a different peak annual wind velocity is recorded and in turn is itself a random variable. If the underlying distribution has an exponentially decaying upper tail, then the cumulative distribution function (CDF) and probability density function (PDF) of the distribution of the extreme, respectively, are [49]

$$F_{Y_n}(y) = [F_X(y)]^n \quad (13)$$

$$f_{Y_n}(y) = n[F_X(y)]^{n-1} f_X(y) \quad (14)$$

which is to say that the final distribution of the extreme values is a function only of the initial

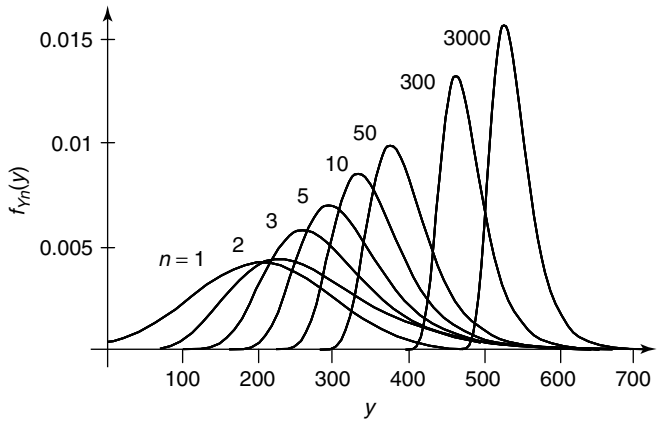


Figure 13. Transformation of a normal distribution to the Type I asymptotic form (Gumbel) for $n = 1$, 2, 3, 5, 10, 50, 300, and 3000. [Reproduced with permission from Ref. 48. © Taylor and Francis, 2007.]

distribution and the sample of size n. This implies that if an extreme phenomenon can be defined in any space (sample size, n, or timeframe), it can be transformed to another desired space. In a physical context, if peak annual wind velocities can be well-defined (given enough yearly observations), then a simple transformation yields the distribution of the 50-year peak wind velocity appropriate for design. Figure 13 depicts the transformation of a normal distribution into the Type I asymptotic (Gumbel) distribution. Each PDF represents the distribution of the extreme value Y_n for the associated sample size.

Using the statistics of extremes can prove useful to monitoring in two ways: (i) once a live load distribution is defined, it can be transformed to an extreme distribution given the number of times sampled and (ii) if extreme values themselves are observed (e.g., maximum daily monitoring values), the extreme value distribution can be defined directly and transformed to the desired time interval.

3.2.1 Live load based upon projected sample size

Given a distribution with an exponentially decaying upper tail, the probability of encountering a value from the extreme tail increases with the number of times the distribution is sampled. Hence, once a sampling rate is established, the distribution of the maximum expected occurrence can be defined for any future time t.

Example 6 The Lehigh River Bridge described in the previous example has an estimated daily crossing rate of 300 trucks. What is the safety level with respect to a 75-year live load requirement?

Answer Assuming a linear relationship between the truck live load and the recorded strain, extreme value statistics can be applied directly to the recorded strain. For a monitored phenomenon with an underlying normal distribution, the distribution for the extreme value for any sample size n is defined by

$$\mu_{Y_n} = \sigma_X \mu_n + \mu_X + \frac{\gamma \sigma_X}{\alpha_n} \tag{15}$$

and

$$\sigma_{Y_n} = \frac{\pi \sigma_X}{\sqrt{6}\alpha_n} \tag{16}$$

where μ_X and σ_X refer to the underlying distribution, μ_n is termed the characteristic value, α_n is a shape factor of the transformed distribution, and $\gamma = 0.5772$ is Euler's number. The shape factor and characteristic value are dependent only on the sample size and are given by

$$\alpha_n = \sqrt{2\ln(n)} \tag{17}$$

and

$$\mu_n = \alpha_n - \frac{\ln[\ln(n)] + \ln(4\pi)}{2\alpha_n} \tag{18}$$

These equations are particular to an underlying normal distribution and are detailed in [49, 50]. Here, for bridge SR-33, with an estimated 300 trucks crossing daily,

$$n = \frac{300 \text{ trucks}}{\text{day}} \times \frac{365 \text{ days}}{\text{year}} \times 75 \text{ years} = 8\,212\,500 \text{ trucks} \quad (19)$$

Applying the above equations first for the shape factor and characteristic value, and then for the mean and standard deviation of the extreme distribution (using $\mu_X = 22.75$ *and* $\sigma_X = 5.91$ from the crawl test) results in

$$\mu_{Y_n} = 53.98 \text{ MPa} \quad (20)$$

$$\sigma_{Y_n} = 1.34 \text{ MPa} \quad (21)$$

for the 75-year live load stress. Combining stresses as before,

$$\mu_L = 61 + 48 + 53.93 = 162.93 \text{ MPa} \quad (22)$$

and

$$\sigma_L = \sqrt{3^2 + 2.4^2 + 1.34^2} = 4.07 \text{ MPa} \quad (23)$$

and solving for an *approximation* of the reliability index, β

$$\beta \approx \frac{\mu_R - \mu_L}{e_s\sqrt{\sigma_R^2 + \sigma_L^2}} = \frac{386 - 162.93}{1.04\sqrt{42.5^2 + 4.07^2}} = 5.02 \quad (24)$$

This is an approximation because error is introduced by the fact that the distribution of the 75-year live load is no longer normal, but Gumbel. This error could be removed through the use of a limit state equation and FORM/SORM analysis.

This approach is an improvement on the previous example with the incorporation of live load effects, which include the consideration of time or multiple truck passages. It finds application in the use of monitoring data obtained by stationary park tests or slow-moving crawl tests. Not yet addressed is the consideration of load combinations. These combinations explore the impact of different loads occurring simultaneously.

3.2.2 Live load based upon in-use monitoring data

Perhaps the best use of monitoring in the assessment of civil infrastructure is the study of long-term data trends. Although this requires some type of continuous monitoring approach, it provides significantly more information. By screening data for peak values, the statistics of extremes can be used to assess live loads. If observed over time, these recordings capture load combinations. Once the extreme distribution is defined in any particular space, it can be transformed to the desired time frame for inclusion in a reliability analysis.

Example 7 An analysis of the performance of connections on the I-39 bridge (Figure 14) spanning the Wisconsin River is desired. Strain gauges are emplaced and maximum daily peak strain values are recorded. Monitoring is conducted for 83 days and strain values are converted to stress via Hooke's law, as reported in Table 3. What is the appropriate live load to assess the structural safety of this connection in a reliability analysis?

Answer Here, sensors were affixed on an existing structure and therefore measurements represent only live loads. Because each reading is the maximum value of hundreds of truck crossings plus environmental effects, it is reasonable to assume that these values follow an extreme value distribution. Type I (Gumbel) and Type II (Lognormal) maximum value distributions are appropriate candidates. Goodness-of-fit tests such as the chi-squared test or the Anderson–Darling test [29] can be used to validate and compare these models. Treating the above values as a univariate data set of extreme values, the mean and standard deviation of these values are calculated as

$$\mu_{Y_n} = 24.67 \text{ MPa} \quad (25)$$

and

$$\sigma_{Y_n} = 4.66 \text{ MPa} \quad (26)$$

For a Type I distribution, the shape factor α_n and characteristic value μ_n can be calculated as [49].

$$\alpha_n = \frac{\pi}{\sqrt{6}\sigma_{Y_n}} = \frac{3.14}{\sqrt{6} \times 4.66} = 0.275 \quad (27)$$

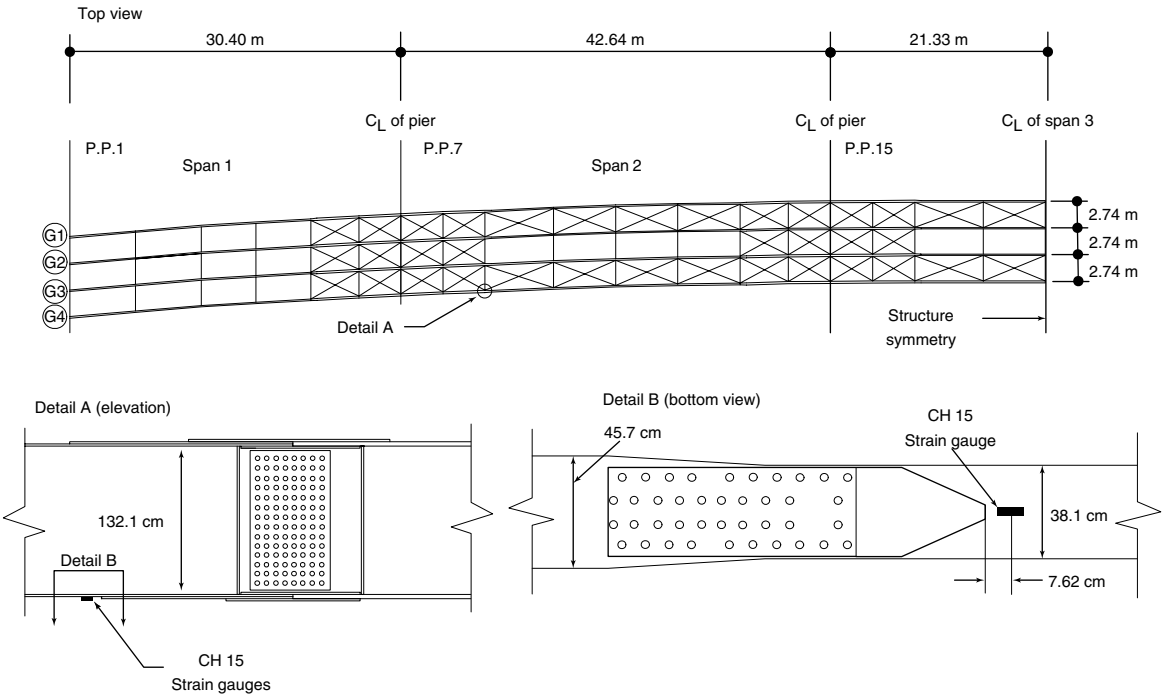


Figure 14. Monitoring of a connection on the I-39 Wisconsin River Bridge. [Adapted from Ref. 51.]

Table 3. Daily recorded peak stresses at monitored connection on the I-39 Wisconsin River Bridge (MPa)

25.27	21.72	19.59	20.62	24.53	22.77	26.11	32.72	39.32	21.48	31.57
30.13	20.81	22.69	23.69	17.07	29.30	22.66	23.44	29.12	30.25	21.32
22.10	34.88	30.64	21.74	31.75	29.31	21.75	24.14	21.16	29.46	22.81
25.86	28.64	22.96	21.12	24.47	28.91	20.36	30.10	24.32	23.87	26.98
32.04	25.19	23.14	22.06	33.97	18.23	25.34	25.94	17.25	23.87	12.76
27.62	26.01	24.75	22.23	21.33	25.28	24.42	23.32	22.10	18.93	22.78
29.33	21.66	33.04	22.00	21.25	20.54	16.86	22.53	27.33	20.13	18.94
24.70	30.46	25.02	21.66	29.58	20.50	—	—	—	—	—

$$\mu_n = \mu_{Y_n} - \frac{\gamma}{\alpha_n} = 24.67 - \frac{0.5772}{0.275} = 22.57 \text{ MPa} \tag{28}$$

A Type II distribution does not have a direct solution and requires statistical treatment (best fit empirical distribution function). Here, using the Type I Gumbel, the distribution is defined for daily maximum stress recordings. Owing to the invariance of the asymptotic forms of extremal distributions, this distribution for $t = 1$ day can be transformed to $t = 75$ years using [29].

$$\begin{aligned}\mu_n &= \mu + \frac{\ln(n)}{\alpha_n} \\ &= 22.75 + \frac{\ln(365.25\,\text{days/year} \times 75\,\text{years})}{0.275} \\ &= 59.9\,\text{MPa}\end{aligned} \tag{29}$$

where μ can represent either the mean or characteristic value. In this example, and most commonly, the characteristic value is transformed as it is used along with the shape factor to define the PDF, which takes the double exponential form [49]

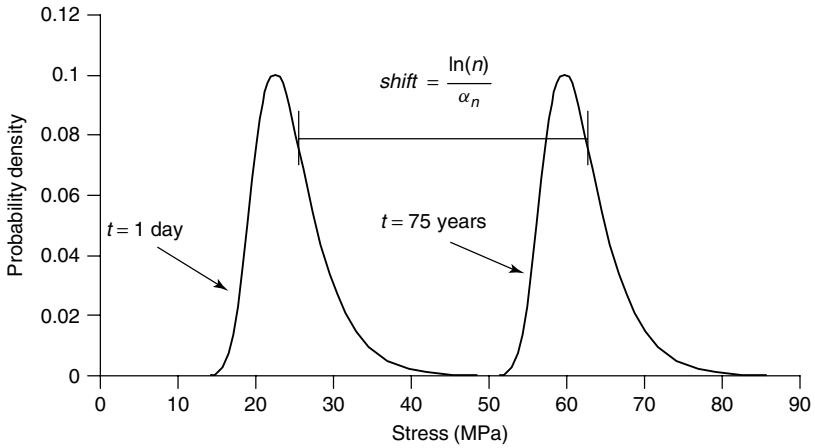


Figure 15. Transformation of a type I (Gumbel) extreme value distribution.

$$f_{Y_n}(y) = n[F_X(y)]^{n-1} f_X(y)$$
$$= \alpha_n e^{-\alpha_n(y-\mu_n)} \exp[-e^{-\alpha_n(y-\mu_n)}] \quad (30)$$

Figure 15 shows the type I (Gumbel) PDFs defined in this example with the associated transformation.

This approach provides the appropriate live load distribution (75-year) to be utilized in a reliability analysis and is useful for monitoring-based safety assessments that consider time effects. Load combinations are also considered if the structure is monitored for a representative period of time as the data will reflect combinations of load demands the structure experiences (e.g. temperature, trucks, wind). Defining what constitutes a representative period of time depends both upon the structure and upon the failure mechanism of interest. Additional information on combinations of extreme events can be found in [52] (*see* **Loads and Temperature Effects on a Bridge**).

3.3 Updating

Monitoring data may not provide the entire basis for the assessment of a structure. More often, there exists some prior belief about the random variable being monitored from codes, expert experience, or experimental tests. In such cases, new information must be combined with existing information in a reasonable manner. When the analysis being conducted is probabilistic in nature and the data sets in question contain variability, Bayesian updating provides a methodology to combine this data.

Assume that $f'(\theta)$ defines an existing (prior) belief of a random variable. The use of monitoring provides sample data that defines this same random variable with a different distribution $f(x)$. The combined (posterior) distribution $f''(\theta)$ is defined as in [50].

$$f''(\theta) = kL(\theta) f'(\theta) \quad (31)$$

The likelihood $L(\theta)$ expresses the conditional probability of observing $f(x)$ given $f'(\theta)$

$$L(\theta) = f(x/\theta) \quad (32)$$

and k is a normalizing factor

$$k = \int_{-\infty}^{\infty} \frac{1}{L(\theta) f'(\theta)\, d\theta} \quad (33)$$

to ensure the area under the PDF is 1. When the information collected is used for updating, two types of information should be distinguished—the equality and the inequality types [47]. Information of the equality type indicates that some basic strength, load, or response variable is measured. Measurement error should be modeled as separate random variable with a mean value of zero (unbiased estimates) and with an appropriate assigned standard deviation. Information of the inequality type refers to observations where it is only known that the observed variable is greater than or less than some limit or threshold: whether

corrosion has initiated, a certain crack width is greater than a prescribed threshold, or a limit state of elastic yielding is reached. In this type of investigation, the uncertainty of the threshold value is taken into account through the use of a probability of detection curve (POD curve). Modeling of POD curves can be found in [53].

Applications for updating in the assessment of civil structures are many. Several include the updating of the expected corrosion initiation time based on chloride penetration levels or presence (inequality), updating the performance of a network based upon success or failure readings at multiple locations (inequality), or including new information about material properties, geometry, or load demands (equality). A number of closed-form solutions for the posterior distribution $f''(\theta)$ can be found in [54–56]. In cases where no analytical solution is available, FORM/SORM techniques [39] may be required or Monte Carlo Simulation may be utilized. A commonly used closed-form solution is the case where both the prior and observed distributions are normal. When this is the case, the resulting distribution of the mean is also normal with the descriptors defined as in [57].

$$E(\theta|x) = \mu'' = \frac{\tau^2}{\tau^2 + \sigma^2}x + \frac{\sigma^2}{\sigma^2 + \tau^2}\mu \quad (34)$$

$$SD(\theta|x) = \sigma'' = \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}} \quad (35)$$

where, μ represents the prior mean, τ^2 the prior variance, x the monitoring-based mean, and σ^2 the monitoring-based variance. In the updating process, if the monitoring results are scattered or offer little new information, the effect on the prior distribution will be minor. In contrast, if the monitoring data shows only minor amounts of dispersion, the effect on the prior distribution will be large [58].

Example 8 An engineer is involved in a disaster relief reconstruction effort. Wood is extremely difficult to come by and the quality is often poor. Only two stockpiles of fir and pine boards of varied qualities and sizes are available. The engineer does not know exactly what type of fir and pine the boards are and can only estimate a low moisture content based on dry local conditions. An estimate of the modulus of elasticity is obtained for each wood by examining material property tables using likely supplier locations and an estimated moisture content of 10%. The fir modulus of elasticity is believed to be normally distributed with $\mu = 11\,000$ MPa and $\tau = 1500$ MPa. The pine modulus of elasticity is estimated as normally distributed with $\mu = 8000$ MPa and $\tau = 1000$ MPa. Using a strain gauge and known load, the engineer sets up a fairly crude three-point bending test sampling 20 boards of each wood type. The results are both fit to normal distributions and are surprisingly better than anticipated. The sample-based modulus for the fir is $x = 13\,000$ MPa and $\sigma = 3000$ MPa and the sample-based modulus for the pine is $x = 10\,000$ MPa and $\sigma = 500$ MPa. What material properties should the engineer utilize in design?

Answer Because the information available is probabilistic and contains varying degrees of uncertainty, a Bayesian approach is appropriate. For the fir, the posterior distribution is characterized by

$$E(\theta|x) = \mu'' = \frac{1500^2}{1500^2 + 3000^2} \times 13\,000 + \frac{3000^2}{3000^2 + 1500^2} \times 11\,000 = 11\,400 \text{ MPa} \quad (36)$$

$$SD(\theta|x) = \sigma'' = \sqrt{\frac{3000^2 \times 1500^2}{3000^2 + 1500^2}} = 1342 \text{ MPa} \quad (37)$$

For the pine, the posterior distribution is characterized by

$$E(\theta|x) = \mu'' = \frac{1000^2}{1000^2 + 500^2} \times 10\,000 + \frac{500^2}{500^2 + 1000^2}8000 = 9600 \text{ MPa} \quad (38)$$

$$SD(\theta|x) = \sigma'' = \sqrt{\frac{500^2 \times 1000^2}{500^2 + 1000^2}} = 447 \text{ MPa} \quad (39)$$

Figure 16 shows the prior, monitored, and posterior distributions for each type of wood. It should be noted that the posterior distribution is most influenced by the data set containing the best-quality information (least amount of variability). For the pine, because the monitored information contained a high degree

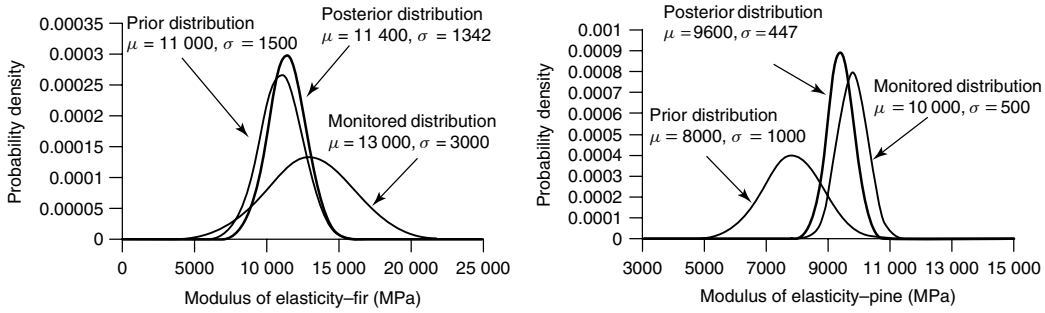


Figure 16. Bayesian updating of the modulus of elasticity of two types of wood.

of variability, it has little effect on the posterior distribution. This is reversed with respect to the fir.

3.4 Time-dependent reliability

Although assessment is a critical step in the management of civil infrastructure, predicting future performance is what allows the efficient scheduling of inspections, maintenance, repairs, and replacements. In fact, such predictions are the basis of LCM methods, durability-based design approaches, and whole-life cost approaches (nearly equivalent terminology). In general, the capacity (resistance) of a structure decreases over time as the structure deteriorates and the load demand increases. Load demand increase in a time-dependent reliability analysis has both a physical and a mathematical sense. In the physical sense, many of the loads being placed on structures have increased, such as the weight of present day trucks compared to those decades ago [43]. Mathematically, as a distribution is repeatedly sampled, the probability of encountering an extreme value increases as previously discussed for live load effects. The deterioration of various materials is a widely researched topic. Enright *et al.* [59] compiled a survey of deterioration models for concrete structures. Deterioration of concrete structures is typically divided into two categories, chemical and physical deterioration mechanisms. Chemical deterioration includes chloride attack, carbonation, acid attack, and alkali-aggregate reaction, while physical deterioration involves freeze–thaw, leaching, erosion, and cracking [26]. For transportation networks, chloride penetration of the concrete and subsequent corrosion of the reinforcing steel is the primary cause of capacity degradation. For steel structures, corrosion models predict where and when corrosion will result in section loss of the steel shape. Albrecht and Naeemi [60] first developed a model for steel girders that was later improved by Thoft-Christensen to include values for random variables [61]. Vulnerability of steel structures to corrosion is largely determined by regional/environmental conditions, proximity of a specific member to exposure, and the presence of aggressive agents. The durability of wood depends largely on environmental conditions, specific wood type, and preservation efforts. Generally, the performance of wood structures decreases over time due to natural weathering, decay, biological attack (fungus or insects), or chemical attack and service life can range from a short period up to 500 years [1].

Much work has been done in the development of reliability-based models to predict structural performance over time [62–64] and in the implementation of structural health monitoring [65–70]. In the most basic sense, these models determine a structure's initial state, model its deterioration, model future demands, and predict future performance. A reliability curve showing structural safety over time as shown in Figure 17(a) is the most common product of such analysis. Because there is uncertainty in (i) initial modeling parameters (ii) rates of deterioration, and (iii) future loads, the performance of the structure at any point of time is also uncertain and requires probabilistic treatment. Monitoring can be introduced at any point in time to reduce uncertainty by improving the accuracy of model input parameters resulting in a change of the reliability index β. Figure 17(b) shows a case in which an initial downward revision in the reliability index would result in the minimum safety threshold, β_{MIN}, being reached sooner.

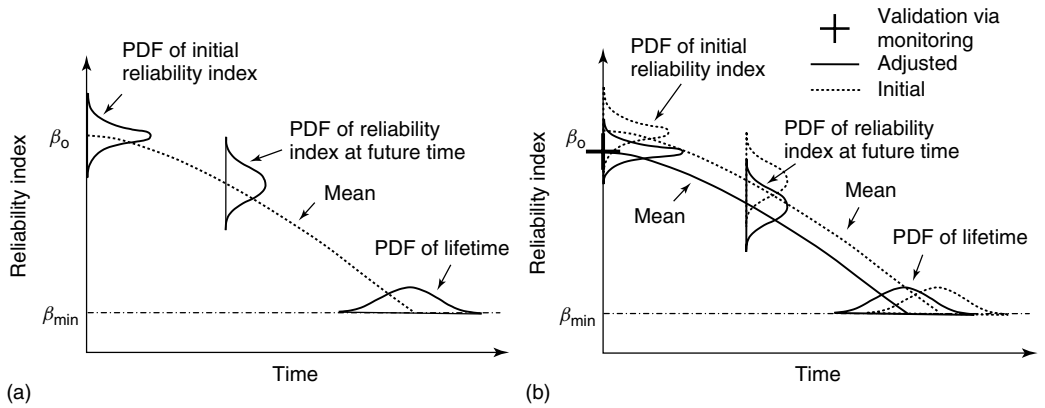


Figure 17. Predicted reliability index (a) without monitoring and (b) after an initial monitoring-based assessment. [Reproduced with permission from Ref. 71. © Taylor and Francis, 2006.]

A natural extension of monitoring at a point in time is monitoring continuously. There is significant benefit to such an approach as (i) the effects of uncertainty compound as one projects further into the future and (ii) permanent monitoring allows for the establishment of performance thresholds that allow a signal to be sent to the asset manager when violated [71]. A framework to introduce continuous monitoring into a reliability-based analysis is shown in Figure 18(a). Figure 18(b) and 18(c) illustrate how multiple or continuous monitoring assessments allow for capturing the rate of deterioration and validation of maintenance actions or repair activities.

3.5 Life-cycle costs with the inclusion of monitoring

The development of a time-dependent reliability-based performance prediction allows for the planning of inspections, maintenance, and repairs. Different combinations of inspection scheduling, preventive or routine maintenance actions, and repair or replacement strategies are typically compared using the metric of life-cycle cost as in [72]:

$$C_{ET} = C_T + C_{PM} + C_{INS} + C_{REP} + C_F \quad (40)$$

where C_{ET} = expected total cost, C_T = initial design/construction cost, C_{PM} = expected cost of preventive or routine maintenance, C_{INS} = expected cost of performing inspections, C_{REP} = expected cost of repairs, and C_F = expected cost of failure. Inclusion of monitoring into this general form results in [30]

$$C^0_{ET} = C^0_T + C^0_{PM} + C^0_{INS} + C^0_{REP} + C^0_F + C_{MON} \quad (41)$$

where C_{MON} = expected cost of monitoring. Because any monitoring solution will need its own management, maintenance, and repair, it is also best treated with respect to life-cycle costs as

$$C_{MON} = M_T + M_{OP} + M_{INS} + M_{REP} \quad (42)$$

where M_T = expected initial design/construction cost of the monitoring system, M_{OP} = expected operational cost of the monitoring system, M_{INS} = expected inspection cost of the monitoring system, and M_{REP} = expected repair cost of the monitoring system. The operational cost of the monitoring system would include the cost of power (battery or electricity), as well as the costs associated with data processing and data management. The benefit, or utility, of the monitoring system, B_{MON}, is then captured through a comparison of the expected life-cycle total cost with and without monitoring:

$$B_{MON} = C_{ET} - C^0_{ET} \quad (43)$$

Close attention must be given to calculating and communicating the utility of monitoring. Because facility managers must prioritize limited available resources, it must be assumed that monitoring would not be utilized unless shown to be cost effective.

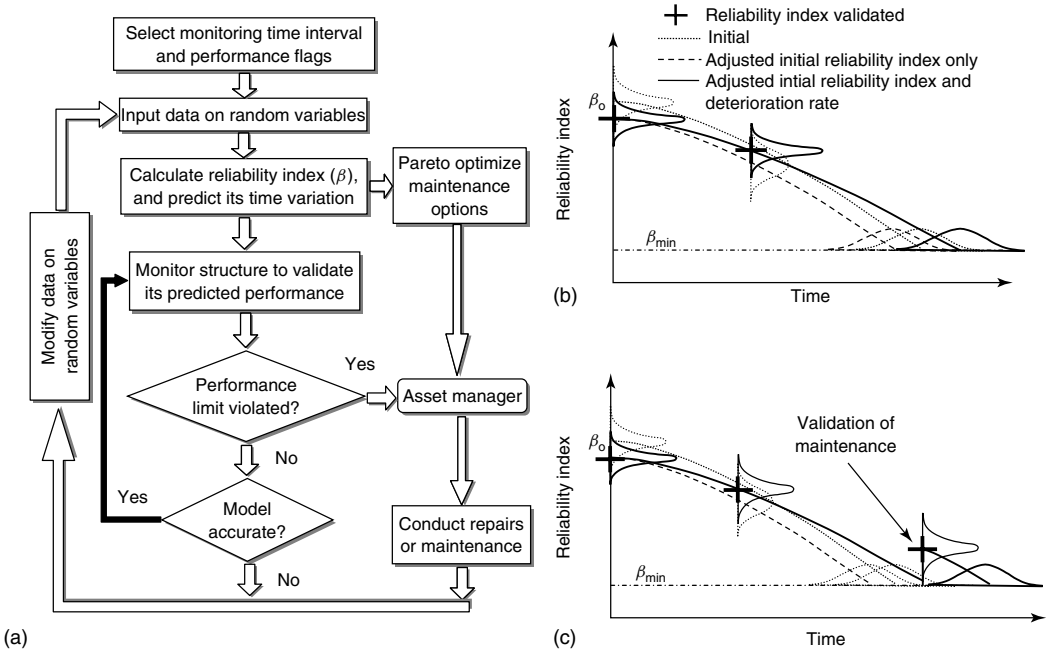


Figure 18. Framework to implement periodic or continuous monitoring. (a) Continuous monitoring framework, (b) validation of deterioration, and (c) validation of maintenance. [Reproduced with permission from Ref. 71. © Taylor and Francis, 2006.]

Example 9 A reliability analysis of a typical short-span highway bridge under corrosion (Figure 19a) yields the reliability index profile shown in Figure 19(b). The engineer believes the model would be significantly improved through a reduction in the uncertainty associated with the section modulus Figure 19(c). In order to investigate this improvement, the standard deviation is reduced, and the analysis was repeated for several alternatives (Figure 19d). Consulting with a monitoring expert, it is agreed that a 50% reduction in the uncertainty associated with the corrosion of the steel girder is feasible via monitoring. Girders are scheduled for replacement when a reliability index threshold of 2.0 is reached at a cost of \$100 000. What is the potential cost benefit associated with the 50% reduction?

Answer The reliability analysis is reconstructed imposing a performance threshold of 2.0 and by projecting the resulting girder replacements when required as shown in Figure 20. In this case, the reduction of model uncertainty results in one less projected repair.

Calculating net present values and assuming a discount rate of 4%,

$$C_{\text{REP}} = \frac{\$100\,000}{(1+0.04)^{25}} + \frac{\$100\,000}{(1+0.04)^{45}} + \frac{\$100\,000}{(1+0.04)^{60}} = \$64\,138 \tag{44}$$

$$C^0_{\text{REP}} = \frac{\$100\,000}{(1+0.04)^{35}} + \frac{\$100\,000}{(1+0.04)^{60}} = \$34\,848 \tag{45}$$

meaning that there is an expected benefit of \$29 290 (i.e., \$64 138 – \$34 848) with respect to the cost of repairs, which can be considered in the design or development of the monitoring solution.

3.6 The incorporation of risk

The answer to the previous analysis is fairly inadequate and can be improved significantly through the incorporation of risk. One of the principle advantages to monitoring should be an increased level of safety through a reduction of uncertainty. Quantifying

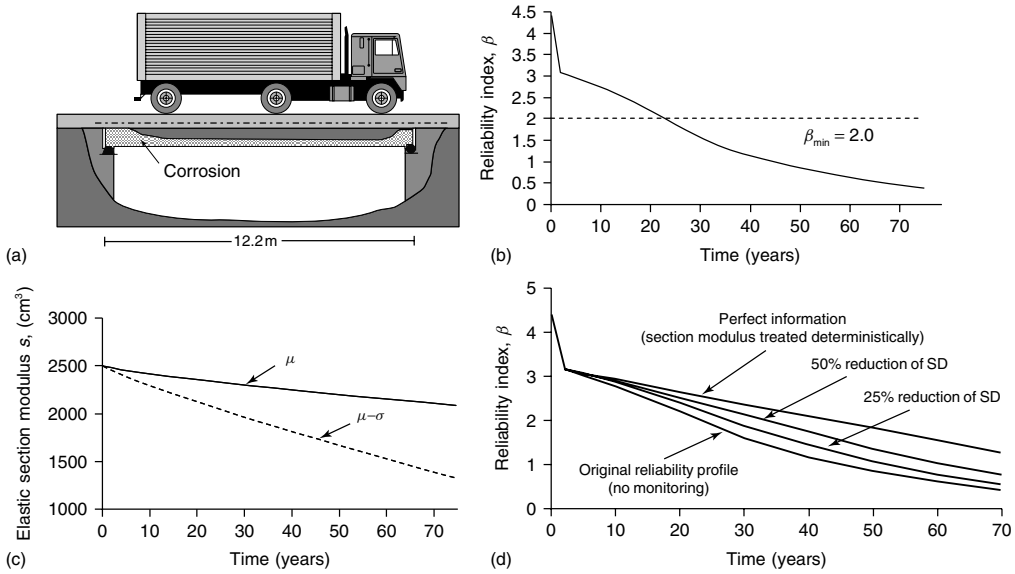


Figure 19. Reliability analysis with respect to flexure of a short-span steel beam highway bridge subjected to live load effects and corrosion over time. (a) Short-span steel girder concrete deck highway bridge, (b) reliability curve with respect to girder flexure, (c) girder average section modulus and deviation, and (d) reliability curve with reduction in section modulus uncertainty. [Reproduced with permission from Ref. 48. © Taylor and Francis, 2007.]

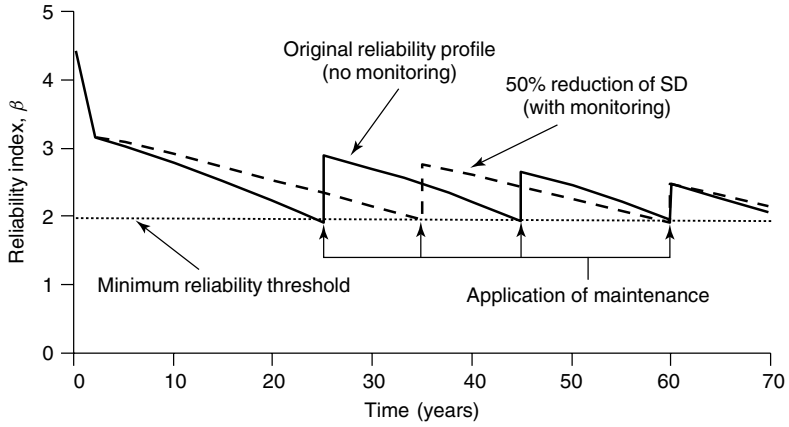


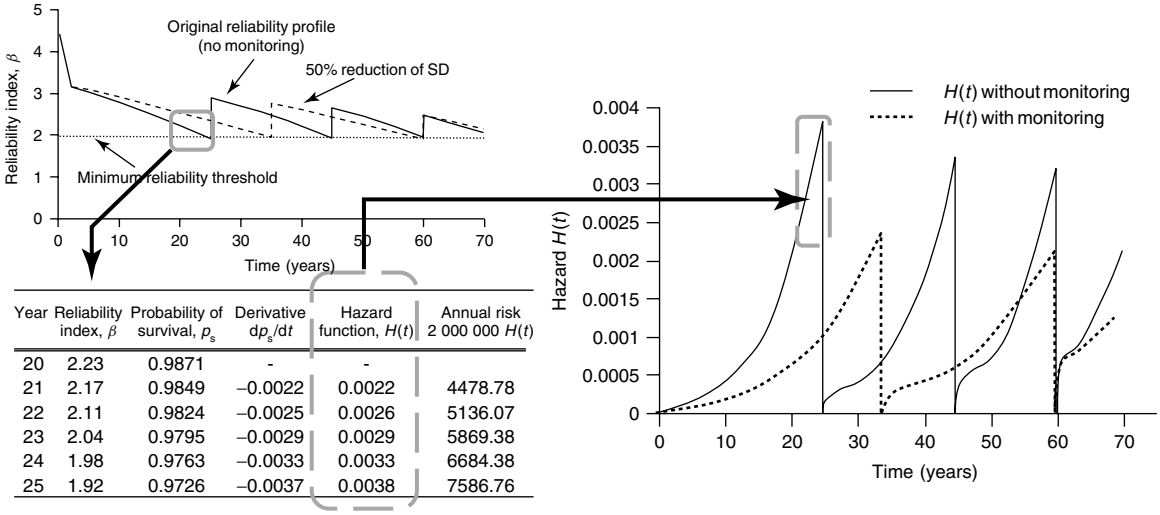
Figure 20. Reliability profile and associated maintenance actions with and without monitoring

this increased safety in dollars is necessary to appropriately communicate the utility of any monitoring solution. Risk, or the expected cost of failure C_F, can be calculated as the product of the likelihood of an event and the associated consequences given the event occurs as

$$Risk = R = C_F = p_f C \tag{46}$$

Introducing risk into a time-dependent reliability analysis requires the use of a hazard function $H(t)$. This function expresses the conditional probability of failure in time $(t, t + dt)$, given that failure has not already occurred [73].

$$H(t) = -\frac{dp_s(t)}{dt} \times \frac{1}{p_s(t)} = -\frac{S'(t)}{S(t)} \tag{47}$$



Year	Reliability index, β	Probability of survival, p_s	Derivative dp_s/dt	Hazard function, $H(t)$	Annual risk 2 000 000 $H(t)$
20	2.23	0.9871	-	-	
21	2.17	0.9849	–0.0022	0.0022	4478.78
22	2.11	0.9824	–0.0025	0.0026	5136.07
23	2.04	0.9795	–0.0029	0.0029	5869.38
24	1.98	0.9763	–0.0033	0.0033	6684.38
25	1.92	0.9726	–0.0037	0.0038	7586.76

Figure 21. Calculation of the hazard function and associated annual risk. [Adapted from Ref. 73.]

where $p_s(t)$ is the probability that an element is safe at any time t, which is also referred to as the survivor function $S(t)$ and $S'(t)$ its derivative.

Example 10 Assuming a consequence of failure of \$2 000 000 for cleanup, liability costs, design, and reconstruction of a new bridge, calculate the associated cost of failure, C_F, for both options considered in example #8.

Answer The hazard function is created by taking the reliability index β, calculating the corresponding probability of survival p_s, and by calculating the derivative of the probability of survival at each time step for use in the above equation of the hazard function $H(t)$. These calculations are detailed along with the corresponding annual risk for years 20–25 for the option without monitoring (Figure 21).

Calculating the net present value of the annual risk each year and then summing these net present values across the life of the structure for both the options—with and without monitoring yields

$$\begin{aligned} C_F &= \$44\,204 \\ C_F^o &= \$27\,371 \end{aligned} \tag{48}$$

meaning there is an expected benefit of \$16 833 (i.e., \$44 204 − \$27 371) with respect to risk. Combining this result with those in Example 9 yields an expected total benefit of \$46 123 (i.e., \$29 290 + \$16 833) of incorporating a monitoring solution that reduces the uncertainty associated with the section modulus by 50%. Such a monitoring approach would only be cost effective if its life-cycle costs were less than this amount. Such information can be used not only to quantify the utility of monitoring but for use as a monitoring design constraint.

4 CONCLUSIONS

The incorporation of monitoring technologies into the design and management of civil infrastructure introduces a number of challenges often not considered in current practice. Probabilistic modeling, rates of deterioration, predicting costs, sensor placement, information management, and linking this information to accepted design and assessment standards are some of the issues, to name a few—many which have yet to be resolved. This chapter used examples to highlight some of these challenges as an introduction to the topic. Optimization of different monitoring alternatives, treatment of networks of sensors, and acceptance criteria for information obtained from SHM are left for further investigation. Notably present in this chapter, however, is a sharp contrast between SHM being utilized as a response to a discovered structural defect, and SHM being part of an LCM process, where, of course, the latter is preferred. Because of the importance and value of civil infrastructure to the

society it supports, designing better and managing more efficiently is necessary and timely, especially as focus turns to managing existing aging structures. Within this context, monitoring has the potential to significantly aid engineers and managers in making optimal resource allocations by providing site-specific real-time information. To justify the initial and follow-up costs of employing monitoring solutions, the development of metrics that quantify and communicate the utility of monitoring systems are crucial to their acceptance and more frequent use. To achieve this end, engineers and managers must become more comfortable in the assessment and design of civil structures within a performance-based, life-cycle, risk inclusive context.

ACKNOWLEDGMENTS

The support by three grants from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance is gratefully acknowledged. Also, the support of the National Science Foundation through grants CMS-0638728 and CMS-0639428 to Lehigh University is gratefully acknowledged. The assistance of Sunyong Kim of Lehigh University in the preparation of figures for this article is also acknowledged. The opinions and conclusions presented in this article are those of the authors and do not necessarily reflect the views of the sponsoring organizations.

REFERENCES

[1] Bijen J. *Durability of Engineering Structures: Design, Repair and Maintenance*. Woodhead Publishing Limited: Abington Hall, Cambridge, 2003.

[2] Estes AC, Frangopol DM. Life-cycle evaluation and condition assessment of structures. In *Structural Engineering Handbook*, *Second Edition*, Chen W-F, Lui EM (eds). CRC Press, 2005; Chapter 36, pp. 36-31–36-51.

[3] Faravelli L (ed). *Structural Health Monitoring*. Sage Journals: ISSN 1475–9217, http://shm.sagepub.com/.

[4] Frangopol DM (ed). *Structure and Infrastructure Engineering: Maintenance, Management, Life-cycle Design and Performance*. Taylor & Francis: ISBN 1573–2479, http://www.tandf.co.uk/journals/titles/15732479.asp.

[5] Chang FK (ed). *Structure Control & Health Monitoring*. Wiley & Sons: ISBN 1545–2255, http://www.interscience.wiley.com/cgi-bin/jhome/106562744.

[6] Frangopol DM, Liu M. Life-cycle cost and performance of civil structures. *McGraw-Hill 2006 Yearbook of Science and Technology*. McGraw-Hill: New York, 2006; pp. 183–185.

[7] Chong KP, Carino NJ, Washer G. Health monitoring of civil infrastructures. *Smart Materials and Structures* 2003 **12**:483–493.

[8] Frangopol DM, Liu M. Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. *Structure and Infrastructure Engineering* 2007 **3**(1):29–41.

[9] Bureau of Transportation Statistics (BTS). *Transportation Statistics Annual Report*, Report BTS. Department of Transportation: Washington, DC, 2003.

[10] Tolliver D. *Highway Impact Assessment: Techniques and Procedures for Transportation Planners and Managers*. Quorum Books: Westport, CT, 1994.

[11] Federal Highway Administration, National Bridge Inventory, available online at http://www.fhwa.dot.gov/bridge/nbi.htm (dated accessed May 2007).

[12] Peil U. Life-cycle prolongation of civil engineering structures via monitoring. *Proceedings of the 4th International Workshop on Structural Health Monitoring 2003*. DEStech Publications: Stanford, CA, 2003; pp. 64–78.

[13] Fujino Y, Abe M. Structural health monitoring—current status and future. *Proceedings of the Second European Workshop on Structural Health Monitoring 2004*. DEStech Publications: Munich, 2004, pp. 3–10.

[14] ASCE, Report card for America's Infrastructure, American Society of Civil Engineers, Reston, VA. Available online at www.asce.org/reportcard/2005/index.cfm (dated accessed May 2007).

[15] Matousek M, Schneider J. *Untersuchungen zur Struktur des Sicherheitsproblems bei Bauwerken,* (in German). Basel: BirkhaÈuser, 1976.

[16] Stewart MG, Melchers RE. *Probabilistic Risk Assessment of Engineering Systems*. Chapman & Hall: London, 1997.

[17] Blind H. The safety of dams. *Water Power and Dam Construction* 1983 **35**:17–21.

[18] Bertrand A, Escoffier L. IFP databanks on offshore accidents. In *Reliability Data collection and Use of Risk and Availability Assessment*, Colombari V (ed). Springer-Verlag: Berlin, 1987.

[19] Andersen T, Misund A. Pipeline reliability: an investigation of pipeline failure characteristics and analysis of pipeline failure rates for submarine and cross-country pipelines. *Journal of Petroleum Technology* 1983 **35**:709–717, http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=6192657.

[20] Scott RL, Gallaher RB. Review of safety-related events at nuclear power plants as reported in 1979. *Nuclear Safety* 1981 **22**(4):505–515.

[21] Joint Task Committee on Maintenance Engineering, *Infrastructure Maintenance Engineering, Japan Society of Civil Engineers*. University of Tokyo Press, 2004 (in Japanese).

[22] Akgul F, Frangol DM. Rating and reliability of existing bridges in a network. *Journal of Bridge Engineering* 2003 **8**(6):383–392.

[23] Liu M, Frangopol DM. Time-dependent bridge network reliability: novel approach. *Journal of Structural Engineering, ASCE* 2005 **2**:329–337.

[24] Liu M, Frangopol DM. Probability-based bridge network performance evaluation. *Journal of Bridge Engineering* 2006 **11**(5):633–641.

[25] Liu M, Frangopol DM. Optimizing bridge network maintenance management under uncertainty with conflicting criteria: life cycle maintenance, failure, and user costs. *Journal of Structural Engineering, ASCE* 2006 **11**:1835–1845.

[26] Rafiq MI. *Health Monitoring in Proactive Reliability Management of Deteriorating Concrete Bridges, PhD Thesis*. University of Surrey: 2005.

[27] Catbas FN, Susoy M, Frangopol DM. Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data, *Engineering Structures*, Elsevier (in press), 2008.

[28] Moon FL, Aktan AE, Structural identification of constructed systems and the impact of epistemic uncertainty. *Proceedings of the Third International Conference on Bridge Maintenance and Safety*. Taylor & Francis: Porto, Portugal, 2006, p. 8 on CD-ROM.

[29] Ang AHS, Tang WH. *Probability Concepts in Engineering Planning and Design Volume II, Second Edition*. John Wiley & Sons: New York, 2007.

[30] Frangopol DM, and Messervey TB. Integrated life-cycle health monitoring, maintenance, management and cost of civil infrastructure. *Proceedings of the International Symposium on Integrated Life-Cycle Design and Management of Infrastructures*. Tongji University: Shanghai, 2007 (keynote paper).

[31] Aktan AE, Ellingwood BR, Kehoe B. Performance-Based Engineering of Constructed Systems. *Proceedings of ASCE Structures 2007*, Long Beach, CA, 2007.

[32] Madsen HL, Krensk S, Lind NC. *Methods of Structural Safety*. Prentice-Hall: Englewood Cliffs, NJ, 1986.

[33] Rubinstein RY. *Simulation and the Monte Carlo Method*. John Wiley & Sons: New York, 1981.

[34] Akgul F, Frangopol DM. Computational platform for predicting lifetime system reliability profiles for different structure types in a network. *Journal of Computing in Civil Engineering, ASCE* 2004 **18**(2):92–104.

[35] Fießler B, Rackwitz R, Gollwitzer S. *Theoretical, Technical and Users Manual*, STRUREL Version 6.1. RCP-GmbH: Munich, 1998.

[36] Liu PL, Lin HZ, Der Kiureghian A. *CALREL User Manual. Report No. UCB/SEMM-89/18. Berkeley, CA: Department of Civil and Environmental Engineering*. University of California: Berkeley, 1989.

[37] DNV, Sesam. *Theory Manual*, DNV Research Report No. 93–2056, PROBAN Version 4, Hovik, Norway, 1993.

[38] Estes AC, Frangopol DM. RELSYS: a computer program for structural system reliability analysis. *Structural Engineering and Mechanics* 1998 **6**(8):901–919.

[39] Connor RJ, McCarthy J. *Report on Field Measurements and Uncontrolled Load Testing of the Lehigh River Bridge (SR-33)*. Center for Advanced Technology for Large Structural Systems (ATLSS): Phase II Final Report, 2006.

[40] Ghosn M, Moses F. Reliability calibration of a bridge design code. *Journal of Structural Engineering* 1986 **112**(4):745–763.

[41] Nowak AS. Live load model for highway bridges. *Structural Safety* 1993 **13**(1–2):53–66.

[42] Nowak AS, Yamani AS. A reliability analysis for girder bridges. *Structural Engineering Review* 1995 **7**(13):251–256.

[43] Gindy M, Nassif H. Effect of bridge live load based on 10 years of WIM data. *Proceedings of the Third International Conference on Bridge Maintenance and Safety*. Taylor & Francis: Porto, Portugal, 2006, p. 9 pages on CD-ROM.

[44] Faber MH, Val DV, Stewart MG. Proof load testing for bridge assessment and upgrading. *Engineering Structures* 2000 **22**:1677–1689.

[45] Moses F, Lebet JP, Bez R. Applications of field testing to bridge evaluation. *Journal of Structural Engineering, ASCE* 1994 **120**(6):1745–1762.

[46] Fu G, Tang J. Risk-based proof-load requirements for bridge evaluation. *Journal of Structural Engineering, ASCE* 1995 **121**(3):542–556.

[47] Faber MH. Reliability based assessment of existing structures. *Progress in Structural Engineering Materials* 2000 **2**:247–253.

[48] Messervey TB, Frangopol DM. Updating the time-dependent reliability using load monitoring data and the statistics of extremes. In *Life-Cycle Performance and Cost of Civil Infrastructure*, Cho H-N, Frangopol DM, Ang A-HS (eds). Taylor & Francis Group: London, 2007; pp. 269–276.

[49] Ang AH, Tang WH. *Probability Concepts in Engineering Planning and Design Volume II*. John Wiley & Sons: New York, 1984.

[50] Estes AC, Frangopol DM. Reliability-based condition assessment. In *Structural Condition Assessment*, Ratay RT (ed). John Wiley & Sons: Hoboken, New Jersey, 2005; Chapter 2, pp. 25–66.

[51] Mahmoud HN, Connor RJ, Bowman CA. *Results of the Fatigue Evaluation and Field Monitoring of the I-39 Northbound Bridge over the Wisconsin River*. Center for Advanced Technology for Large Structural Systems (ATLSS), 2005.

[52] Ghosn M, Moses F, Wang J. *Design of Highway Bridges for Extreme Events*, NCHRP TRB Report 489. Transportation Research Board: Washington, DC, 2003, http://www.trb.org/news/blurb_detail.asp?id=1884.

[53] Thomson DO, Chimenti DE. *Review of Progress in Quantitative Non-Destructive Evaluation*, Col. 2a. Plenum Press: New York, 1983.

[54] Raiffa H, Schalifer R. *Applied Statistical Decision Theory*. Harvard University Press: Cambridge, MA, 1961.

[55] Aitchison J, Dunsmore IR. *Statistical Prediction Analysis*. Cambridge University Press: Cambridge, MA, 1975.

[56] Rackwitz R, Schrupp R. Quality control, proof testing, and structural reliability. *Structural Safety* 1985 **2**:239–244.

[57] Casella G, Berger RL. *Statistical Inference*. Duxbury, Thompson Learning: Pacific Grove, 2002.

[58] Melchers RE. *Structural Reliability Analysis and Prediction*, *Second Edition*, John Wiley & Sons: Chichester, 1999.

[59] Enright M, Frangopol DM, Hearn G. Degradation of reinforced concrete structures under aggressive conditions. In *Materials for the New Millennium*, Chong KP (ed). ASCE: New York, 1996; Vol. 2.

[60] Albrecht P, Naeemi AH. *Performance of Weathering Steel in Bridges*, National Cooperative Highway Research Program Report 272. Transportation Research Board, 1984.

[61] Thoft-Christensen P, Jensen FM, Middleton CR, Blackmore A. Assessment of the reliability of concrete slab bridges. In *Reliability and Optimization of Structural Systems*. Frangopol DM, Corotis RB, Rackwitz R (eds). Pergamon, Elsevier: 1997; pp. 321–328.

[62] Frangopol DM, Neves LC. Life-cycle maintenance of structures by condition, reliability and cost oriented probabilistic optimization. In *Innovation in Computational Structures Technology*, Topping BHV, Montero G, Montenegro R (eds). Saxe-Coburg Publications: Stirling, Scotland, 2006; Chapter 5, pp. 95–110.

[63] Frangopol DM, Liu M. Multiobjective optimization of risk-based maintenance and life-cycle cost of civil infrastructure. In *System Modeling and Optimization* (Text of Plenary Lecture, Ceragioli E, Dontchev A, Furuta H, Marti K, Pandolfi L (eds). Springer: Boston, 2006; pp. 123–136.

[64] Frangopol DM, Miyake M, Kong JS, Gharaibeh ES, Estes AC. Reliability and cost–oriented optimal bridge maintenance planning. In *Recent Advances in Optimal Structural Design*, Burns S (ed). Reston, VA, ASCE: 2002; Chapter 10, pp. 257–270.

[65] Glisic B, Inaudi D. *Fibre Optic Methods for Structural Health Monitoring*. John Wiley & Sons: West Sussex, 2007.

[66] Cheung MMS, Noruziann B, Yang C-Y. Health monitoring data in assessing critical behaviour of bridges. *Structure and Infrastructure Engineering* 2007 **3**(4):325–342.

[67] Chang S-P, Kim S, Lee J, Bae I. Health monitoring system of a self-anchored suspension bridge (planning, design and installation/operation). *Structure and Infrastructure Engineering* 2008 **4**(3):193–205.

[68] Wang Y, Lynch JP, Law KH. A wireless structural health monitoring system with multithreaded sensing devices: design and validation. *Structure and Infrastructure Engineering* 2007 **3**(2):103–120.

[69] Strauss A, Frangopol DM, Kim S. Statistical, probabilistic and decision analysis aspects related to the efficient use of structural monitoring systems. *Beton- und Stahlbetonbau (Concrete and Reinforced Concrete Structures)* Ernst & Sohn Verlag, 2008 **103**:23–28.

[70] Frangopol DM, Strauss A, Kim S. Bridge reliability assessment based on monitoring. *Journal of Bridge Engineering, ASCE*, **13**(3):2008 pp. 258–270.

[71] Messervey TB, Frangopol DM, Estes AC. Reliability-based life-cycle bridge management using structural health monitoring. In *Bridge Maintenance, Safety, Management, Life-Cycle Performance and Cost*, Cruz PJS, Frangopol DM, Neves LC (eds). Taylor & Francis Group: London, 2006; pp. 545–546.

[72] Frangopol DM, Lin KY, Estes AC. Life-cycle cost design of deteriorating structures. *Journal of Structural Engineering, ASCE* 1997 **123**(10): 1390–1401.

[73] Frangopol DM, Messervey TB. Risk assessment for bridge decision making. *Proceedings of the Fourth Civil Engineering Conference in the Asian Region*, CECAR 4 (invited paper); In *ASCE Tutorial and Workshop on Quantitative Risk Assessment*, Taipei, Taiwan, June 25–28, 2007, 37–42.

Chapter 92

Usage Management of Military Aircraft Structures

Rolf H. Neunaber

Industrieanlagen Betriebsgesellschaft mbH, Ottobrunn, Germany

1 INTRODUCTION

An essential requirement in the aircraft (A/C) design is the minimization of the structure weight. Therefore, the structure components are designed only to withstand the cyclic loading of the specified usage profile without any damage (safe-life design) or with damage that can be discovered during in-service inspections before becoming critical (fail-safe design). The structure will be qualified within these boundary conditions for a specific number of allowable flight hours.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

On the other hand, this design enforces that a deviation from the certified usage spectrum during in-service usage has a draw back to the allowable flight hours. This is especially true for aluminum structures, which are particularly damaged by structural fatigue.

For A/C in civil usage, deviations from the certification profile are within narrow limits only (e.g., flight time per flight). In contrast to civil A/C, the missions and outboard store configurations for military A/C change quite often, which results in a high scatter of usage for individual A/C. For a proper assessment of the individual life consumption, the usage of an adapted structural health monitoring (SHM) system is mandatory.

The SHM usage is thereby not restricted to flight safety only. The SHM system is increasingly used for the usage management of the A/C structure. An essential subtask of this is the optimization of maintenance, e.g., inspections and modifications.

2 A/C STRUCTURE CERTIFICATION

Development specifications for military A/C generally require a fatigue test of the entire airframe, or at least its principal components, as proof of structural

integrity. Although test speeds allow the simulation of a large number of daily flights, several years are necessary for the completion of the fatigue test, allowing for interruptions for repair, modification, and inspection. Thus, for reasons of time and money, series production of A/C begins before all results of the fatigue tests on the various fatigue critical areas (FCAs) are known and the necessary modifications made. Individual batches of series production A/C, therefore, still have structural FCAs, albeit with decreasing frequency, whose inherent fatigue life in all probability will not last throughout the planned operational usage.

Hence, necessity to retrofit and repair structural components is unquestionable. The question is, which A/C requires which repair, and when. There are several reasons for asking this question, for example,

- restrictions on operational usage for unmodified structures and
- cost-effective modification.

In order to answer these questions satisfactorily, it is necessary to know the fatigue life ratio of the operational stress spectrum on FCAs to the stress spectrum applied during the fatigue test, i.e., how many operational flight hours are equivalent to a simulated flight hour during testing. Depending on the structural point in question, such spectra comparisons range from

- either a higher or lower operational stress spectrum compared to the test to;
- the stress spectrum simulated during the test not being representative in detail for operational usage.

Thus, it is both useful and essential for the A/C user to have an instrument, which answers the complex questions mentioned above and enables necessary activities to be planned. This task is dealt with by A/C SHM (Figure 1).

3 SHM SYSTEM DESIGN

3.1 System concept

An SHM system is normally composed of the following two essential parts:

- **Onboard**

Load and/or condition data acquisition and processing:

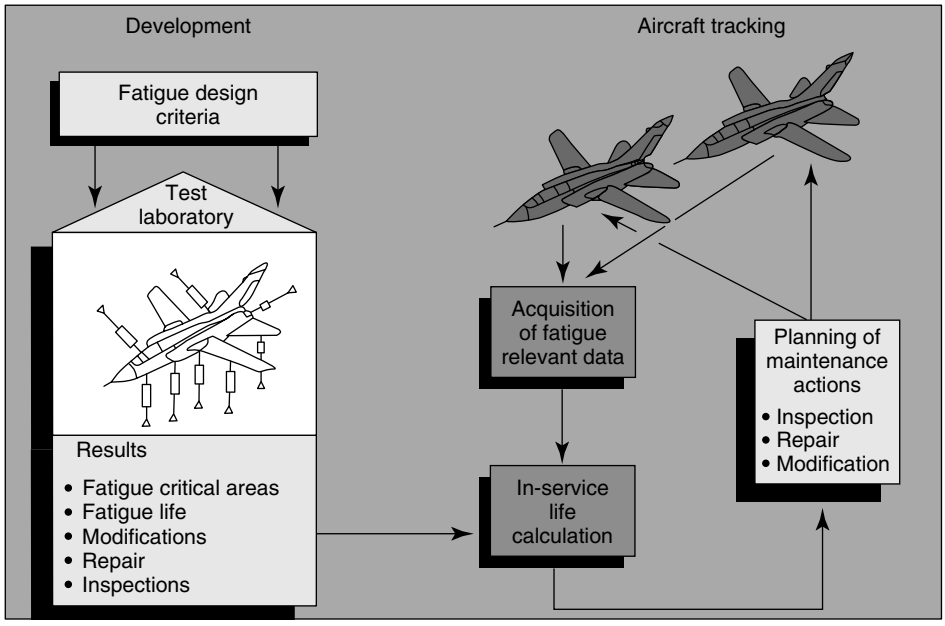


Figure 1. Aircraft structural health monitoring.

The primary function of this part is the acquisition of the load and/or condition data of the FCAs. This part also comprises data processing to a certain extent and data recording for later evaluations on ground.

- **On-ground**

Data processing, evaluation, and fleet management support:

After several input data checks, the milked A/C data are processed with adapted software tools with the aim to generate the load/stress spectra and the accumulated damages for the specific flight of a tail number for each FCA. The processed data are stored together with the relevant identification data for that flight (e.g., flight hours, configuration, and number of landings). This SHM data set forms the basis for the follow-on evaluations required for fleet management.

The onboard/on-ground data transfer is carried out with type-specific designed hand-held terminals. The on-ground station is normally a commercial PC with adapted software tools. Furthermore, the SHM systems can be differentiated according to the following two design variants:

- SHM system is designed and developed together with the A/C itself
  - The SHM system has to be designed to be able to monitor the complete primary structure of the A/C. This is because the FCAs are unknown at this time.
  - If possible, the A/C structure itself can be designed partly as smart structure with the aim to support SHM.
- SHM system has to be retrofitted
  - In case that no FCAs are known, the SHM system has to be designed to be able to monitor the whole A/C primary structure.
  - If the FCAs are known from test or service experience, the SHM system can be designed for the monitoring of those FCAs. In this case, the cost/benefit relation will reach an optimum.

3.2 SHM system types and evaluation

From the acquisition point of view, two types of SHM systems can be differentiated (Table 1).

3.2.1 Load-acquisition systems

- **Flight parameter data**

The available flight control system parameter data (altitude, velocity, normal acceleration, etc.) are

Table 1. Load-acquisition system versus condition-acquisition system

		Load acquisition system	Condition acquisition system
SHM design criteria	– Location of critical area – Type of damage – Stress level/stress spectrum	Selection of relevant flight parameters/strain	Selection of suitable condition acquisition system sensors
	Data acquisition	– Flight parameter/strain versus time – Flight parameter/strain spectra	Sensor and structure condition
In-service usage	Data processing (onboard/on ground)	– Stress spectra generation – Damage calculation	—
	Type of results	– Degree of fatigue degradation – Damage initiation and propagation	Damage initiation and propagation

acquired and stored for stress and damage calculations later on. The data processing is carried out in the following conversion steps:

- parameter history to stress history using specific templates for each FCA;
- stress history to frequency of stress occurrence matrix using a rainflow algorithm and
- frequency of stress occurrence matrix to damage using adapted damage algorithms.

- **Strain data**

The strain data of each local measurement area are acquired and stored for processing later on, analogous to the flight parameter procedure above. The data processing is carried out in the following conversion steps:

- strain history of the measurement point to stress history of the associated FCA using adapted strain to stress algorithms;
- stress history to frequency of stress occurrence matrix as before and
- frequency of stress occurrence matrix to damage as before.

The biggest advantage of the strain sensors is that no templates are necessary for the derivation of the FCA stress history.

In case the fleet is monitored with a flight parameter–based SHM system, additional strain-gauge equipment in a few A/C can be used to derive the necessary FCA templates directly. The equipment mentioned before allows the simultaneous acquisition of flight parameter and strain data as a basis for a correlation analysis according to Figure 2. The results of this analysis are transformed into an FCA template.

3.2.2 Condition-acquisition systems

The condition sensors acquire a change in the structural behavior directly in contrast to the load sensor types. The following condition sensors are examples used in test application and partially in-service:

- comparative vacuum monitoring (CVM)
- fiber Bragg gratings (FBG)
- acousto ultrasonic (AU)
- acoustic emission (AE)
- eddy-current fail sensors (ECFS).

Condition data acquisition during the flight is not necessarily required and/or possible depending on the specific sensor type. Often the after-flight data acquisition is sufficient for the monitoring of a change in the structure's behavior. In this case, the SHM installation equals permanently mounted non destructive inspection (NDI) equipment.

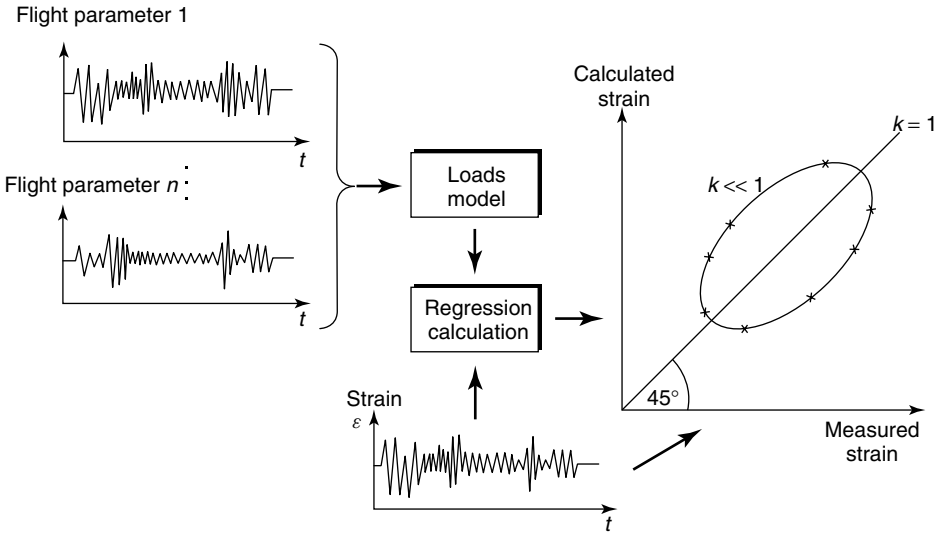


Figure 2. Flight parameter/strain correlation analysis.

3.2.3 Comparison of the SHM systems

- Load-acquisition systems allow better monitoring of the structural degradation especially before crack initiation. Thereby, these systems are used primarily for the planning of preventive measures (inspections or modifications).
- Condition-acquisition systems have the advantage of registering a crack or even crack initiation directly. Compared to load-acquisition systems, no additional safety factor or conservative element needs to be included. This allows the usage of the total individual life capacity.
- Condition-acquisition sensors need to be placed more or less directly on the FCA, which is in many cases not possible.
- Load-acquisition systems are more flexible in usage. This is especially true for flight parameter systems where the necessary FCA parameters (stress templates) need to be developed only once for the whole fleet. In addition, if a new FCA has been uncovered, the generation of one template is sufficient for the monitoring of the whole fleet.

Summarizing the pros and cons, it has to be accepted that the specific application dictates which method is preferable.

4 TORNADO SHM SYSTEM

The Tornado A/C is built mainly out of metallic materials and thereby prone to structural fatigue. The SHM system is designed as a retrofit system primarily used for means of material conservation. The value of this monitoring system thus varies by the ratio of its performance, including such aspects as

- number and importance of FCAs;
- precision of damage calculation per FCA; and
- scope of application for all FCAs of the A/C structure.

to its costs, which include

- investment costs for data acquisition and processing systems and
- recurring costs for data evaluation and report preparation.

Taking the point of view that a number of FCAs must be monitored rather than the entire A/C, the question of optimization of the procedure must be related to the individual FCAs themselves. Most of the Tornado FCAs are located at the center fuselage and wing, where the relevant structure loads are highly influenced by the A/C's normal acceleration.

The tracking concept of the WS Tornado is divided into the three sectors (Figure 3):

- individual aircraft tracking (IAT)
- selected aircraft tracking (SAT)
- temporary aircraft tracking (TAT).

Monitoring is based essentially on flight parameters, which are available on the existing data-acquisition unit of the crash recorder. The following four parameter data sets have been defined according to the concept requirements:

- **Recorder parameter set (RPS)**

The direct acquired flight parameters form the RPS, which is registered by a flight recorder.

- **RPS + strain-gauge data**

During TAT, the strain-gauge data are acquired simultaneously with the RPS flight parameter for the extended data set.

- **Full parameter set (FPS)**

Through differentiations and conversions of several flight parameters, the RPS + data set was transformed to the FPS.

- **Pilot parameter set (PPS)**

To reduce the expense of data acquisition for each individual A/C, an additional severely reduced PPS was defined, containing only the essential flight parameter for stress on the most vital FCAs.

4.1 Temporary aircraft tracking

The key monitoring elements of TAT (Figure 4) are the strain gauges in the various FCAs. They are evaluated by regression techniques to produce a realistic correlation between operational strain in

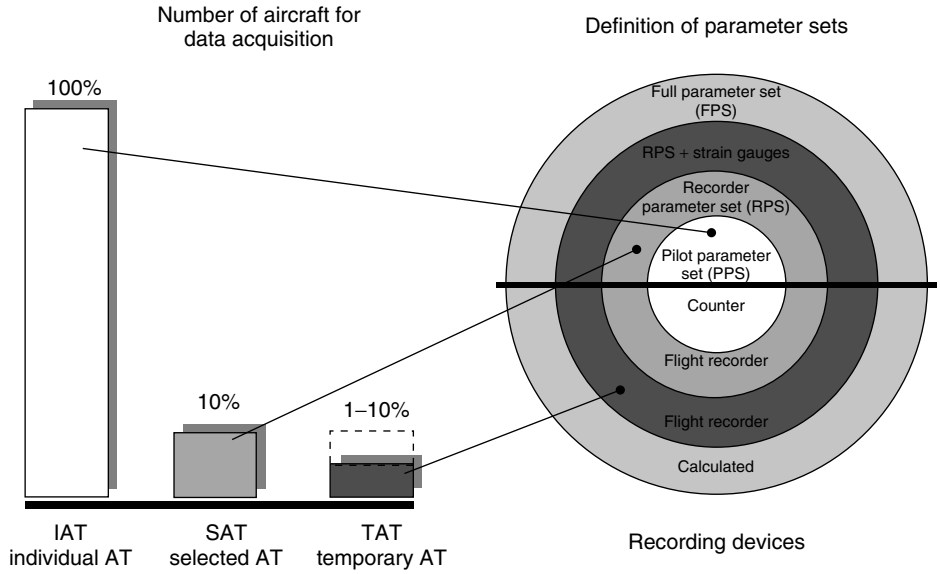


Figure 3. Tracking concept.

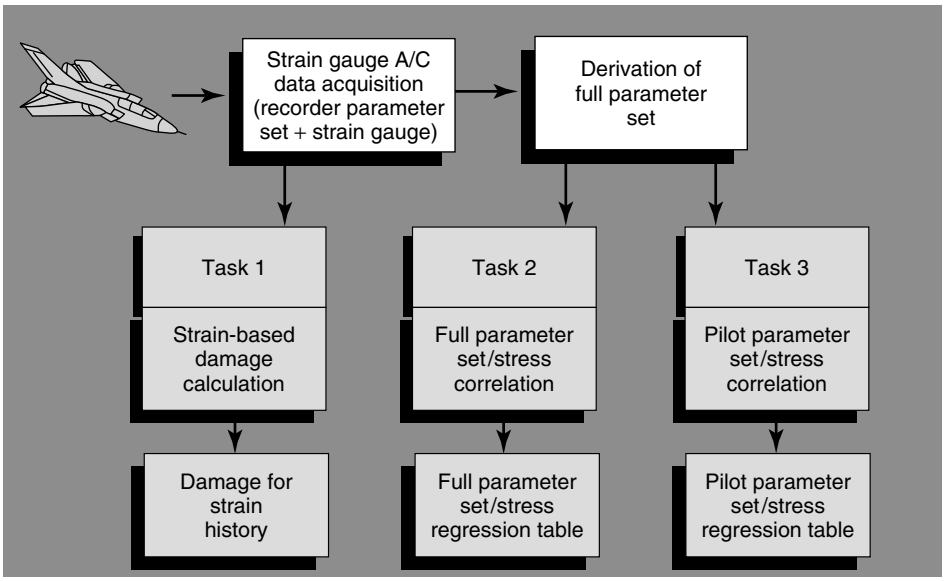


Figure 4. Tornado, temporary aircraft tracking (TAT).

the structure and the flight parameters that cause it. As the strain/parameter correlation is only partially deterministic because of the limited parameter sets, it is cyclically repeated within the TAT segment for the same FCA on several A/C. Hence, mismatches in the parameter algorithms, due to alterations in the A/C configurations and in the operational missions are covered simultaneously.

After transformation of the RPS + strain-gauge data set, the resulting FPS contains all essential influencing variables for mechanical strain (e.g., accelerations and weight).

- **Task 1: Strain-based damage calculation**

Direct conversion of the measured strain history into a frequency of occurrence matrix and calculation of the accumulated damage. The validity is restricted to the A/C observed throughout the measuring period.

- **Task 2: FPS/stress correlation**

Correlation of the FPS with the stress at FCAs obtained from local strain measurement. The correlation analysis extends to a representative number of flights (>100) for a specific A/C configuration and is repeated cyclically.

- **Task 3: PPS/stress correlation**

Correlation of the pilot parameters utilized for IAT with stress at FCAs, calculated from the measured local strain. The correlation analysis extends, as in Task 2, over a representative sample.

4.2 Selected aircraft tracking

The flight recorders are distributed on a statistically representative basis throughout the individual squadrons and register within the RPS the flight parameter spectrum of selected A/C (Figure 5). The derivation of the FPS takes place similarly to TAT, although the recordings do not contain strain-gauge parameters. The generation of the appropriate stress patterns for all FCAs is carried out via the FPS/stress regression table (Task 2 in TAT).

- **Task 4: Flight hour-related damage**

Conversion of the strain history calculated into a frequency of occurrence matrix and, based on that, the calculation of accumulated damage. The applicable safe damage rate per flight hour is calculated on a statistical basis for each squadron.

- **Task 5: Usage spectra survey**

The frequency of occurrence matrices, in a similar manner to Task 4 for selected parameters and stresses, is stored as a basis for the evaluation of further correlated components and structural areas.

- **Task 6: Damage comparison of parameter sets**

The first step simulates IAT, where a stress spectrum is generated with the PPS/stress regression table. In the second step, for the same data sequence, a stress frequency of occurrence matrix is generated with the FPS/stress regression table. In the third step, the

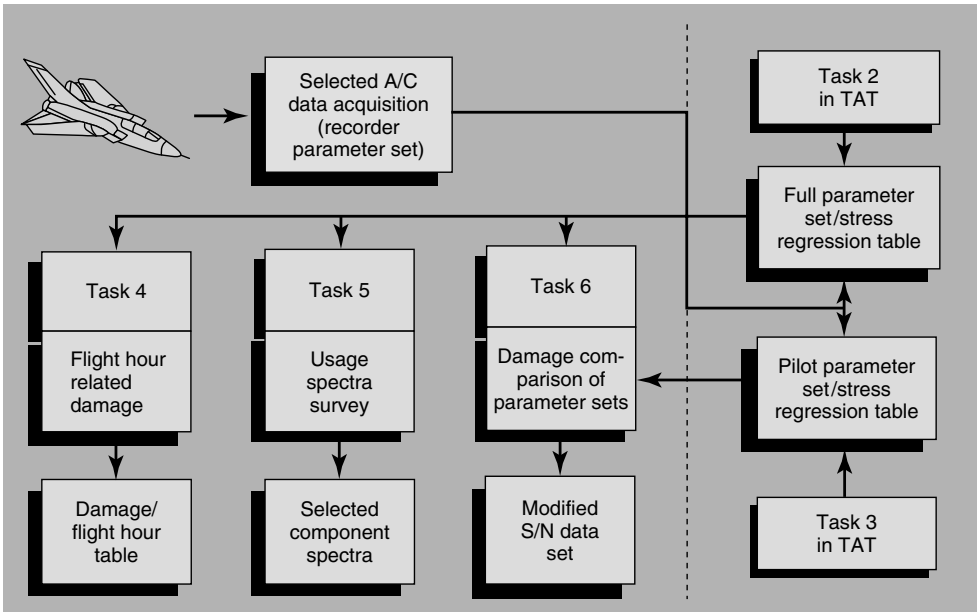


Figure 5. Tornado, selected aircraft tracking (SAT).

resulting damages are compared at individual stress levels. Out of that procedure, the modified stress vs. number of cycles (S/N) data set is generated in order to make the FPS knowledge available for the individual A/C.

Task 4 corresponds to the classic form of SAT, i.e., if the A/C cannot be individually monitored with reference to a specific FCA. This necessarily occurs when the IAT pilot parameters are blind to the stress variation of the FCA. In this case, the scatter of the SAT damage/flight hour is lower compared to the damage calculated with the pilot parameters.

4.3 Individual aircraft tracking

From the conceptual point of view, the IAT permits optimum utilization of the inherent structural life. This naturally presupposes that appropriate sensors (pilot parameters) exist in the individual A/C for the acquisition of stress data. The pilot parameter signals acquired by IAT are converted via the appropriate stress regression table (Task 3 in TAT) into FCA stress spectra and transformed into the individual damage index via the modified S/N data set (Task 6 in SAT).

5 TORNADO USAGE MANAGEMENT

5.1 LEDA Tornado

The Tornado usage management is supported by a software tool named LEDA. LEDA itself is a database, which contains all necessary major inspection actions (including depot inspection) and modification actions for each A/C. The LEDA database is frequently updated with the SHM information from each A/C, which comprises

- flight hours
- A/C configurations
- flown missions
- number of landings
- overall load spectra
- accumulated damage for each FCA.

With the LEDA-internal algorithms, the long and midterm planning of the necessary inspections and modifications for each A/C can be carried out. Several measures can thereby be synchronized for an optimized maintenance schedule.

Beside the LEDA internally processed statistics about the SHM, input data and other calculated data support the fleet management in several decisions. In addition, the LEDA database contains the planning data for other associated programs (Section 6).

5.2 Management actions

As described before, the main aim for A/C monitoring is the optimization of several fleet maintenance tasks, which includes the following:

- Nonperiodic inspection planning (Figure 6)

Planning of usage-dependent inspections of the A/C primary structure. This affects mainly FCAs for which a preventive modification is too expensive at this time and the repair risk is low. The structural safety is thereby established by inspection. The planning of the inspections comprises the initial inspection (threshold) and the follow-on inspections (interval).

For the midterm planning of the nondestructive testing (NDT) inspections at wing box lower panel station XY, Figure 6 shows the relevant A/C tail numbers per calendar year.

- Modification planning (Figure 7)

This planning comprises the necessary preventive modifications of A/C primary structure. Such preventive modifications need to be planned in the long term, because the embodiment requires extensive prerequisites. This includes actions like

  - planning the necessary budget
  - procurement of mod kits
  - adaptation of A/C deployment
  - synchronization with other maintenance actions
  - supply of the required maintenance resources.

Owing to the high number of modification prerequisites, the above planning has to be carried out years before the life of the unmodified structure has been consumed. Bearing that in mind, a load monitoring system has, compared to a condition-monitoring system, the capacity to support long-term planning.

LEDA Tornado

	Midterm actions (calendar year)																			
A/C type: GS/GT	Squadron: XX		Component: R/H wing				Part: Wing box lower panel station XY										Action: NDT inspection			
Number of A/C																				
16																				
15																				
14																				
13																				
12																				
11																				
10																				
9																				
8																				
7																				
6																				
5																				
4	A/C 1																			
3	A/C 2																			
2	A/C 3		A/C 5									A/C 7								
1	A/C 4		A/C 6															A/C 8	A/C 9	
Quarter	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
A/C per quarter	4	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
Year	2008				2009				2010				2011				2012			
A/C per year	6				0				1				0				2			

Figure 6. Midterm actions (calendar year).

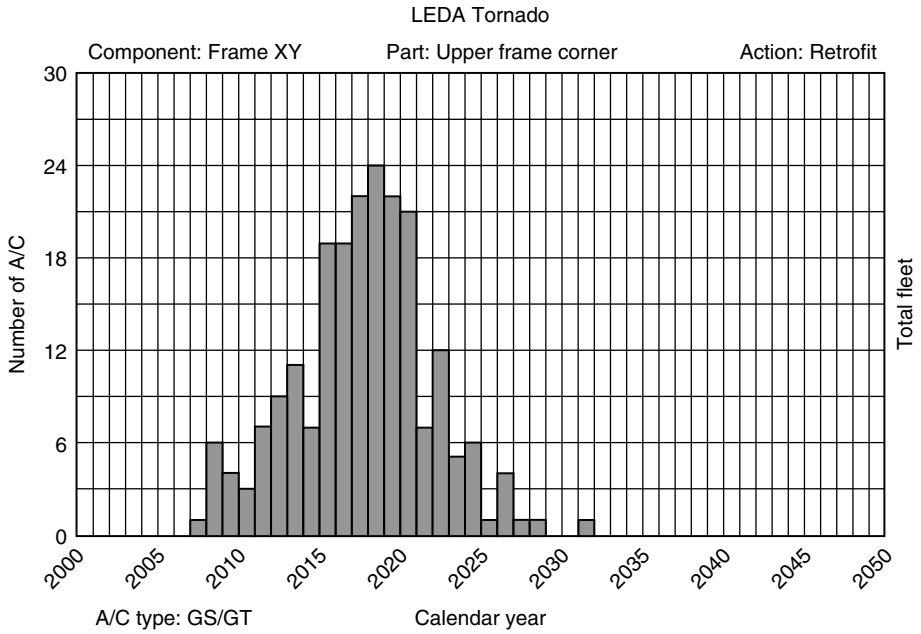


Figure 7. Long-term actions (calendar year).

As an example, Figure 7 shows the long-term planning for the frame XY, upper frame corner retrofit action. The number of retrofit actions per calendar year is presented by graphical methods.

- Fleet management
  - With A/C degradation information, the fleet manager is able to synchronize the operational planning with the maintenance requirements.
  - The A/C degradation information and the maintenance schedule support the fleet manager in the decision on which A/C should be used for specific missions, especially for foreign deployment.
  - If possible, from the A/C configuration point of view, the A/C are rotated between squadrons from time to time based on their accumulated damage. The reason behind that is to reach a balanced degradation over the fleet.

- Out-of-service planning

If for some reasons, A/C have to be retired, then the accumulated damage of the FCAs is one of the selection criteria.

6 ASSOCIATED TASKS

6.1 Analytical condition inspection (ACI)

In addition to the normal periodic inspections, the analytical condition inspection (ACI) program has been defined for the structure and associated equipment to

- find damage in areas, which are normally not inspected due to the required inspection depth or accessibility, sufficiently early, and
- collect information on the technical condition of the fleet leader A/C and to extrapolate the inspection results for follow-on decisions about necessary activities (e.g., modifications) on all A/C.

The ACI program comprises a specific number of A/C per year. The ACI inspection areas are associated with the main structure components (front fuselage, center fuselage, aft fuselage, wing, fin, taileron, and undercarriage).

The LEDA database is the source for the in-depth knowledge of the structural degradation of each A/C. On the basis of this, the ACI candidates are selected (Figure 8).

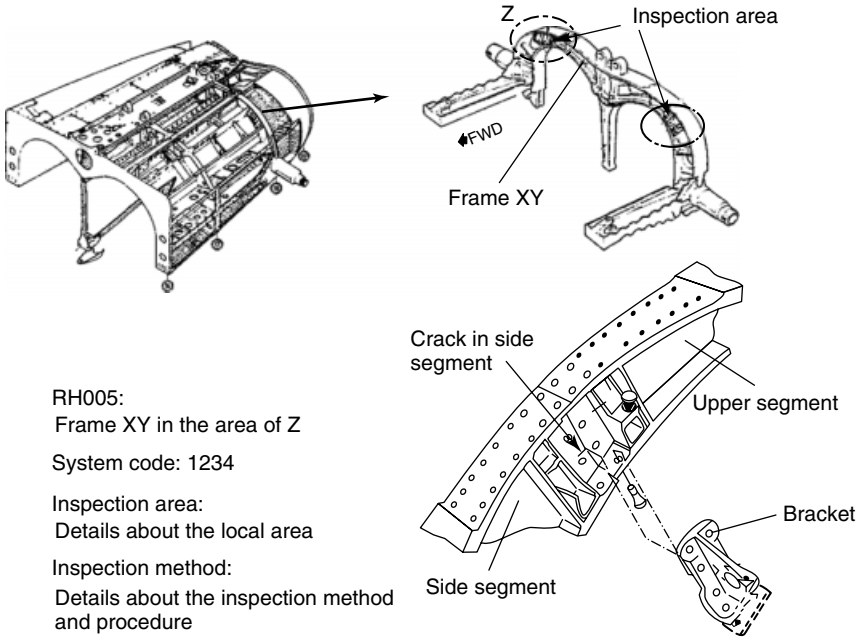


Figure 8. Tornado, analytical condition inspection (ACI).

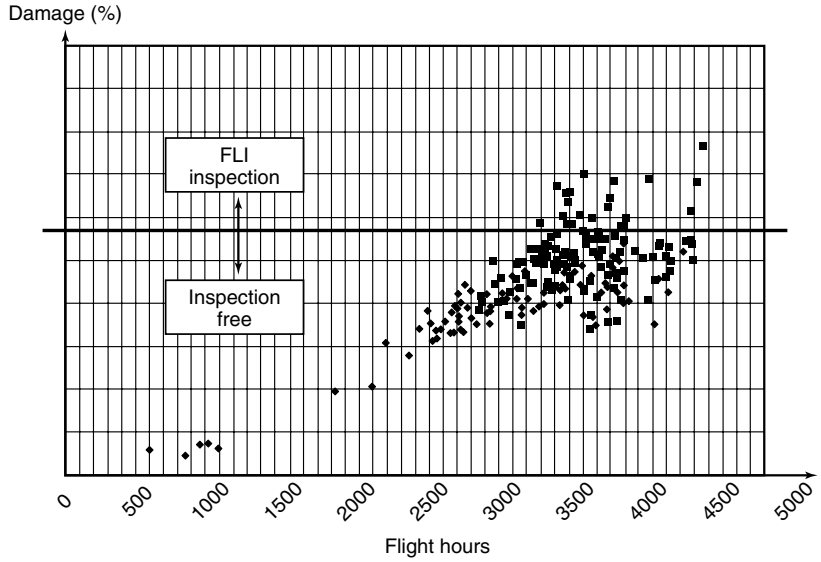


Figure 9. FLI—fleet leader inspection.

6.2 Fleet leader inspection (FLI)

Specific FCAs and their surrounding structure are inspected in the fleet leader inspection (FLI) fleet leader sampling program, which has to be carried out in predefined intervals together with other periodic inspections. The selection of the fleet leader A/C for a specific FCA is based on the in-depth knowledge of the individual FCA damages generated by the SHM system.

To incorporate the influence of several scatter factors, the FLI inspections will be carried out on the most damaged 10% of all A/C. If these A/C are crack-free, it is assumed that the whole fleet is crack-free in this area and the LEDA algorithms are on the safe side (Figure 9).

6.3 Life extension

Germany has decided to extend the Tornado usage from 4000 to 8000 flight hours. This decision was based on the detailed information in the LEDA database about the structure degradation of each individual A/C.

On the basis of LEDA, the necessity for equipment requalification has been derived from the comparison of the actual service usage spectra with the original qualification spectra. If necessary, a requalification has been carried out.

With this knowledge and supported future usage, the main body of the fleet can reach this target with a minimum maintenance overhead.

7 SHM—COST AND BENEFIT COMPARISON

As explained before, the Tornado SHM is carried out primarily for reasons to economize the maintenance effort, e.g., for inspections and modifications. Figure 10 shows a good example for the necessary modification effort of the wing and center fuselage over the complete usage period. The figure differentiates between the costs "without SHM" and "with SHM".

- **Without SHM**

According to the results of fatigue testing, the structural components had to be modified very early within the original design life of 4000 flight hours.

- **With SHM**

Owing to the fact that the in-service FCA stress spectra of the already mentioned structural components are lower than the design spectra, the A/C usage

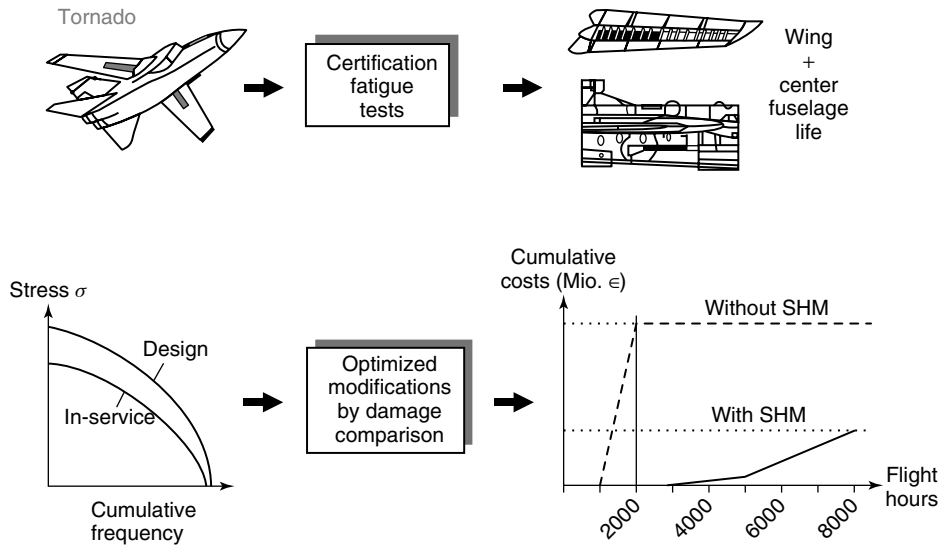


Figure 10. SHM: cost/benefit comparison.

could be extended significantly with very limited modification costs. Even with the life extension to 8000 flight hours, the necessary modifications can be reduced to required packages for the A/C concerned.

This particular example illustrates that modification costs for the case "with SHM" will probably be only one-third of the "without SHM" case. Even if the nonrecurring and recurring costs are considered in addition to the "with SHM" modification costs, the final cost is still considerably less than that for the "without SHM" case.

RELATED ARTICLES

Lamb Wave-based SHM for Laminated Composite Structures

Damage Detection Using Piezoceramic and Magnetostrictive Sensors and Actuators

Damage Measures

Principles of Structural Degradation Monitoring

Loads Monitoring in Aerospace Structures

Military Aircraft

Usage Management of Civil Structures

Commercial Fixed-wing Aircraft

Agile Military Aircraft

Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft

SHM and Lifetime Management of Industrial Piping Systems

FURTHER READING

Ahrensdorf K. *Tests on Details and Components*, Agardograph No. 241, 1978.

Chang FK. *Structure Health Monitoring Current Status and Perspectives*. Technomic Publishing Company: Lancaster, 1997.

Department of Air Force, *Military Specification – Airplane Strength and Rigidity, Flight Loads*. MIL-A-008861 B (USAF), December 1994.

Department of the Air Force, *Military Specification – Airplane Strength and Rigidity Reliability, Requirements, Repeated Loads and Fatigue*, MIL-A-008866 C (USAF), May 1960.

Ministry of Defence Fatigue, *Safe Life Substantiation*, Def Stan 00-970, Part 1/4, Section 3 Leaflet 35, January 2006.

Schütz W. *Calculation Methods for Fatigue Life and Crack Propagation*, Agardograph No. 231, 1978.

Ward AP. *The Development of Fatigue Management Requirements and Techniques*, AGARD-CP-506, 1991.

Watson P, Dabell BJ. Cycle counting and fatigue damage. *Journal of the Society of Environmental Engineers* 1976 **9**:3–8.

Wheeler OE. *Crack Growth under Spectrum Loading*, Report FZM-5602. General Dynamics, 1970.

Zgela MB, Madley WB. *Durability and Damage Tolerance Testing and Fatigue Life Management: A CF18 Experience*, AGARD CP-506, 1991.

Chapter 93
Usage Management of Civil Structures

Ayaho Miyamoto
Graduate School of Science and Engineering, Yamaguchi University, Ube, Japan

1 INTRODUCTION

In bridge structures, there are many different unforeseen conditions on which we do not have enough information, because bridges are larger and often need to serve longer—more than 100 years—compared with the other products, such as those in the fields of electrical, mechanical, and systems engineering. And they are subjected to diverse types of deterioration mechanisms such as corrosion, fatigue, carbonation, alkali-aggregate reactions, etc. Although in design code we are forced to consider some parameters or factors affecting structural behavior, there are even more items that we cannot consider them practically. Therefore it is probable to have some risk for not fulfilling the complete standards of safety, so some failure probability is possible. Structural aging, environmental conditions, and reuse are examples of circumstances that could affect the reliability and the life of a structure. Engineers have been visually inspecting, monitoring, and proof testing bridges for centuries. However, presently health and performance are described based on subjective measures that are not universal. In addition, defects, deterioration, and damage are not discovered until it is possible to visually observe the signs they exhibit at which time these would have taken their toll on health. These shortcomings impact the timeliness, effectiveness, and the reliability in any management decision irrespective of any sophistication in the management process. Moreover, even experienced engineers may find visual signs of defects, deterioration, and damage and still not be able to diagnose the causative mechanisms, or their impact on the reliability of the system and global health. The global health of an entire bridge as a system, inclusive of the performance criteria corresponding to each of the limit states, is actually what is needed for effective management decisions. There are needs of periodic inspections to detect deterioration resulting from normal operation and environmental

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

attack or inspections following extreme events, such as strong-motion earthquakes or hurricanes. To quantify these system performance measures, there should be some means of monitoring and evaluating the integrity of civil structures while in service [1]. Then, we need to develop a practical health monitoring system for detecting deterioration phenomenon as early as possible for bridge maintenance.

2 DEFINITION OF HEALTH, PERFORMANCE, AND HEALTH MONITORING OF BRIDGE SYSTEMS

Health can be defined as the reliability of a bridge structure to perform adequately for the required functionalities. Some of these functionalities are as follows [2]:

- utility
- serviceability and durability
- safety and stability of failure at ultimate limit states and
- safety at conditional limit states.

If P_f denotes the probability that the bridge may fail to perform successfully under any likely demand,

$$\text{Health} = (1 - P_f) \text{ for all limit} - \text{state demands/ throughout the life cycle/given as-is condition and symptoms/given the operational and maintenance management} \quad (1)$$

It is not possible to quantify health and reliability of a bridge system for many of the limit states without extensive data that we often do not have. On the basis of a general definition, monitoring is the frequent or continuous observation or measurement of structural conditions or actions [3]. There is another definition that gives more detail: structural health monitoring (SHM) is the use of *in situ*, nondestructive sensing, and analysis of structural characteristics, including the structural response, for detecting changes that may indicate damage or degradation [4]. Figure 1 shows the basic components of a typical structural health monitoring system.

Monitoring is usually carried out in order to achieve one or several goals [2]. The most important goals are as follows:

- structural management
- increase of safety and
- knowledge improvement.

3 STRUCTURAL (BRIDGE) MANAGEMENT

The most safe and durable structures are usually those that are well managed. Measurement and monitoring have an essential role in structural management. The data resulting from the monitoring program is used to optimize the operation, maintenance, repair, and replacing of the structure based on reliable and objective data. Detection of ongoing damage can be used to detect deviations from the design performance. Monitoring data can be integrated in structural management systems and increases the quality of decisions by providing reliable and unbiased information. Many structures are in much better conditions than expected. In these cases, monitoring allows to increase the safety margins without any intervention on the structure. Taking advantage of better material properties, overdesign, and synergetic effects, it is possible to extend the lifetime or load-bearing capacity of structures. A small investment at the beginning of a project can lead to considerable savings by eliminating or reducing overdesigned structural elements.

A few structures might present deficiencies, which cannot be identified by visual inspection or modeling. In these cases, it is possible to increase safety and to decrease managing costs by taking actions before it is too late. Repair will be cheaper and will cause less disruption to the use of the structure if it is done in time. Monitoring can also reduce insurance costs.

The economic impact of structural deficiency is twofold: direct and indirect. The direct impact is reflected by costs of reconstruction, while the indirect impact involves losses in the other branches of the economy. The complete collapse of historical monuments, such as old stone bridges and cathedrals, represents an irretrievable cultural loss for society.

Malfunctioning of civil structures often has serious consequences. The most serious is an accident

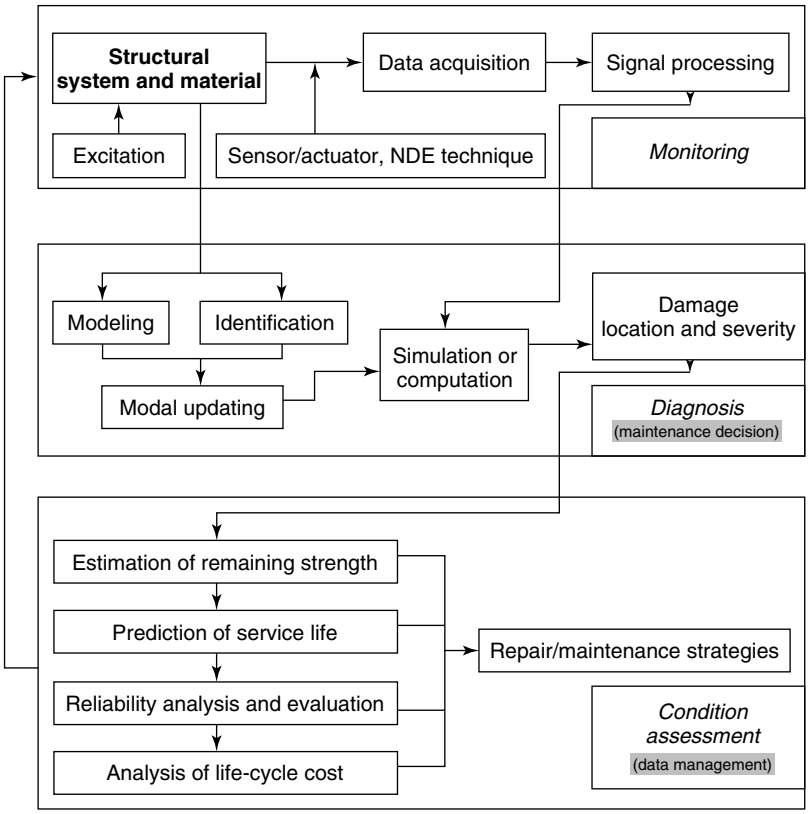


Figure 1. Basic components of a typical structural health monitoring system.

involving human victims. Even when there is no loss of life, populations suffer if infrastructure is partially or completely out of service. Collapse of certain structures, such as nuclear power plants, may provoke serious ecological pollution.

Having permanent and reliable monitoring data from a structure can help to guarantee the safety of the structure and its users.

Learning how a structure performs in real or laboratory conditions will help to design better structures for the future. This can lead to cheaper, safer, and more durable structures with increased reliability and performance. Structural diversity due to factors such as geographical region, environmental influences, soil properties, loads, etc., makes absolute behavioral knowledge impossible: there are no two identical structures.

A good way to enlarge knowledge of structural performance is to monitor their behavior. That is why monitoring during the complete lives of structures, from construction to the end of service, is of interest from the theoretical point of view as well as from the point of view of structure management. Theories need to be tested, and have a better understanding of their health conditions.

In bridge structures, a strategic life-cycle management aided by the latest information technologies, such as network-based database system, intelligent health monitoring, multimedia virtual reality, artificial intelligence, etc., has to be developed to increase their safety and have a better understanding of their health conditions.

4 TECHNOLOGY ISSUES FOR SHM

At present, there are many themes for doing research on SHM. The most important topics are as follows [5, 6]:

- Optic fiber sensing
  - distributed sensors
  - point sensors
  - smart and advanced sensors
- Remote sensing and wireless sensor networks
- Data acquisition
- Data analysis
- Signal processing
- Structural/system identification (SI)
- Damage detection/diagnosis and damage localization
  - forced vibration-based damage detection
  - wind-induced vibration-based damage detection
  - ambient vibration-based damage detection
- Neurocomputing methods for SI and damage detection
  - artificial neural networks (ANNs)
  - fuzzy logic
- Nondestructive structural condition methods and tests
- Structural modeling and simulation by finite-element analysis methods
- Model updating, safety evaluation, reliability, and decision making
- Intelligent materials and structures
- Database and management systems
- System integration
- Health monitoring for special bridges
  - medium-span bridges
  - long-span bridges
    - suspension
    - cable-stayed
- Inclusion of health monitoring system in bridge management system
- Health monitoring of existing and/or new bridge structures
- Life-cycle performance design
- Health monitoring of strengthened bridges with new material or methods like Fiber Reinforced Plastics (FRP).

As is obvious, SHM is a very vast domain of research. It provides benefits to design engineers and practitioners who will have a better understanding of structural behavior and can make feasible decisions. SHM is mainly done to achieve the following objectives:

- detecting the existence of damage;
- finding the location of damage;
- estimating the extent of damage; and
- predicting the remaining fatigue life.

The practical method to investigate the structural condition is getting some results from dynamic responses by using dynamics-based methods. The attractiveness of dynamic responses is due to the fact that we are able to detect and locate damage using them. Dynamic responses are directly related to the global behavior of structure and they can provide rapid inspection of large structural systems. The dynamics-based methods can be divided into four groups:

1. spatial-domain methods
2. modal-domain methods
3. time-domain methods and
4. frequency-domain methods.

Spatial-domain methods use changes of mass, damping, and stiffness matrices to detect and locate damage. Modal-domain methods use changes of natural frequencies, modal damping ratios, and mode shapes to detect damage. In the frequency-domain method, modal quantities such as natural frequencies, damping ratio, and model shapes are identified. In time-domain method, system parameters were determined from the observational data sampled in time.

Moreover, one can use model-independent methods or model-referenced methods to perform damage detection using dynamic responses presented in any of the four domains. The literature shows that model-independent methods can detect the existence of damage without much computational efforts, but they are not accurate in locating damage. On the other hand, model-referenced methods are generally more accurate in locating damage and require fewer sensors than model-independent techniques, but they require appropriate structural models and significant computational efforts. Although time-domain methods use original time-domain data measured using conventional vibration measurement equipment, they require certain structural information and massive computation and are case sensitive. Furthermore, frequency- and modal-domain methods use transformed data, which contain errors and noise

due to transformation. Moreover, the modeling and updating of mass and stiffness matrices in spatial-domain methods are problematic and difficult to be accurate. There are strong development trends that two or three methods are combined together to detect and assess structural damages. For example, several researchers combined data of static and modal tests to assess damages. The combination could remove the weakness of each method and check each other. It suits the complexity of damage detection.

SHM is also an active area of research in aerospace engineering, but there are significant differences in the methods used in aerospace engineering, mechanical engineering, and civil engineering in practice. For example, because bridges, as well as most civil engineering structures, are large in size, and have quite low natural frequencies and vibration levels, at low amplitudes, the dynamic responses of bridge structure are substantially affected by the nonstructural components, and changes in these components can easily to be confused with structural damage. Moreover, the level of modeling uncertainties in reinforced-concrete bridges can be much greater than in the single beam or a space truss. All these make the damage assessment of complex structures such as bridges a still challenging task for bridge engineers.

One of the important problems facing SHM is that very little is known about the actual stress and strains in a structure under external excitations. For example, the standard earthquake recordings comprise recordings of motions of the floors of the structure and no recordings of the actual stresses and strains in structural members. There is a need for special sensors to determine the actual performance of structural members.

SHM requires integrated sensor functionality to measure changes in external environmental conditions, and signal-processing functionality to acquire, process, and combine multisensor and multimeasured information. Individual sensors and instrumented sensor systems are then required to provide such multiplexed information.

4.1 Sensor types

1. Sensor types in structural control and their applications in civil engineering are summarized. In general, there are different kinds of sensors for on-line monitoring, such as piezoelectric transducer, optical fiber sensors, and embedded Brag grating sensors. The reliability and durability of these sensors have been testified in many large bridges and tall buildings. Unlike many mechanical systems, typical civil engineering structures are often large in size and therefore have very low natural frequencies. In addition, the vibration level of the structural responses is quite low very often except under strong earthquake. Therefore, the sensors of a monitoring system must be able to work in a very low-frequency range and they must have a large dynamic measurement range. The industry has made great achievements in developing sensor and is still working forward [4].
2. ISIS Canada, a pioneer research center in the SHM field, has published a report that includes useful information about sensors and sensing systems. It is a very good reference about sensors [7].
3. Now, a new type of sensor, maximum and cumulative displacement memorizing sensor, has been introduced. This sensor is able to measure the maximum, minimum, present, and cumulative displacement without any battery. The sensor system consists of linear potentiometer, rotation potentiometer, transmitter, and noncontact reader (receiver). This sensor system is suitable for memorizing the deformation of aseismic dampers. This system has been installed in several buildings, and its effectiveness is being verified [8].

4.2 Optimum sensor arrangement

The estimation of the parameter values involves uncertainties due to limitations of the mathematical models used to represent the behavior of the real structure, the presence of measurement error in the data, and insufficient excitation and response bandwidth. Although intensive development continues on innovative sensor systems, there is still considerable uncertainty in deciding on the number of sensors required and their location in order to obtain adequate information on structural behavior. In particular, the choice of the number and the location of the sensor in the structure has a major influence on the quality, or

equivalently the uncertainty, of the model parameter estimation. Because complete modal data is impossible for a large flexible structure and measurements yield only partial mode shapes with respect to the total degree of freedom (DOF) corresponding to the finite-element model (FEM), a common practice to bridge the gap is to expand the measured mode shapes or the reduction of freedom in FEM.

Unfortunately, this process unavoidably introduces consequential errors and increases the difficulty in damage detection. One alternative method is to use the measured incomplete mode shapes to detect damage. Then the additional information collected can be used advantageously for damage detection.

1. Methods have been developed to place sensors in an optimal fashion to address the identification and control of dynamic structures by some researchers [9].
2. An effective independence algorithm based on the contribution of each sensor location to the linear independence of the identified modes is proposed. The initial candidate set of sensor locations was quickly reduced to the number of available sensors [10].
3. The effective independence method is extended in an algorithm where sensor placement is achieved in terms of the strain-energy contribution of the structure [11].
4. A methodology for optimum sensor locations is proposed for parameter identification in dynamic systems [12].
5. A rational statistical-based approach is developed for the optimal location of sensor based on Fisher's information matrix (FIM) for the model parameters [13].
6. Using the expected Bayesian loss function involving the trace of the inverse of the FIM, the above-mentioned rational statistical-based approach is extended to treat the case of large model uncertainties expected in model updating [14].
7. An optimal sensor placement is reported for the purpose of detecting structural damage. The prioritization of sensor locations is based on an eigenvector sensitivity analysis of a FEM of the structure [15, 16].
8. Two methods are proposed to determine the optimal or near optimal positioning of sensors. The optimal sensors location is proposed on the FEM associated to the structure to be tested. The first method of location of sensors emphasizes the minimization of the noise effect, and the estimate of the modal coordinates is found in a least-squares sense. The second method is based on the observability Gramian and the optimal sensors location had to ensure observability requirements [17].
9. A method is presented in which the sensor locations are prioritized according to their ability to localize structural damage based on the eigenvector sensitivity method. Numerical examples and test results show that this approach is effective for detecting structural damage directly using optimum and incomplete test modes [18].
10. Measurement selection is proposed in terms of two factors, namely the sensitivity of a residual vector to the structural damage and the sensitivity of the damage to the measurement noise. The advantage of the proposed technique is that it is based on the undamaged state of structure and thus independent of the damaged configuration. Therefore, it is applicable in practice to determine the measurement selection prior to field testing and damage identification analysis [19].
11. A probabilistic advancing cross-diagnosis method is proposed in diagnosis decision making for SHM. It is experimented in the laboratory using a coherent laser radar system and a Charge Coupled Devices (CCD) high-resolution camera. Results show that this method is promising for field application [20].
12. Another new idea is that neural network (NN) techniques are used to place sensors. In an attempt the NN and methods of combinatorial optimization are used to locate and classify faults [21].
13. It has been shown that relatively low-resolution (as compared with the acceleration data) displacement data is now becoming available through Global Positioning System (GPS) technologies, and also potentially through optical technologies [22].
14. A new sensor system is designed for measuring damage indexes and memorizing the data using mechanical memory. Theoretical and experimental studies exhibited excellent performance of the proposed sensor [23].

15. By extensive simulation and experiments for sensor network, it is investigated and verified that the higher the oversampling frequency, the broader the bandwidth and the higher the signal-to-noise ratio (SNR) [24].
16. The problem of locating sensors on a bridge structure is considered with the aim of maximizing the data information so that structural dynamic behavior can be fully characterized. Six different optimal sensor placement techniques, three based on the maximization of the FIM, one on the properties of the covariance matrix coefficients, and two on energetic approaches, are investigated. Mode-shape displacements are taken as the measured data set and two comparison criteria are employed. The first criterion is based on the mean square error between the FEM and the cubic spline interpolated mode shapes. The second criterion measures the information content of each sensor location to investigate on the strength of the acquired signals and their ability to withstand the noise pollution keeping intact the information relative to the structure properties. The results show that the effective independence driving-point residue (EFI-DPR) method provides an effective method for optimal sensor placement to identify the vibration characteristics of the studied bridge. The variance method (VM) gives results very close to the EFI-DPR technique, in terms of the capability to capture the vibration mode shape and signal strength. However, the VM presents unique characteristics in the world of the Optimal Sensor Placement (OSP) techniques, which is the indication of the optimal number of sensors (ONS) [25].

The static and dynamic data are collected from all kinds of sensors that are installed on the measured structures. And these data are processed and usable information is extracted. So the sensitivity, accuracy, locations, etc., of sensors are very important for damage detection. The more information is obtained, the damage identification will be conducted more easily, but the price should be considered. Thus the sensors are determined in an optimal or near optimal distribution. In a word, the theory and validation of optimum sensor locations are still being developed.

5 WHOLE LIFETIME MONITORING OF BRIDGES

The importance of whole life-span monitoring is highlighted in this section [26].

5.1 Monitoring during construction of a new bridge

Construction is a very delicate phase in the life of structures. For concrete structures, material properties change through aging. It is important to know whether or not the required values are achieved and maintained. Defects (e.g., premature cracking) that arise during construction may have serious consequences on structural performance. Monitoring data help engineers to understand the real behavior of the structure and this leads to better estimates of real performance and more appropriate remedial actions.

Important information obtained through monitoring during construction includes the following: Estimation of hardening time of concrete in order to estimate when shrinkage stresses begin to be generated; deformation measurements during early age of concrete in order to estimate self-stressing and risk of premature cracking; and when structures are constructed in successive phases, measurement can help to improve the composition of concrete when necessary. In case of prefabricated structures, sensors may be useful for quality control; optimization between two successive phases of pouring due to evaluation of cure in previous phases; for prestressed structure, deformation monitoring of cables helps to adjust prestressing forces and determine the relaxation; monitoring of foundation settlement helps to understand the origins of built-in stresses; damage caused by unusual loads such as thunderstorm or earthquakes during construction may influence the ultimate performance of structures; optimal regulation of structural position during erection; and knowledge improvement and recalibration of models. The installation of a monitoring system during the construction phases allows monitoring to be carried out during the whole life of the structure. Since most structures have to be inspected several times during service, the best way to decrease the costs of monitoring and inspection is to install the monitoring system from the beginning.

5.2 Monitoring after refurbishing, strengthening, or enlargement of bridge

Material degradation and/or damage are often the reasons for refurbishing existing structures. Also, new functional requirements for the bridge (e.g., enlarging) lead to requirements for strengthening. If strengthening elements are made of new concrete, a good interaction of new concrete with the existing structure has to be assured. Early age deformation of new concrete creates built-in stresses and bad cohesion causes delamination of the new concrete, thereby erasing the beneficial effects of the repair or strengthening efforts.

Since new concrete elements observed separately represent new structures, the reasons for monitoring them are the same as for new structures, presented in previous subsection. The determination of the success of refurbishment or strengthening is an additional justification.

5.3 Monitoring during testing of bridge

Bridges have to be tested before service for safety reasons. At this stage, the required performance levels of structures have to be reached. Typical monitored parameters (such as deformation, strain, displacement, rotation of section, and cracks opening) are measured. Tests are performed in order to understand the real behavior of the structure and to compare it with theoretical estimates. Monitoring during this phase can be used to calibrate numerical models describing the behavior of structures.

5.4 Monitoring during service of bridge

The service phase is the most important period in the life of a structure. During this phase, construction materials are subjected to degradation by aging. Concrete cracks and creeps, and steel oxidizes and may crack due to fatigue loading.

The degradation of materials is caused by mechanical (loads higher than theoretically assumed) and physicochemical factors (corrosion of steel, penetration of salts and chlorides in concrete, freezing of concrete, etc.). As a consequence of material degradation, the capacity, durability, and safety of structure decrease.

Monitoring during service provides information on structural behavior under predicted loads, and also registers the effects of unpredicted overloading. Data obtained by monitoring are useful for damage detection, evaluation of safety, and determination of the residual capacity of structures. Early damage detection is particularly important because it leads to appropriate and timely interventions. If the damage is not detected, it continues to propagate and the structure no longer guarantees required performance levels. Late detection of damage results in either very elevated refurbishment costs or, in some cases, the structure has to be closed and dismantled. In seismic areas monitoring is very critical. Subsequent auscultation of a structure that has not been monitored during its construction can serve as a basis for understanding of present and for prediction of future structural behavior.

5.5 Monitoring during dismantling of bridge

When the structure no longer responds to the required performances and the costs of reparation or strengthening are excessively high, the ultimate life span of the structure is attained and the structure should be dismantled. Monitoring helps to dismantle structures safely and successfully.

6 ANALYSIS TECHNIQUES

The most important issue in practical SHM is, how do we make syntheses assessment for existing bridge condition state based on the collected monitoring data, as the final goal of monitoring?

Recent research and implementation of SHM and damage assessment can be summarized in two main methods:

1. laboratory testing methods and
2. field-testing researches.

Bridge testing methods are mainly static and/or dynamic. Dynamic testing includes ambient vibration testing and forced vibration testing. In ambient vibration testing, the input excitation is not controllable.

The loading could be any load from environment like wind or passing vehicles among others. This type of testing is of interest to the researchers or engineers owing to the convenience of measuring the vibration response while the bridge is being used and also because of the increasing availability of robust data acquisition and storage systems. Since the input is unknown, certain assumptions have to be made. Forced vibration testing involves application of input excitation of known force level at known frequencies. The excitation manners include electrohydraulic vibrators, force hammers, vehicle impact, etc. The static testing in the laboratory may be conducted by actuators and by standard vehicles in the field testing. Here are some recent findings from laboratory and field-testing research on the damage assessment of bridge structures:

1. Slab-on-girder bridges are stiffer than the corresponding calculated values. The floor systems of steel truss bridges may contribute substantially to the combined stiffness of the structure. In most cases, the actual load-carrying capacities are higher than those from calculations [27].
2. Prestressing the concrete deck slab in the vicinity of the pier supports eliminated transverse cracking of the slab, enhanced the natural frequencies, and increased the fatigue life as well as the ultimate load-carrying capacity [28].
3. The ambient vibration method provided approximately the same resonant frequencies and mode shapes as those used in modal analysis [29].
4. Changes in modal frequency and damping (except mode shapes) can be good damage indicators [30].
5. Modal flexibility provided a relevant/reliable measure of structural state [31–34].
6. Changes in the flexibility matrix can be effectively used in detecting and locating damages [35].
7. Ambient excitations were used for the modal testing, which is generated by pulling a model car along the bridges. The results showed that the energy transfer ratio (ETR) is a good indicator of structural damage [36].
8. A modal testing package by using simplified experimental modal analysis and time-domain identification method is also beneficial for monitoring and damage detection system [37].
9. This work represented the first attempt in using the wavelet estimation technique directly on transient data and nor on the impulse response estimates obtained via the random decrement technique [38].
10. The test results are transformed to both strain influence lines and modal flexibility, which have been demonstrated to be a conceptual, quantitative, comprehensive, and damage-sensitive signature [39].
11. A comparison of the ultrasonic pulse velocity (UPV) damage index with the normalized acoustic emission counts revealed that the two methods had different sensitivities at different stages of loading and could potentially complement each other as a hybrid damage assessment tool [40].
12. The integration of a nondestructive damage detection method with an on-site data-acquisition system is done to remotely monitor a conventional concrete slab bridge and a composite bridge utilizing Carbon Fiber Reinforced Plastics (CFRP) and Glass Fiber Reinforced Plastics (GFRP) and evaluate their performance [41].
13. Monitoring system is feasible for the *in situ* monitoring of the stayed cable tension [42].
14. Inclusion of virtual reality and Internet for the development of a platform for SHM shows some benefits in reduction of labor in bridge management systems by automation [43].
15. A flexible videogrammetric technique for structural dynamic measurement by commercial-grade digital cameras. It can capture structural 3-D dynamic trajectories at low cost, to be used for some SI problems. The laboratory results show that this method is capable of providing accepted accuracy for dynamic response measurement, and proved to be a good complement to traditional sensors especially in 3-D vibration measurement [44].

There are many other research works that are similar in some senses to the above-mentioned points. The above-mentioned points are only some result examples that have been found recently. Studying these points show the following results:

- The models in the laboratory are mainly beams, columns, truss, and/or frame structures.

- The location and severity of damage in the models are determined in advance.
- The testing has demonstrated lots of performances of damage structures.
- The field testing and damage assessment of real bridges are more complicated than the models in the laboratory.
- The correlation between the damage indicator and damage type, location, and extent still needs further improvement.
- Advanced numerical methods are capable to help the health monitoring system evaluations.
- There are new technologies emerging in health monitoring systems such as introducing new sensors and sensor networking.
- Multimedia technologies can be useful in health monitoring systems.

The bridge health monitoring and damage detection are both concerned with two fundamental criteria of the bridges, namely,

- the physical condition and
- the structural function.

In structural dynamics, these fundamental criteria can be treated as mathematical models, such as response models, modal models, and physical models. Instead of taking measurements directly to assess bridge condition, the bridge monitoring system and damage detection evaluate these conditions indirectly by application of mathematical models. The health monitoring and damage detection are active areas of research in recent years. For example, numerous papers on these topics appear in the Proceedings of International Modal Analysis. In many journals and conferences, there are many papers and reports about analytical development for bridge health monitoring systems. Some of the methods that are reported in them are as follows:

6.1 Signature analysis and pattern recognition approach

A modal model is characterized by a set of modal parameters, which can be extracted from response model by modal testing techniques. Traditionally, the major modal parameters are natural frequencies, damping ratios, and mode shapes. The modal model can be also used as a vibrational signature. For example, in mechanical engineering, condition monitoring of rotating equipment is typically done by looking for signature changes in a power spectrum of the measured vibrations. The same nonparametric approach could be used for civil structures, but it is more typical to use identified modal parameters to provide the signature characterizing the structure. In order to not only detect damage but also locate its position, observed changes in the signature must be compared with a database of possible changes and the most likely change must be selected. This is a type of pattern recognition where the database of "pattern" is generated by analyzing various damage scenarios or "failure modes". The representative researches on damage detection through a modal model are briefly summarized below.

One approach to detect damage has been to use changes in the modal frequencies. With fiber-reinforced plastics, damage can be detected from a decrease in natural frequencies and from an increase in damping [45]. Performed experiments on a highway bridge demonstrate that the decrease in natural frequencies can be used to detect the presence of damage [46]. The problem of understanding when it is sufficient to measure and use only natural frequencies, thus avoiding the need to measure modal shapes in vibration beams, or beam systems is addressed. The identification procedure was based on the minimization of an objective function that accounts for the difference between the analytical and experimental quantities [47]. Further study demonstrated that the observed changes in natural frequencies, especially the changes in fundamental natural frequencies, were unable to determine the location of crack damage. This occurs because a certain amount of damage at two different locations may produce the same amount of frequency change [48].

Sensitivity analysis has been proposed to improve the sensitivity of natural frequency change to the structural damage. The basic idea behind this was to compare the frequency changes obtained from experimental data collected on the deteriorated structure with the sensitivity of the modal parameters obtained from an analytical FEM of the structure. Accuracy of sensitivity-based methods is dependent on the quality of the FEM used to computer the sensitivities. It should be kept in mind that obtaining an accurate analytical model in itself remains a difficult task.

The uncertainties of analytical model may influence the results of damage detection [30]. Results from some experimental and numerical studies have suggested that the lower vibration modes would probably be suited for damage detection. Using the information from the mode shapes, a method to localize damage by using the pattern-recognition method is reported. The study was on a beam model with known mode shapes, and then generated mode shapes at any location using interpolation. The location of damage compared fairly well with finite-element (FE) analysis.

Finally, this method was applied to the real bridges including 49.7 m plate girder bridges and a two-span simply supported truss bridges (the length of each span is approximately 61.3 m). It was concluded that the method can accurately locate damage though the damage pattern was not quite distinctive [49].

The combination of different modal parameters, especially the combination of natural frequency and mode shapes, has been used by several researchers. It is found that crack propagation in a beam can cause substantial shifts in certain frequencies and mode shapes can be used to locate the damage [50]. With the help of analytical beam models, it is possible to show that the use of changes in the curvature mode shapes is capable to detect and locate damage [51]. A damage indicator called *curvature damage factor* is introduced, in which the difference in curvature mode shape for all modes can be summarized in one number for each measured point. It is applied to a real prestressed concrete bridge, named *Z24*, which crosses the highway A1 between the Bern and Zurich in Switzerland [52]. Another combination in terms of natural frequencies, mode shapes, and modal assurance criteria (MAC) was employed on a scale bridge model test. It is concluded that natural frequencies should be used to detect damage, and mode shapes and MAC values can be further used to identify damage locations [53]. It is possible to compare the transfer function parameter change of the testing system to detect damage and locate the position by using a few sensors [54].

Transmittance functions (TFs) and the sensor–actuator system to detect, locate, and assess damages on a composite beam are introduced. Further work is now under way by using sequential TFs to detect damage on large panel and blade structures using a dense pattern of measurements formal scanning laser Doppler vibrometer (LDV) [55]. Closely spaced frequencies, which cannot be distinguished if the traditional peak picking technique is used, can be identified successfully by eigensystem realization algorithm (ERA). But user expertise is indispensable in this type of analysis [56]. An algorithm to construct a proportional flexibility matrix (PFM) is presented by introducing a dummy structure in relevancy with the real structure.

The PFM is established within a scalar to the flexibility matrix, and the scalar is shown to be the first modal mass. Vectors method is implemented with the PFMs and a multidamage scenario is successfully identified for a simple system. For the algorithm developed, only a very small number of modes are needed to calculate the PFMs with sufficient accuracy and only one reference degree with unchanged mass after damage is needed to make PFMs of predamaged and postdamaged structures to be comparable [57]. The residual wavelet force (RWF), which is expressed in terms of the wavelet transform coefficients of structural-free vibration responses of the damaged structure and the undamaged baseline mass and stiffness matrices, is defined, and then a damage location indicator based on the residual wavelet force (DLIRWF) is presented. The results of the numerical simulations suggest that the proposed method has the ability to localize multiple damages with various extents occurred in structures. Furthermore, the method is well suited to noise-contaminated structures and thus is very promising in practical applications. Moreover, good damage localization results can be obtained when different decomposition scales or different mother functions are used [58].

Changes in the power spectral density (PSD) curvature due to the presence of structural damage have been investigated. The experimental results obtained from a steel bridge model and bookshelf structure demonstrate the usefulness of the changes in PSD curvature as a diagnostic parameter in detecting the damage and locating its position [59]. The results of modal testing present that global low-order natural frequencies of frame structure are insensitive to the local change structural stiffness. The position of stiffness change of frame structure can be detected by the maximum energy high-order mode method proposed in this article with a good accuracy [60].

A damage assessment algorithm is developed for mass density and Young's modulus varied system.

The Latin hypercube sampling (LHS) technique for Monte Carlo simulation is an effective way to deal with variations [61]. Since an optimal and unique algorithm cannot be proposed for damage detection depending on the variety of applications, a multi-algorithm procedure seems to be the only reliable approach for damage detection in the field of continuous static monitoring. For example wavelet transforms identity damage initiation only and not its persistency in time. But proper orthogonal decomposition (POD) procedure overcomes this limitation, because it is able to detect correlation changes between sensors in adjacent sections [62].

A new signal-energy-based damage index is suggested, which releases the limitation of accurate measurement requirement of structural modal parameters involved in the traditional damage indicators. The results demonstrate that the new damage index is capable to identify the existence and location of structural damage of both simple damage (one damage location) scenario and complicated damage (multiple damage locations) scenario [63, 64].

This kind of vibrational signature analysis has been proven to be successful in localizing damage. However, it is not sensitive to most types of damage that occur in bridge structures. Model testing and field testing have shown that the changes of natural frequencies due to local damage are very small, and that mode shapes (especially higher mode shapes) are sensitive to the changes of local stiffness but it is very difficult to measure them accurately. There are similar problems in other vibration signatures, such as mode-shape curvature (MSC), modal flexibility, MAC, etc. None of these can provide sufficient information for the detection of both small and large defects. The successful applications of these modal model methods may rely on the development of test techniques and new findings of model-based approaches.

6.2 System identification (SI) approach

System identification (SI) is the process of constructing or updating an accurate mathematical model of a system based on input and output (I/O) observations. Among other applications, SI can be applied to SHM and damage assessment, e.g., by determining the structural stiffness values and comparing them with previously determined values or originally intended values. Research interest in this subject area has increased steadily over the years.

In the context of civil engineering structures, the first attempt was to carry out SI study by means of a recursive least-square algorithm [65]. Two case studies of state estimate are used to illustrate the use of the Kalman filter and the extended Kalman filter (EKF) [66]. Two SI algorithms, namely, the EKF and iterated linear filter smoother are applied to identify the hydrodynamic coefficient matrices for an offshore structure problem [67]. A weighted global iteration algorithm is proposed to improve the convergence characteristics of the EKF process [68]. The above-mentioned method is subsequently applied in the study of a running load on a beam [69]. In another research, the structural parameters of a damage bridge structure are identified by the EKF [70]. In another investigation, a bridge structure is studied by the EKF. It is also done by another filter-based SI approach considering the incorporation of a memory fading function [71]. Most of the SI studies in structural engineering have dealt with few DOFs and few unknown structural parameters. In practice, however, modeling of engineering structures often requires the contrary. The difficulty and the computational effort required for convergence increase drastically when the numbers of DOFs and unknowns increase. To this end, various means have been proposed in recent years to tackle the numerical problems generally associated with SI of large systems. A substructural identification method is formulated to improve the convergence performance by decomposing the structural system into several smaller subsystems [72–74]. A two-step damage detection and health monitoring approach has been developed for large and complex structures with a limited number of measurements. The first step is initial damage detection, based on the optimal-updating techniques and changes of stiffness. The second is detailed damage detection by the design sensitivity method and linear perturbation theory [75, 76].

In all the above-mentioned works, classical SI techniques were used, such as EKF, recursive least squares, instrumental variable, and maximum likelihood methods; these methods, in one way or another, search the optimal solution by exploiting the previous solution. Treating the problem as an inverse problem, many classical methods require the use of secant, tangent, or higher-order derivatives of

the objective function. As the system of unknowns grows in size, the numerical difficulty increases and often the convergence becomes extremely difficult, if not impossible. Such "exploitation" methods perform point-to-point search and have the danger of converging to local optima. On the other extreme, a random search (e.g., trial and error) may be used to explore the entire search space. To overcome one trial solution with another, an error norm has to be defined as a measure of deviation of the estimated response (computed based on the estimated parameters) from the actual (measured) response. The search continues until the error norm is deemed to be small. Such a blind "exploration" strategy is obviously too time consuming for large systems owing to huge number of possible combinations.

For instance, if there are 10 unknowns to be identified and each unknown is divided into 100 discrete values within its search range, there will be a total of 10^{20} possible combinations—an astronomical figure to work with even for today's powerful computers.

In this regard, a worthwhile attempt is to employ evolutionary algorithms, which have proved in the last decade to be a powerful search and optimization tool. The main features of these algorithms are that they attempt to imitate living things and are stochastic in nature. There are presently four main approaches, namely, genetic algorithms (GAs), evolutionary programming, evolutionary strategies, and simulated annealing. By far the most widely known approach in engineering is perhaps GA. This approach was developed to solve discrete or integer optimization problems as opposed to continuous parameter optimization problems. In the case of parameter identification, this can be tuned into an advantage of controlling the resolution of identified parameters through the (integer) length of the chromosome (number of bits) as explained below.

A GA search is conducted in the modal domain of a much smaller dimension than the physical domain. The objective function was defined based on the estimated modal response in time domain and the corresponding modal response was transformed from the measured response. This method had been shown to work well in terms of mean error (10–15%) for a fairly large system with 50 DOFs and 52 unknown parameters [77]. A GA with real number encoding is applied to identify the structural damage by minimizing the objective function, which directly compares the changes in the measurements before and after damage. Three different criteria are considered, namely, the frequency changes, the mode-shape changes, and a combination of two. A laboratory tested cantilever beam and a frame are used to demonstrate the proposed techniques; numerical results show that the damage elements can be detected by GA, even when the analytical model is not accurate [78].

Structural SI within the linear regions has been well developed and many techniques have been applied to structural damage assessment. However, the question of whether a structure is still linear after the damage remains. This is very important because the dynamical behavior of a nonlinear system can be quite different from those of its associated linear system.

Also if the structural system becomes nonlinear after damage, its dynamical characteristics cannot be estimated by using the linear SI methods.

An attempt has been made to develop methods for the identification of highly localized structural damage in weak nonlinear structures. The damage was defined as either a reduction of stiffness or a change of restoring force characteristics. The location vector method (LVM) is applied to identify the location and type of damage. The fast Fourier transform (FFT) and the least-squares method are used to quantify the damage [79]. The NN is employed to detect the changes in nonlinear systems [80–82]. A frequency-domain modal analysis technique is formulated to apply to weaker nonlinear multidegree of freedom (MDOF) systems. One of the advantages of the method is the ability to determine the response of the nonlinear system at any level once its variable modal parameters has been identified at some reference force level [83]. An adaptive on-line parameter identification algorithm based on the variable trace approach for the identification of nonlinear hysteretic structures is presented. At each time step, this recursive least-square-based algorithm upgrades the diagonal elements of the adaptation gain matrix by comparing the value of the estimated parameter between two consecutive time steps. The effectiveness and efficiency of the proposed algorithm is shown by considering the effects of excitation amplitude, the measurement units, larger sampling time interval, and measurement noise [84]. The vibrations of a clamped beam for two different kinds of nonlinearity are investigated. First, the beam shows

a nonlinear behavior characterized by a piecewise linear stiffness and second, the nonlinearity comes from a bilinear stiffness. The performance of the restoring force surface method together with both the numerical and experimental results is demonstrated [85].

Obviously, the nonlinear SI will be developed by many researchers in the not too far future. When performing vibration tests on civil engineering structures, such as bridges, it is often unpractical and expensive to use artificial excitation (shakers and drop weights). Ambient excitation on the contrary is freely available (wind and traffic). This output-only SI now becomes more and more important [86–88].

Although the regularization increased the popularity of parameter identification owing to its capability of deriving a stable solution, the significant problem is that the solution depends upon the regularization parameters chosen as mentioned below.

A generalized model of differential hysteresis that contains 13 control parameters is developed. Three identification algorithms are developed to estimate the control parameters for different classes of inelastic structure. These algorithms are based upon the simplex, EKF, and generalized reduced gradient method. Novel techniques such as global search and internal constraints are incorporated to facilitate convergence and stability. Effectiveness of the devised algorithms is demonstrated through simulations of two inelastic systems with both pinching and degradation characteristics in their hysteretic traces [89]. A GA with real number encoding is applied to identify the structural damage by minimizing the objective function, which directly compares the changes in the measurements before and after damage. Three different criteria are considered, namely, the frequency changes, the mode-shape changes, and a combination of the two. The technique does not seek to tune the analytical FEM to obtain an improved one in the undamaged and damaged state, but rather to update the FEM so that its model data changes equal the measured modal data changes as closely as possible. Therefore, an accurate analytical model is not needed in the analysis. With proper weights to frequency and mode-shape data, it also accurately detects damages in the symmetric portal frame by using both the frequency changes and the mode-shape changes [90].

In structural SI, different mathematical models introduce different explanations on the result of identification even with the same set of input/output data. The model inaccuracy in structural SI can be categorized into two items:

- the uncertainty due to nonlinear model and
- the completeness of model description (or extract description).

Selecting the exact model becomes one of the important issues for identification. Changes of modal parameters contain the information of structural variations. Thus the sensitivities of four modal parameters, namely, natural frequency, modal shape, the slope, and the curvature of modal shape to flexibility increase are important parameters that should be analyzed [91–93].

A common theme in using SI for SHM and damage diagnosis is to use a model updating approach. Usually, highly accurate and detailed FEMs are required to analyze and predict the dynamical behavior of complex structures during analysis and design. Once the FEM of a physical system is concentrated, its accuracy is often tested by comparing its modes of vibration and frequency response with those obtained from the physical system. If the correlation between the two is poor, then assuming that the experimental measurements are correct, the analytical model must be adjusted so that the agreement between the analytical predictions and the test results is improved. The updated model may then be considered a better representation of the physical structure than the initial analytical model. Any observed local decrease in the stiffness of the model is assumed to indicate the location and severity of damage in the monitored structure. The updated model can subsequently be used with reasonable accuracy to assess the stability and control characteristics and to predict the dynamical responses of the structure. The above process of correcting the system matrices is known as *model updating*.

The methods for FEM update that are used for Nondestructive Evaluation (NDE) can be divided into in the following major categories: mode flexibility methods, optimal matrix update methods, sensitivity-based matrix update methods, eigenstructure assignment methods, changes in measured stiffness methods, combined modal parameters methods, etc. All of these FEM update techniques require that

the user select a subset of the measured modes to be correlated with the corresponding modes of the FEM. Normally, the first few modes of the structure are used in the FEM correlation because they are generally the best identified modes. However, in some situations the higher frequency modes are critical to the location of structural damage, and so it is necessary to include them in the set of modes for FEM correlation. Many modes that are lower in frequency do not undergo significant modification as a result of the damage, so that they contribute to the computational burden without contributing significantly to the location of the damage. The number of modes is limited not only by the computational burden but also by the inherent ill-conditioning and statistical bias associated with large-order update problems. Because of this limit, it is important to have systematic criteria for selecting the modes that are most indicative of the structural damage [94–97].

7 DAMAGE DETECTION METHODS

7.1 Statistical analysis methods

FE modeling provides a complete set of analytical and theoretical modal parameters for a structure, but these parameters are usually not accurate. The experimental data is accurate to some extent, but incomplete, and also intervolved by the noise. Any method to do modal updating must address the mismatch between the level of information in the detailed analytical FEM and the relatively sparse information.

A general Bayesian statistical approach is presented, which treats the uncertainties that arise from measurement noise, modeling error, and possible nonuniqueness in the problem of updating the stiffness distribution [98, 99]. This approach is extended to multiple damage locations [100, 101]. The above approach is used for on-line monitoring, wherein specified modal parameters are identified on a regular basis and the probability of damage for each substructure is continually updated [102].

New approaches that use two set of measured frequency response data to update the analytical system mass and stiffness parameters in order to improve the agreement between the dynamical behaviors of the analytical and actual systems are developed. The algorithm adjusted model without iteration [103]. A similar method to identify multiple damage locations of multistory frame structures and reinforced-concrete bridge column is employed [100, 101].

A reliable damage detection algorithm is presented for framed structures, whose stiffness properties can be explicitly expressed with those of members, by introducing a regularization technique for SI, a parameter grouping technique for locating damaged members and overcoming the sparseness of measured data, a data perturbation method for obtaining statistical distributions of system parameters with a set of noise-polluted measured data, and a statistical approach by a hypothesis test for damage assessment [104].

Unlike most references that focus on the different methods for extracting damage-sensitive features from vibration response measurements, a statistical pattern-recognition paradigm to quantifying the observed changes in these features is introduced. Various projection techniques such as principal component analysis (PCA) and linear and quadratic discriminate operators with the Statistical Process Control (SPC) are employed in an effort to enhance the discrimination between features from the undamaged and damaged structures [105].

A general methodology for structural fault detecting using fuzzy logic is presented, based on the monitoring of the static, eigenvalue, and dynamic responses. Fuzzy logic coupled with principles of continuum damage mechanics is used to identify the location and extent of the damage. This methodology represents a unique approach to damage detection that can be applied to a variety of structures used in civil engineering, machine, and aerospace applications [106]. A probabilistic approach is presented, which examines the eigenvalue problem from a statistical standpoint by considering eigenvalue and eigenvector uncertainty, along with a correlated analytical stochastic FEM to assess the damage. The effectiveness of the proposed technique is illustrated using simulated data on a 3 DOF spring-mass system and on an Euler–Bernoulli cantilever aluminum beam [107].

A Bayesian probabilistic framework for modal updating is adopted and a new probabilistic approach that use the statistic properties of an estimator of the spectral density to obtain expressions for the updated probability density function (PDF) of the modal

parameters is proposed. Examples of single degree of freedom (SDOF) systems and MDOF systems using simulated data are presented to illustrate the proposed method [108, 109].

A statistical method with combined uncertain frequency and mode-shape data for structural damage identification is proposed. The FEM is updated by comparing the measured vibration data before and after damage occurs. The effects of uncertainties in both the measured vibration data and the FEM are considered as random variables in model updating. The statistical variations of the updated FEM are derived with perturbation method and Monte Carlo technique. The probabilities of damage existence in the structural members are then defined. The comparison of results between the calculation and testing show that all the damages are identified correctly with high probabilities of damage existence [110].

The vibration-based SHM problem is addressed as the double task of detecting damages modeled as changes in the eigenstructure of a linear dynamic system and localizing the detected damages within (a FEM of) the monitored structure. The proposed damage detection algorithm is based on a residual generated from a stochastic subspace-based covariance-driven identification method and on the statistical local approach to the design of detection algorithms. This algorithm basically computes a global test, which performs a sensitivity analysis of the residuals to the damages, relative to uncertainties and noises. Damage localization is stated as a detection problem. This problem is addressed by plugging aggregated sensitivities of the modes and mode shapes with respect to FEM structural parameters in the above setting. This results in directional tests, which perform the same type of damage-to-noise sensitivity analysis of the residual as for damage detection [111].

A damage detection method is proposed for SHM under varying environmental and operational conditions. The method is based on PCA applied to vibration features identified during the monitoring of the structure. The advantage of the method is that it does not require to measure environmental parameters because they are taken into account as embedded variables. The number of principal components of the vibration features is implicitly assumed to correspond to the number of independent environmental factors. Since the environmental effects may be effectively eliminated by the proposed procedure, the residual error of the PCA prediction model remains small if the structure is healthy, and it increases significantly when structural damage occurs. Novelty analysis on the residual errors provides a statistical indication of damage. The environmental conditions are assumed to have a linear (or weakly nonlinear) effect on the vibration features, and the PCA-based damage detection method is illustrated using computer-simulated and laboratory testing data [112].

A method based on vector autoregressive volatility (ARV) models is proposed. These models accurately capture the predictable dynamics present in the response. They leave the unpredictable portion, including the component resulting from unmeasured input shocks, in the residual. An estimate of the autoregressive model residual series standard deviation provides an accurate diagnosis of damage conditions. Additionally, a repeatable threshold level that separates damaged from undamaged conditions is identified, indicating the possibility of damage identification and localization without explicit knowledge of the undamaged structure. Similar statistical analysis applied to the raw data necessitates the use of higher-order moments that are more sensitive to disguised outliers, but are also prone to false indications resulting from overemphasizing rarely occurring extreme values [113].

A damage detection method of mechanical system based on subspace identification concepts and statistical process techniques is presented. The aim is to propose a method that is sensitive to small-sized structural damages and suitable for on-line monitoring. Measured time responses of structures subjected to artificial or environmental vibrations are assembled to form the Hankel matrix, which is further factorized by performing singular-value decomposition to obtain characteristic subspaces. It may be demonstrated that the structural responses are mainly located in the active subspace defined by the first principal components, which is orthonormal to the null subspace defined by the remaining principal components. If no structural damage occurs, the orthonormality relation between the subspaces remains valid with small residues when consecutive data sets are compared, and these residues may be evaluated by

the proposed damage indicators. The method is validated using an experimental mock-up of an airplane subjected to different levels of damages simulated. It is also applied in environmental vibration testing of a street lighting device to monitor structural fatigue evolution [114].

A stochastic output error (OE) vibration-based methodology for damage detection and assessment (localization and quantification) in structures under earthquake excitation is introduced. The methodology is intended for assessing the state of a structure following potential damage occurrence by exploiting vibration signal measurements produced by low-level earthquake excitations. It is based upon (i) stochastic OE model identification, (ii) statistical hypothesis testing procedures for damage detection, and (iii) a geometric method (GM) for damage assessment. The methodology's advantages include the effective use of the nonstationary and limited duration earthquake excitation, the handling of stochastic uncertainties, the tackling of the damage localization and quantification subproblems, the use of "small" size, simple and partial (in both the spatial and frequency bandwidth senses) identified OE-type models, and the use of a minimal number of measured vibration signals. Its feasibility and effectiveness are assessed via Monte Carlo experiments employing a simple simulation model of a six story building. It is demonstrated that damage levels of 5 and 20% reduction in a story's stiffness characteristics may be properly detected and assessed using noise-corrupted vibration signals [115].

A novel analytical tool for early detection of fatigue damage in polycrystalline alloys that are commonly used in mechanical structures is presented. Time series data of ultrasonic sensors have been used for anomaly detection in the statistical behavior of structural materials, where the analysis is based on the principles of symbolic dynamics and automata theory. The performance of the proposed method has been evaluated relative to existing pattern-recognition tools, such as NNs and PCA, for detection of small changes in the statistical characteristics of the observed data sequences. This concept is experimentally validated on a special-purpose test apparatus for 7075-T6 aluminum alloy specimens, where the anomalies accrue from small fatigue crack growth [116].

Model updating within a statistical framework appears to be a promising general approach to damage diagnosis and SHM of large civil structures in view of the inescapable data and modeling uncertainties. But many aspects require further research, including optimal location of sensors, the type of damage, which can be reliably detected and located using a given array of sensors on a structure, and strategies for making decisions about possible damage and determining the corresponding probabilities of false alarm and missed alarms, etc.

7.2 Damage index methods

There are some research works using damage index methods. Some of them are as follows:

It is proved that the ratio of the model frequency change between any two models is the function of the damage location only. The ratios were then used as damage indicators, which were calculated from a candidate set of assumed possible damage scenarios. The structural damage was then localized by comparing the predicted ratios with the ratios computed based on measured modal frequencies [117–121].

The modal flexibility involves functions of both the natural frequencies and mode shapes. Some researchers have found experimentally that modal flexibility can be a more sensitive parameter than natural frequencies or mode shapes for structural monitoring and damage detection in bridges [35, 122, 123].

The sensitivity is studied theoretically by comparing the use of natural frequencies, mode shapes, and modal flexibilities for monitoring. The results demonstrate that modal flexibilities are more likely to indicate damage than either natural frequencies or mode shapes [124, 125].

Using the ratio of change in model strain energy in each element can be considered as another damage indicator. The approach requires only the elemental stiffness matrix, the analytical mode shapes, and the incomplete measured mode shapes. The effect of analytical mode truncation, incomplete measured mode, and measurement noise in the damage detection are discussed. Results from the modal simulation and experiment with a two-story partial steel frame indicate that the presented method is effective in localizing damage, but it is noise sensitive in the damage quantification to some extent [18, 126–132].

A sensitivity and statistical-based method is presented to localize structural damage by direct use of incomplete mode shapes. The method is an extension of the multiple damage location assurance criterion (MDLAC) reported in [133], by using incomplete mode instead of model frequency. In general, the damage detection strategy localizes the damage sites first by using incomplete mode shapes, and then detects the damage sites and extent again by using measured natural frequencies, which have a better accuracy than mode shapes [134].

A new modal parameter, the ETR, based on the complex damping theory, is presented and it is proved theoretically that ETR indicator can be much more sensitive to structural damage [135, 136].

A new process of modal parameter identification based on complex modal energy measurement (including the ETR index) is proposed. The damage growth measurement is performed by using the proposed diagnostic technique based on ETR in large-scale structures. The ETR index has been investigated through real steel bridge as a sensitive damage indicator, but it has not been applied on the concrete bridge structures [137].

A method to identify the location of damage in civil engineering structures, which is based on changes in the component transfer functions of the structure, or the transfer functions between the floors of a structure, is proposed. Multiple damage locations can be identified and qualified using the proposed approach. Experimental verification of this approach using a four-story frame structure in the Washington University Structural Control and Earthquake Engineering Lab is also provided [54, 138].

A comparative study of applying various mode-based indices to the structural damage detection of the Tsing Ma suspension bridge with a main span of 1377 m and an overall length of 2160 m is presented. Five mode-based damage indices, including coordinate modal assurance criterion (COMAC), enhanced coordinate modal assurance criterion (ECOMAC), MSC, and modal strain-energy index (MSEI), and modal flexibility index (MFI) are applied respectively for the damage location identification of various simulated damage scenarios in the bridge by 3-D finite-element method. The numerical simulation results show that the applicability and the performance of each index depend on the damage type concerned. On the basis of the performance evaluation, the preferred damage indices in accordance with different damage types are recommended [79].

Structural damage from a known increase in the fundamental period of a structure after an earthquake or prediction of degradation of stiffness and strength for a known damage is estimated. A modified Clough–Hohnston SDOF oscillator is proposed to establish reliable correlations between the response functions in the case of a simple elastic-plastic oscillator. The proposed model has been used to demonstrate that ignoring the effects of aftershocks in the case of impulsive ground motions may lead to unsafe designs [139].

A damage identification technique at an element level is proposed. The element damage equations have been established through the eigenvalue equations on the basis of the changes in frequencies and mode shapes of vibration. Several solution techniques are discussed and compared. Numerical results show that the nonnegative least-squares method can lead to satisfactory results in most cases. An experimental program of the reinforced-concrete beam under static and dynamic loading is used to demonstrate the identification scheme. The adaptation of the FEM is required in this technique [140].

A technique to localize damage in structures is presented. Central to the approach is the computation of a set of vectors, designated as damage locating vectors (DLVs) that have the property of inducing stress field whose magnitude is zero in the damaged elements. The DLVs are associated with sensor coordinates and are computed systematically as the null space of the change in measured flexibility. Numerical simulations carried out with realistic levels of noise and modeling error illustrate the robustness of the technique [141].

A numerical study of the relationship between damage characteristics and the changes in the dynamic properties is presented. It is found that the rotation of mode shape is a sensitive indicator of damage. The numerical results show that the rotation of mode shape has the characteristic of localization at the damaged region even though the displacement modes are not localized. Also, the results illustrate that the rotations of modes are robust in locating multiple damage locations with different sizes in a structure. Furthermore, using the changes in the rotation of

mode shape does not need very fine grid of measurements to detect and locate damage effectively [142].

A methodology to nondestructively locate and estimate the size of damage in structures, for which a few natural frequencies or a few mode shapes are available, is presented. First, a frequency-based damage detection (FBDD) method is outlined. A damage-localization algorithm to locate damage from changes in natural frequencies and a damage-sizing algorithm to estimate crack size from natural frequency perturbation are formulated. Next, a mode-shape-based damage detection (MBDD) method is outlined. A damage index algorithm to localize and estimate the severity of damage from monitoring changes in modal strain energy is formulated. The FBDD method and the MBDD method are evaluated for several damage scenarios by locating and sizing damage in numerically simulated prestressed concrete beams for which two natural frequencies and mode shapes are generated from FEMs. The result of the analyses indicates that the FBDD method and the MBDD method correctly localize the damage and accurately estimate the sizes of the cracks simulated in the test beam [143].

A method to identify and quantify damage in structures, called *residual error method* in the movement equation, is evaluated by a numerical analysis to verify its efficiency when applied to continuous beams and frame structures. This method is based on the alteration, produced by damage, in the dynamic properties of the structures. The location of the damage is performed by observing the error in the movement equation of the intact structures. The structures are discretized in finite elements and the damage is introduced by a stiffness and area reduction of the elements' cross sections. The observation of the obtained results and then its comparison with some other damage detection methods demonstrated that the residual error method in the movement equation is efficient in the damage location and quantification of the studied structures [144].

A procedure for locating variability in structural stiffness is presented. For some types of structures, this variability is directly related to manufacturing defects and/or in-service damage. Unlike many published damage detection methods, the procedure presented here uses only data obtained from the damaged structure. Baseline data and theoretical models of the undamaged structure are not used during the analysis presented here. The procedure locates regions in a structure where the stiffness varies. Only if it is known that the structure, in its undamaged state, is homogeneous with respect to stiffness, the procedure will detect the areas of inhomogeneity that are caused by the incipient damage. For nonhomogenous structures, some knowledge of the structural details (for example, engineering drawings or a baseline test) is required in order to discriminate damage. The procedure is a two-dimensional generalization of a previously published one-dimensional gapped smoothing method, whereby local features in vibration curvature shapes are extracted using a localized curve fit (i.e., smoothing). A variability index is generated for each test point on the structure. Increased variability is due either to structural stiffness features or damage. A statistical treatment of the indices enables discrimination of areas with significant stiffness variability. Provided that the damaged areas are sufficiently small compared to the total surface area, their indices will be statistical outliers. The procedure can either analyze mode-shape data or frequency-dependent operating displacement shape data. The procedure is demonstrated with a FEM of a plate, and by performing experiments on composite plates with deliberately induced multiple delaminations. Finally, the method is demonstrated on data taken from a large composite hull structure. In all cases the procedure successfully located the damaged regions [145].

A methodology to identify damage in a structure is presented. The method utilizes a new form of damage index based on the changes in the distribution of the compliance of the structure due to damage. The changes in the compliance distribution are obtained using the mode shapes of the predamaged and the postdamaged state of the structure. The validity of the method is demonstrated using numerically generated data from beam structures and experimental data from a free–free beam structure with inflicted damage. In the numerical and experimental examples, the damage identification performance of the proposed method is compared with that of the existing strain-energy-based method. The results of the numerical and experimental studies indicate that the proposed compliance-based damage index method can be used in damage identification of the structure [146].

The wavelet packet transform (WPT) is a mathematical tool that has a special advantage over the

traditional Fourier transform in analyzing nonstationary signals. It adopts redundant basis functions and hence can provide an arbitrary time-frequency resolution. A damage detection index called *wavelet packet energy rate index* (WPERI) is proposed for the damage detection of beam structures. The measured dynamic signals are decomposed into the wavelet packet components and the wavelet energy rate index is computed to indicate the structural damage. The proposed damage identification method is first illustrated with a simulated simply supported beam and the identified damage corresponds with assumed damage. Afterward, the method is applied to the tested steel beams with three damage scenarios in the laboratory. Despite the presence of noise factor in the real measurement data, the identified damage pattern is comparable with the tests. Both simulated and experimental studies demonstrated that the WPT-based energy rate index is a good candidate index that is sensitive to structural local damage [147].

A vibration-based damage evaluation method that can detect, locate, and size damage utilizing only a few of the lower mode shapes is proposed. The proposed method is particularly advantageous for beamlike structures with uncertain applied axial load, mass density, and foundation stiffness. On the basis of a small damage assumption, a linear relationship between damaged and undamaged curvatures is revealed in the context of elasticity. It turns out that the resulting damage index equation inherently suffers from singularities near inflection nodes. The transformation of the problem into the multiresolution wavelet domain provides a set of coupled linear equations. With the aid of the singular-value decomposition technique, the solution to the damage index equation is achieved in the wavelet space. Next, the desired physical solution to the damage index equation is reconstructed from the one in the wavelet space. The performance of the proposed method is compared with two existing damage detection methods using a set of numerical simulations. The proposed method attempts to resolve the mode selection problem, the singularity problem, the axial force problem, and the absolute severity estimation problem, all of which remained unsolved by earlier attempts [148].

A vibration-based damage monitoring scheme to give warning of the occurrence, the location, and the severity of damage under temperature-induced uncertainty conditions is proposed. Experiments on a model plate-girder bridge, for which a set of modal parameters were measured under uncertain temperature conditions, are performed. Then a damage warning model is selected to statistically identify the occurrence of damage, by recognizing the patterns of damage-driven changes in natural frequencies of the test structure and by distinguishing temperature-induced off-limits. A frequency-based damage index method based on the concept of modal strain energy is implemented in the test structure to predict the location and the severity of damage. In order to adjust the temperature-induced changes in natural frequencies that are used for damage detection, a set of empirical frequency correction formulae are derived from the relationship between temperature and frequency ratio [149].

A dynamic-based damage detection method for large structural systems using the Hilbert–Huang transform (HHT) has been proposed. The proposed method has been verified numerically by implementing the scheme on a model of wingbox. In the implementation process, the following steps have been identified as being important: (i) axis-symmetry signal extension methods in order to solve the end effect problem of empirical mode decomposition (EMD), (ii) finite-element modeling for the purpose of establishing the base line, and (iii) a feature index vector for structural damage detection. The obtained results show that the damage feature index vector is more sensitive to small damage. The effect of noise is also considered. Examination of the results confirms that the proposed damage detection method is very robust [150].

There may be other damage indices to indicate the locations and extent of damage. For real civil structures, only one damage index may not be enough. Until now, the relationships between damage type and damage index are not clear. A lot of further research is needed in this area.

7.3 Static data methods

Static parameter estimation is based on measured deformations induced by static loads such as a slowing moving track on a bridge. There are many instances in which static loadings are more economical than dynamic loading. Many applications require

only element stiffness for condition assessment. In these cases, static testing and analysis can prove simple and more cost-effective [151–159].

Since the natural frequencies, mode shapes, and static responses of a structural system are functions of structural parameters, these parameters may be identified by comparing the dynamic and static characteristics predicted from the mathematical model with those values determined by test. One of the consequences of the development of damage is the decrease in local stiffness, which in turn results in changes in some of the responses. It is therefore necessary that the dynamic and static characteristics of the structure be monitored for damage detection and assessment.

On the basis of the above-mentioned concept, an improved method that can identify a FEM of a structure capable of providing structural characteristics that are consistent with those measured in static and dynamic tests (e.g., the curvature of mode and the static displacement data) is proposed. The detection of damage in a member with stronger influence on the higher modes is more difficult. Thus, the use of static displacements obtained by a loading condition that simulate higher modes is proposed as a solution to this problem [160].

A new method for parameter identification is presented based on the strain and displacement data from static testing, in which Gauss–Newton, gradient, and Monte Carlo formulas are compositely employed to solve the ill-condition and uncertainties. Furthermore, based on the formula of the algorithm of static responses, a complex approach is also proposed, where combined static strain and displacement with dynamic response (e.g., mode shape) are used to localize damage and identify the severity of damage. Several algorithms are compositely applied to improve the sensitivity of parameter identification and enhance the reliability of solution process. The static and dynamic responses are utilized to calibrate the confidence of identification [161].

A structural damage identification algorithm using both the static test data and changes in natural frequencies is proposed. A proper definition of measured damage signature (MDS) and predicted damage signature (PDS) is presented and matched to detect the location of damage. After obtaining the possible damage location, an iterative estimation scheme for solving nonlinear optimization programing problems, which is based on the quadratic programing technique, is used to predict the damage extent. A remarkable characteristic of the approach is that it can be directly applied in the cases of incomplete measured data. Two examples are presented and the results show that the algorithm is efficient for the damage identification [162].

The SI method is used to identify structural parameters in a FEM by minimizing the error between measured and analytical computed responses. A regularization scheme is applied to alleviate the ill-posedness of an inverse problem by adding a regularization function to the primary error function. Two different algorithms depending on the type of measured response have been developed to assess damage. Static displacements from static loading and modal data from impact vibration are measured through laboratory experiments on a grid-type model bridge. Damage is simulated by saw-cutting the cross section with various depths and identified as the reduction in the structural stiffness of the elements around the crack. Through the experimental works, the applicability of the SI-based damage assessment algorithms has been rigorously investigated [163].

A structural damage detection algorithm using static test data is presented. Changes in the static response of a structure are characterized as a set of nonlinear simultaneous equations that relate the changes in the static response to the location and severity of damage. Damage is considered as a reduction in the structural stiffness parameter (axial and/or flexural). An optimality criterion is introduced to solve these equations by minimizing the difference between the load vector of the damaged and the undamaged structure. The overall formulation leads to a nonlinear optimization problem with nonlinear equality and linear inequality constraints. A method based on stored strain energy in elements is presented for selecting the loading location. Measurement locations are selected based on the FIM. Numerical results for a plane truss demonstrate the ability of this method to detect damage in a given structure with the presence of noise in the measurements [164].

The estimation of nonlinear autoregressive moving average with exogenous input (NARMAX) models for vibrating multilayer composite plates is presented and used to assess internal delamination in the plates. Both static and dynamic tests are carried out to investigate the nature of the nonlinearity of the composite system. The nonlinear nature and order

of nonlinearity are then approximated, and used to restrict the search space of the potential NARMAX model. After the structure and terms of the NARMAX model are identified, all coefficients are calculated for the intact plate. Delamination in the plate is assessed according to the coefficient variations in the model using the measured input/output data for the damaged system. Results show that the model output prediction is in agreement with the test data. The proposed method is available to discover and assess severity and location of delamination in the composite plates [165].

An on-line and real-time detection system is developed through the concept of inverse analysis. In this system, the detectors are selected based on natural frequencies and static strains whose relations with material properties can be obtained from analytical solution or commercial finite-element software or experimental data. By transferring their relations into training patterns of ANNs, the elastic properties of composite wing structures can be determined on-line with frequency and strain sensors embedded into structures. Test results show that the material properties determined through this on-line system well agree with the values obtained from the conventional testing methods. The difference is that the present method determines the properties on-line and in real time without any specimen being obtained from the structures and tested in the laboratory [166].

Modal updating by finite-element method is often used to identify the changes of damage using static testing data. Because the errors caused by FEM may be greater than changes of damage, the FEMs should be first calibrated using the measured modal properties and experimental data. Only the FEMs are reliable, and the results from modal updating by finite-element methods are valuable.

7.4 Substructure analysis methods

In the model updating approach, it is common only to update stiffness correction factors for selected substructures rather than for individual structural members. The goal is to reduce the number of stiffness parameters to be updated so that the ill-conditioning and nonuniqueness are kept within tolerable levels. Considering smaller substructures where damage has occurred is desirable so that better localization and assessment of its severity can be performed as follows:

A substructural approach to estimate the stiffness and damping coefficients from the measurements of dynamic responses is proposed. The structures are decomposed into several smaller subsystems for which state and observation equations are formulated and solved by EKF method with a weighted global iteration algorithm [72]. A research work is reported on the substructural identification in frequency domain for the identification of frequency-dependent systems such as soil–structure interaction systems [167]. A substructural identification method is proposed using autoregressive and moving average with stochastic input NARMAX model and the sequential prediction error method. Since the damage locations are not known *a priori*, adaptive substructuring is useful [74]. A damage detection and assessment algorithm is developed based on the parameter estimation with an adaptive parameter grouping scheme from static response [159]. A SI-based approach for analysis and diagnosis of structures under operating conditions is developed. Of interest in this work is the separation of diagnosis into global damage alarm and damage detection. A simplified algorithm is presented for measurement of the statistical likelihood of damage. This statistical test does not attempt to quantify potential damage, but only provides an intelligent alarm, which takes into account all individual changes of modal frequencies and shapes and compares them to their confidence domain to evaluate whether the changes might be significant. The global alarm concept is perhaps more achievable than damage detection for complex and uncertain civil structures [168]. Two complementary methods are reviewed for model-based damage detection with applications, i.e., the substructural flexibility method and the substructural transmission zeros method [169]. A computational procedure for extracting substructure-by-substructure flexibility from global frequencies and mode shapes is presented. The proposed procedure appears to be effective for structural applications such as damage localization and FEM reconciliation [170, 171]. A damage identification algorithm termed as *constrained submatrix factor adjustment* is proposed. Further this algorithm is extended by using both static and dynamic measurements [172].

For damage detection and condition assessment of large and complex structural systems, substructural identification may be an effective way.

7.5 Neural network approaches

The model updating approach described in the last subsection is based on a parametric structural model. Health monitoring techniques may rely on nonparametric SI approaches, in which *a priori* information about the nature of the model is not needed.

Nonparametric models can be used to detect damage, although it is more difficult to use them for localization of damage. NNs are one of the nonparametric identification approaches that have been receiving growing attention recently. NNs do not require information concerning the phenomenological nature of the system being investigated, and they also have fault tolerance, which makes them a robust means for representing model-unknown systems encountered in the real world. NNs do not require any prior knowledge of the system to be identified. It can treat both linear and nonlinear systems with the same formulation. A number of investigators have evaluated the suitability and capabilities of these networks for damage detection purposes.

NNs are trained to recognize the frequency response characteristics of healthy and damaged structures in which the properties of individual members are adjusted to reflect varying levels of damage [173, 174]. A FEM is used to develop failure patterns that are used to train a NN so that it can later diagnose damage in the reference structure [175]. A NN approach is presented based on mapping the static equilibrium requirement for a structure in a finite-element formulation, with the assumption that structural damage is reflected in terms of stiffness reduction. All of these exploratory studies indicate that NNs offer a powerful tool for assessing the condition of structures with inherent damage [176].

Another study complements the work of other investigators by concentrating on a class of problems where knowledge of the failure states is not available. In other words, the potential failure modes of the test structure are so varied and so unpredictable that it is not feasible to train the NN by furnishing it with pairs of failure states and corresponding diagnostic response. Without postulating or searching among limited set of expected failure modes, the approach of this study can be applied equally well to determine whether the underlying structural response is linear or not. However, such an approach has the disadvantage that detectable changes in the signature of the analyzed response measure of the structure are not directly attributable to a specific failure mode, but simply indicate that damage has been sustained by an element or unit of a structure that has a dominant contribution to the response measure being analyzed [81].

A new method of dynamic FEM updating is proposed using NNs. Because all practical experimental data will contain noise, so it is desirable to develop an updating method that is resistant to noise. It is widely known that NNs tend to be robust in the presence of noise and are able to distinguish between these random errors and the desired systematic outputs. Hence, it seems natural and appropriate to apply NNs to this field. In this method, the experimental data are first prepared by using modal analysis on the frequency response functions (FRFs), and then the resulting model shapes and natural frequencies are assembled into an experimental vector. Another advantage of the proposed approach is the avoidance of the common problem of coordinate incompleteness; i.e., the NN updating method is capable of working with a limited number of experimentally measured DOFs and modes. The proposed updating method is tested on a simple cantilever beam, with promising results. The main drawback is that this method is computationally expensive, and it will fail if FEM has repeated modes. However, it would seem that there is significant potential for this model updating method to work with practical structures [177].

In another research FRFs are implemented to identify faults in FEMs. Modal properties and FRFs are implemented simultaneously to identify faults [178, 179]. An adaptive NN method is proposed for model updating and damage detection. The NN model is first trained off-line and then is retrained during iteration if needed. Numerical simulation of suspension bridge model updating demonstrates the effectiveness of the proposed method [180]. The measured FRFs are used as an input data to ANNs to detect structural damage. The results show that, in all cases considered, it is possible to distinguish the changed states with good accuracy and repeatability [181].

A committee of NN techniques, which employs FRFs, modal properties (natural frequencies and model shapes), and wavelet transformation (WT) data simultaneously, is established to identify damage in structures. The committee approach assumes that the errors given by the three individual approaches are uncorrelated, a situation that becomes more apparent when using measured data rather than simulated data. The committee approaches are used in parallel to diagnose faults on a 3 DOF structure and a population of cylindrical shell. It is demonstrated that the committee procedure is more reliable than using each procedure individually. The disadvantage of the committee is that it requires more than one trained network [181, 182].

In another work the main focus is on evaluating the efficiency of model-unknown identification approaches such as NNs for detecting modifications in the characteristics of the underlying physical systems. Such methods would be particularly useful in assessing intricate mechanical systems whose internal states are not accessible for measurements. In particular, these methods address the issue of low-sensor spatial resolution, unknown system topology, and measurement noise well. The system is tested in its "virgin" state as well as in "damaged" states corresponding to different degree of parameter changes. It is shown that the proposed method is a robust procedure and a practical tool for the detection and overall quantification of changes in nonlinear structures whose constitutive properties and topologies are not known [82].

Multinovelty indices are developed to detect the damage region based on vibration measurement. First, a bridge is partitioned into a set of structural regions and it is assumed that there are vibration transducers at each region. For each region, a NN based on novelty detector is formulated by using the global natural frequencies and the localized modal components. Then the modal flexibility values are used to train an autoassociative NN and to obtain a novelty index. The applicability of the proposed method for structural damage region identification is demonstrated by taking the Tsing Ma Bridge and the Ting Kou Bridge as examples [183].

To diagnose faults in engineering structures in the situations where the excitation signals are unavailable or inaccessible, response-only data, transmissibility function, are utilized to train NNs. The technique is verified with two examples based on two different structural systems. The NN classifiers clearly deliver the diagnostic indications of the faults introduced into the structural systems, which suggests that the transmissibility function is a sensible response-only data source for structural fault diagnosis [184]. ANN is used as a tool in soft computing to determine the dynamic characteristics of a structural system from its dynamic responses. Then the modal parameters of the structural system are directly estimated from the weighing matrices in the NN [185].

A novel NN-based strategy is proposed and developed for the direct identification of structural parameters (stiffness and damping coefficients) from the time-domain dynamic responses of an object structure without any eigenvalue analysis and extraction and optimization process that is required in many identification algorithms for inverse problems. The proposed strategy is extremely efficient in computation and thus has potential of becoming a practical tool for near-real-time monitoring of civil infrastructures [186]. A NN-based damage detection method using the modal properties is presented, which can effectively consider the modeling errors in the baseline FEM from which the training patterns are to be generated. The differences or the ratios of the mode-shape components between before and after damage are used as the input to the NNs in this method, since they are found to be less sensitive to the modeling errors than the mode shapes themselves. Results of laboratory test on a simply supported bridge model and field test on a bridge with multiple girders confirm the applicability of the present method [187].

A damage detection procedure, using pattern recognition of the vibration signature, is assessed using a FEM of a real structure—a suspension bridge more than 100 years old. Realistic damage scenarios are simulated and the response under moving traffic is evaluated. Feature vectors generated from the response spectra are included in two unsupervised NNs for examination. It is shown that the sensitivity of the NNs may be adjusted so that a satisfactory rate of damage detection may be achieved even in the presence of noisy signals [188].

A Bayesian probabilistic approach is presented for smart structure monitoring (damage detection) based on the pattern matching approach utilizing dynamic data. ANNs are employed as tools for

matching the "damage patterns" for the purpose of detecting damage locations and estimating their severity. It is obvious that the selection of the class of feed-forward ANN models, i.e., the decision on the number of hidden layers and the number of hidden neurons in each hidden layer, has crucial effects on the training of ANNs as well as the performance of the trained ANNs. This approach presents a Bayesian probabilistic method to select the ANN model class with suitable complexity, which is usually overlooked in the literature. An example using a five-story building is used to demonstrate the proposed methodology, which consists of a two-phase damage detection method and a Bayesian ANN design method [189].

By using the advanced modeling method of element stiffness matrix modification, the order of the global stiffness matrix can be kept invariable in establishing the model of intact and damaged structures. Then, eigenvalue perturbation theory is introduced to obtain the eigenvalues and eigenvectors of the damaged structure for reducing the computation load. Two ANNs are trained based on the response data simulated using finite-element modal (FEM) and perturbation theory enhanced finite-element modal (PFEM), respectively. The damage identification capability of these two ANNs is compared. Results show that the PFEM using the first-order eigenvalue perturbation theory provides enough precision for detecting small structural damage and the computational requirement is greatly reduced. Typically, the eigensolution computational time for obtaining the train sample data using PFEM is only 1% of that using the traditional FEM [190].

The effectiveness of NN methods is determined by the completeness of original data library and the reliability of algorithms. The NN method may be effective for the on-line monitoring of large structures, such as cable-stayed bridges and suspension bridges.

8 EXAMPLES OF HEALTH MONITORING IMPLEMENTATION

In order for the technology to advance sufficiently to become an operational system for the maintenance and safety of civil infrastructures, it is of paramount importance that new analytical developments are ultimately verified with appropriate data obtained from monitoring systems, which have been implemented on civil infrastructures, such as bridges.

1. The Sunshine Skywey Bridge in Florida has been instrumented with more than 500 sensors to verify design assumptions, and monitor construction quality and the conditions in service [191].
2. A monitoring program has been initiated to study the dynamic properties of the 1543-m cable-stayed Tampico Bridge in Mexico. The main span is 360 m in length; 21 servo accelerometers have been installed and ambient and pull-back tests conducted. The resulting frequencies are in good agreement but the damping values are still being estimated [192].
3. Continuous monitoring of two steel bridges over the Conrail mainline tracks in Rochester, NY, is being done. These bridges were built in 1963. The monitoring system is included as part of rehabilitation contracts. Altogether 5 inclinometers, 22 accelerometers, and 5 strain gauges have been installed in these two bridges. All these were monitored to connect with circular to a remote host computer. Natural frequencies, mode shapes, damping ratios, MAC, etc., are then computed for use in condition monitoring and assessment [193, 194].
4. Structural monitoring system is designed and evaluated by solid-state sensors for installation in several bridges and buildings. It is reported this system is used in a project by installing fully automatic and telemetered strain sensors on 10 bridges in Georgia [195].
5. The feasibility of health monitoring of a 720-m span Hakucho Suspension Bridge in Japan is studied by ambient vibration measurement. An identification scheme that makes use of cancellation of randomness in data by shaking is employed to use the ambient vibration measurements with high accuracy [196].
6. The measurement and documentation of construction and service effects for a three-span continuous steel stringer bridge in Cincinnati, OH, is addressed. A total of 642 channels of sensor data are available for bridge monitoring. The measurements are used to check the design and the project is ongoing [197].

7. A long-term continuously operating health monitoring system has been designed and implemented for the Commodore Barry Bridge, with spans of $822 + 1644 + 822$ ft. Over 80 channels of different sensor types have been installed to collect data such as temperature, wind speed and direction, strains, acceleration, etc. Many types of sensors for health monitoring have also been installed in several long-span suspension bridges in China, such as Tsingma Birdge, Humen Birdge, and Jiangyin Yangzi Bridge [198].
8. The programs within European nations and European collaboration programs are introduced for SHM [199].
9. The research and development of SHM systems at the Bridge and Structural Lab of the University of Tokyo is introduced. The ambient vibration-based approaches for LDV and the applications in the long-span suspension bridges are also presented. The extraction of the measured data is extremely difficult because it is hard to separate changes in vibration signature due to damage from changes, normal usage, changes in boundary conditions, or the release of the connection joints [196].
10. There are more comprehensive examples for the following bridges mainly in United States [200].

The examples comprise the following:

1. Akashi Kaikyo (Japan).
2. Hong Kong Bridges (Hong Kong, China).
3. Woodrow Wilson Bridge (Virginia-Maryland, USA).
4. Benicia-Martinez Bridge (California, USA).
5. Commodore Barry Bridge (Pennsylvania, USA).
6. Colle D'Isarco Viaduct (Italy).
7. Hamilton Avenue Bridge (Ohio, USA).
8. Reading Road Bridge (Ohio, USA).
9. Sandusky Bridge (Ohio, USA).
10. Toledo Bridges (Ohio, USA).
11. Fleet Health Monitoring of T-Beam Bridges (Pennsylvania, USA).
12. A good review about the current state of SHM in Japan is reported [201].
13. State of the art for SHM in China covers some advanced methods [202].
14. Applications and researches in bridge health monitoring systems are reported [203].

New bridges offer opportunities for developing complete SHMs for bridge inspection and condition evaluation of the bridges from "cradle to grave".

Existing bridges provide challenges for applying state of the art in SHM technologies to determine the current conditions of the structural element, connections and systems, to formulate model for estimating the rate of degradation, and to predict the existing and the future capacities of the structural components and systems. Advanced health monitoring systems may lead to better understanding of structural behavior and significant improvements of design, as well as the reduction of the structural inspection requirements. Great benefits due to the introduction of SHM are being accepted by owners, managers, bridge engineers, etc.

9 RESEARCH AND DEVELOPMENT NEEDS

Most damage detection theories and practices are formulated based on the following assumption that failure or deterioration would primarily affect the stiffness and therefore affect the modal characteristics of the dynamic response of the structure. This is seldom true in practice, because of the following reasons:

1. Traditional modal parameters (natural frequency, damping ratio, mode shapes, etc.) are not sensitive enough to identify and locate damage. The estimation methods usually assume that structures are linear and proportional damping systems.
2. Most currently used damage indices depend on the severity of the damage, which is impractical in the field. Most civil engineering structures, such as highway bridges, have redundancy in design and are large in size with low natural frequencies. Any damage index should consider these factors.
3. Scaled modeling techniques are used in current bridge damage detection. A single beam/girder models cannot simulate the true behavior of a real bridge. Similitude laws for dynamic simulation and testing should be considered.

4. Many methods usually use the undamaged structural modal parameters as the baseline compared with the damaged information. This will result in the need of a large data storage capacity for complex structures. But in practice, there are majority of existing structures for which baseline modal responses are not available. Only one developed method, which tried to quantify damage without using a baseline, may be a solution to this difficulty [204]. There is a lot of research work to be carried out in this direction.
5. None of the methods has the ability to distinguish the type of damages on bridge structures.

It is not easy to establish the direct relationship between various damage patterns and the changes of vibrational signatures. Health monitoring requires clearly defined performance criteria, a set of corresponding condition indicators and global and local damage and deterioration indices, which should help diagnose reasons for changes in condition indicators. It is implausible to expect that damage can be reliably detected or tracked by using a single damage index. We note that many additional localized damage indices, which relate to highly localized properties of materials or the circumstances that may indicate a susceptibility of deterioration such as the presence of corrosive environments around reinforcing steel in concrete, should be also integrated into the health monitoring systems.

There is now a considerable research and development effort in academia, industry, and management department regarding global health monitoring for civil engineering structures.

Several commercial structural monitoring systems currently exist, but further development is needed in commercialization of the technology. We must realize that damage detection and health monitoring for bridge structures by means of vibration signature analysis is a very difficult task. It contains several necessary steps, including defining indicators on variations of structural physical condition, dynamic testing to extract such indication parameters, defining the type of damages and remaining capacity or life of the structure, and relating the parameters to the defined damage/aging. Unfortunately, to date, no one has accomplished the above steps. Many further studies are needed.

9.1 Where to go from here?

In order for the model-based damage detection methods to be adopted eventually for on-line health monitoring, the following should be addressed:

1. The accurate definition of damage and new sensitive damage indices should be developed. These indices could distinguish not only the place and the extent of damage but also the types of damage in a structure.
2. Fast algorithms for SI are needed, if possible, on a real-time basis. First, the identification of the basic characteristics of existing bridges must be accurate and reliable. This is the basis of the structural damage identification. Secondly, the objects of sensor arrangement must be clear, because there are large amounts of data to be treated and the noise should be filtrated. Furthermore, a method for localized structural identification is highly desirable so that not all of the sensor output is processed for on-line monitoring purposes. In this way, substructure identification method may be the direction.
3. There are still considerable uncertainties in the testing, analysis, and environment for the purpose of damage detection. Sometimes, it is very difficult to sort out the uncertainties and pinpoint whether the lack of reliable results from modal analysis was due to the damage or due to an error in the considerably complicated procedures of modal analysis. At this point, it is concluded that a controlled study of a physical model in the laboratory would be an excellent method to understand the sources of uncertainties and limits of confidence when modal analysis was used as a technology for condition assessment and damage identification. More research is needed on the analytical techniques for damage identification using available and realistic structural monitoring data, including the combination of static data and vibration testing information. This research should consider the uncertainties inherent in the materials and construction, the variability of structural properties due to environmental conditions, unknown modeling errors and assumption, etc.
4. Although we have not discussed any non-model-based damage detection methods, a

robust on-line health monitoring system would require a hybridization of both nonmodel and model-based methods. Studies are needed to develop sensible hybrid damage detection methods that are easy to implement and robust. According to the existing study, one of the realistic methods may be the stochastic subspace identification method by using the environmental excitation data.

5. Unfortunately, nearly all of the existing systems are not instrumented to get the responsible data. The data-acquisition systems with multichannel and signal-processing system are being developed for SHM. Economical sensor placement and data collection methodologies, both onsite and remote, are needed in order for on-line health monitoring technologies to have practical benefits for nation's existing infrastructure and transportation systems.
6. The reliability and durability of the entire SHM system should be studied. The interrelationship of structural behavior and the effect of all components of the monitoring system on the overall safety should be studied, including sensors and their optimal placements, communication, data acquisition, etc.
7. The evaluation of serviceability and load-carrying capacity for existing highway bridges based on the damage identification and reliability theory should be studied. It is very important for the load rating, condition assessment, and decision making of repair, strengthening, and rehabilitation of existing highway bridges.
8. Information techniques and imaginative systems are needed to integrate field, theoretical, and laboratory research for solving large SI and condition assessments problems.
9. New and innovative construction materials will enhance the strength and durability of the infrastructure system in the twenty-first century. Testing and evaluation methodologies need to be developed specifically for characterization of newer and high performance materials. Advanced condition monitoring technologies will enable detection of cracks, onset of failure, extent of degradation, and location of damaged zones in structural elements.
10. Standards and code for SHM, and how SHM can be better applied in practice—that is, Philosophies, cost, devices, efficiency. The health monitoring system of bridge will be included in the scope of bridge management system [205].

REFERENCES

[1] Wang T.-L, Zong Z. *Final Report: Improvement of Evaluation Method for Existing Highway Bridges*. Department of Civil & Environmental Engineering, Florida International University, March 2002.

[2] Aktan AE, Pervizpour M, Catbas N, Grimmelsman K, Barrish R, Curtis J, Qin X. *Information Technology Research for Health Monitoring of Bridge Systems*. Drexel University Intelligent Infrastructure and Transportation Safety Institute: Philadelphia, PA, 2002.

[3] Helmut Wenzel and Hiroshi Tanaka; *Samco Monitoring Glossary: Structural Dynamics for Vbhm of Bridges*, Austria 2006.

[4] Housner GW, *et al.* Structural control: past, present, and future. *Journal of Engineering Mechanics, ASCE*, 1997 **123**(9):897–971.

[5] Ou JP, Li H, Duan ZD. Structural health monitoring and intelligent infrastructure. *Proceedings of the Second International Conference*. Shenzhen, 2006.

[6] Wu ZS, Abe M. Structural health monitoring and intelligent infrastructure. *Proceedings of the First International Conference*. Tokyo, 2003.

[7] National Research Council Canada, *Guidelines for Structural Health Monitoring, The Canadian Network of Centers of Excellence on Intelligent Sensing for Innovative Structures*, Design Manual No. 2, September 2001.

[8] Okada K, Shiraishi M, Takeuchi K. Structural health monitoring system using displacement memorizing sensor (Part 1 and 2). *Proceedings of Annual Meeting of Architectural Institute of Japan(AIJ)*. September 2005, (in Japanese).

[9] Udwadia FE, Garba JA. Optimal sensor locations for structural identification. *Proceedings of the, JPL Workshop on Identification and Control of Flexible Space Structures*, Tokyo, 1985; pp. 247–261.

[10] Kammer DC. Sensor placement for on orbit modal identification and correlation of large space structures. *AIAA Journal* 1991 **26**(1):104–112.

[11] Hemez FM, Farhat C. An energy based optimum sensor placement criteria and its application to structural damage detection. *Proceedings of the 12th International Modal Analysis Conference, Society of Experimental Mechanics*. Honolulu, HI, 1994; pp. 1568–1575.

[12] Penny JET, Friswell MJ, Garvey SD. Automatic choice of measurement locations for dynamic testing. *AIAA Journal* 1994 **32**(2):407–414.

[13] Udwadia FE, Methodology for optimum sensor locations for parameter identification in dynamic systems. *Journal of Engineering Mechanics, ASCE* 1994 **120**(2):368–390.

[14] Heredia-Zavoni E, Esteva L. Optimal instrumentation of uncertain structural systems subject to earthquake motions. *Earthquake Engineering and Structural Dynamics* 1998 **27**:343–362.

[15] Cobb RG, Liebst BS. Sensor location prioritization and structural damage localization using minimal sensor information. *AIAA Journal* 1997 **35**(2):369–374.

[16] Cobb RG, Liebst BS. Structural damage identification using assigned partial eigenstructure. *AIAA Journal* 1997 **35**(1):152–158.

[17] Reynier M, Hisham AK. Sensors location for updating problems. *Mechanical Systems and Signal Processing* 1999 **13**(2):297–314.

[18] Shi ZY, Law SS, Zhang LM. Optimizing sensor placement for structural damage detection. *Journal of Engineering Mechanics, ASCE* 2000 **126**(11):1173–1179.

[19] Xia Y, Hao H. Measurement selection for vibration-based structural damage identification. *Journal of Sound and Vibration* 2000 **236**(1):89–104.

[20] Fu GK, Moosa AG. Health monitoring of structures using optical instrumentation and probabilistic diagnosis. In *Condition Monitoring of Materials and Structures*, Ansari F (ed), Taylor & Francis/Balkema, 2000; pp. 190–201.

[21] Worden K, Burrows AP. Optimal sensor placement for fault detection. *Engineering Structures* 2001 **23**:885–901.

[22] Smyth AW. The potential of GPS and other displacement sensing for enhancing acceleration sensor monitoring array data by solving low frequency integration problems. *Proceedings of the Second International Conference on Bridge Maintenance, Safety, Management and Cost, IABMAS04*. Kyoto, October 2004; pp. 533.

[23] Mita A, Takahira S. Damage index sensor for smart structures. *Department of System Design Engineering, Keio University, Structural Engineering and Mechanics* 2004 **17**(3–4):331–346.

[24] Mita A, Yoshikawa S. Digital sensor network using delta-sigma modulation for health monitoring of large structures. In *II ECCOMAS Thematic Conference on Smart Structures and Materials*, Mota Soares CA *et al.* (eds), A A Balkema Publishers: Lisbon, July 18–21, 2005.

[25] Meo M, Zumpano G. On the optimal sensor placement techniques for a bridge structure. *Engineering Structures* 2005 **27**:1488–1497.

[26] Glisic B, Inaudi D, Vurpillot S. Whole lifespan monitoring of concrete bridges. *First International Conference on Bridge Maintenance, Safety and Management, IABMAS'02*, Barcelona, July 2002.

[27] Bakht B, Jaeger LG. Bridge testing—a surprise every time. *Journal of Structural Engineering, ASCE*, 1990 **116**(5):605–611.

[28] Kennedy JB, Grace NF. Prestressed continuous composite bridges under dynamic load. *Journal of Structural Engineering, ASCE*, 1990 **116**(6):1660–1678.

[29] Mazurek DF, DeWolf JT. Experimental study of bridge health monitoring technique. *Journal of Structural Engineering, ASCE*, 1990 **116**(9): 2532–2549.

[30] Hearn G, Testa RB. Modal analysis for damage detection in structures. *Journal of Structural Engineering, ASCE* 1991 **117**(10):3042–3063.

[31] Aktan AE, Farhey DN, Helmicki DJ, Grimmelsman KA. Structural identification for condition assessment: experimental arts. *Journal of Structural Engineering, ASCE* 1997 **123**(12):1674–1684.

[32] Aktan AE, Catbas FN, Turer A, Zhang ZF. Structural identification, analytical aspect. *Journal of Structural Engineering, ASCE* 1998 **124**(7): 817–829.

[33] Aktan AE, Tsikos CJ, Faust D. Challenge and opportunities in bridge health monitoring. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 461–473.

[34] Aktan AE, *et al.* Integrated field, theoretical and laboratory research for large systems identification problems. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). The Hong Kong Polytechnic University: Hong Kong, December 2000; pp. 1–18.

[35] Pandey AK, Biswas M. Damage detection in structures using changes in flexibility. *Journal of Sound and Vibration* 1994 **169**(1):3–17.

[36] Lee GC, Liang Z. Development of a bridge monitoring system. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 349–358.

[37] Haritos N. Dynamic testing techniques for structural identification of bridge superstructures. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). The Hong Kong Polytechnic University: Hong Kong, December 2000; pp. 1013–1020.

[38] Piombo BAD, Fasana A, Marchesiello S, Ruzzene M. Modeling and identification of the dynamic response of a supported bridge. *Mechanical Systems and Signal Processing* 2000 **14**(1):75–89.

[39] Liao WH, Wang DH, Huang SL. *Wireless Monitoring of Cable Tension of Cable-Stayed Bridges Using PVDF Piezoelectric Films*, Hong Kong, 2000.

[40] Mirmiran A, Wei YM. Damage assessment of FRP-encased concrete using ultrasonic pulse velocity. *Journal of Engineering Mechanics, ASCE* 2001 **127**(2):126–135.

[41] Sikorsky C, Stubbs N, Bolton R, Seible F. Performance evaluation of a composite bridge using structural health monitoring. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Stanford University, 2001.

[42] Liao WH, Wang DH, Huang SL. Wireless monitoring of cable tension of cable-stayed bridges using PVDF piezoelectric films. *Journal of Intelligent Material Systems and Structures* 2001 **12**(5):331–339.

[43] Alaghehbandian A, Abe M, Fujino Y. An Internet oriented platform for structural health monitoring. *Proceedings of the First International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Tokyo, November 2003; pp. 339–344.

[44] Chang CC, Ji YF. Sensing of low-frequency vibration using photogrammetric technique. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 325–333.

[45] Adams RD, Cawley P, Pye CJ, Stone BJ. A vibration technique for nondestructively assessing the integrity of structures. *Journal of Mechanical Engineering Science* 1978 **20**:93–100.

[46] Biswas M, Pandey AK, Samman MM. Diagnostic experimental spectral/modal analysis of a highway bridge. *The International Journal of Analytical and Experimental Modal Analysis* 1990 **5**(1):33–42.

[47] Capecchi D, Vestroni F. Monitoring of structural systems by using frequency dada. *Earthquake Engineering and Structural Dynamics* 1999 **28**:447–461.

[48] Casas JR, Aparicio AC. Structural damage identification from dynamic-test data. *Journal of Structural Engineering, ASCE*, 1995 **120**(8):2437–2450.

[49] Stubbs N, Kim JT, Farrar CR. Field verification of a nondestructive damage localization and severity estimation algorithm. *Proceedings of the 13th International Modal Analysis Conference (IMAC)*. Nashville, TN, 1995; pp. 210–218.

[50] Mazurek DF, DeWolf JT. Experimental study of bridge monitoring techniques. *Journal of Structural Engineering, ASCE* 1990 **116**(9):2532–2549.

[51] Pandey AK, Biswas M, Samman MM. Damage detection from changes in curvature mode shapes. *Journal of Sound and Vibration* 1991 **145**(2): 321–332.

[52] Wahab MMA, Roceck GD. Damage detection in bridges using modal curvatures: application to a real damage scenario. *Journal of Sound and Vibration* 1999 **226**(2):217–235.

[53] Alampalli S, Fu G, Aziz IA. Modal analysis as a bridge inspection tool. *Proceedings of the 10th International Modal Analysis Conference (IMAC)*. San Diego, CA, 1992; pp. 1359–1366.

[54] Lee JS. Using transfer function parameter changes for damage detection of structures. *AIAA Journal* 1995 **33**(11):2189–2193.

[55] Zhang H, Schulz MJ, Ferguson F. Structural health monitoring using transmittance functions. *Mechanical Systems and Signal Processing* 1999 **13**(5):765–787.

[56] Areemit N, Yamaguchi H, Matsumoto Y, Ibi T. Modal identification of a four-story reinforced concrete building under renovation by using ambient vibration measurement. *Proceedings of the First International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Tokyo, November 2003; pp. 513–520.

[57] Duan ZD, Yan GR, Ou JP, Spencer BF. Damage localization in ambient vibration by constructing proportional flexibility matrix. *Proceedings of the First International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Tokyo, November 2003; pp. 561–565.

[58] Yan GR, Duan ZD. Damage localization based on the residual wavelet force. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 777–784.

[59] Beskhyroun S, Oshima T, Mikami S, Tsubota Y, Takeda T. Damage identification of steel structures based on changes in the curvature of power spectral density. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 791–797.

[60] Xu L, Wu GL, Yi WJ, Yi ZH. The research of structural damage detection using high order local modes. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 813–820.

[61] Lin RJ, Cheng FP. A damage detection approach considering the stiffness and mass variations. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 839–845.

[62] Del Grosso A, Lanata F. Damage detection and localization algorithm for continuous static monitoring of structures. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 855–861.

[63] Chen J, Li J. Signal energy based index for structural damage detection. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 945–950.

[64] Yan ZH, Miyamoto A. SRM and experimental study on bridge condition assessment method. *Lifetime Engineering of Civil Infrastructure*. Research Center for Environmental Safety (RCES), Yamaguchi University: Ube, Vol. 3, August 2008; pp. 267–282.

[65] Caravani P, Waston ML, Thomson WT. Recursive least-square time domain identification of structural parameters. *Journal of Applied Mechanics* 1977 **44**:135–140.

[66] Carmichael DG. The state estimate problem in experimental structural mechanics. *Proceedings of the 3rd Conference on Application of Statistics and Probability in Soil and Structural Engineering*. Sydney, 1979; pp. 802–815.

[67] Yun CB, Shinozuka M. Identification of nonlinear structural dynamic system. *Journal of Structural Mechanics* 1980 **8**(2):187–203.

[68] Hoshiya M, Statio E. Structural identification by extended Kalman filter. *Journal of Engineering Mechanics, ASCE* 1984 **110**(12):1757–1770.

[69] Hoshiya M, Maruyama O. Structural identification by extended Kalman filter. *Journal of Engineering Mechanics, ASCE* 1987 **110**:1757–1770.

[70] Yun CB, Kim WJ, Ang AHS. Damage assessment of bridge structures by system identification. *Proceedings of Korea-Japan Joint Seminar on Emerging Technologies in Structural Engineering and Mechanics*, Seoul, 1988; pp. 182–193.

[71] Sato T, Qi K. Adaptive H-infinity filter: its application to structural identification. *Journal of Engineering Mechanics, ASCE* 1998 **124**(11): 1233–1240.

[72] Koh CG, See LM, Balendra T. Estimation of structural parameters in time domain: a substructure approach. *Earthquake Engineering and Structural Dynamics* 1991 **20**:787–801.

[73] Oreta AWC, Tanabe TA. Element identification of member properties of framed structures. *Journal of Structural Engineering, ASCE* 1994 **120**(7):1961–1967.

[74] Yun CB, Lee HJ. Substructural identification for damage estimation of structures. *Structural Safety (Amsterdam)* 1997 **19**(1):121–140.

[75] Kim HM, Bartkowics TJ. A two-step structural damage detection approach with limited instrumentation. *Journal of Vibration and Acoustics* 1997 **119**(2):258–264.

[76] Kim HM, Bartkowics TJ. An experimental study for damage detection using a hexagonal truss. *Computer and Structures* 2001 **79**:173–182.

[77] Koh CG, Hoon B, Liaw CY. Parameter identification of large structural systems in time domain. *Journal of Structural Engineering, ASCE* 2000 **126**(8):957–963.

[78] Hao H, Xia Y. Vibration-based damage detection of structures by genetic algorithm. *Journal of Computing in Civil Engineering, ASCE*, 2002 **16**(3):222–229.

[79] Wang BS, *et al.* Comparative study of damage indices in application to a long-span suspension bridge. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). the Hong Kong Polytechnic University: Hong Kong, December 2000; pp. 1085–1092.

[80] Masri SF, Chassiakos AG, Caughey TK. Identification of nonlinear dynamic systems using neural

networks. *Journal of Applied Mechanics, Transactions of the ASME*, 1993 **60**:123–133.

[81] Masri SF, Nakamura M, Seed RB. A neural network approach to the detection of changes in structural parameters. *Journal of Engineering Mechanics, ASCE* 1996 **122**(5):442–448.

[82] Chassiakos AG, Caughey TK, Hunter NF. Application of neural network for detection of changes in nonlinear systems. *Journal of Engineering Mechanics, ASCE* 2000 **126**(7):666–676.

[83] Chong YH, Imregun M. Variable modal parameter identification for non-linear MDOF system, Part I: formulation and numerical validation; Part II: experimental validation and advanced case study. *Shock and Vibration* 2000 **7**:217–240.

[84] Lin JW, Betti R, Smyth AW, Longman RW. Online identification of non-linear hysteretic structural system using a variable trace approach. *Earthquake Engineering and Structural Dynamics* 2001 **30**:1279–1303.

[85] Kerschen G, Golinval JC. Theoretical and experimental identification of a non-linear beam. *Journal of Sound and Vibration* 2001 **244**(4):597–613.

[86] Peeters B, Roeck GD. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing* 1999 **13**(6):855–878.

[87] Masri SF, *et al.* Application of a Web-enabled real-time structural health monitoring system for civil infrastructure systems. *Smart Materials and Structures* 2004 (13):1269–1283.

[88] Huang CS, Liu HL. Modal identification of structures from ambient vibration, free vibration, and seismic response data via a subspace approach. *Earthquake Engineering and Structural Dynamics* 2001 **30**:1857–1878.

[89] Zhang HC, Foliente GC, Yang YM, Ma F. Parameter identification of inelastic structures under dynamic loads. *Earthquake Engineering and Structural Dynamics* 2002 **31**:1113–1130.

[90] Hao H, Xia Y. Vibration-based damage detection of structures by genetic algorithm. *Journal of Computing in Civil Engineering* 2002 **16**(3):677–683.

[91] Shi YN, He B, Zhu HP. Damage identification of multi-storey buildings based on modal sensitivity analysis and genetic algorithm. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 1015–1020.

[92] Casciati F, Faravelli L, Marazzi F, Rossi R. Damage detection via genetic algorithm. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 1049–1053.

[93] Peng JY, Li H. Parameter identification and qualitative sensitivity analysis of hysteretic model by particle swarm optimization. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 1009–1014.

[94] Doebling SW, Hemez FM, Peterson LD, Farhat C. Improved damage location accuracy using strain energy-based on mode selection criteria. *AIAA Journal* 1997 **35**(4):693–699.

[95] Law SS, Chan THT, Wu D. Efficient numerical model for the damage detection of large scale. *Engineering Structures* 2001 **23**:436–451.

[96] Lardies J, Larbi N. A new method for model order selection and modal parameter estimation in time domain. *Journal of Sound and Vibration* 2001 **245**(2):187–203.

[97] Koh BH, Ray LR. Feedback controller design for sensitivity-based damage localization. *Journal of Sound and Vibration* 2002 **251**(1):59–78.

[98] Beck JL, Katafygiotis LS. Probabilistic system identification and health monitoring of structures. *Proceedings of the 10th World Conference on Earthquake Engineering*, Madrid, 1992.

[99] Beck JL, Katafygiotis LS. Updating structural dynamic models and their uncertainties: statistical system identification. *Journal of Engineering Mechanics, ASCE* 1997 **123**(11):1170–1179.

[100] Sohn H, Law HW. Bayesian probabilistic approach for structural damage detection. *Earthquake Engineering and Structural Dynamics* 1997 **26**: 1259–1281.

[101] Sohn H, Law HW. Bayesian probabilistic damage detection of a reinforced-concrete bridge column. *Earthquake Engineering and Structural Dynamics* 2000 **29**:1131–1152.

[102] Vanik MW, Beck JL, Au SK. Bayesian **probabilistic** approach to structural health monitoring. *Journal of Engineering Mechanics, ASCE* 2000 **126**(7):738–745.

[103] Cha PD, Tuck-Lee JP. Updating structural system parameters using frequency response data. *Journal of Engineering Mechanics, ASCE* 2000, **126**(12): 1240–1246.

[104] Yeo I, Shin S, Lee HS, Chang SP. Statistical damage assessment of framed structures from static responses. *Journal of Engineering Mechanics, ASCE* 2000 **126**(4):414–421.

[105] Sohn H, Czarnecki JA, Farrar CR. Structural health monitoring using statistical process control. *Journal of Structural Engineering, ASCE* 2000 **126**(11): 1356–1363.

[106] Sayyer JP, Rao SS. Structural damage detection and identification using Fuzzy logic. *AIAA Journal* 2000 **38**(12):2328–2335.

[107] Papadopoulos L, Garcia E. Structural damage identification: a probabilistic approach. *AIAA Journal* 1998 **36**(11):2137–2145.

[108] Katafygiotis LS, Yuen KV, Chen JC. Bayesian modal updating by using of ambient data. *AIAA Journal* 2001 **39**(2):271–278.

[109] Fugate ML, Sohn H, Farrar CR. Vibration-based damage detection using statistical process control. *Mechanical Systems and Signal Processing* 2001 **15**(4):707–721.

[110] Xia Y, Hao H, Brownjohn MW, Xia PQ. Damage identification of structures with uncertain frequency and mode shape data. *Earthquake Engineering and Structural Dynamics* 2002 **31**(5):1053–1066.

[111] Basseville M, Mevel L, Goursatc M. Statistical model-based damage detection and localization: subspace-based residuals and damage-to-noise sensitivity ratios. *Journal of Sound and Vibration* 2004 **275**:769–794.

[112] Yan AM, Kerschen G, De Boe P, Golinval JC. Structural damage diagnosis under varying environmental conditions—Part I: a linear analysis. *Mechanical Systems and Signal Processing* 2005 **19**:847–864.

[113] Mattson SG, Pandit SM. Statistical moments of autoregressive model residuals for damage localization. *Mechanical Systems and Signal Processing* 2006 **20**:627–645.

[114] Yan AM, Golinval JC. Null subspace-based damage detection of structures using vibration measurements. *Mechanical Systems and Signal Processing* 2006 **20**:611–626.

[115] Sakellariou JS, Fassois SD. Stochastic output error vibration-based damage detection and assessment in structures under earthquake excitation. *Journal of Sound and Vibration* 2006 **297**:1048–1067.

[116] Gupta S, Ray A, Keller E. Symbolic time series analysis of ultrasonic data for early detection of fatigue damage. *Mechanical Systems and Signal Processing* 2007 **21**:866–884.

[117] Cawley P, Adams RD. The location of defects in structures from measurement natural frequencies. *Journal of Strain Analysis* 1979 **14**(2):49–57.

[118] Friwell MI, Penny JET, Wilsonm DAL. Using vibration data and statistical measures to locate damage in structures. *The International Journal of Analytical and Experimental Modal Analysis* 1994 **9**(4):239–254.

[119] Kaouk M, Zimmerman DC. Structural damage assessment using a generalized minimum rank perturbation theory. *AIAA Journal* 1994 **32**(4): 836–842.

[120] Lim TW, Kashangaki TAL. Structural damage detection of space truss structures using best achievable eigenvector. *AIAA Journal* 1994 **32**(5): 1049–1057.

[121] Wahab MMA, Roceck GD, Peeters B. Parameterization of damage in reinforced concrete structures using modal updating. *Journal of Sound and Vibration* 1999 **228**(4):717–730.

[122] Raghavendrachar M, Aktan AE. Flexibility of multireference impact testing for bridge diagnostics. *Journal of Structural Engineering, ASCE* 1992 **118**(8):2186–2203.

[123] DeWolf JT, Zhao J. *Dynamic Vibration Techniques in Highway Bridge Monitoring*, Report to Department of Transportation of Connecticut State. Report No. CEE-98-01. University of Connecticut, Storrs: Connecticut, CT, 1998.

[124] Zhao J, DeWolf JT. sensitivity study for vibration parameters used in damage detection. *Journal of Structural Engineering, ASCE* 1999 **125**(4): 410–416.

[125] Ivanovic SS, Trifunac MD, Todorovska MI. On identification of damage in structures via wave travel times. *Proc. Nato Advanced Research Workshop on Strong-motion Instrumentation for Civil Engineering Structures*. Kluwer Publisher: Istanbul, June 2–5 1999.

[126] Shi ZY, Law SS, Zhang LM. Structural damage localization from model strain energy change. *Journal of Sound and Vibration* 1998 **218**(5): 825–844.

[127] Shi ZY, Law SS, Zhang LM. Structural damage detection from modal strain energy change. *Journal of Engineering Mechanics, ASCE* 2000 **126**(12):1216–1223.

[128] Shi ZY, Law SS, Zhang LM. Improved damage quantification from elemental modal strain energy change. *Journal of Engineering Mechanics, ASCE* 2002 **128**(5):521–529.

[129] Mak PS, Law SS. Structural damage assessment by elemental modal strain energy changes. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). The Hong Kong Polytechnic University: Hong Kong, December 2000.

[130] Law SS, Shi ZY, Zhang LM. Structural damage detection from incomplete and noisy modal test data. *Journal of Engineering Mechanics, ASCE* 1998 **124**(11):1280–1288.

[131] Law SS, *et al.* Modal strain energy changes in neural network damage assessment. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). The Hong Kong Polytechnic University: Hong Kong, December 2000; pp. 1037–1044.

[132] Law SS, Chan THT, Wu D. Efficient numerical model for the damage detection of large scale structure. *Engineering Structures* 2001 **23**(5):436–451.

[133] Messina A, Williams EJ, Contursi T. Structural damage detection by a sensitivity and statistical-based method. *Journal of Sound and Vibration* 1998 **216**(5):791–808.

[134] Shi ZY, Law SS, Zhang LM. Damage localization by directly using incomplete mode shapes. *Journal of Engineering Mechanics, ASCE* 2000 **126**(6):656–660.

[135] Liang Z, Lee GC. Damping of structures. *National Center for Earthquake Engineering Research*. State University of New York at Buffalo: Buffalo, NY, 1991, NCEER 91-0004.

[136] Kong F. *The Application of Energy Transfer Ratio in the Bridge Condition Assessment*, Ph.D. dissertation. State University of New York at Buffalo, 1996.

[137] Huang TJ. *Damage Probes in Structural and Mechanisms Utilizing Energy-Based Modal Parameter Identification*, Ph.D. dissertation. State University of New York at Buffalo: New York, November 1997.

[138] Caicedo J, Dyke SJ, Johnson EA. Health monitoring based on component transfer functions. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). The Hong Kong Polytechnic University: Hong Kong, December 2000; pp. 997–1004.

[139] Gupta VK, Nielsen SRK, Kirkegaard PH. A preliminary prediction of seismic damage-based degradation in RC structures. *Earthquake Engineering and Structural Dynamics* 2001 **30**:981–993.

[140] Ren WX, De Roeck G. Structural damage identification using modal data: I: simulation verification; II: test verification. *Journal of Structural Engineering, ASCE* 2002 **128**(1):87–104.

[141] Bernal D. Load vectors for damage localization. *Journal of Engineering Mechanics, ASCE* 2002 **128**(1):7–14.

[142] Abdo MA-B, Hori M. A numerical study of structural damage detection using changes in the rotation of mode shapes. *Journal of Sound and Vibration* 2002 **251**(2):227–239.

[143] Kim JT, Ryu YS, Cho HM, Stubbs N. Damage identification in beam-type structures: frequency-based method vs. mode-shape-based method. *Engineering Structures* 2003 **25**:57–67.

[144] Brasiliano A, Doz GN, Luis V, de Brito J. Damage identification in continuous beams and frame structures using the residual error method in the movement equation. *Nuclear Engineering and Design* 2004 **227**:1–17.

[145] Yoon D, Heider MK, Gillespie JW Jr, Ratcliffe CP, Crane RM. Local damage detection using the two-dimensional gapped smoothing method. *Journal of Sound and Vibration* 2005 **279**:119–139.

[146] Choi S, Park S, Stubbs N. Nondestructive damage detection in structures using changes in compliance. *International Journal of Solids and Structures* 2005 **42**:4494–4513.

[147] Han JG, Ren WX, Sun ZS. Wavelet packet based damage identification of beam structures. *International Journal of Solids and Structures* 2005 **42**:6610–6627.

[148] Kim BH, Park T, Voyiadjis GZ. Damage estimation on beam-like structures using the multi-resolution analysis. *International Journal of Solids and Structures* 2006 **43**:4238–4257.

[149] Kim JT, Park JH, Lee BJ. Vibration-based damage monitoring in model plate-girder bridges under uncertain temperature conditions. *Engineering Structures* 2006 **28**(9):1286–1297.

[150] Chen HG, Yan YJ, Jiang JS. Vibration-based damage detection in composite wingbox structures by HHT. *Mechanical Systems and Signal Processing* 2007 **21**:307–321.

[151] Hajela P, Soerio FJ. Structural damage detection based on static and model analysis. *AIAA Journal* 1990 **28**(9):1110–1115.

[152] Sanayei M, Onipede O. Damage assessment of structure using static test data. *AIAA Journal* 1991 **29**(7):1174–1179.

[153] Sanayei M, Scampoli SF. Structural element stiffness identification from static test data. *Journal of Engineering Mechanics, ASCE* 1990 **117**(5): 1021–1036.

[154] Sanayei M, Onipede O, Babu SR. Selection of noisy measurement location for error reduction in static parameter identification. *AIAA Journal* 1992 **30**(9):2299–2309.

[155] Sanayei M, Imbaro GR, McClaim JAS, Brown LC. Structural model updating using experimental static measurements. *Journal of Structural Engineering, ASCE* 1997 **123**(6):792–798.

[156] Sanayei M, Saletnik MJ. Parameter estimation of structures from static strain measurements. I: formulation; II: error sensitivity analysis. *Journal of Structural Engineering, ASCE* 1997 **123**(5):555–572.

[157] Banan MR, Hjelmstad KD. Parameter estimation of structure from static response. I: computational aspect; II: numerical simulation studies. *Journal of Structural Engineering, ASCE* 1994a,b **120**(11):3243–3283.

[158] Sanayei M, Saletnik MJ. Parameter estimation of structures from static strain measurements. I: formulation; II: error sensitivity analysis. *Journal of Structural Engineering, ASCE* 1996a,b **122**(5):555–572.

[159] Hjelmstad KD, Shin S. Damage detection and assessment of structures from static response. *Journal of Engineering Mechanics, ASCE* 1997 **123**(6):568–576.

[160] Oh BH, Jung BS. Structural damage assessment with combined data of static and modal tests. *Journal of Structural Engineering, ASCE* 1998 **124**(8):956–965.

[161] Cui F. *Parameter Identification and Load-Carrying Capacity Evaluation for Bridge*, Ph.D. Dissertation. Department of Bridge Engineering, Tongji University: China, January 2000.

[162] Wang X, Hu N, Fukunaga H, Yao ZH. Structural damage identification using test data and changes in frequencies. *Engineering Structures* 2001 **23**(6):610–621.

[163] Jang JH, Yeo I, Shin S, Chang SP. Experimental investigation of system-identification-based damage assessment on structures. *Journal of Structural Engineering, ASCE* 2002 **128**(5):673–682.

[164] Bakhtiari-Nejad F, Rahai A, Esfandiari A. A structural damage detection method using static noisy data. *Engineering Structures* 2005 **27**:1784–1793.

[165] Wei Z, Yam LH, Cheng L. NARMAX model representation and its application to damage detection for multi-layer composites. *Composite Structures* 2005 **68**:109–117.

[166] Cheng SH, Hwu C. On-line measurement of material properties for composite wing structures. *Composites Science and Technology* 2006 **66**:1001–1009.

[167] Zhao Q, Sawada T, Hirao K, Nariyuki Y. Localized identification of MDOF structures in the frequency domain. *Earthquake Engineering and Structural Dynamics* 1995 **24**:325–338.

[168] Abdelghani M, Basseville M, Benveniste A. Model-based monitoring and diagnostics of vibrating structures under operating conditions. *6th International Conference on Recent Advances in Structural Dynamics*. ISVR: Southampton, July 1997.

[169] Park KC, Reich GW. Model-based health monitoring of structural systems: progress, potential, and challenges. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 82–95.

[170] Alvin KF, Peterson LD, park KC. Method for determining minimum-order mass and stiffness matrices from modal test data. *AIAA Journal* 1995 **33**(1):128–135.

[171] Alvin KF, park KC. Extraction of substructural flexibility from global frequencies and mode shapes. *AIAA Journal* 1999 **37**(11):1444–1451.

[172] Zhang QW. *Modal Updating and Damage Detection for Bridge Structures*, Ph.D. Dissertation. Department of Bridge Engineering, Tongji University: China, February 1999.

[173] Gahboussi J, Garrett JH, Wu X. Knowledge-based modeling of material behavior with neural networks. *Journal of Engineering Mechanics, ASCE* 1991 **117**(1):132–153.

[174] Wu X, Gahboussi J, Garrett JH. Use of neural networks in detection of structural damage. *Computer and Structures* 1992 **42**(4):649–659.

[175] Elkordy MF, Chang KC, Lee GC. Neural networks trained by analytically simulated damage states.

Journal of Computing in Civil Engineering, ASCE 1993 **7**(2):130–145.

[176] Szewezyk P, Hajela P. Damage detection in structures based on feature-sensitive neural networks. *Journal of Computing in Civil Engineering, ASCE* 1994 **8**(2):163–178.

[177] Levin RI, Lieven NAJ. Dynamic finite element model updating using neural network. *Journal of Sound and Vibration* 1998 **210**(5):593–607.

[178] Atalla MJ, Inman DJ. On model updating using neural networks. *Mechanical Systems and Signal Processing* 1998 **12**(1 or 2):135–161.

[179] Marwala T, Hunt HEM. Fault identification using finite element models and neural networks. *Mechanical Systems and Signal Processing* 1999 **13**(3):475–490.

[180] Chang TCP, Chang CC, Xu YG. Updating structural parameters: an adaptive neural network approach. *Proceedings of the 2nd International Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 379–389.

[181] Zang C, Imregun M. Structural damage detection using artificial neural networks and measured FRF data reduced via principal component projection. *Journal of Sound and Vibration* 2001 **242**(5):813–827.

[182] Marwala T. Damage identification using committee of neural networks. *Journal of Engineering Mechanics, ASCE* 2000 **126**(1):43–50.

[183] Ni YQ, *et al.* Vibration-based damage localization in Ting Kau Bridge using probabilistic neural network. In *Proceedings of the International Conference on Advances in Structural Dynamics*, Ko JM, Xu YL (eds). The Hong Kong Polytechnic University: Hong Kong, December 2000.

[184] Chen Q, Chan YW, Worden K. Structural fault diagnosis and isolation using neural networks based on response-only data. *Computers and Structures* 2003 **81**:2165–2172.

[185] Chen CH, Yang YB. Structural system identification using a neural network algorithm. *Proceedings of the First International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Tokyo, November 2003; pp. 587–592.

[186] Xu B, Wu Z, Chen G, Yokoyama K. Direct identification of structural parameters from dynamic responses with neural networks. *Engineering Applications of Artificial Intelligence* 2004 **17**:931–943.

[187] Lee JJ, Lee JW, Yi JH, Yun CB, Jung HY. Neural networks-based damage detection for bridges considering errors in baseline finite element models. *Journal of Sound and Vibration* 2005 **280**:555–578.

[188] Yeung WT, Smith JW. Damage detection in bridges using neural networks for pattern recognition of vibration signatures. *Engineering Structures* 2005 **27**:685–698.

[189] Yuen KV, Lam HF. On the complexity of artificial neural networks for smart structures monitoring. *Engineering Structures* 2006 **28**:977–984.

[190] Yu L, Cheng L, Yam LH, Yan YJ. Application of eigenvalue perturbation theory for detecting small structural damage using dynamic responses. *Composite Structures* 2007 **78**:402–409.

[191] Overman TR, Shiu KN, Weinmann TL, Morgan BJ, Schultz DM. Evaluation and surveillance of concrete bridge structures. In *Bridge and Transmission line Structures*, Tall L (ed). ASCE: New York, 1987; pp. 224–236.

[192] Muria-Vala D, Gomez R, king C. Dynamic structural properties of cable-stayed Tampico bridge. *Journal of Structural Engineering, ASCE* 1991 **117**(11):3396–3416.

[193] Alampalli S, Fu G. *Remote Bridge Monitoring Systems for Bridge Condition*, Client Report 70. Engineering Research and Development Bureau, New York State Department of Transportation: Albany, NY, 1994.

[194] Alampalli S. Effects of testing, analysis, damage, and environment on modal parameters. *Mechanical Systems and Signal Processing* 2000 **14**(1):63–74.

[195] Westermo BD, Thompson LD. Design and evaluation of passive and active structural health monitoring systems for bridges and buildings. *Proceedings of the, UCII'95*, California, 1995.

[196] Abe M, Fujino Y, Kajimura T, Yanagihara M, Sato M. Monitoring of long span suspension bridge by ambient vibration measurement. *Proceedings of the 2nd international Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 400–407.

[197] Helmicki A, Hunt V, Wiklo M. Multidimensional performance monitoring of a recently constructed steel-stringer bridge. *Proceedings of the 2nd international Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 408–416.

[198] Catbas FN, Grimmelsman KA, Susoy M. Structural Identification and health monitoring of a

long-span bridge. *Proceedings of the 2nd international Workshop on Structural Health Monitoring*. Stanford University: Stanford, CA, September 1999; pp. 417–429.

[199] Foote PD. Structural health monitoring—tales from Europe. *Proceedings of the 2nd international Workshop on Structural Health Monitoring*. Stanford University, Stanford, CA, September 1999; pp. 24–35.

[200] Aktan AE, Catbas FN, Grimmelsman KA, Pervizpour M. *Development of a Model Health Monitoring Guide for Major Bridges*, Report 292. Report Submitted to: Federal Highway Administration Research and Development, Drexel Intelligent Infrastructure and Transportation Safety Institute, September 2002.

[201] Wu ZS Structural health monitoring and intelligent infrastructures in Japan. *Proceedings of the First International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Tokyo, November 2003; pp. 153–167.

[202] Ou JP, Li H. The state-of-the-art and practice of structural health monitoring for civil infrastructures in the mainland of China. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 69–93.

[203] Koh HM, Choo JF, Kim S, Kil HB. Applications and researches in bridge health monitoring systems and intelligent infrastructures in Korea. *Proceedings of the Second International Conference on Structural Health Monitoring and Intelligent Infrastructure*. Shenzhen, 2006; pp. 151–162.

[204] Stubbs N, Kim JT. Damage localization in structures without baseline model parameters. *AIAA Journal* 1996 **34**(8):1644–1649.

[205] Ryall MJ. *Bridge Management*. Butterworth-Heinemann Publisher: Great Britain, 2001.

Chapter 142
Monitoring Marine Structures

Liming W. Salvino[1] and Matthew D. Collette[2]

[1] *Structures and Composites, Carderock Division, Naval Surface Warfare Center, West Bethesda, MD, USA*
[2] *SAIC Advanced Systems and Technology Division, Bowie, MD, USA*

1 INTRODUCTION

The development of novel structures, continued existence of substandard ships, and the occurrence of structural damage at sea all demonstrate the importance and the need to implement an effective structural health monitoring (SHM) system for ship hull and local structures. These problems acknowledged in the early 1990s [1] have still to be solved or even fully addressed at the present time. There are significant differences between older, existing marine structures where limited data is available and newer structures where the design process is largely dependent upon numerical modeling. The inspection and monitoring needs and the required procedures and technologies of structural health diagnosis and prognosis are also greatly differing for commercial and military ships. Although general marine structural monitoring background is discussed as needed, this article will emphasize ship hull structure monitoring applications in US navy. An overview of marine SHM systems is presented in the introduction, followed by a description of typical ship structures and loads. The application of SHM systems throughout the vessel's lifetime is reviewed, and then health-monitoring systems and diagnostics are discussed. Finally, an example application of SHM systems to a current vessel is given.

A major objective of future US navy ship design is the development of high-performance and high-speed vessels. These designs will include many novel lightweight structural features, as well as possessing multimission adaptability and allowing for reduced levels of manning. For example, to reduce weight as much as possible to allow higher speeds and increased cargo capacity, high-speed vessels often employ novel and aggressive structural designs, using composite, aluminum alloys or high-strength steel, with innovative arrangements and fabrications to

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

maximize lightship weight reduction. These structural features and high-speed operating profiles may increase possible fatigue, buckling, and vibration problems as well as crew discomfort from increased wave slams and acceleration levels, and higher working stresses in the structure. To minimize the risk of operating such vessels in an unrestricted manner will require the ability to monitor operational loads, detect structural damage and performance degradation in the earliest possible stage, i.e., the ability to implement structural diagnosis in real time. This structural diagnostic information can then be used to predict the time to potential structural failure, and to provide strategies for corrective actions to support future navy operation and maintenance.

Despite great interest in the development of an SHM system in many industries, almost all damage detection methods currently in use are either visual or localized experimental methods, such as acoustic or ultrasonic nondestructive evaluation (NDE). These local detection methods are not only time consuming to perform, but they also require the general location of the damage to be known in advance and the structure to be taken out of service for inspection. In many cases, the visual or local inspection is impossible to perform owing to the inaccessibility of major portions of the structure. For example, it would be completely impractical on most naval vessels since the internal surfaces are almost entirely covered with insulation. This is true for both steel and aluminum ships. Long-term monitoring and early detection of damage via SHM systems would therefore present a significant benefit for the management of naval ship structures. In addition, inspections using visual or localized NDE methods do not attempt to provide a quantitative value for the remaining strength of the structure; in other words, these approaches do not intend to provide future progression of the damage and assessment of the residual performance capability of the structure.

As defined in [2], monitoring involves constant measuring or surveillance of ship structures to give actual time histories. The primary purpose is to be aware of what is happening to a structure, to:

- assess structural degradation at the time of inspection and in the future, and to manage the gradually degrading condition as it approaches a stage at which the level of risk or the degree of unreliability become unacceptable [3];
- verify design assumptions, in particular, for innovative structures where design loads and structural responses are based on advanced calculations and model tests, but where there is no experience available [4, 5];
- assess potential failures due to gross errors in the design, fabrication, and operation;
- assess the operational utilization of the structure.

These goals of a ship monitoring system define an integrated SHM process that combines structural diagnosis and prognosis components. The primary objective of a diagnostic task is to assess structural degradation by detecting, locating, and quantifying material or structural component damage using measured data (from past and present) and to establish state awareness through a diagnostic algorithm. Diagnosis covers sensing the current loading and condition of the structure. Although this information is useful on its own, load and structural response measurement system are just a subset of SHM and diagnosis functionality. Prognosis, on the other hand, aims to predict the future capability of structural systems using up-to-date diagnostic information and structural models in addition to estimating expected future loading. Prognostic information with some level of statistical confidence can be developed into decision-making tools to allow the authority to make intelligent deployment and maintenance decisions. This vision of model-based SHM approach is discussed recently in [6].

The development of monitoring, diagnostic and prognostic capabilities for marine structures, which is similar to other application areas, is uneven. Basic load and responses measurement systems are mature and used in applications today. Processing of these data into useful information to aid ship operations, such as providing feedback to the crew for the current loading on the vessel and for maintaining its proper speed to reduce slamming, is also fairly well advanced. However, damage detection capabilities and any prognosis capabilities are currently only developed to the basic science level.

To address the SHM technology in marine structural application, the remainder of this article provides a brief overview of typical ship structural configurations and a variety of loading sources at sea. It will also discuss the importance and the roles of an effective SHM system throughout the life cycle of a ship. Previous and current US navy SHM development efforts as well as required technologies for

health monitoring of ship structures will be reviewed and further discussed. To focus on getting information about the health state of ship structure on demand in real time, we address vibration-based damage assessment approaches and discuss the use of onboard sensors, advanced signal-processing methods to extract information from online data. In particular, examples of wave loadings at high speeds will be given and possible effects on local structures will be discussed. The choice of implementing appropriate technologies for ship SHM application is ultimately to address critical need in operation and maintenance, as well as a compromise between cost and benefit considerations.

2 OVERVIEW OF SHIP STRUCTURES AND SERVICE LOADING

2.1 Description of typical ship structures

A complete description of the wide variety of ship structures is beyond the scope of the current article; however, several established reference books contain extensive descriptions of structural configuration and analysis techniques [7–9]. For the purposes of this article, ship structures constructed out of metallic (steels and aluminum alloys) and composite material will be considered. Metallic ship structures are well developed and broadly similar over a wide variety of ship types and sizes [7, 8]. Steel metallic structures are used on all sizes of vessel, from small harbor craft to the largest naval combatants and cargo ships. The use of aluminum is typically restricted to higher-speed craft and to lengths less than roughly 140 m. Metallic structures typically consist of stiffened plate panels oriented with their long axis parallel to the vessel's keel. These panels are supported periodically by larger frames running transversely across the vessel to form larger grillages. Such grillages are in turn supported by a system of girders and pillars, or may be placed back-to-back to form a double-hull construction. These larger structures are shaped and combined to create the overall hull shape of the vessel. Typical structural configurations are shown in Figure 1. Connections between structural members are typically made by welding, although extrusions are commonly used for certain components in aluminum vessels. In weight-sensitive applications, plate-stiffener combinations may be replaced by a variety of cored panels assembled via welding for steel structures, or extruded in the case of aluminum.

Composite structures follow many of the same principles as metallic structures; however, the structural details differ to allow efficient construction with composite laminates. Composite construction has proven popular in smaller vessels, where reductions in

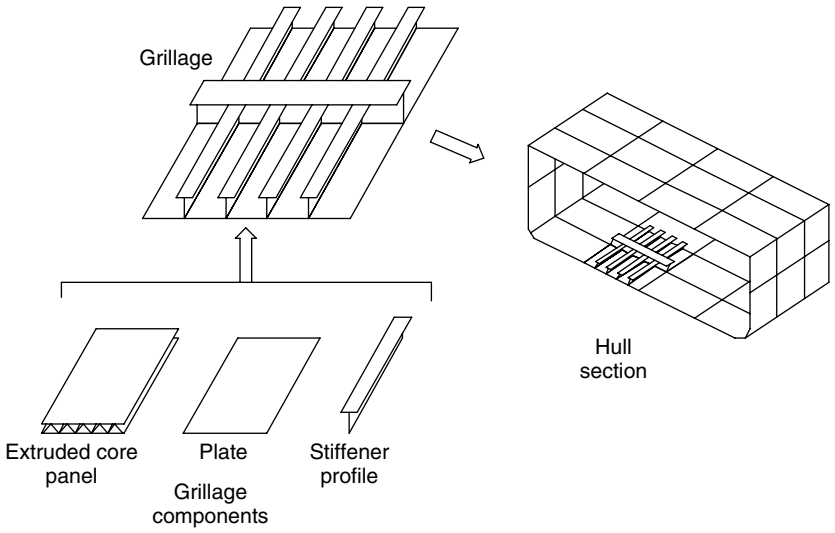


Figure 1. Typical metallic structural components.

weight, maintenance costs, and manufacturing costs for serial production are all achievable if composite construction is used in place of metallic structures. Composites may also be used for certain components of a larger metallic vessel, including deckhouses, masts, and independent tanks. Greene [9] gives a comprehensive overview of the uses of composites to date. In weight-sensitive applications, typically the outer skin of the vessel will be of cored construction, with two composite outer layers separated by a wood, honeycomb, or foam core. Such coring material is also used to build up stiffener and frame shapes. Major sections of the hull are joined by lap construction. A sample composite structure is shown in Figure 2.

2.2 Source of loading

When operating in a marine environment, a typical vessel structure is subjected to a wide variety of loads. Typically, these loads are classified into categories to facilitate load determination and analysis. A common breakdown is to use the time scale associated with the load application; this approach is taken in Principles of Naval Architecture [10], where four time scales are used. Some of the more common loads are summarized below, but this list is not exclusive and typically a vessel-specific list of loads would be generated for each design.

Static loads

- Local pressure and global bending moments, torsion, and shear forces from integration of calm-water buoyancy pressures and forces.

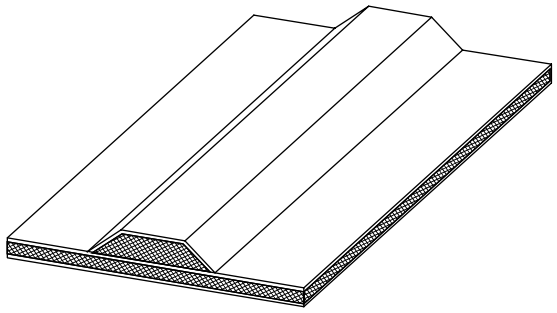


Figure 2. Example composite construction.

- Self-weight of the ship and its cargo.
- Thermal loads.
- Drydocking loads.

Low-frequency dynamic loads

- Hydrodynamic and global bending moments, torsion, and shear forces from integration of hull pressures from passing waves.
- Inertial cargo equipment and vessel self-weight reactions from ship's accelerations in waves.

High-frequency dynamic loads

- Main engine, waterjet or propeller vibratory loads.
- Hydro-elastic response of the overall hull girder.
- Hull girder whipping response from impact loads listed below.

Impact loads

- Slamming pressures from wave impacts.
- Green water impacts on deck.
- Internal tank sloshing impacts.
- Shock, blast, and other weapon effects.
- Aircraft and helo landings.

2.3 Degradation of marine structures

The marine environment is very demanding and all vessel structures suffer from age-related degradation that reduces their strength. A central goal of an SHM system is to monitor and identify such damage, and to quantify its effects on structural safety. Metallic structures suffer from three major categories of degradation [11]: corrosion wastage, fatigue cracking, and local denting and deformation of the structure. Corrosion is a central concern for steel vessels of all sizes, and significant effort is put into maintenance and inspections to minimize the danger of failure from reduced component thickness. While aluminum does not generally suffer wide-scale material loss from corrosion in the marine environment, certain alloys can be sensitive to intergranular and exfoliation corrosion. Fatigue cracking is a common occurrence in aging ship structures. A typical vessel can experience on the order of 10^8 wave loading cycles over its lifetime [8], leading to both fatigue-crack initiation and growth. This is further complicated by

the fact that a large vessel may have tens of thousands of structural connection details where fatigue cracks may initiate. Finally, cargo loading, docking and undocking, handling boats over the side, and a wide variety of impact loading discussed above can lead to local structural dents and deformations. Such local damage, and the associated out-of-plane deformations of the structure, can reduce the structure collapse strength. Additionally, the high local loading accompanying such deformations has the potential to cause yielding or cracking at welded connection details and other locations of high stress.

Composite structures are subject to the same types of loading; however, they suffer from different degradation mechanisms. Corrosion is generally not a concern for composites; however, water absorption can be an issue for certain types of composite construction. Fatigue failures of composites are complex, and may result in many different types of structural degradation, including debonding, delamination, cracking, or fiber breakage [9]. Impact loading may lead to a wide variety of local structural damages, especially in thin, cored construction, and thus, local denting and deformation is an important degradation mechanism in composites structures as well as metallic structures.

3 STRUCTURAL HEALTH MONITORING THROUGHOUT THE VESSEL'S LIFE CYCLE

The application of SHM systems to vessel structures is still an area of active development; however, it is possible to identify several common patterns in the application of SHM systems to date. Although the industry is still in the process of moving toward a complete through-life support of the vessel's structure with a comprehensive monitoring system, there is significant benefit in applying these systems in four main areas. The first area is design validation through sea trials. The second area is through-life load and usage monitoring, the third is damage detection and diagnostic systems, and the final area is prognosis, which has been mainly developed for fatigue life predictions to date. Each of these areas will be reviewed in turn.

3.1 Design validation through sea trials

Modern ship design involves a large amount of engineering modeling and idealization, especially in the areas of loading predictions and structural response. This is especially true for complex vessels or vessels whose design is at the edge of previous engineering experience. In these cases, it is useful to validate the engineering models used in design by testing the completed vessel in several sea conditions. Such an approach is especially valuable for lightweight high-speed vessels whose operation may be restricted to wave conditions below the limit threshold. In these cases, it is useful to validate that the vessel's responses are so modeled as to prevent enforcing needlessly restrictive sea condition limitations, as well as minimizing the probability of in-service damage from responses not anticipated during the design phase. Seakeeping trials may also be carried out after a significant conversion or change in the service environment of the vessel, or in response to a particular operational concern that has arisen in service. The process of conducting seakeeping sea trials to validate motions and loading is well established, especially for naval vessels [12]. The vessel is typically tested in several combinations of sea condition, vessel heading, and vessel speed. The sea conditions are captured as accurately as possible, often using monitoring components, such as deployed wave buoys, wave height meter (TSK) or a wave radar attached to the vessel.

In the most basic trials, the vessel's response may only be captured by accelerometers, which can then be used to estimate both motions and related global loads on the hull. However, more comprehensive trials will use strain gauges, local pressure transducers, and other additional monitoring components also to resolve more accurately the vessel's structural response. An example of a recently conducted seakeeping trial for HSV-2 Swift, a 98-m long, high-speed all-aluminum, wave-piercing catamaran built by Bollinger/Incat USA and delivered to the US navy in 2003, shown in Figure 3, will be discussed later in this article. Both global vibration modes and the stress state in selected local areas can be investigated. Thomas *et al.* [13] present an example of such a trial conducted during the delivery voyage of a high-speed ferry, whereas Pegg *et al.* [14] give an overview of

Figure 3. All aluminum high-speed vessel HSV-2, Swift.

conducting a sea trial and finite element model verification through natural frequency predictions of a SWATH vessel. Such design verification is a fairly routine practice today; however, both the sea condition and structural response monitoring equipment are typically removed or deactivated at the completion of a trial.

3.2 Continued load and usage monitoring

Although a typical sea-trial structural response measurement equipment is removed at the end of the trial period, there is growing interest in continuing to monitor motion and loads throughout the life of the vessel. High-speed ferries and other vessels that operate only in a restricted set of sea conditions are often fitted with permanent accelerometers to warn the crew as sea conditions and resulting accelerations approach the limiting value for the vessel. Such systems have also been extended to include pressure and strain monitoring equipment, as well as data processing and storage facilities so that the long-term loading on the vessel can be recorded for further structural analysis. Similar systems have been developed for the new generation of very large container ships, where limited visibility of the bow region from the bridge (ship navigation room) makes it difficult to accurately monitor bow slamming or shipping of green water on deck by purely visual means. For such vessels, monitoring systems can warn the bridge when large loads or structural responses are occurring. Such systems have also been developed by some vendors to include long-term load recording for fatigue life predictions and predictive modules that can anticipate likely change in motions and loads if the vessel changes heading or speed [15, 16]. Similar systems have also been developed to support transportation and installation of offshore structures, where real-time feedback on experience acceleration and loads is important to ensure that operational restrictions are observed. At the current time, some standards and guidance on developing such systems are starting to appear [17].

The experience with continued load and usage monitoring is currently growing, although most ships operate without any such system. Preventing damage to the ship structure by providing real-time feedback to the operating crew of the current structural loads and responses is one of the key advantages of fitting a continued load and usage monitoring system. Another key advantage of continued load and usage monitoring is that it provides real-time data for early damage identification and a real-world quantitative basis for fatigue life prediction. As reviewed previously, fatigue cracking is one of the primary degradation mechanisms for ship structures. The ability to correlate applied loading and service usage with fatigue damage found in the hull girder opens up the possibility of better planning fatigue inspections and estimating the remaining economic life of the structure, and also developing operational plans that minimize the exposure of the structure to the most damaging conditions.

Although the research and development communities of the navy envision this role of an SHM system application, especially for large high-speed vessels where quantitative measurements are needed to avoid unintentional damage, the continued load and usage monitoring has not been practiced for navy ship structures at the present time. However, there are some current and ongoing programs that will enable the continued load and usage monitoring technology for navy ships in service. This will be further discussed and an example will be given later in this article.

3.3 Structural damage detection and diagnostic systems

Early detection of structural damage via an SHM system would represent a major advance in marine

structure life-cycle management. As discussed in the introduction session of this article, most structural inspections for ship structures are conducted periodically by a trained surveyor, relying principally on visual inspection, backed up by local thickness measurements and NDE techniques in areas of interest. These local techniques are highly accurate but time consuming and can only be employed on a small subset of the structure. Visual inspection is able to detect corrosion and local deformations; however, for fatigue cracks, detection by visual inspection is typically assumed to become dependable only when the crack has grown to several inches in length [18]. Some vessels prone to structural damage may also have problem areas inspected more frequently by the crew using purely visual inspections. The current approaches are often time consuming, especially on vessels where much of the structure is normally covered by structural fire protection or other internal joiner work. Developing structural diagnosis capabilities that can supplement visual inspection is a key technical task for future marine SHM systems.

Various damage detection techniques and methodologies have been developed over the past several decades for aerospace, mechanical, civil, and marine structures. Numerous papers, books, and specialized conference proceedings have been published such as those in [19–23]. Damage detection and structural diagnosis will remain to be an area of active research. Some of these damage detection methods and diagnostic techniques have demonstrated feasibility in laboratory and controlled testing environments. However, we are still a long way from a shipboard SHM system that contains robust damage diagnosis capabilities. More detailed discussions and scientific and technological requirements for a comprehensive damage detection system for ship structures will be given in the next session.

3.4 Prognosis and fatigue life predictions

Along with structural damage detection, prognosis of the future state of the structure is an active area of research [24]. At the present time, most prognosis approaches in use are computationally quite simple. This is perhaps a reflection on the state of the art in structural damage detection, as limitations in knowledge of the structure's current state prevent the application of complex prognosis techniques. Fatigue life prognosis can be made in conjunction with continued load and usage monitoring. Typically, the fatigue life of a vessel's structure is estimated at the design stage by using the S–N fatigue approach and the Miner–Palmgren cumulative damage model [25]. Thus, the loading experienced and recorded by the SHM system in service can be converted into a Miner–Palmgren damage summation for various locations on the structure. This is often done via Rainflow cycle counting of recorded strains, with a suitable conversion to local fatigue stresses [26]. These service damage summations can be used to forecast when the fatigue life of the structure will be consumed. The S–N approach is typically used to predict fatigue-crack initiation and does not give any information on resulting crack growth or the risk of fracture. This limits such prognosis techniques to forecasting when fatigue cracking is expected to become prevalent on the vessel without directly determining quantitative measures of risk of structural collapse. Additionally, if the rate of recorded fatigue damage can be correlated to vessel speed, heading and sea conditions present when the recording was made, it is also possible to use this data to examine changing the vessel's operational restrictions in an effort to extend the fatigue life of the structure. The fatigue prognosis process has seen limited extension to remaining life assessments in the presence of a detected crack (e.g. [27, 28]). Such approaches could also be incorporated into an SHM system. As structural damage detection methods advance, creating more complex prognosis capabilities is likely to become more attractive to vessel owners and operators.

4 HEALTH MONITORING AND DIAGNOSIS

To reduce the risk of operating high-performance and high-speed vessels in an unrestricted manner will require the development and application of structural risk assessment and decision-making tools using innovative, time-efficient health-monitoring technologies. For example, applications of SHM system include long-term embedded sensors, wireless communications, *in situ* global and local damage detection methodologies, and software-based decision

tools. Many emerging technologies developed in the multidisciplinary field of SHM for a broad range of civil and aerospace structures can be utilized in marine structure applications. There are several components required for a general SHM capability, regardless of applications in a specific engineering system. For example, the three required components are

1. accurate knowledge of the structure's failure modes and related material's corrosion and fatigue properties;
2. diagnosis to know how the structure is actually being used via monitoring and detect damage to determine the actual current state of the structure;
3. a prognosis capability based on an engineering model capable of predicting the future condition of the structure based on the diagnosis data.

All three basic components require specific science and engineering knowledge of the system. This is the key for a successful implementation of an SHM system.

4.1 Marine structure monitoring considerations

The definition of marine structure failure modes that may affect each level of the structure is complicated by the continuous nature of the ship construction. Accounting for the interdependencies requires simplifying assumptions to allow a focus on one component or subsystem at a time, and its associated failure modes. Common-cause failures are failures of multiple components due to the same loading event. Each component may be subject to several failure modes that might combine to cause the component to fail depending on the form of the failure mechanism. Each failure mode for each component in the ship structure may have different levels of significance, or consequence of failure. The potential failure locations for a typical ship structure number in thousands. These facts are compounded by the uncertainty of loads that a typical ship experiences. As discussed previously, seaway load effects on the vessel hull are transformed by way of interaction with the vessel structure. Primary loads affect the entire hull, whereas secondary loads affect localized regions of the hull structure. Typically, numerically derived seaway loads are transformed into stresses similar to the manner in which strain gauge information is transformed into stress. The stress values are then input into failure mechanism models at the material and structural level. Prediction of realistic seaway loads is done using probabilistic models, and is a significant challenge for any surface vehicle, but particularly difficult for vessels that travel at high speed leading to great uncertainty as to the accuracy of the probabilistic predictions. In addition, the structural configuration is a challenge for modeling and simulation efforts as there are differences between the design and as-built, structural geometry and dimensions. Production methods for ships structures are not as controlled as other vehicle production such as for airframes. Misalignments, distortions, and residual stresses are introduced into the structure during the welding and fit-up, all of which are difficult to model accurately using a sophisticated numerical modeling and simulation process due to the scale of the structure and the lack of information at a sufficient level of detail. Introduction of new material, details and welding stresses into the structure requires an update to the structural model. It is important to be mindful of the basic engineering system characteristics as well as specific risk and economic concerns for marine structure health monitoring applications.

4.2 Effects of loading from random seaway and slamming

In addition to continued seaway load, one of the challenges facing ship hull structures is overloading that can occur during normal operations at high sea state (slamming) or when encountering an extreme event such as explosion or collisions. More importantly, the challenge is the ability to assess the adverse effects of the ship with accumulative loading demand throughout the service life, in other words, the effects of structural fatigue due to global and local load. An SHM system for ship structure applications needs to address both overloading and fatigue problems since each overloading incidence contributes to overall ship structural fatigue. At the present time, structural response due to local pressure created by slamming is not well understood. For example, HSV-2 encountered many slamming events during

rough water trial runs in winter/spring 2004 [29]. The hull response is measured by two groups' strain gauges that are identified by: global strain T1, stress concentrations related to local strain, T2. The global T1 strain gauge locations were chosen on the basis of a full-ship finite element model to capture primary load response. The local T2 strain gauge locations were chosen to indicate the level of structural response in known or suspected areas of high stress. Examples of T1 and T2 are shown in Figure 4.

Other measurements such as accelerometers are used to record wave impacts that can be used to correlate high strains at a given time. An example of recorded acceleration and global strain is shown in Figure 5.

An example of structural response due to wave slamming is shown in Figure 6. This figure displays

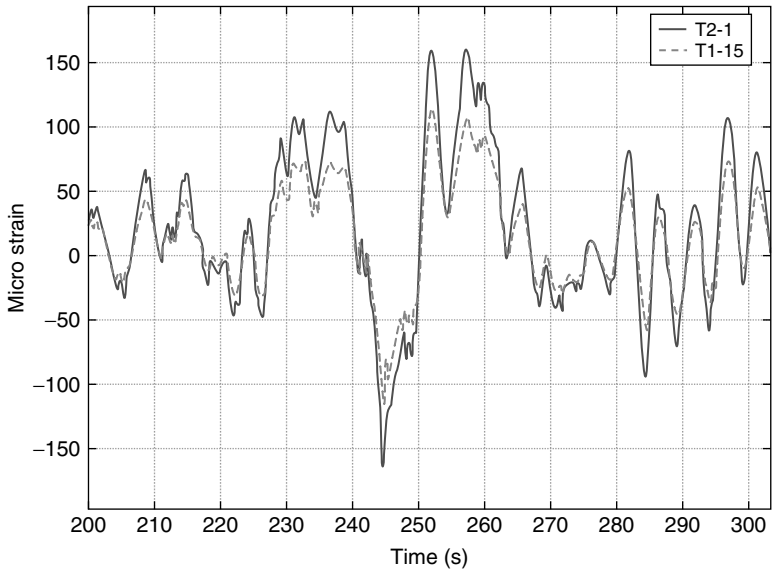


Figure 4. Global (dash line) and local (solid line) strains recorded in one of the seakeeping trial runs.

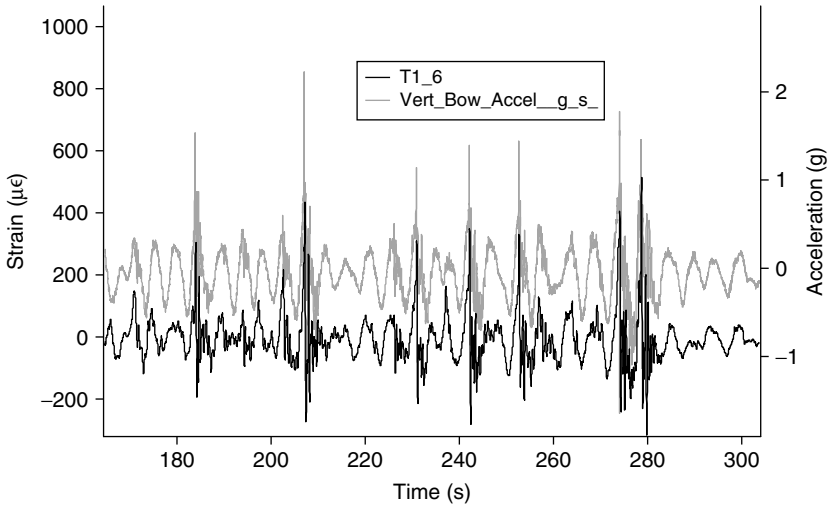


Figure 5. Recorded global strain and vertical acceleration in the bow area in one of the seakeeping trial runs.

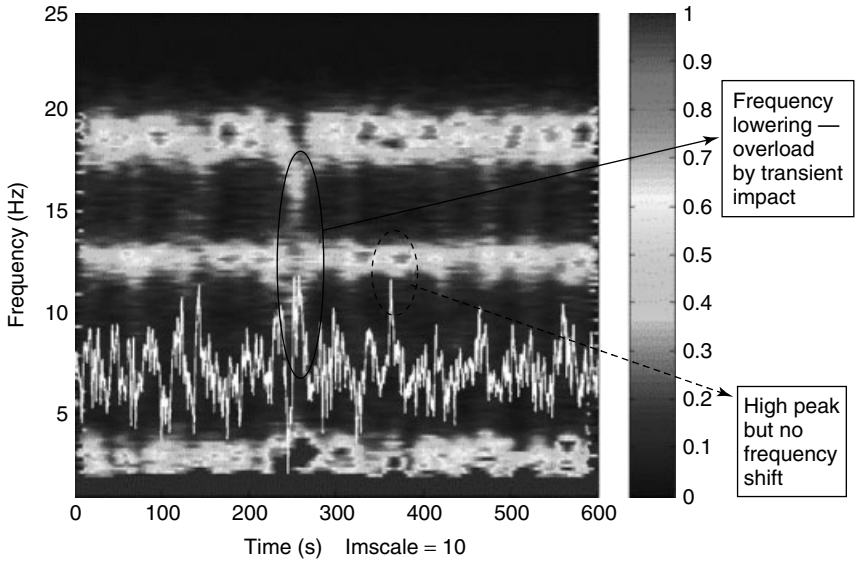


Figure 6. An example of time–frequency analysis of local strain.

the time–frequency analysis of measured strain time history using empirical mode decomposition and Hilbert–Huang transform [30]. The measured local strain T2 (solid white line) is at the deck-head on cross bracing near the side hull joint of the HSV-2 during one of the initial sea-trial runs. The peak-to-peak value of the data is approximately 400 microstrains displayed with the normalized time–frequency distributions.

At about 260 s, noticeable frequency downshift (indicating temporary stiffness loss) occurred for local responses at about 19 and 13 Hz. There are also high global structural responses between 2 and 4 Hz at the same-time window. This is likely due to significant overloading by the slamming event. On the other hand, about 120 s later, a similar peak strain is recorded from another wave slamming. Although the amplitude of the strain from this later slamming is similar to the first one, there is no frequency change observed. These results indicate that wave slamming magnitude and duration both play an important role on local structures.

The current methodologies to calculate accumulative fatigue due to wave impact do not take time durations and structural response differences into considerations. Lightweight ships will encounter more frequent slamming during high-speed operations, and therefore, it may be important to account for these local response differences to correctly assess the overall structural fatigue. This analysis result may lead to requirements such as more complex signal analysis that is capable of identifying time durations and determining the details of structural responses or other adjustments to the diagnosis component.

4.3 Real-time monitoring and damage detection

Global damage detection methods, also known as *vibration-based* (or *dynamic-based*) *methods*, are a promising technique for real-time shipboard SHM applications. This classification is based on the relative relationship of the signal wavelength with respect to the defect size, as well as to the overall structural dimensions. Typically, the local approach uses the wavelengths smaller than the size of damage to be detected and more sensitive to incipient damage, e.g., high-frequency guided wave, but the sensor and data handling cost can be overwhelming. In contrast to local NDE technologies and other local methods, global damage identification approach offers distinct advantages to damage detection of large and complex structures. For example, these methods require simultaneous measurements of much fewer locations to

provide global coverage of large area of the structure and they are capable of working at a "system level". It is not necessary that the sensors are located close to the damage site, which is an advantage that is significant for overall structural state assessment. Furthermore, they tend to be time effective and less expensive than most alternatives. The fundamental principal upon which vibration-based methods are founded is that small changes in a structure cause behavioral discrepancies in its vibration response [31, 32], and they can use the measured time traces and the traveling sequence of abnormality or sudden change in the time trace to locate damage. However, the dynamic-based approach may require more complex signal analysis and diagnostic algorithms that are capable of identifying pertinent features to infer the corresponding structural anomaly.

Many recent studies in global damage detection approach are stimulated by the new developments in advanced signal processing, as reported in [33]. To date, most damage detection methods use Fourier analysis as the primary signal-processing tool. From the resulting spectra, modal properties or damage-sensitive features are extracted to detect change in the signal properties. However, often the most important aspect of damage is in how these characteristics change in time as well as whether the system is nonlinear. Examination of the modal properties of a structure can neither typically identify a sudden change in system properties due to complete lack of time information nor can it provide information about whether a system is nonlinear. The nonstationary nature of measured signals produced by mechanical faults or structural damage suggests that time varying method of signal processing can be used to detect such faults as an alternative approach to the classical, Fourier-based methods. In particular, time-variant modeling and nonlinear signal feature extraction can offer a powerful solution to many nonlinear and nonstationary problems, especially when combined with time–frequency adaptive methods. (Some of these methods applied to SHM are discussed in greater detail in **Time–frequency Analysis**; **Wavelet Analysis**; **Damage Detection Using the Hilbert–Huang Transform**; **Nonlinear Features for SHM Applications**.) The development of real-time damage detection algorithm for large and complex structural systems such as ship structures will need to utilize these advanced signal process tools.

A robust damage detection system for ship structures includes capabilities to identify the presence of damage, to determine the location and extent of the damage, to establish the nature of damage, and to assess the cause of the damage. This is similar to four levels of damage assessment scale proposed in the civil engineering community about 15 years ago [34]. In addition, the system should be sensitive to critical damage mechanisms (e.g., impact and fatigue) and capable of characterizing the ensuing damage with an acceptable level of accuracy and reliability. In addition to the requirement of having to be sensitive to damage features, the advance signal-processing method also needs to be able to distinguish the vibrational characteristics change of a structure due to damage versus the change from other environmental and operational effects. To consistently identify these differences presents a technical challenge in the development of robust damage detection algorithms. Moreover, the most challenging fact is that damage is typically a local phenomenon and generally does not significantly modify the lower frequency response characteristics or global response of a given structure. In practical applications of real-time diagnosis, these dynamical response changes also need to be detected under normal operational conditions.

5 CURRENT HULL MONITORING SYSTEM AND REMAINING CHALLENGES OF MARINE STRUCTURE APPLICATIONS

5.1 An example of current hull monitoring system

The development of high-speed and high-performance vessels greatly accelerated the interest in equipping navy ships with an operational real-time monitoring system. Currently, there are several navy acquisition programs that contain ship and hull monitoring aspects. An example of such is a hull condition monitoring project for the Joint High Speed Vessel Program. This program develops a hull monitoring system that is composed of stand-alone commercially available software and hardware. It provides data

measurement, data collection and conditioning, data processing and evaluation, as well as results presentation and storage. An overview of the main contents of this hull monitoring system is shown in Figure 7.

The hardware system utilizes sensors and infrastructure such as accelerometers, resistance strain gauges, and wave measurements available from previous seakeeping trials [29]. These sensors, distributed throughout the hull structure, are tied back into a data recorder and acquisition systems.

It is important to note that there are many more practical issues that need to be addressed for continued shipboard data acquisition. Generally speaking, the available space and weight of a given vessel is very limited. Long-term load and usage monitoring system must efficiently utilize hardware size, weight, and space. If not addressed appropriately in the early system implementation, the whole system may be turned off or taken apart owing to other needs and priorities. The hardware system, such as computer, cable, and data acquisition, is significantly reduced in size and streamlined in design in the current hull monitoring system development effort compared with the measurement system used in seakeeping trials. Software developments include real-time indicators to alert the operators of structural overload and running fatigue damage estimates based on global wave impact. It can indicate the exposure of the ship by recording wave impact amplitude and nominal high stresses, and display a trend of the above indicators. These real-time display indicators for ship navigation and engineering spaces are intended to provide instantaneous information and knowledge for the operators. In addition, background processes that include statistical calculations, load, and fatigue damage accumulative estimation are periodically updated by running an "analysis engine" [35]. More details on the hull monitoring system development can be found in [36].

The hull monitoring system example discussed above is leveraging experiences gained from previous monitoring efforts on navy vessels as well as routine practice from design verification sea trials. Although it does not have structural diagnosis and prognosis capabilities, it delivers improved data recording, motion, load, and usage monitoring system. This project provides valuable experience toward the development of integrated SHM system that ultimately includes structural evaluation and diagnostic and prognostic capabilities. These capabilities will provide intelligent strategies and automated processes to effectively manage future naval ship structures.

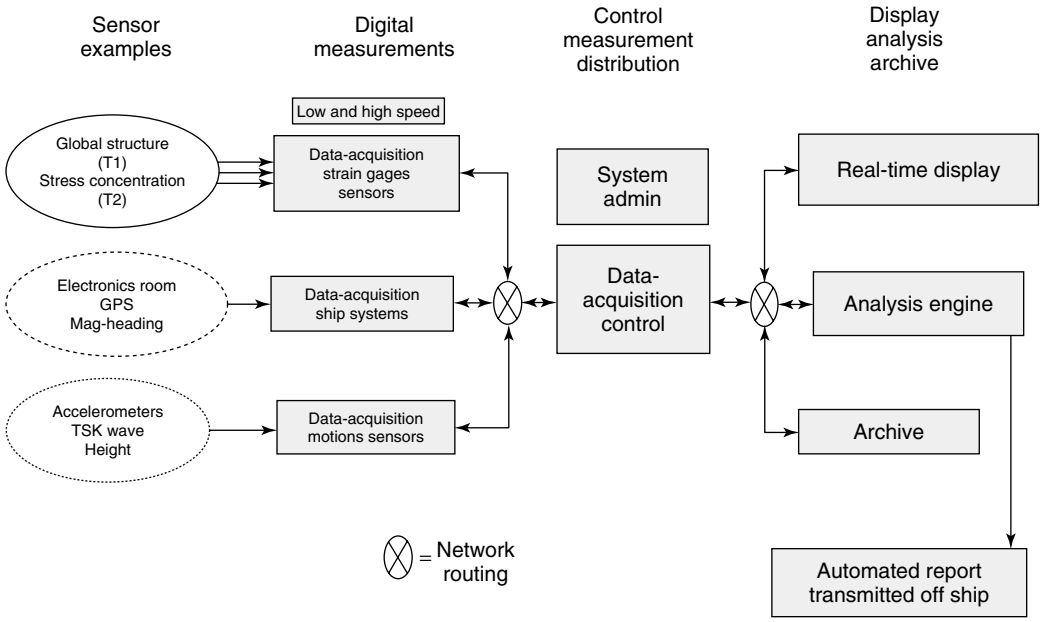


Figure 7. Overview of hull monitoring system for Joint High Speed Vessel Program.

5.2 Remaining challenges of implementing SHM systems for marine structures

As discussed previously in this article, SHM systems for marine applications are still largely developmental. Systems for sea-trial design validation have been widely used to date. Continued load and usage monitoring systems for fatigue prognosis exists on some commercial high-speed vessels and are being implemented on naval vessels in recent years. Real-time monitoring and reporting of vessel motions and structural responses are also increasingly common. Often these measurements are reported to the crew in comparison to predetermined threshold values, and used to warn the crew about dangerous loadings. Although these systems provide useful real-time feedback to the crew and are growing in complexity, most of them do not make any attempt to diagnose the current health of the structure or detect and localize early-stage damage such as fatigue-crack propagation. Likewise, prognosis applications of the current generation system tend to revolve around recording strain cycles and estimating the percentage of a predetermined fatigue budget that has been consumed to date.

Within the marine community, there is keen interest in more advanced diagnosis and prognosis capabilities that would allow for more targeted inspection and repair actions as well as support for handling accident situations, where the vessel has suffered damage from collision, grounding, or enemy actions in the case of naval vessels. However, several challenges still remain to be overcome to achieve such capabilities. In addition to the typical challenges related to sensor and signal processing common across many disciplines of SHM, marine structures suffer from a multitude of potential damage locations, i.e., "hot spots", including fatigue cracking and local structural collapse. At the current time, it is not economically feasible to monitor all of these locations. Prioritizing locations is also difficult as it is unheard of to test a full-scale prototype of a design for fatigue life as is common in the aircraft industry. Thus, ranking of hot spots must be done by analytical means and engineering judgment and the resulting system may miss some of the important damage locations. In addition to material, dimension, geometry, structural detail, and construction complexities of marine structures, difficulties of implementing SHM systems are also compounded by the uncertainty of seaway loads and their effects on the hull structure by way of local interactions as well as the scales of structure failure modes. Common-cause failures can be damages of multiple components due to the same loading event or several failure modes that might combine to cause the component to fail over time. Well-planed programs and collective efforts from many agencies and resources are needed to implement an effective and comprehensive SHM system for shipboard applications.

6 SUMMARY AND DISCUSSION

Ship structure and material, whether composite or metallic, degrade over time due to operational loads and the challenging ocean environmental conditions. Accounting for degradation, damage and repair in determining structural health is a significant challenge to ship operation and maintenance. The development of integrated SHM system with monitoring, diagnosis, and prognosis capabilities is critical for assessing the true performance and safety of the ship structural system. The current state-of-the-art marine SHM is not yet advanced enough to produce complete SHM system for shipboard field deployment. Both diagnosis and prognosis capabilities are still active research area with significant hurdles to overcome before practical systems are ready for widespread application. However, incremental development of continued load and usage monitoring aspects of SHM systems is actively being perused by both the US navy and the commercial marine industry. These ship hull monitoring systems can provide valuable experience toward the development of integrated SHM system that ultimately includes structural evaluation, diagnostic and prognostics capabilities.

It may be necessary to implement a diagnostic strategy in multiple stages for ship hull and local structural assessment. Initially, a real-time, onboard sensor network combined with dynamic-based damage detection algorithm is used to pinpoint possible problems and identify their locations. Then, further evaluations are justified using more localized techniques as well as incorporating sensor and inspection information into fracture-based fatigue models to evaluate details with known or suspected flaws to mitigate the risk of fracture and support

remaining life or time-to-repair prediction. This multiple stage approach can enhance the potential for finding damage early in the damage progression as well as perform the assessment for the entire ship to preventing catastrophic failure. As technologies become more sophisticated, such an approach could also provide feedback on how to operate the vessel for a given damage condition and provide the opportunity to greatly improve operability, maintenance, and repair strategies.

REFERENCES

[1] ISSC, In *Report of Committee*, Version 1, Hsu P, Wu Y (eds). Elsevier Applied Science, 1991.

[2] ISSC, In *Report of Committee*, Version 2, Mansour A, Ertekin R (eds). Elsevier Applied Science, 2003.

[3] Brooking M, Barltrop N. *Ship Structural Management, Shipbuilding Technology International*, 1993; pp. 168–171.

[4] Iaccarino R, Monti S, Sebastiani L. *Evaluation of Hull Loads and Motion of a Fast Vessel Based on Computations and Full Scale Measurements*. NAV 2000: Venice, 2000.

[5] Grossi L, Dogliani M. *Load and Seakeeping Assessment of HSC Based on Full Scale Monitoring*. NAV 2000: Venice, 2000.

[6] Hess P. Structural health monitoring for high-speed naval ships. In *Structural Health Monitoring 2007: Quantification, Validation, and Implementation*, Chang FK (ed). EDStech Publications, 2007, pp. 3.

[7] Lamb T (ed). In *Ship Design and Construction*. SNAME: Jersey City, NJ, 2004.

[8] Hughes O. *Ship Structural Design*. SNAME: Jersey City, NJ, 1988.

[9] Greene E. *Marine Composites*, *Second Edition*. Eric Greene Associates: Annapolis, MD, 1999.

[10] Paulling J. In *Strength of Ships, Principles of Naval Architecture*, Lewis E (ed). SNAME: Jersey City, NJ, 1988; Vol. I.

[11] ISSC, Report of committee V. 6, condition assessment of aged ships. In *Proceeding of the 16th International Ship and Offshore Structures Congress*, Frieze PA, Shenoi RA (eds). University of Southampton Press: Southampton, 2006.

[12] Lloyd ARJM. *Seakeeping: Ship Behaviour in Rough Weather*. Lloyd: Gosport, 1998.

[13] Thomas GA, Davis MR, Holloway DS, Wayson NL, Roberts TJ. Slamming response of a large high-speed wave-piercer catamaran. *Marine Technology* 2003 **40**(2):126–140.

[14] Pegg NG, Gilroy LE, Kumar R. Full scale verification of finite element modeling of a 75 tonne SWATH vessel. *Marine Structures* 1995 **8**(3):211–228.

[15] Lyngsø latest in automation. *The Naval Architecture*, April 2007: 27.

[16] SRA to be extended to passenger ships. *The Naval Architecture*, October 2005: 34.

[17] American Bureau of Shipping (ABS), *Guide for Hull Condition Monitoring Systems*. ABS: Houston, TX, 2003.

[18] Demsetz L, Cabrera J. *Detection Probability Assessment for Visual Inspection of Ships*, Ship Structure Committee Report SSC-408. Ship Structure Committee: Washington, DC, 1999.

[19] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Report No. LA-13976-MS. Los Alamos National Laboratory, 2003.

[20] The Royal Society, Structural Heath Monitoring. *Philosophical Transactions of the Royal Society A* 2006 **365**; special issue.

[21] Chang FK (ed). In *Structural Health Monitoring 2003: From Diagnostics and Prognostics to Structural Health Management*. EDStech Publications, 2003.

[22] Chang FK (ed). In *Structural Health Monitoring 2005: Advancements and Challenges for Implementation*. EDStech Publications, 2005.

[23] Chang FK (ed). In *Structural Health Monitoring 2007: Quantification, Validation, and Implementation*. EDStech Publications, 2007.

[24] Inman DJ, Farrar CR, Steffan V, Lopes V (eds). In *Damage Prognosis for Aerospace, Civil and Mechanical Systems*. John Wiley & Sons: Chichester, 2005.

[25] Miner MA. Cumulative damage in fatigue. *Journal of Applied Mechanics* 1945 **12**:A159–A164.

[26] Downing SD, Socie DF. Simple rainflow counting algorithms. *International Journal of Fatigue* 1982 **4**(1):31–40.

[27] Xu T, Bea R. Fatigue of cracked ship critical structural details: cracked S-N curves and load shedding. *International Journal of Offshore and Polar Engineering* 1998 **8**(2):130–139.

[28] Okawa T, Sumi Y, Mohri M. Simulation-based fatigue crack management of ship structural details applied to longitudinal and transverse connections. *Marine Structures* 2006 **19**(4):217–240.

[29] Brady TF. *Global Structural Response Measurement of SWIFT (HSV-2) from Jlots and Blue Game Rough Water Trials*, NSWCCD-65-TR-2004/33, December 2004.

[30] Salvino LW, Pines DJ, Todd M, Nichols JM. EMD and instantaneous phase detection of structural damage. In *Hilbert_Huang Transform and Its Applications, Interdisciplinary Mathematical Sciences*, Huang N, Shen S (eds). World Scientific, 2005; Vol 5.

[31] Doebling SW, Farrar CR, Prime MB, Shevitz DW. A summary review of vibration-based damage identification methods. *The Shock and Vibration Digest* 1998 **30**:91–105.

[32] Farrar CR, Doebling SW, Nix DA. Vibration-based structural damage identification. *Proceedings of the Royal Society of London, Series A* 2001 **359**:131–149.

[33] Staszewski WJ, Worden K. Signal processing for damage detection. In *Health Monitoring of Aerospace Structures*, Staszewski WJ, Boller C, Tomlinson GR (eds). John Wiley & Sons: Chichester, 2003.

[34] Rytter A. *Vibration Based Inspection of Civil Engineering Structure*, Ph.D. Dissertation, Department of Building Technology and Structural Engineering, Aalborg University, 1993.

[35] Hildstrom GA. *JHSV Analysis Engine*, NSWCCD-65-TR–2006/15, June 2007.

[36] Salvino LW, Brady TF. Hull structure monitoring for high-speed naval ships. In *Structural Health Monitoring 2007: Quantification, Validation, and Implementation*, Chang FK (ed). EDStech Publications, 2007.

Chapter 143

Diagnosing Offshore Machines and Power Plants Using Vibration Methods

Andrzej Grzadziela

Faculty of Mechanical and Electrical Engineering, Naval University of Gdynia, Gdynia, Poland

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 OFFSHORE TECHNOLOGY

The offshore industry uses some of the most advanced and extremely risk-laden technologies in the world. The basic industry structure is an oil rig, consisting of a high steel tower floating on large tanks, full of air. The tanks are underwater, making the oil rig mobile, and are called *submersible*. Sometimes, vessels are used for offshore technology and they are called *drillships*. When the sea is not very deep, rigs stand on tall legs on the sea bottom—modern types are called *guyed tower platform* (GTP). Sometimes, rigs stand permanently, but most often the legs retract and the platform moves from one position to another—these are known as *jack-up* oil rigs.

Oil rigs have a stainless steel screw capped by cemented carbide called a *drill*, which bores a narrow hole, 6–20 in. (0.1524–0.508 m). wide into the sea bottom. Then the machine pushes a strong, steel pipe into the borehole. If oil is near the surface of the sea bottom, it flows into the pipe at high pressure. The deeper the drilling required, the more advanced the technology that is needed.

The oil that comes from the borehole is generally a thick, sticky, dark brown liquid. It is a mixture of many kinds of hydrocarbons, both liquids like petrol and kerosene and gases such as propane, butane, etc. A refinery located in the tower separates oil into different kinds of liquids and gases. The oil is

carefully heated to a temperature higher than that of boiling water, so the petrol, which boils easily, turns into vapor. The kerosene, which boils at a higher temperature, remains liquid. The vapor of petrol is cooled, and the two liquids are stored separately. Gases are purified and pressured into a compressor, which carries them out into tanks. Sometimes, gases are carried out into pipes placed at the sea bottom, which connect the oil rig to the mainland [1].

2 THE OIL RIG-SPECIFIC MACHINES AND EQUIPMENT

The offshore industry is a combination of mechanical, electrical, and chemical technology. The harsh sea environment demands that all machines have to fulfill specific requirements. They cover the following aspects:

- marine classification certificates;
- fire resistance;
- reliability;
- EEx certificate;
- ISO requirements for all types of machines.

The main machines using diagnostic procedures in the production process on the oil rig are

- oil pumps with propulsion systems;
- gas compressors with propulsion systems;
- electric power station;
- propulsion systems of auxiliary vessels like tugs, tankers, etc.;
- automation of production processes.

An oil pump is a machine that carries out oil from the main pipe into separators; it is a basic constituent of the production chain. The pump, usually a centrifugal type, is capable of handling sand- and gravel-laden oil without undue clogging or wearing. Horizontal and vertical oil pumps are designed for heavy-duty applications such as offshore pipelines. A pump has a motor unit enclosed by clamshell housing sections or it stands free. It is powered by electric motors, diesel engines, or gas turbine engines (GTEs). The pump moves a working fluid through the housing to an outside load. Each section has intake walls defining a portion of an intake chamber and outlet walls defining a portion of an outlet chamber. Pumps work sometimes 24 h per day for 14 days.

A gas compressor is a mechanical device that increases the pressure of the mixture of gases coming from the separator by reducing its volume. Compression of the mixture of gases increases its temperature, so the cooling system is an important part of gas compressors. Reciprocating machines are mainly used as gas compressors on rigs. They are commonly powered by electric motors.

The electric power station is an important part of an oil rig. It supplies energy for technological processes, life support systems, and logistics. Electric generators are commonly propelled by gas turbines or diesel engines. Gas turbine propulsion units of electric power stations need gear boxes because of the high rotational speed of power turbines. At least one electric generator works 24 h per day, so reliability of the energy system is the main purpose of operational policy—oil rigs are not connected with the mainland's energy system.

The propulsion systems of an auxiliary vessel are subjected to specific sea loads due to wave and dynamic factors associated with the technical purpose of the vessel. Tugs and tankers play an important role in offshore technology. They ferry petroleum and kerosene, and supply food, hydroengineering materials, technical devices, and crew. Moreover, tugs support diving works and can be used as fire vessels, an important role due to the seriousness of fire and eruption alerts on the rigs.

3 HEALTH-MONITORING OIL RIG MACHINES—REQUIREMENTS

Currently, vibration procedures for diagnosing machines are very popular in offshore technology. They provide lots of information about the dynamic behavior of rotating machines and their structure [2]. The main disadvantages of the vibration method of diagnosing are the requirement for an *on-line* procedure and the need for knowledge of construction, materials, and forces reacting inside the machine. However, an *off-line* procedure is applied mainly in ISO standards' tests, which are sometimes quite useful. The off-line vibration method of diagnosing

needs statistical analysis of vibration signals to create operational characteristics.

Every monitoring system working on the rig has a declaration of conformity of *EEx* requirements (working in potentially explosive atmospheres) and ensures reliability of production. The principal requirements are:

- monitoring system—global or separated into subsystems;
- monitoring system should cover all strategic machines;
- guarantee redundant data transmission for a global system;
- guarantee shutdown of machines in fire and eruption alerts;
- recording all alerts to the memory, including operational parameters;
- connection to acoustic and visual signals alerts.

4 MECHANICAL VIBRATION MEASUREMENTS—PROCEDURES AND EQUIPMENT

Vibration monitoring systems demand specific, high-quality gauges (mainly accelerometers), cables, amplifiers, and analyzers. A typical, simplified configuration is presented in Figure 1.

There are three quantities that are of interest in vibration measurements: the displacement, D; the velocity, V; and the acceleration of vibration, A. Signals are analyzed as time waveforms, shocks, and spectra. Diagnosing systems can be made *on-line* (real-time monitoring systems) or *off-line* systems.

On-line monitoring systems are recommended for offshore technology, but *off-line* vibration measurements also perform an important role. Drillships often do not need a global monitoring system, since it can be very expensive and is only used rarely. They use *off-line* systems adapted to the specific operational requirements.

Simple periodic vibration may be analyzed as a time waveform, but unfortunately, it can be applied only to the very simple machines. The basic parameters of time waveform parameters are zero peak, peak-to-peak, root mean square values, etc. Because much more complicated machines usually work on rigs, this procedure is seldom used.

Monitoring systems generally perform spectral analysis, bandwidth filtration, and advanced statistical analysis of recorded signals. The following methods of vibration signals analysis are adapted:

1. FFT analysis without tracking;
2. analysis based on ordinary FFT spectrum analysis, where FFT spectra and slices, are shown as function of RPM measured by a tachometer;
3. order tracking, where the frequency range of the order analyzer changes in accordance with rational speed of the machines shaft—tracking;
4. autotracking, where Bayesian statistics is used for parameter estimation [3].

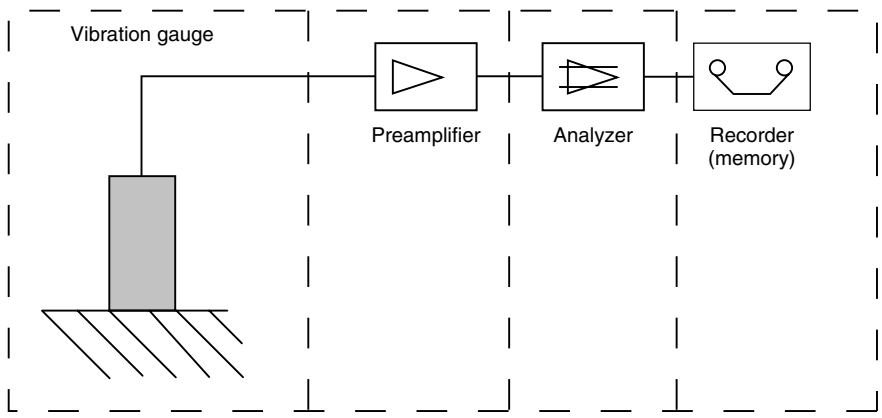


Figure 1. Typical vibration configuration.

Vibration and thermodynamic analysis requires the following transducer standardization (vibration and others):

- relative vibration, axial position, and keyphasor measurement;
- absolute vibration—accelerometers should be correctly suited to the type of machine;
- thermocouples or thermoresistor for temperature measurements;
- pressure transducer (static and dynamic measurements).

5 RECIPROCATING GAS COMPRESSORS— PRINCIPLE OF VIBRATION MONITORING SYSTEM

Reciprocating gas compressors are monitored by the manufacturer's system or an individually prepared system in accordance with ISO recommendations—ISO 10816-6:1995(E). Every manufacturer's monitoring system is a result of its experiences and knowledge, so the configurations of gauges and analysis types are specially declared for all types of machines [4, 5].

Sometimes, compressors do not have monitoring systems. *On-line* or *off-line* monitoring systems should fulfill the manufacturers' vibration parameter limits or the limiting values of overall vibration measured on the machine structure. This is determined by many parameters. Classification of compressors depends on their application, size, power, configuration, mounting, speed, etc. Limiting values of the vibration parameters (D, V, and A) are presented for a few classification numbers of machines.

Compressors with the same level of power reciprocation can be used in different machines. In this case, the limits of rigid vibration values can shut down a compressor that is in good technical condition. To avoid this situation, classifications that may be agreed upon between manufacturer and customers are made on the basis of experience or as a result of operations. Application of computer simulation for monitoring systems and finally diagnosing the technical state of gas compressors should be done during calculation and project process. This is a very useful tool for model analysis. There is a problem when the manufacturer does not include this kind of know-how in the technical specifications given to the customer. This situation occurs quite often for gas compressors used in offshore technology. The main problem is to assess the technical state of a machine and find the source of the problem. One of the potential solutions is to apply the vibration method for diagnosing the actual technical state [6, 7]. When the compressor is not identified as an object for *FE* (*final elements*) calculations, the situation requires application of another operational tool [8, 9]. The only way then is through periodical vibration research and tests accomplished through preferred standardization. Tests are accomplished according to ISO 18061-6, in three axes (Figure 2).

Research assumes periodic vibration tests as a function of crankshaft rotation and operational time. Accelerometers are mounted on the cylinder tie-bolt nut and to the bedplate in three axes. Every measured point is analyzed and each vibration spectra provides "fingerprints" of that particular machine. Every point of the spectra is stored in the data base and analyzed as a time function (Figure 3). Permanent measurement is defined for the *on-line* diagnosing system. A scheme of analyzing parameters is presented in Figure 4 and the measured point on the compressor cylinder is shown in Figure 5.

Off-line and portable diagnostic systems are inexpensive, but they have two major disadvantages:

- they fulfill only guarantee period requirements;
- all analyses are sustained.

On-line systems avoid these disadvantages and allow continuous monitoring. The main points are:

- the monitoring system makes managing the compressor's operation time much more rational, especially before maintenance;
- the vibration method is noninvasive and does not require taking the compressor out of service [10];
- the monitoring system leads to important economic benefits, especially in reliability improvement.

Finally, requirements for reciprocating compressors demand the following measurements (vibration and others):

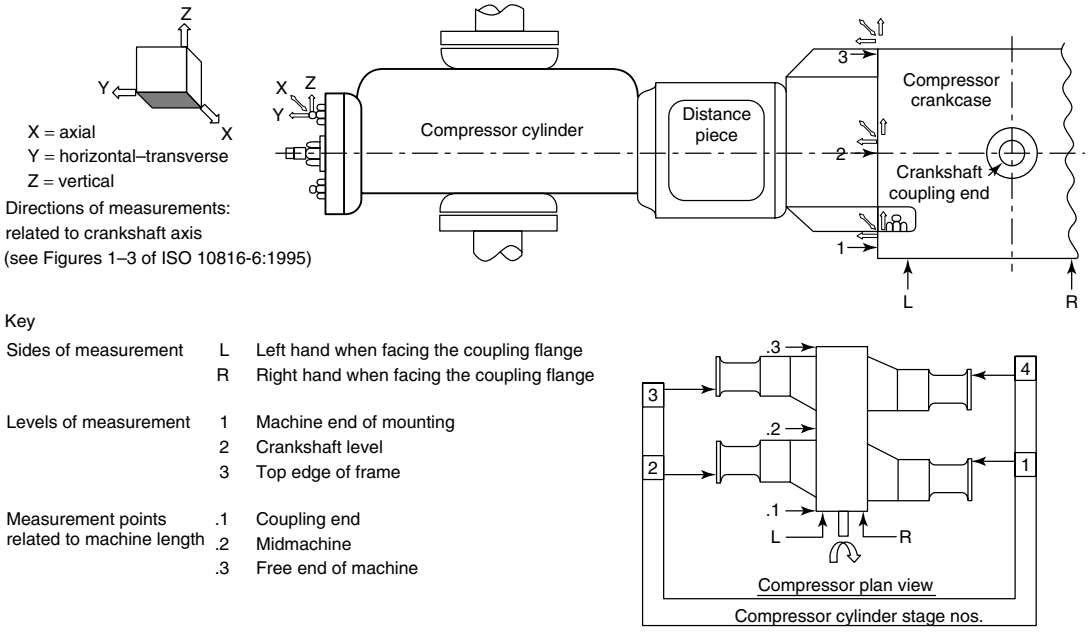


Figure 2. Location of measured points of a compressor.

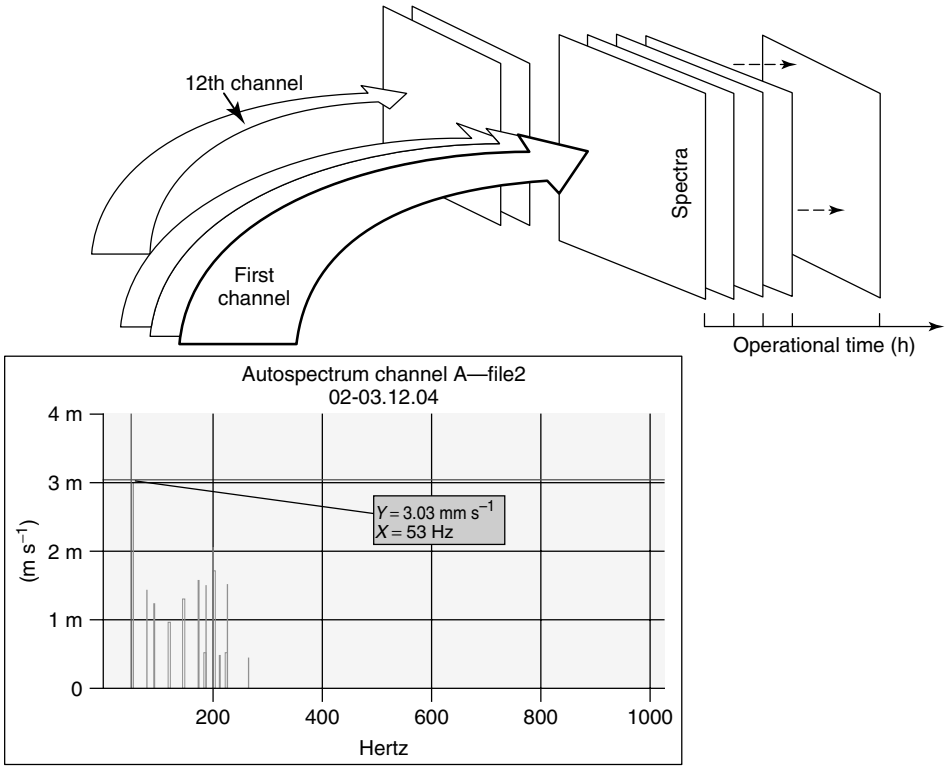


Figure 3. Structure of spectra data base.

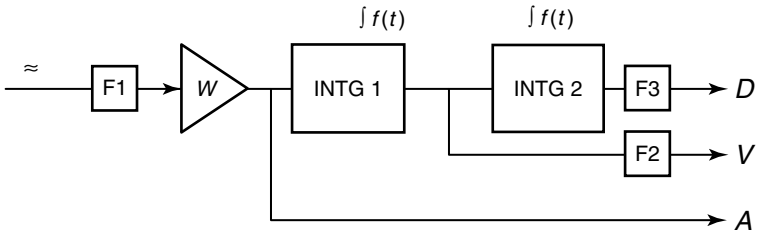


Figure 4. Scheme for measuring D, V, and A parameters.

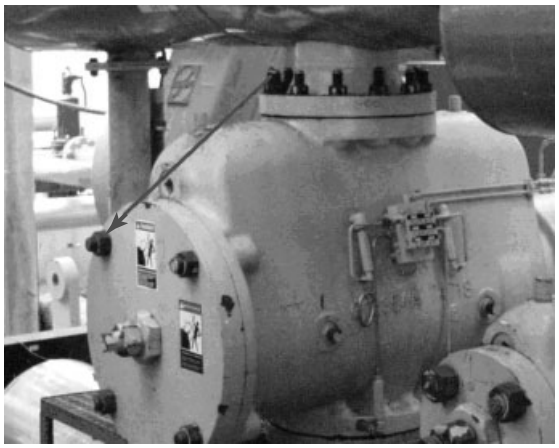


Figure 5. Measured point—cylinder tie-bolt nut.

- relative shaft vibration (two proximity probes on each slide bearing of the compressor and the driver machine);
- revolution of shaft;
- absolute vibration on each bearing housing of the electrical motor and the piston compressor;
- absolute vibration on the cylinder and the frame of the piston compressor;
- measurement of rod drop on each rod (two proximity probes in two directions—horizontal and vertical);
- temperature of the slide bearings (two redundant transducers mounted in the bearing sleeves);
- temperature of the windings (for electrical motor);
- temperature of suction and discharge valves;
- pressure measurement on each cylinder, damping device, and pipe [11];
- compressor suction and discharge pressure, and temperature measurement;
- lubricating oil pressure and temperature measurement;
- piston rod seal temperature measurement;
- piston rod seal leak—flow and pressure measurement;
- other important measurements: pressure, pulsation, temperature, power, etc., agreed upon between the vendor and the user of the machines.

6 PROPULSION SYSTEMS OF ELECTRIC POWER STATION AND OIL PUMPS—ON-LINE MONITORING SYSTEMS AND OFF-LINE MEASUREMENTS

Modern electric power stations that work on rigs are powered by GTEs and they are called *gas turbine generators* (GTGs; Figure 6). Application GTEs enable a fast start and the ability to take on a full load in a relatively short time interval (about 1 min). This is why GTEs can be adapted as the propulsion units for sand pumps and gas compressors

Figure 6. GTE enclosure on a rig.

on rigs [12]. Moreover, GTGs and GTEs consume one of the liquids produced—petrol [13]. Diesel engines are generally used for auxiliary generators.

Contemporary GTEs used for aircraft or shipboard propulsion have a wireless *on-line* monitoring system. The system operates time–frequency-based signal processing, probabilistic analysis, neural network classifiers, fuzzy logic decision support analysis, and Bayesian networks, just a few of the algorithmic approaches currently being implemented to provide for better awareness and prediction of equipment health state. The following mechanical faults should be recognized by the monitoring system:

- unbalance
- shaft interaction
- eccentricity
- squeeze film malfunction
- blade rub
- rotor instability
- oil in rotor
- flange/joint slip
- looseness
- misalignment
- swashed track.

To obtain reliable data on diagnostic parameters, monitoring of the gas turbines installed in electric power stations is carried out by means of the multi-symptom diagnostic model, a main feature of which is recording and analyzing vibration signals [14]. The measurements are mainly aimed at determination of permissible in-service imbalance and appropriate assembly of turbine rotors on the basis of selected vibration parameters, and, finally, determining their permissible operation time resources. Another task of monitoring vibration parameters is to shut down the engine in case vibrations parameters exceed their limits.

Another important problem is shaft misalignment between engines, reduction boxes, and generators (or pumps). Dynamic reactions, resulting from exceeding the allowable alignment deviations of the torque transmission elements, are liable to cause failure of the propulsion system in a relatively short time.

The dynamic problems of GTE and GTG are related to basic elements such as rotors, bearings, struts of bearings, engine body, type of substructure, hydro- and meteorological conditions (sea environment), and gas-flow parameters inside the engine. Dissipation of energy in rotating machines displays as a torque, revolutions, temperature, gas flow, and vibration. Vibrations are related to rotor unbalance, oversize of tolerated axis slope of shafts, abrade of blade tips with the inner roller, wear of axis and radial bearings, asymmetry of spring stiffness and damping characteristics of the rotor and their parts, and irregularity of gas-flow forces [15, 16]. Emission of vibration brings a lot of information including knowledge of the technical state. Measurements of vibrations, their identification, classification, and mathematical analysis, including trend function, bring information on the actual technical state and allow predicting the wear process in the future.

Every rigid body like the rotor has six degrees of freedom; however, a deformation body has unlimited degrees of freedom. Rotating machines like GTE have total degrees of freedom equal to the sum of all degrees of freedom of the engine's parts reduced by the number of rigid nodes connecting these elements of the engine. Each rotating part of the engine can be represented by physical characteristics obtained from vibration measurements or from modeling of geometry and material—a rigidly joined structure. The application of a specified model of a rigid body gives ordinary differential equations. The deformation body needs portative differential equations. The second assumption is much more complicated, but it can be closer to the real object, especially when it works over a wide range of rotary speed, like the sand pump propulsion system. This is the reason for choice of the second model for modeling. The scheme for a diagnostics model GTE is presented in Figure 7.

Residual imbalance occurs at each and every stage of the rotor. Two vectors of imbalance at both ends of the shaft can represent the imbalance effects. The vectors have different values and phase shifts. This FE individual and average model creates responses of unbalancing that can be compared with the reports of vibration measurements. The most sensitive imbalance response point at the GTE is the front frame over a vertical strut. It is an effect of minimal thermal expansion for radial and axis vibration at this point. The previous model can be linear, so it is clear that response is directly proportional to the amount of imbalance. It should be noted, that the real GTE response is unlikely to be linear over a wide range

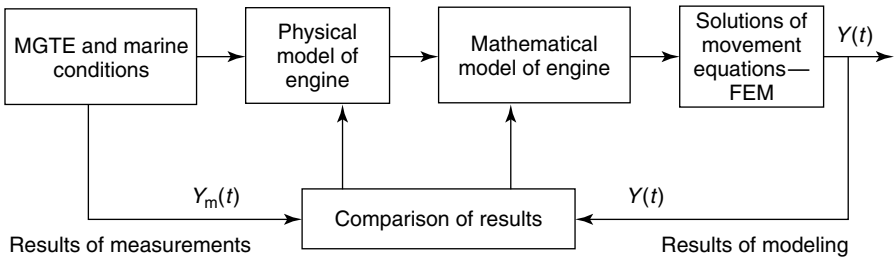


Figure 7. Scheme of diagnostics model GTE (FEM, finite element method).

(revolution of shaft). Furthermore, the result can be accepted only as a statistical approximation of the dynamic engine response.

The construction of a rotor has forces to consider. There is not only unbalancing or gas-flow forces but also vibration from cooperated machines like pumps, drills, enclosure modules, generators, and gearboxes as well. These forces form the focus during modeling and investigation of real systems. Losses of material also influence changes of moments of inertia of rotating parts. They cause the displacement of the main axis of inertia, which is not coincident to the axis of rotation. Finally, a main source of large vibration is low damping. Mathematical models are difficult because of assessments of damping and stiffness coefficients of struts and bearings. The shape of the axis deflection is defined as discrete sets:

- set of static deflections, $\mathbf{u}_s$;
- set of dynamic deflections, $\mathbf{u}_d$.

Both sets depend on the actual technical state of the rotor and the geometry, which are changed through cracks and wanes of engine parts.

$$\mathbf{u}(\omega t) = \mathbf{u}_s + \mathbf{u}_d(\omega t) \tag{1}$$

This equation is a discrete set of displacement value points of the axis of a rotor. Taking into account the damping and stiffness of a bearing's supports, it can be posited that they are functions of the temporary positions,

$$k_{ik} = f(u) \quad c_{ik} = f(u) \tag{2}$$

To simplify the problem, it can be assessed that for constant rotation these values are also constant. Using the FE method, the model presents a three-dimensional discrete model. Rotors of GTE, because of circular symmetry, have been described by one-dimensional, two hatches balk—rod symmetry FE, which have six degrees of freedom. All parts have geometrical and material characteristics. Movement parameters of a discrete model have been found by solving the following equation:

$$\mathbf{Ku} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{M}\ddot{\mathbf{u}} = \mathbf{F}(t) \tag{3}$$

where $\mathbf{K}$ is the matrix of the structure's stiffness, $\mathbf{C}$ is the matrix of the structure's damping, $\mathbf{M}$ is the matrix of the structure's inertia, $\mathbf{F}$ is the vector of forces, and $\mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}}$ is the displacement and its derivatives (velocity and acceleration).

The issue can be solved as a linear problem, but in the GTE's rotor, one has to allow for changes of stiffness and damping, which are functions of the movement parameters. In this case, equation (3) can be expressed as

$$\mathbf{K}(\mathbf{u}, \dot{\mathbf{u}})\mathbf{u} + \mathbf{C}(\mathbf{u}, \dot{\mathbf{u}})\dot{\mathbf{u}} + \mathbf{M}\ddot{\mathbf{u}} = \mathbf{F}(t) \tag{4}$$

The main purpose is to find sensitive vibration symptoms representing residual imbalance of the rotors and forces from misalignment of shafts. FEA (finite element analysis) is used successfully for a wide range of problems and it may also be used for the modeling and analysis of a rotor system (Figures 8 and 9).

Currently, monitoring systems commonly use FE models and rotordynamics in conjunction with vibration analysis for detection and identification of unbalancing and misalignment of shafts. A linear model obeys the basic principle of linear superposition. Applied to a structure, this means that displacement resulting from a combination of structural loads is the sum of the displacements due to each individual load making up the combination.

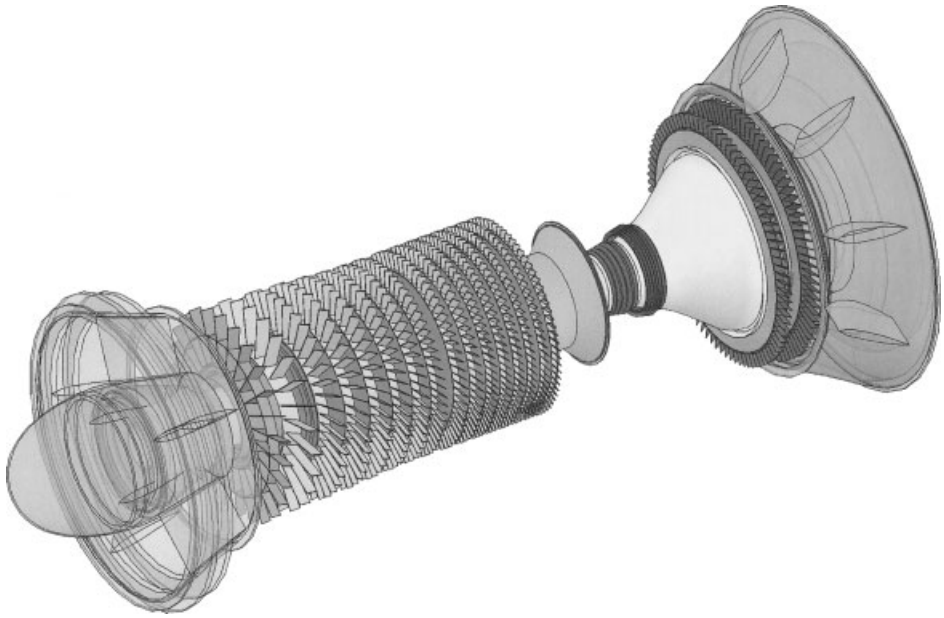


Figure 8. 3D model of a rotor system GTE.

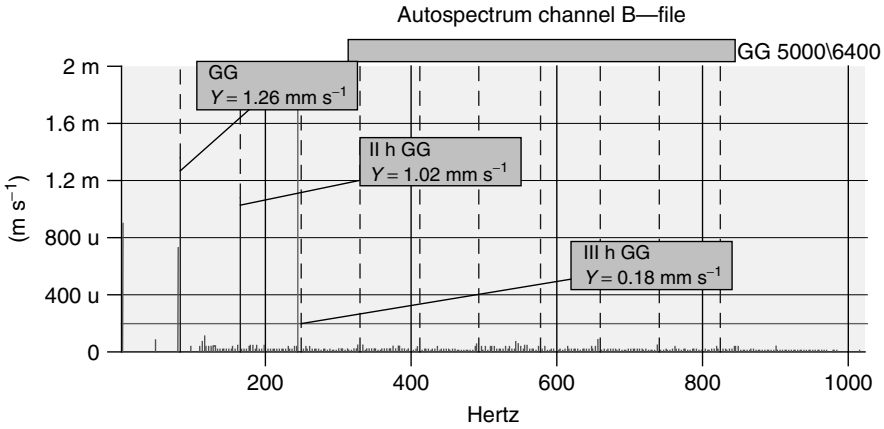


Figure 9. Example of a GTE spectrum.

Measuring transducers (generally accelerometers) are fixed to steel cantilevers located on the flange of the compressors and power turbines. The fixing accelerometers' cantilevers are characterized by a vibration natural resonance frequency value, sufficiently different from harmonic frequencies due to the rotation speed of the rotors and their harmonics. The measurements are taken perpendicular to the rotation axes of the rotors over the main bearings. Such a choice is made on the basis of theoretical consideration of excitations due to unbalanced shaft rotation [17]. As signals, usable for the "defect-symptom" relation, the following magnitudes are usually selected:

- the first harmonic rms value of vibration velocity (or displacement) amplitude related to the rotor of the compressor and the power turbine;

- the rms value of vibration velocity amplitude within the individual range of the bandwidth filter;
- the dimensionless parameters characterizing imbalance assessment of the turbine rotors:
  - S1—the ratio of the mean vibration velocity amplitude of a given rotor (first harmonic) and the velocity component relevant to the second harmonic excitation frequency of the rotor in question;
  - S2—the ratio of the mean vibration velocity amplitude of a given rotor (first harmonic) and the velocity component relevant to the third harmonic excitation frequency of the rotor in question.

Analyzing steady states of the compressor needs FFT analysis with tracking. Another problem is resonances, which require order tracking or autotracking analysis. Sometimes, changes in vibration symptoms are also subject to analysis of their trends as functions of the service time of the engines.

7 DIAGNOSING POWER PLANTS ON OFFSHORE TUGS AND TANKERS–CALL OF RELIABILITY

The propulsion systems of tugs and tankers are subjected to specific sea loads due to waving and dynamical impacts associated with the mission of a given vessel. Sea waving can be sufficiently exactly modeled by means of statistical methods. Many more problems arise from modeling impacts due to contact with floating objects, the hull of another vessel, or a rig's legs. In the operation of contemporary technical objects, including offshore vessels, greater and greater attention is paid to notions such as time of serviceability, repair time, maintenance, and diagnosing costs. Diagnosing process has now become a standard procedure performed during technical maintenance. Out of those mentioned above, the notions of time of serviceability and maintenance costs seem to be crucial for the diagnosing process of a vessel's power plant. Knowledge of a character of loading, which affects ship shaft line, engines, gear boxes, or propellers, can make it possible to identify potential failures by means of on-line vibration measuring systems [18]. This way, elimination of costly and time-consuming overhauls on dock leads to lowering operational costs and increasing vessel reliability.

Usual measurement methods of misalignment parameters of the propulsion system, like the laser method, require disassembling protection covers of shafting between engines and reduction gears. Measurement conditions make it necessary to suspend operation of the propulsion system for a few days and this is, of course, an intrusive method. A vibration method allows assessing permissible values of the alignment parameters without stopping operational use of the vessel.

Appropriate assembly of the main engines and the other torque transmission elements, inclusive of propellers, is practically determined by a set of tolerated dimension and geometrical location requirements, called the *geometrical dimension assembling chain*. Power plants are prone to coaxiality deviation from permissible values and, in consequence, to possible failure of one or more elements of the propulsion system. The excessive deviation can lead to the loads on bearings and gear teeth much higher than calculated and result in their premature failure [19]. The usual control methods of the coaxiality deviations do not always fulfill the user's expectations. Difficult access to flange connections, long control time, organizational difficulties, and lack of qualified personnel create hazards of taking measurements with errors exceeding allowable values. The vibration method, instead of the usual coaxiality control methods, is preferred on the basis of the result of an analysis of the earlier mentioned operational hazards.

The energy emitted as a result of a change of technical state of the flange connection is reflected in the recorded vibration signal—the second harmonic of the velocity. Archive charts of the investigation results of a typical offshore tug are exemplified in Figure 10. Application of the order tracking analysis in the power transmission system of offshore vessels for identification of resonances is an efficient method (Figure 11). Especially, in the case of an enclosed power pack, the autotracking method should be adapted as a preferable method.

In static calculation procedures of shaft lines, no analysis of dynamic excitations, except torsional vibrations, is taken into consideration. In certain circumstances, the adoption of static load criterion

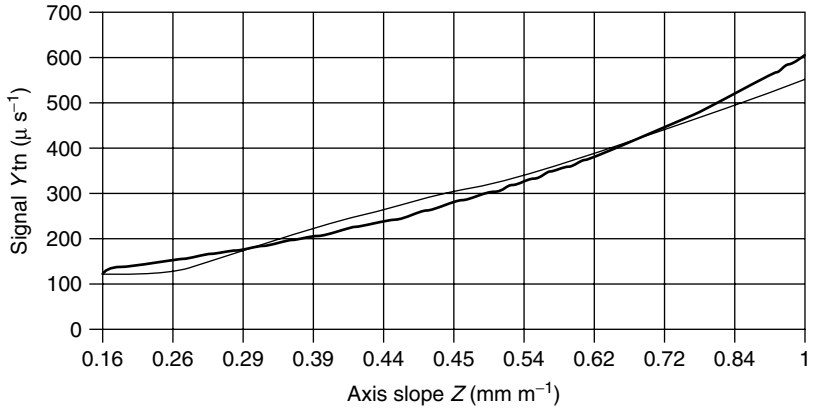


Figure 10. Regression functions of the change in vibration symptoms at different slope values.

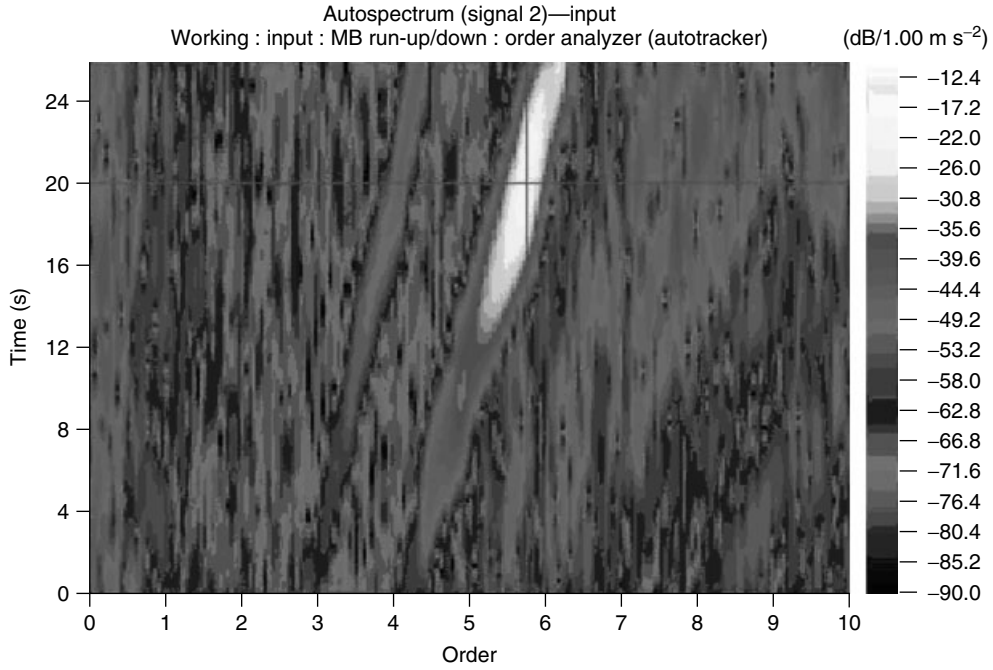


Figure 11. An example of identification of the resonance during the acceleration of shaft lines.

may be disastrous, especially in the case of resonance between natural vibration frequencies and those of external forces due to dynamic impacts. Theoretical analysis indicates that shaft bending deformation continuously accumulates as part of shaft torque. However, the quantity of torque nonuniformity is rather low since shaft-line eccentricity is low. It results from manufacturing tolerance, nonhomogeneity of material, propeller weight, and the permissible assembly clearances of bearing foundations. This condition makes it possible to predict that run-out of propeller shaft may also happen at rotational speeds other than the critical first kind calculated during the design process. For vessel propulsion systems, the torque pulsation expressed by means of the Fourier series is much more

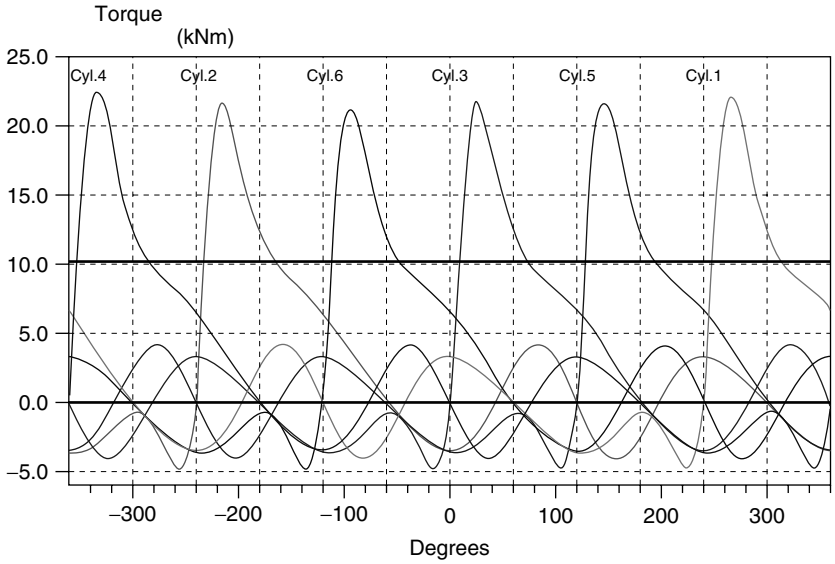


Figure 12. Time waveform of torque simulations for a six-cylinder four-stroke marine engine as a function of crankshaft rotation.

complex. It also contains components resulting from a number of propeller blades, kinematical features of reduction gear, as well as disturbances from the main engine and neighboring devices. In general, occurrence of only one harmonic case does not change reasoning logic. For a better understanding of the theoretical background, the virtual model of a propulsion system's excitation are presented in Figure 12.

Theoretical analysis of operational conditions of intermediate and propeller shafts indicates that static and dynamic loads appear. In a more detailed analysis of dynamic excitations of all kinds, the following factors should also be taken into consideration:

- disturbances coming from the ship's propellers (torsional, bending, and compressive stresses);
- disturbances from the propulsion engine (torsional and compressive stresses);
- disturbances from the reduction gear (torsional stresses);
- disturbances from other sources like a hull contact.

The wide range of stochastic dynamic loads acting on vessels during their lifetime mean that in the nearest future, the application of *on-line* diagnostic techniques to all types of offshore propulsion systems based on analyzing vibration signals will constitute an obvious economic and technical necessity.

8 CONCLUSIONS

Application of an *on-line* monitoring of vibration parameters of offshore machines and propulsion systems makes possible technical diagnostic tests of torque transmission systems, fluid and gas transportation, and identification of possible defects. In practice, vibration analyses are accomplished by two different procedures:

- *on-line* — in real time;
- *off-line* — periodic or single measurements.

Both procedures have their advantages and disadvantages. *On-line* systems give permanent control of vibration parameters in real time. They allow for monitoring vibration parameters, holding memory, and shutting down the machine in a critical state. The data preview of memory can activate the trend functions and it shows changes of frequency or time parameters of vibration as a function of operational time. The disadvantages of this system are linked with the cost of software and hardware, which are most often individually prepared.

The monitoring systems consist of few vibration and thermodynamical parameters. The main requirements are as follows:

- should be an autonomous system of modular structure, equipped with monitors for continuous measurements (*on-line* mode);
- should provide buffered signal outputs;
- should automatically store, archive, and analyze each monitored variable;
- should automatically perform basic vibration analysis like synchronous time–wave forms of vibration signals, spectrums, cepstrums, signal envelopes, orbits, shaft center line, waterfall spectrums, cascade spectrums, bode plots, polar plots, indicator plots (for reciprocating compressors), etc.;
- should register and archive measuring events and system events;
- should be equipped with an automatic self-control and self-diagnostic system for self-diagnosis of each of its own measuring loops;
- the cables should be specified by producers of machine monitoring systems;
- the cables should have minimum connections (without any additional junction boxes other than that on the skid);
- the cables should be situated away from any other power and supply cables;
- all equipment should be correctly grounded.

To summarize the difference between *on-line* and *off-line* systems, it should be noted that traditional *on-line* monitoring is either a simple safety switch system with no analysis or database functionality or a huge and multichannel stationary installation, featuring full capability described above. Such solutions can also be used for diagnostics; however, for less expensive machines it never brings an economic rationale that can be implemented. Also, monitoring systems provide a wide range of basic diagnostic tools, but are never flexible enough with regard to in-depth investigations of difficult and unique problems.

On the other hand, the typical *off-line* instrumentation is robust, handy, and portable, also featuring quite an acceptable cost of implementation. The instruments are compact analyzers and data loggers with usually one vibration channel (rarely, two) and some extra inputs for process parameters (like temperatures) and tachometers. They quite satisfactorily cover the periodic check procedures for less critical machines and can supplement big monitoring systems.

This classification naturally appears with some gray areas, which can be discovered by asking the question—then what about a multichannel, portable *off-line* and advanced analysis capability? Fortunately, recent developments bring such instruments, with multichannel capability, still portable and battery operated, featuring the recording functionality and performing standard, as well as very advanced analysis, both on site and as postprocessing. When combined with a 48-bit wide-input dynamic range, these become suitable for less-skilled operators.

RELATED ARTICLES

Civil Infrastructure Load Models for Structural Health Monitoring

Modal–Vibration-based Damage Identification

Data Preprocessing for Damage Detection

Statistical Time Series Methods for SHM

Time–frequency Analysis

Monitoring Marine Structures

Ship and Offshore Structures

Gas Turbine Engines

REFERENCES

[1] Tonolli A, Borovik V, Volpini M, Voronov A. Offshore part of Russia—Turkey pipeline system Blue Stream: operation experience. *2nd International Conference Gas Transportation Systems: Present and Future (GTS-2007)*. Moscow, 2007.

[2] Steihausen J. Dynamic design of the foundation of reciprocating machines for offshore-installations—case study. *5th European Forum for Reciprocating Compressors*. Prague, 2007.

[3] Pedersen TF, Gade S, Harlufsen H, Konstantin-Hansen H. *Order tracking in Vibro-acoustic Measurements: A Novel Approach Eliminating the Tacho Probe*. Technical Review No. 1. Brüel & Kjær, 2006, pp. 15–28.

[4] Drewes E. Condition monitoring for piston compressors—state of art. *2nd European Forum for Reciprocating Compressors*. The Hague, 2001.

[5] Lenz J. Diagnostics methods for condition monitoring of reciprocating compressors. *1st European Forum for Reciprocating Compressors*. Dresden, 1999.

[6] Eijk A. Large reciprocating compressors on the glas dowr FPSO: concept selection and design considerations. *4th European Forum for Reciprocating Compressors*. Antwerp, 2005.

[7] Koop LGM. Performance monitoring on reciprocating compressors, a rational extend on condition monitoring. *2nd European Forum for Reciprocating Compressors*. The Hague, 2001.

[8] Cyklis P. CFD simulation of the two phase pulsating flow in the reciprocating compressor installation. *4th European Forum for Reciprocating Compressors*. Antwerp, 2005.

[9] Harper Ch. Dynamic analysis of reciprocating compressors on fpso topside modules. *5th European Forum for Reciprocating Compressors*. Prague, 2007.

[10] Lenz J, Brümmer A. Investigation of vibration problems at 7 MW recip. *2nd European Forum for Reciprocating Compressors*. The Hague, 2001.

[11] Peters MCAM. Evaluation of low frequency pulsation damping devices. *2nd European Forum for Reciprocating Compressors*. The Hague, 2001.

[12] Boiko A, Smreka B, Titov A. Reciprocating compressors with gas turbine drivers in gas industry of the Russian federation. *2nd European Forum for Reciprocating Compressors*. The Hague, 2001.

[13] Osipov MI, Tumashev RZ, Molyakov VD, Ivanov VL. High-performance gas turbine units for offline production of heat and electric energy. *2nd International Conference Gas Transportation Systems: Present and Future (GTS-2007)*. Moscow, 2007.

[14] Bachschmid N, Pennacchi P, Vania A. Experimental results in simultaneous identification of multiple faults in rotor system. *14th International Congress COMADEM*, Huston, 2000.

[15] Maxwell JH, Rosario DA. Using modeling to predict vibration from the shaft crack. *14th International Congress COMADEM*, Manchester, 2001; pp. 243–250.

[16] Strackeljan J, Behr D. Vibration monitoring of non-stationary rotor systems. *Conference ISROMAC-7*, Honolulu, 1998; pp. 135–144.

[17] Grzadziela A. Vibration analysis of unbalancing of marine gas turbines rotors. *XIII International Conference Noise Control '04*. Gdynia, Poland, 2004.

[18] Bruski S, Korczewski Z. Spectrum analysis methods of shafting torsional vibration for the injection fuel valves failures identification of marine diesel engines. *Explo—Diesel & Gas Turbine '05*, Copenhagen, 2005; pp. 43–48.

[19] Charchalis A, Grzadziela A. Diagnosing the shafting alignment by means of vibration measurement. *7 International Congress on Sound and Vibration, 4.7.07*. Garmisch-Partenkirchen, 2000.

FURTHER READING

Roemer MJ. Health monitoring of gas turbine engines. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, 2008.

Chapter 147
Wind Turbines

Goutham R. Kirikera[1], Mannur Sundaresan[2], Francis Nkrumah[2], Gangadhararao Grandhi[2], Bashir Ali[2], Sai L. Mullapudi[3], Vesselin Shanov[4] and Mark Schulz[3]

[1] *Center for Quality Engineering and Failure Prevention, Northwestern University, Evanston, IL, USA*
[2] *Department of Mechanical and Chemical Engineering, North Carolina A&T State University, Greensboro, NC, USA*
[3] *Department of Mechanical Engineering, University of Cincinnati, Cincinnati, OH, USA*
[4] *Department of Chemical and Materials Engineering, University of Cincinnati, Cincinnati, OH, USA*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

With increasing costs associated with fossil-based energy, there is large interest in deploying small and large wind turbines (WTs) for generating distributed power for residential and rural areas [1]. It is estimated that a small but significant part of the energy needs of the United States could be obtained through large-scale deployment of WTs [2]. WTs are deployed in rural areas where access to these structures is sometimes not feasible because of the elevated positions. Damage to the WT could cause large financial losses including loss of energy tapped during this outage; see Migliore *et al.* [1]. Structural health monitoring (SHM) of WTs is needed to ensure safety and to avoid overdesign of components. Overdesign will not allow maximum power from the wind to be captured. This article is a review of the current literature describing techniques for health monitoring of WTs. The main focus of the article is on monitoring the turbine blades, which may be the most critical component of

the turbine. Continuous health monitoring of the drive train (gearbox and bearings) and the wind turbine blade (WTB) will ensure proper performance of the WT. Bently Nevada Inc. has developed a commercially available module for continuous monitoring of the drive train [3] of WTs. Monitoring bearings and other components may be done using accelerometers. Since the frequencies of vibration and operation are low, the data-acquisition problem is simplified as a consequence of which allows a larger number of accelerometers or other sensors to be used. Similarly, continuous health monitoring of the WTB is needed. The design of the WTB is a critical factor in the performance (power production) and reliability of WT systems. The trend in blade design is toward improving the aerodynamic efficiency of the blade thus necessitating higher strength-to-weight-ratio materials such as carbon composites owing to their thinner airfoil profiles. A low-cost continuous health monitoring system could provide critical information about the location and propagation of damage in WTBs, and provide predictive maintenance information prior to a blade becoming unsafe. If a blade fails, the rotor can become unbalanced and cause the blade to impact the turbine tower, which can severely damage the turbine drive train. It is common for blades made of composites to have sudden audible acoustic emissions (AEs) during damage growth. AE is defined as the class of phenomena whereby transient elastic waves are generated by a rapid release of energy from a localized source or sources of damage; see Wells *et al.* [4]. Acoustic emission techniques (AETs) are often used to locate damage and to detect the growth of cracks during qualification testing of WTBs. Use of AE and other techniques for SHM of blades is discussed in this article.

Several researchers have attempted in the past to monitor WTBs for damage in a laboratory setting using various techniques. The simplest technique is to use strain gauges to detect high strains that could indicate damage. In general, too many strain gauges would be needed to have a high probability to detect small damage before failure could occur. Another approach is to use AE sensors. These are typically heavy barrel-type sensors that would be difficult to use in practice owing to the size and requirement that a large number of sensors and channels of high-rate analog-to-digital (A/D) data conversion would be needed to monitor the blade. Typically triangulation is used to locate damage similar to how the epicenter of an earthquake is located. Strain gauges and AE methods are passive methods wherein no artificial excitation of the structure is used, and the structure must be operating for damage to be detected. Active methods are another approach for damage detection. In active methods, a diagnostic waveform or some other form of artificial excitation is used to probe the structure for damage. Sundaresan *et al.* [5] used an actuator to pulse the WTB intermittently during quasi-static loading and store the received waveforms. The current data was compared with baseline data to identify damage. Migliore *et al.* [1], Dutton *et al.* [6], Joosse *et al.* [7], and other researchers performed testing on WTBs using AETs. AE "threshold-based" arrival systems can be used to approximately locate damage. An advantage of AE methods is that knowledge of the wave speed is not required to detect damage. Wave speed is a function of material properties and the geometry of the structure. WTBs consist of multiple anisotropic materials and have complex geometric features. An advantage of AE methods is that they can detect damage on complex anisotropic structures like the WTB shown in Figure 1. On the other hand, the following disadvantages are inherent in existing conventional AE-based systems: (i) each sensor requires an electrical circuit containing a preamplifier that increases the overall mass of the sensor; (ii) each sensor's output must be converted into a digital format using a high sampling rate A/D converter; (iii) a large amount of data is obtained, which requires rapid real-time data collection, storage, and processing; and (iv) the cost of the overall health monitoring system increases because of the above factors.

With the improvement in the field of microelectromechanical systems (MEMSs), the mass of the sensor and associated electronics can be mitigated. Recently, a structural neural system (SNS) has been developed to overcome some of the limitations of current AETs. The SNS is a signal-processing system that emulates how the human body processes signals from large numbers of highly distributed neurons and receptors (sensors). The SNS uses piezoelectric sensors bonded onto the blade. The surface-mounted sensors are easy to install, and can be repaired, retrofitted, or updated, and the sensor cannot degrade the integrity of the blade. These sensors are highly sensitive and can detect AE stress waves caused by

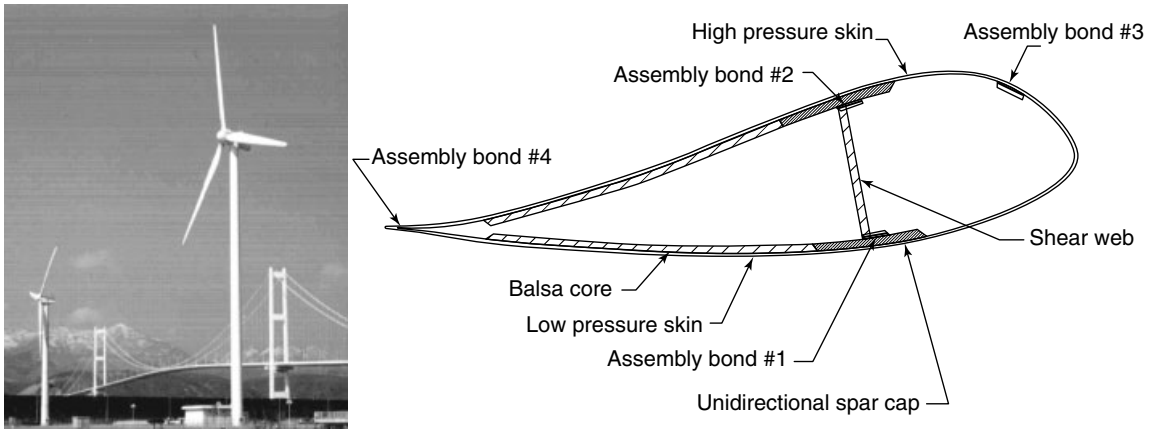


Figure 1. Wind turbines in action and cross section of a typical blade. [Reproduced from Ref. 8. © Derek Berry, 2004.].

small damage propagating in the blade, as shown by Joosse *et al.* [7]. The SNS is based on receiving AE signals and is a passive health monitoring method since no artificial means of exciting the structure is used. The frequency of the AE signal that is sensed is on the order of hundreds of kilohertz depending on the blade material and geometry. The SNS architecture was installed on a 9-m-long WTB that was tested to failure during quasi-static loading. Results of detecting and locating growing damage are summarized in this article. Information on testing of blades using conventional AE methods can be found at the US National Renewable Energy Laboratory (NREL) web site [2]. Information on testing different sensor types on WTBs for health monitoring and control is at present available through Sandia National Labs [8–11]. Nondestructive evaluation (NDE) and SHM of WTs is discussed in [9–36]. WT failures that have occurred over the past 30 years and safety concerns are briefly discussed next to help designers decide how practical SHM techniques should be developed.

1.1 Wind turbine failures and safety

To make WTs widely accepted, the public must be assured that WTs are reliable, safe, and benign [37–40]. This section provides an overview of WT failures to give designers an idea of how SHM systems might improve reliability and safety of WTs. WT failures are not well documented and data is not available for new designs. An unofficial and not comprehensive summary of WT accidents from 1975 to November 30, 2007 is given in [39]. The total number of accidents was 403. These resulted in 49 fatalities, of which about 35 involved wind industry workers. Fourteen accidents were public fatalities. The most common cause of fatalities was falls from turbines. Human injury occurred in a further 18 accidents. Blade failure is the most common failure mode on turbines. A total of 118 separate incidences of blade failure were reported in the study. Blade failure may result in whole blades or pieces of blade being thrown from the turbine. Fire is the second most common accident cause. Fire can arise from a number of sources including electrical malfunction and lightning strikes. Turbine fires cannot usually be stopped because of the turbine height, unless there is an onboard fire suppression system. Burning debris from a turbine may be scattered causing a wider-area fire risk. Structural failure is the third most common accident cause as per the data in [12]. Structural failure is assumed to be a major component failure usually owing to high wind gust exposure. Ice falling or throwing is another mode of danger. Transport accidents involve turbine sections falling from transporters including at sea. Driver distraction by turbines, thrown ice, and blade pieces landing on the road have also caused accidents. Some cases of environmental damage including bird and bat deaths have been reported. Other types of damage and accidents are owing to component malfunction, hail, and lightning strikes. Some of the possible indirect problems caused by WTs include interference with TV or

microwave reception, depreciating property values, WT noise, increased traffic, road damage, rotating shadows from the blades, aesthetics, concerns about electrical danger, and increased lightening strikes.

Failures of WT components [36] on a percent basis are electrical control 13%; gearbox 12%; yaw system 8%; entire turbine 7%; generator 5%; hydraulic 5%; grid 5%; blades 5%; brakes 3%; entire nacelle 1%; mechanical control 2%; air brake 2%; axle/bearing 1%; other components 30%; and the tower, foundation, hub and coupling <1%. Large WTs are equipped with a number of safety devices to ensure safe operation during their lifetime. Existing safety devices/approaches include vibration analysis, oil analysis, component temperature measurement, thermographics, shaft alignment, strain measurement, acoustic analysis, photo/thermo elastic analysis, electrical effects, overspeed protection, aerodynamic braking systems including tip brakes and mechanical braking systems, and visual and aural inspection. Difficulties that are to be overcome in order to install detailed SHM systems on WTs include the following: the requirement that they should last 30 years without causing false positives for failure, the cost to monitor data from the turbine, and the difficulty in servicing the SHM system. Research in SHM of WTs is being conducted in national labs [2, 11, 17, 33, 35–38], industry [3, 8, 36, 39], and universities [4, 7, 13, 21–28, 31, 34]. In particular, the Danish WT industry is a leader in commercialization. Twenty per cent of Danish domestic electricity production comes from wind. Based on the above data, SHM systems could be applied to the major areas where failures occur, such as blades and system components, to increase the safety and reliability of turbines. The SHM system would identify degraded parts in time for replacement to prevent failure, and this could prevent injuries that occur owing to failure of components or repair of failed turbines. Approaches for SHM of WTBs are discussed next.

2 METHODS FOR DAMAGE ASSESSMENT AND PROGNOSIS ON WIND TURBINES

This section discusses possible new methods of SHM of WTs [12, 13, 40]. Several different methods are needed to monitor the entire turbine because of the variety of mechanical components that are present in the turbine.

2.1 General methods for SHM of wind turbines

There are many methods of SHM being developed for a large number of applications. General characteristics of these methods are presented in order to help developers consider candidate techniques for specific WT applications. Monitoring the blades in a rotating system presents the problem of data transfer from the rotating frame to the fixed frame for all the methods discussed. A summary of the main methods is given in Table 1. It is anticipated that a medley of sensors and methods will be the best approach to monitor WTs. A fully integrated SHM system that can use different damage detection techniques and types of sensors would be the ultimate goal. Recent papers in the area of SHM (2002-up to the present) are published in *Structural Health Monitoring: An International Journal* [14]. This journal has a fairly comprehensive list of techniques for SHM. Other references on SHM are given in [9–11, 15–35] and can be found in [41–54].

Most of the SHM techniques and sensor types listed in Table 1 are discussed in other articles of this encyclopedia. However, there is not much information on application of these techniques to SHM of WTs. Several methods that have been applied to WT SHM include a scanning laser vibrometer [32], the fiber-optic method [33], vibration and other methods [34], and the impedance method [35]. The SNS, in particular, was developed for SHM of WTs and is not discussed elsewhere in the encyclopedia. Recent results are also available using the SNS to monitor damage on a WTB in a proof test. Thus, an overview of the SNS method and its testing are given in this article.

2.2 Introduction to the acoustic emission technique as an SHM tool

The AET has been used as a nondestructive testing (NDT) technique for several decades and is fairly well developed with commercially available instrumentation and standardized testing procedures [15].

Table 1. General methods for SHM

Method	Sensor/actuator type	Description of the method
Vibration	Accelerometer, piezo, or MEMS	The natural frequencies of the blade can be monitored for changes indicating damage. Small damage is difficult to detect
Strain	Foil strain gauge or fiber-optic cable	Strain can be monitored at critical points in the blade and other components. It is difficult to measure strain inside the composite and to have enough gauges to detect small damage. Fiber-optic Bragg gratings can provide a large number of low-bandwidth strain measurements. Passive method [33]
Ultrasonic wave propagation	Piezoelectric wafer	Good for monitoring uniform sections or hot spots. Need predamage reference data that may vary owing to environmental changes or sensor aging. Can detect damage when the blade is operating or not operating. Commercial systems are well along in development for several applications. Active method [12, 34]
Smart paint	Piezoelectric or fluorescent particles	Paint changes color when damaged. Visual technique that is low cost but not automated for remote applications. Passive method
Acoustic emission conventional	AE wideband barrel sensor	Can detect damage in complex structures. Fretting can cause false indications of damage and the turbine must be operating. Sensors are large and many are required. Fast multiplexing is simplifying the hardware requirements. MEMS AE sensors are reducing size and weight of the system. Passive method
Structural neural system (SNS)	Piezoelectric wafers or other sensor types	Overcomes problems of conventional AE by using piezoelectric wafer sensors and biomimetic highly distributed massively parallel signal processing. Multistate sensors (nanotube thread, pressure, temperature, etc.) can also be used. Passive method. Active SNS has also been tested and uses a simple method of neuron firing to detect damage in the passive and active systems [13]
Impedance	Piezoelectric wafer	High-frequency method that can detect local damage. Active method that uses an impedance analyzer and diagnostic signal; Inman [34]
Laser vibrometry	Scanning laser Doppler vibrometer	Difficult to measure the rotating blade which is also changing orientation, and the cost is high. Good for characterizing the blade in the lab. Active or passive
Impedance tomography	Carbon nanotube or other conductive filler	This method uses carbon nanotubes or other conductive particles to make the blade material electrically conductive. The impedance between arrays of electrodes is monitored to detect damage. The size of damage the method can detect depends on the size of the electrode patterns, and the method is passive and simple and does not require the structure to be operating.
Thermography	Infrared camera	This method uses an infrared camera to map the temperature of the structure. Damage can produce a local temperature rise owing to crack and delamination breathing. The method is expensive and difficult to use on a rotating system and it is difficult to detect interior damage
Laser ultrasound	Laser	A laser beam pulse excites the structure and also measures the response of the structure. Like the NDE ultrasound technique but noncontact. Difficult and expensive to use in the field and minor damage to the surface of a composite may occur by the laser excitation. Good for automated NDE of metals.

(*continued overleaf*)

Table 1. (*continued*)

Method	Sensor/actuator type	Description of the method
Nanosensors	Electronic particle	This future approach uses small particles embedded in the blade during the fabrication process. The sensors are transceivers that reflect an RF signal which changes owing to very local damage [13]. *See* **Nanoengineering of Sensory Materials** for further information
Nonlinear dyn., condition monitoring, etc.	Accelerometers, strain gauges, others	A wide variety of specialized methods for SHM are described in [14] by Todd, Adams, Zimmerman, Chang, Farrar, Pai, Peairs and Inman, and many others
Buckling health monitoring (BHM)	PZT patches	BHM has received little attention in the literature but is important for wind turbine blades and other structures such as radomes, civil infrastructure, and tower structures. Some work has been done by Frank Pai using the nonlinear finite element code Geometrically Exact Structural Analysis (GESA) and by Sundaresan using wave propagation

The AET relies on the dynamic release of elastic strain energy as damage grows within materials under stress. The released elastic energy propagates through the structure in the form of guided waves. AE sensors suitably located on the structure can detect these signals, as shown in Figure 2. The requirement of loading the structure and the need for the growth of damage for the AET to evaluate the structure separates this technique from other NDE or SHM techniques, which frequently do not need the damage to be propagating under load. This very same requirement also provides the AET the potential to directly quantify the damage severity possibly in terms of crack growth rate or the remaining fatigue life, and makes it suitable for health monitoring of structures in the field. However, one of the major shortcomings of the traditional AET is its susceptibility to false positives. False positives can be triggered by a large number of extraneous signals including mechanically induced noise signals such as friction of mating surfaces (fretting) and hydraulic noise, as well as radio frequency electrical noise. With the availability of new types of sensors including MEMS devices, miniaturized electronics, advanced signal-processing techniques, and pattern recognition, it is likely that false positives will be reduced significantly.

Health monitoring of WTBs is currently practiced only during laboratory tests and certification tests

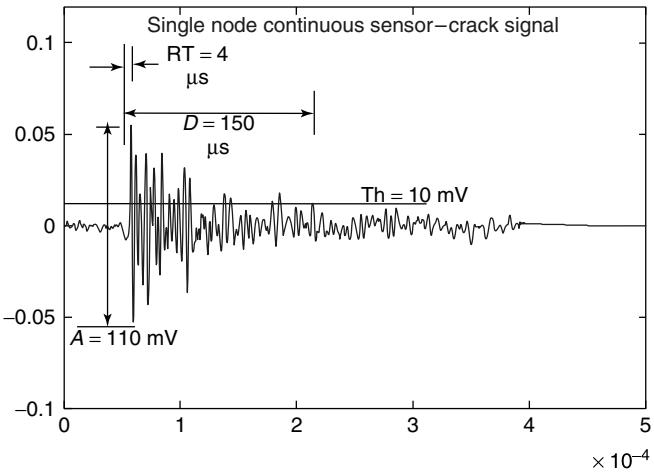


Figure 2. Conventional acoustic emission parameters extracted from a waveform.

[7, 16–18]. Such tests have been monitored using multichannel AE monitoring systems during repeated static loading to meet the certification requirements as well as during fatigue loading. Weak blades have been successfully identified by the AET during repeated ramp loading to the normal operational load of the blade, or a slightly higher level, followed by constant load hold periods and subsequent unloading. Composite materials that make up the wind turbine blades are known to be profuse sources of AE events and the difficulty of evaluating the blade lies in separating AE signals that are related to the critical damage events from the noncritical damage events. During these blade tests, it is assumed that good quality blades would emit AE signals during the load ramp-up and remain quiet during the hold period. When AE signals continue during the hold period, it is assumed to be a sign of active damage growth when none is expected and hence is an indication of a weak blade.

In addition to the AET, other techniques that have been considered for the SHM of WTBs include vibration monitoring using accelerometers and fiber-optic sensors [18]. Since the critical damage sizes in WTBs are relatively large and may introduce measurable changes in the frequency and mode shapes, accelerometers are considered a viable option. Microbend-type fiber-optic sensors are suggested for monitoring bond lines in WTBs. Remote monitoring of rotor blades in large offshore WTBs [18, 33] is considered to be both economical and technically feasible, and it is estimated that the cost of the additional expense related to the SHM instrumentation will be recouped within 3–8 years of the operation of the turbine. The large WTs (with capacities approaching 5 MW for a single turbine) that are being designed now are among the largest composite structures ever built. The initial investment and the reliability requirement of these blades are sufficiently high to justify the development and incorporation of SHM capability in these blades.

The transition of the techniques that have been used to monitor structural tests in the laboratory to actual SHM on operating turbines will require considerable effort. Embeddable sensors, instrumentation, and information transfer are needed. The design of the SHM system will have to consider the unique requirements of the current and next generation WTs including fabrication techniques (often on-site), material failure behavior [18, 34], the load spectrum, environmental conditions, and economics. Figure 3 shows a schematic of a WTB embedded with sensors and electronics that can potentially monitor AE signals, vibration characteristics, and blade deformation. Moreover, there is recent interest in using sensors on blades for control purposes. Thus, sensor systems on turbines may serve dual uses, for SHM and also for feedback control of the turbine.

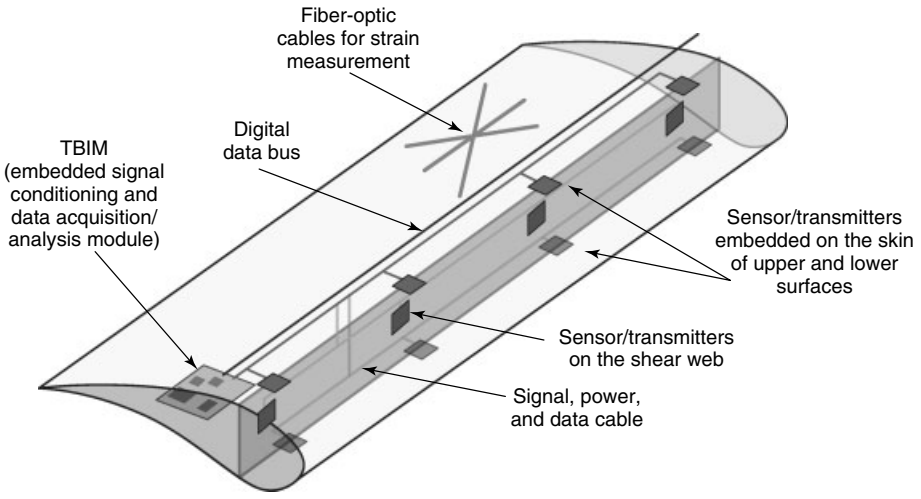


Figure 3. Schematic illustration of a blade with sensors for SHM under field conditions.

2.3 Damage assessment and prognosis using the AE method

Newer WTBs are made from several different materials and are tailored to harvest the maximum energy from the wind through bending-induced twist coupling and other means. Simultaneous achievement of good dynamic properties, light weight, and structural integrity is a challenge. Both gradual as well as sudden changes in the material properties are anticipated during the service life. Gradual reduction of properties could be owing to normal fatigue, ultraviolet (UV) radiation, hygrothermal effects, erosion, etc., while the sudden changes could be owing to lightning strike or gust loads. In addition, blade failure could result from local buckling, which is one of the most common modes of failure. An SHM technique should be able to recognize and address the interaction between these different causes of damage evolution. In addition, the sensors integrated into the blade could be utilized to assess the initial condition of the blade after the fabrication and establish the baseline responses useful for tracking the evolution of damage in the blade. A combination of sensors capable of addressing different aspects of damage evolution is likely to be successful. Embedded wireless resistance strain gauges and accelerometers will be inexpensive approaches to track the gradual degradation of elastic moduli and the blade's natural frequencies. Power harvesting from vibration of the blade is an approach to power the wireless sensors.

The AET offers a real-time SHM capability, which is particularly useful to identify sudden increases in the damage growth and hence is likely to be particularly useful in the prevention of catastrophic failure. AE-based SHM has the advantage of directly assessing the damage growth rate unlike other techniques that attempt to measure the damage size. Recently, bondable "continuous sensors" with individual nodes having wideband characteristics and sensitivities comparable to commercial resonant frequency AE sensors have been developed. This continuous sensor and the array sensor with multiple sensor nodes interconnected and with one signal output offer a means of simple and inexpensive monitoring of large WTBs with minimal hardware and signal processing.

The continuous sensors were shown to perform well in the fatigue life prediction of individual composite specimens and for their life extension. This approach of predicting the fatigue life of individual components has been tested on four different groups of specimens with different combinations of geometry, material properties, and loading conditions. Figure 4 shows a schematic representation of the service life of a structural component experiencing a spectrum loading, gradual aging, and encountering adverse events such as impact or lightning strike. While it is not possible to prevent damage to the blade owing to such unexpected events, it is possible to assess the combined effect of these factors and prevent further steep degradation in structural integrity by limiting the operational envelope of the turbine.

Figure 5 shows a portion of the life of the turbine simulated using laboratory specimens. A group of 30 woven fabric composite tensile specimens were subjected to varying levels of impact damage followed by simulated spectrum loading with load spikes. For these specimens, AE signals were collected only during the load increment. The specimens were subjected to fatigue load with peak stress

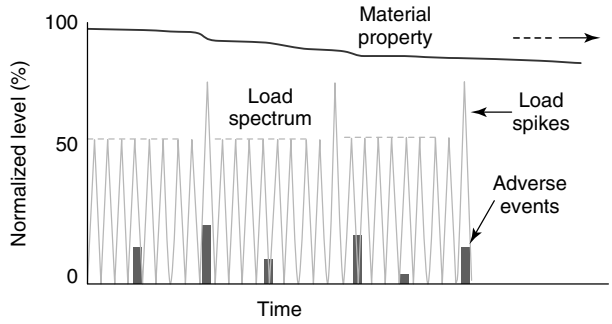


Figure 4. Schematic representation of load, adverse events, and aging of wind turbine blades.

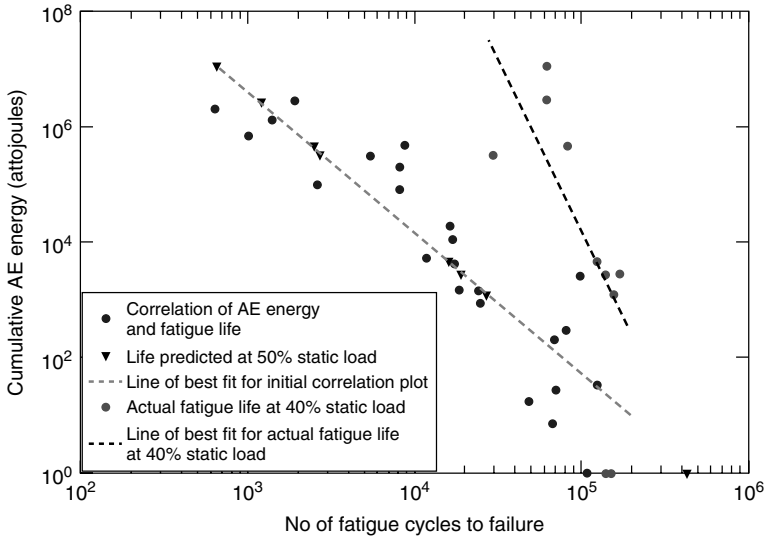


Figure 5. Correlation between AE prediction and actual fatigue life of woven fabric composite specimens subjected to impact load. The left dashed line represent the 50% static load case, the right dashed line represents the 40% static load case.

at 50% of the average static strength of these specimens. As indicated by the circles near the left dashed line in Figure 5, the cumulative AE energy during the incremental load correlated well with the subsequent fatigue durability of these specimens. A subset of this group was selected to verify if this prognosis could be used to identify and separate specimens that are bound to fail prematurely so that they could be used in less demanding services. The predicted lives of these specimens, at the load amplitude employed for the rest of the group, are shown as black triangles in Figure 5. To simulate a less demanding service life, the fatigue load was reduced from 50 to 40% of the average static strength. The distribution of fatigue lives at this reduced load amplitude is shown as circles near the right dashed line in Figure 5. The reasonable correlation between AE data and extended fatigue life for these specimens appears to support both the validity of AE-based prognosis and the opportunity for optimal utilization of available WTBs.

2.4 The structural neural system

Composite structures deteriorate gradually owing to operational effects and aging, and owing to unpredictable events such as impact and lightning strikes, and contamination from the environment. It is important to monitor the condition of large systems to prevent catastrophic failure owing to the combination of all the deterioration effects. Because multiple physical states exist in systems, the use of different types of sensors simultaneously may be a more accurate way to identify degradation, contamination, and damage before it reaches a critical size or level. From a practical standpoint, conventional methods of sensing and signal processing are often too expensive, heavy, and complex for comprehensive *in situ* monitoring of large systems, such as WTs, where degradation or chemical contamination (e.g., owing to rain erosion, acid rain, water ingestion, UV degradation, and volcanic ash in the air) can occur anywhere in the system.

Approaches for structural damage detection that use a large number of individually wired sensors, or storing large sets of predamage data, or using amplifiers to generate diagnostic waves, or performing complex signal processing may not be feasible for WTs. Future sensor architectures ideally would produce information based on ambient conditions, rather than using complex analytical models and predamage data to diagnose degradation and damage. To develop an efficient health monitoring system, we can gain inspiration from the human neural

system [19, 20]. The human body is composed of complex anisotropic heterogeneous materials. To monitor its health, millions of receptors and excitatory and inhibitory neurons are distributed throughout the body, which simultaneously process signals in an efficient hierarchical way. This section describes how the functional capability of the human neural system can be mimicked using electronic components. Continuous sensors attached to a signal processor form a neuron, which is analogous to the biological cell body or soma. The processor performs thresholding, firing, and inhibition to measure waves or strains that are caused by growing or breathing damage in a structure, or the change in impedance of a chemical sensor, or by measuring other physical states of the system. The receptor neurons can be designed to sense many types of physical response thus opening the door for many important applications on WTs and other structures.

2.4.1 Description of the SNS

The SNS is a biomimetic signal-processing system, which uses a multiple state continuous sensor network for health monitoring of large composite and metallic structures and other large systems. Passive sensing based on AE monitoring is used to detect damage such as fiber breaking and delamination in composites. Parallel signal processing inspired by the biological neural system is used to combine continuous sensors in a grid pattern into four channels of data acquisition to locate damage and capture the sensor signal. The SNS is a generic biomimetic signal-processing architecture designed to simplify health monitoring of large structures. In Figure 6(a), each vertical line indicates a column neuron and the horizontal line represents a row neuron. The neurons are not connected to each other. Figure 6(b) shows the general architecture of the SNS. The neurons are a continuous sensor connected to an analog processor at the end [21–28]. A continuous sensor also called a neuron is a series connection of many individual sensors with only one output signal per neuron. The output of the neurons is represented by V1 through V20 in Figure 6. V1–V10 are column neurons and V11–V20 are row neurons. Thus, a large sensor grid is formed wherein a reduction in the required number of channels of data acquisition is achieved by processing the outputs of only the neurons that produce anomalous signals, and keeping a track of which neurons are firing using the structural neural system analog processor (SNSAP). The output of the first 10 channels (column neurons V1–V10) is the input to the SNSAP 1, which reduces the required

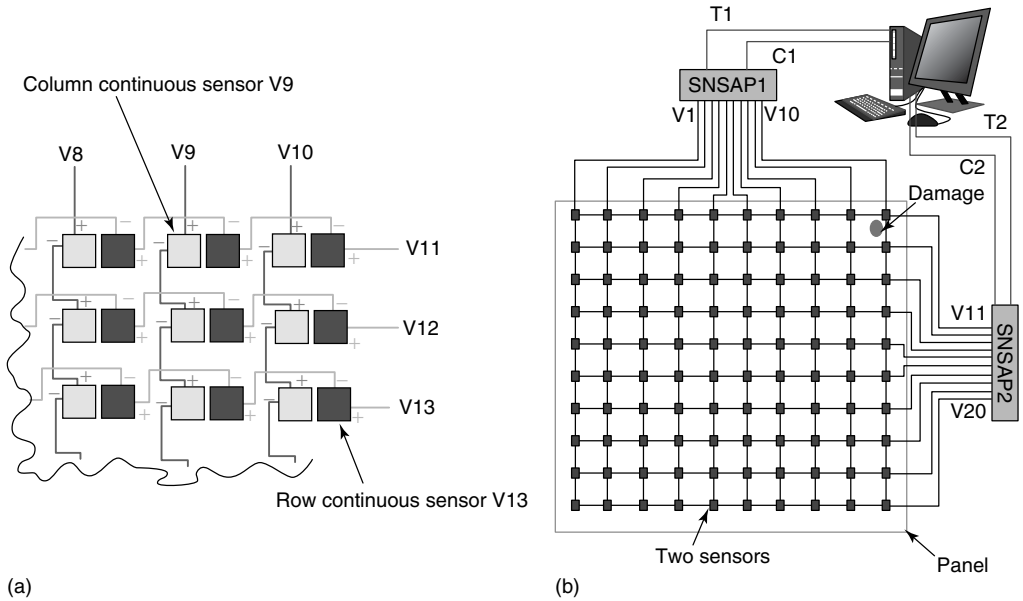


Figure 6. The SNS for WTBs: (a) detail of sensor connections and (b) sensor array.

number of data-acquisition channels to only two using firing of the neurons. One of the channels (C1) predicts the location of damage based on firing of the neurons. An algorithm is used to decode the firing signals and tell which neurons are firing. The other channel (T1) is the actual time response of the neuron and is used to qualitatively predict the severity of the damage. The same procedure follows for the row neurons (V11–V20) using the SNSAP 2 processor and the outputs C2 and T2. Thus, with an arbitrary number of sensor nodes, only four channels of data acquisition are required to predict the location and estimate the severity of the damage within a grid of sensors. Since each neuron can have 10 (or more) piezoelectric sensor nodes, and there are 20 neurons in this example, signals from 200 sensors are monitored using four channels of data acquisition.

The configuration of the electronic processor used to mimic the biological neural system is described in [21–26]. A brief explanation of how the neurons fire is as follows. When a neuron receives an AE signal, the signal is evaluated by the analog processor. If the signal meets set frequency and amplitude characteristics related to damage, the neuron "fires" a unique location signal and also passes the acoustic signal with those of any other neurons that are firing. A combined AE response analog signal and a second analog signal containing information on the locations of damages are sent to a central PC that performs A/D conversion and locates the damage and approximates the magnitude of the damage. Because of the high coverage of continuous sensors (neurons) on the structure, a neuron can be close to any damage site. This will allow the damage signal to be detected before it is attenuated and distorted and this makes signal processing and filtering easier when compared to using conventional AE sensor systems.

An important application of SHM is detecting crack initiation in areas of high feature density such as at joints. It is difficult to place strain gauges or to propagate waves to detect small damage in joints. The SNS is a simple onboard real-time NDE approach that listens for AE signals from damage in joints. There is no other sensor system that enables the simultaneous monitoring of tens or more long continuous sensors using as few as four channels of data acquisition. In comparison to other techniques, the SNS does not need an amplifier, diagnostic waves, or individual wires for large numbers of discrete sensors/actuators, and there is no need for storage of predamage data. The SNS provides *in situ* simultaneous sensing and intelligence at the sensor level that reduces hundreds of signals into easily interpretable damage information. The SNS is applicable for onboard real-time monitoring of any type of large sensor system in which anomalous events must be detected. The SNS is accurate, simple, miniaturized, lightweight, interior or exterior surface mounted, repairable, redundant, passive, requires low power, and is safe. This meets the objectives of reliable, accurate, and cost-effective operation. The SNS characterizes the severity of the damage right from damage initiation and tells if the damage is serious while the structure is operating.

2.4.2 Damage location algorithm

The neurons that are firing are used to locate damage. A combinatorial algorithm in MATLAB is used to decode which neurons are firing at any time. Because of the limited voltage range of the electronics and the limited number of unique combinations of voltages that can be decoded, this approach of identifying which neurons are firing is limited to about 10 neurons. Thus, a digital approach to locate the firing neurons is suggested for future work to allow up to hundreds of neurons to operate on a digital data bus. Four digital data buses and one power supply wire from a computer are expected to be able to transmit signals from hundreds of neurons in the SNS. This approach can make NDE using the AET practical and can allow monitoring of tens of row and column neurons and hundreds of miniature sensors on a large structure.

2.4.3 Developing the SNS

Initially a two-neuron first prototype of the SNS was designed and built. Results of testing the passive SNS two-neuron prototype are given in [24]. Subsequently, an SNSAP was designed with much faster electronic components and also with better electronic shielding. Also, to prove that the system is practical on a larger scale, the two-neuron prototype was extended to a four-neuron prototype. Results of testing the four-neuron prototype on a thin composite plate in a laboratory-controlled environment were reported by Kirikera *et al.* [25]. Kirikera *et al.* [26] implemented the SNS on a WTB for identifying

the location of propagating cracks during the quasi-static testing. Also, to understand the propagation of Lamb waves on a thin plate, a wave simulation algorithm was developed [25, 27]. The wave simulation algorithm was derived using the modal superposition method and is based on classical thin plate theory. The wave simulation algorithm was extended to excite a composite structure using an actuator at a specific frequency. Long continuous piezoelectric sensors were recently modeled in the wave simulation algorithm to study the ability of continuous sensors, which detect acoustic wave propagation [28]. Testing of the SNS is described in the next section.

3 MONITORING A WIND TURBINE BLADE DURING PROOF TESTING

The SNS was used to identify damage initiation and propagation on a 9-m-long WTB during a quasi-static proof test to failure at the NREL test facility in Golden, CO. The spar caps of the blade were constructed using a constant thickness variable width quasi-unidirectional unitary 3WEAVE carbon/Glass hybrid material [55, 56] developed by 3TEX Inc. The shear web is constructed from fiberglass and balsa wood. Balsa wood is also used in the leading and trailing edge and serves as a panel stiffener. A brief summary of identifying damage on the WTB is discussed in this section. For a detailed explanation the reader is referred to the article by Kirikera *et al.* [22, 26].

WTBs are composite structures with complex geometry and sections that are built of different materials. The 3D structure, large size, anisotropic material properties, and the potential for damage to occur anywhere on the blade make damage identification a significant challenge. During this test, 12 piezoelectric sensors were bonded onto the surface of the WTB and were connected to form four continuous sensors, which were used in the SNS to identify damage. Although 12 sensors monitored the WTB, the SNS produced only two output signals; the first signal identified and located damage and the second comprised combined AE waveforms. Figure 7(a) shows the top view of the outline of the blade with the locations of three load saddles. The load saddles are used for quasi-static testing. Figure 7(b) is an enlarged view of the sensor area

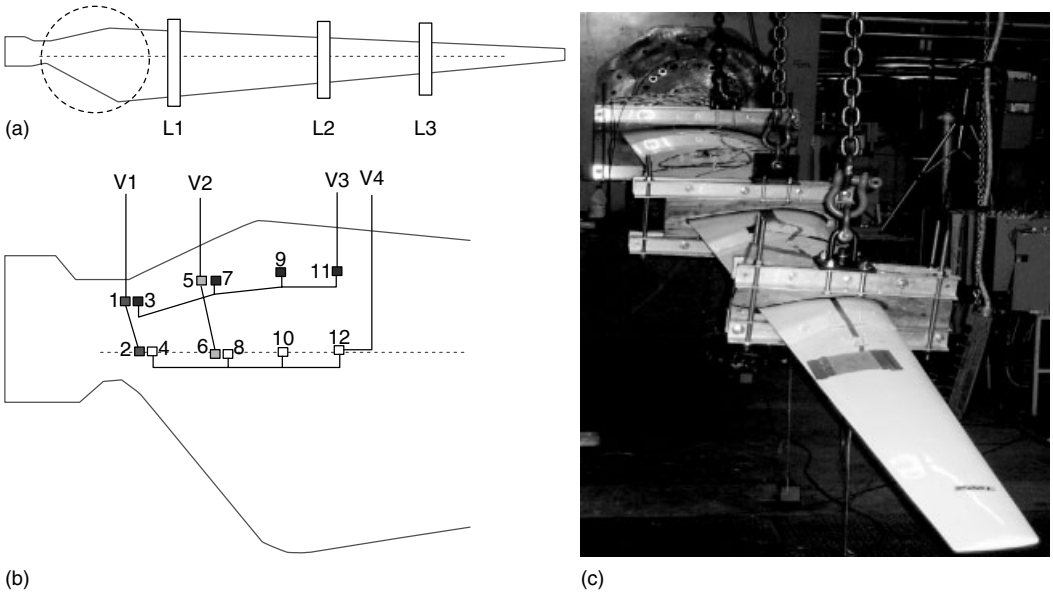


Figure 7. Experimental setup for the blade and SNS sensors: (a) geometry of the blade; (b) sensor locations on the blade; and (c) test setup with three load saddle. (Figure 7c is courtesy of 3TEX Inc. and the National Renewable Energy Laboratory).

on the blade, which is represented by a circle in Figure 7(a). The circled portion has surface-bonded continuous sensors (V1–V4). Twelve piezoelectric wafer sensors are connected in series to form four continuous sensors. Signals from these four continuous sensors are sent as analog inputs to the SNSAP. These signals are further reduced to two channels (T1 and C1). A load profile was applied to the blade using the saddles as shown in Figure 7(c). It is noted that this blade failed at a load above the design load.

Figure 8(a) shows the load profile applied to the WTB and also the response of SNSAP (the T1 and C1 channels in Figure 6). Each continuous sensor was assigned a unique voltage by the response of the C1 channel of SNSAP. A software was used to convert the unique voltages into corresponding neuron numbers, as shown on the ordinate of Figure 8(b). Figure 8(b) indicates the times at which each neuron received the AE signal.

Based on the neuron that receives the AE, multiple damage zones were manually mapped out as seen in Figure 9(a). At the end of the test, multiple damage zones were identified by the SNS (Figure 9a). After the predictions of the damage zones were made based on the SNS results, the blade was cut into sections by the NREL engineers to determine the actual locations of damage. The predicted damage zones were compared with three observed damage locations (Figure 9b) based on cutting the blade into sections. Zones 1A and 1B correspond to damage 1, zones 5, 1A, 1B, and 2 correspond to damage 3, and zones 3A and 3B correspond to damage 2. Zone 4 as predicted by the SNS was not identified in the visual analysis. On subsequent discussions with NREL engineers, it was concluded that damage indeed existed at zone 4 and was missed during visual postfailure observations.

Figure 10 shows the number of AE hits received from each zone. The number of zones is shown in Figure 9(a). On the basis of the number of AE hits, it is concluded that zone 1A has a large portion of the damage compared to zone 1B. Both zones 1A and 1B correspond to the same damage (damage 1 in Figure 9b). Similarly zone 3B has a larger damage than zone 3A as concluded from the visual postfailure analysis of Figure 9(b).

Damages 1 and 2 were visible on the surface of the WTB, damage 3 was not visible on the surface of the WTB. The blade was cut open to understand the cause of failure. The primary mode of failure was panel buckling [29]. The buckling could have started by either the leading edge of the WTB buckling inward or the trailing edge buckling outward leading to a rotation of the spar cap about its axis effectively peeling the spar cap off from the shear web section of the WTB. The rotation of the spar cap was likely a secondary mode of failure [29]. The out-of-plane movement of the panel led to catastrophic damage on the surface of the WTB, Figure 11. The separation of the spar cap from the shear web caused a bond line failure running from 700 to 1650 mm on the WTB, and is shown in Figure 12. Also, a dye penetrant was used to detect brittle gel coat cracks present on the surface of the WTB. The dye pattern did not correlate with the damage locations. The spar cap remained intact indicating that it had performed its

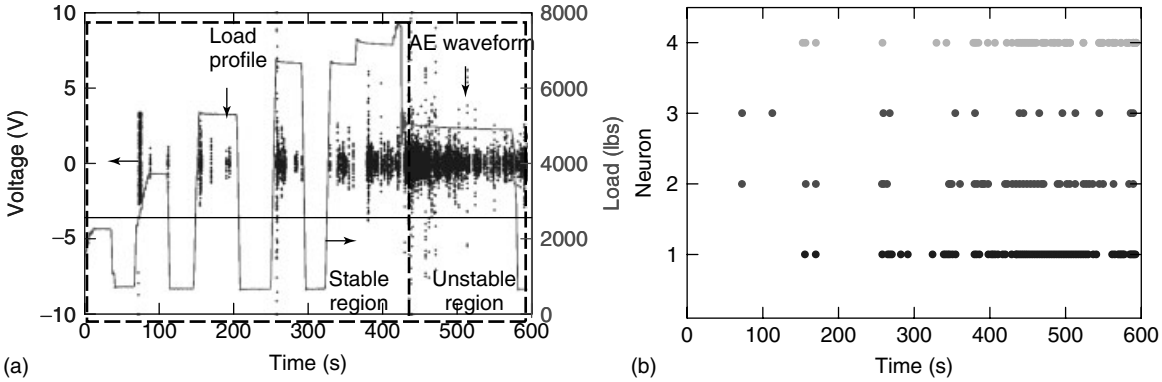


Figure 8. Test data from the SNSAP: (a) T1 channel of the SNSAP superimposed on the load profile and (b) response of the C1 channel of SNSAP used to locate the damages.

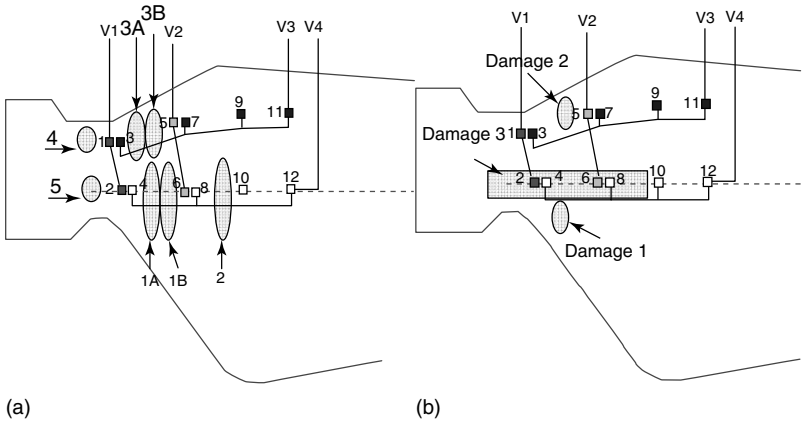


Figure 9. Comparison of the predicted and actual damage locations on the wind turbine blade: (a) damage locations predicted by the SNS before sectioning of the blade and (b) damage locations observed by the NREL engineers after the postfailure sectioning of the blade.

function of being the primary load carrying member, without sustaining catastrophic damage. Also, strain gauges were used on the WTB. The strain data did not indicate the damage progression, but detected the damage just at the onset of buckling failure. Strain gauge locations were not sufficiently close to the failure region to adequately capture the damage.

The SNS indicated the general area where the damage started and how the damage progressed, which is valuable information for verifying and improving the blade design and for verifying the manufacturing procedure. In the future, a grid pattern of sensors (neurons) could be attached inside the blade along the shear web to detect damage inside the blade with greater sensitivity. A major outcome of this testing was to provide confidence that SHM of large anisotropic composite structures that have complex geometry and multiple materials is practical using a simple, low-cost SNS. This test also showed that the SNS can detect cracking that precedes buckling and may be useful for buckling health monitoring (BHM) of large structures.

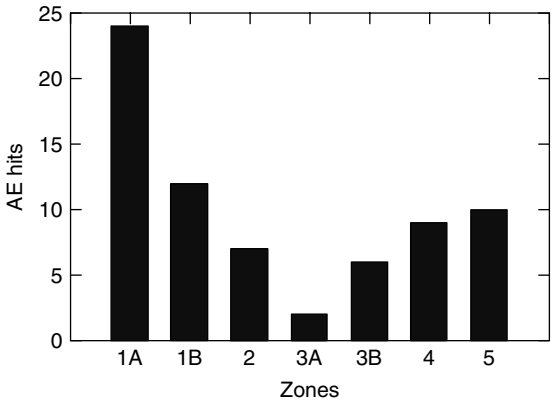


Figure 10. The number of AE hits produced from each zone.

4 BUCKLING HEALTH MONITORING TECHNIQUES

WTBs are hollow tubes designed to have bending twisting coupling and are prone to local buckling. Uncontrolled local buckling can lead to catastrophic failure of individual blades and subsequently the whole turbine. Since buckling initiates as an elastic deformation, the subsequent damage to the structure could be entirely prevented if this tendency is identified in real time and countermeasures are taken. Two different approaches have been examined for identifying the local buckling in WTBs.

The first approach is to propagate low-frequency Lamb waves across the region that has the tendency to undergo local buckling, regions A and B in Figure 13(a), as was used during static testing of a WTB [5]. During this test, a 5-kHz Lamb wave signal was able to indicate the initiation of buckling deformation at 40% length of the blade from the root

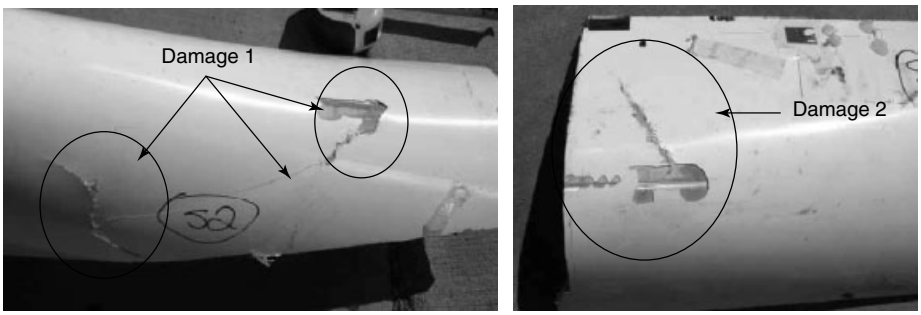


Figure 11. Postfailure sectioning showing visible damage: (a) "damage 1" on the trailing edge of the WTB and (b) "damage 2" on the leading edge of the WTB. [Pictures courtesy of NREL. Reproduced with permission.]

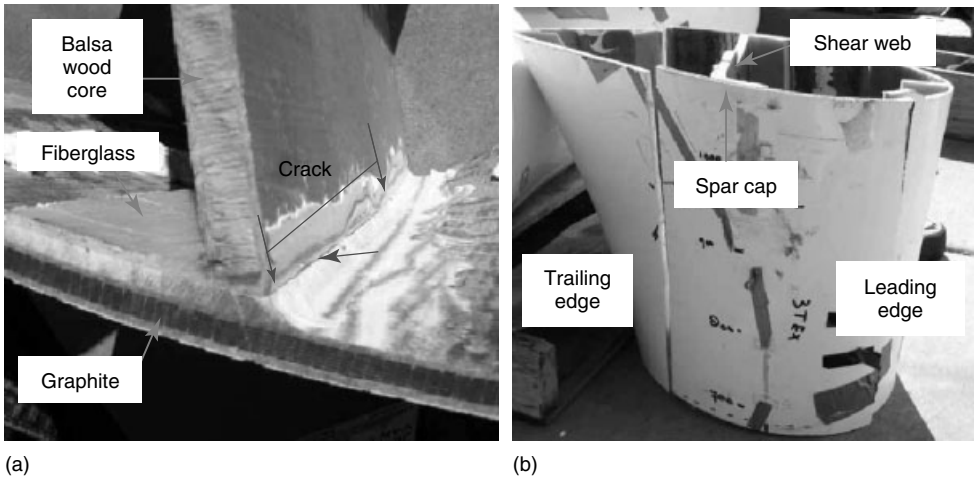


Figure 12. Cut sections of the WTB (a) indicating bond line failure between the spar cap and shear web and (b) showing the interior geometry of the blade. [Pictures courtesy of NREL. Reproduced with permission.]

end. The blade failure at about 4500 lbs (2040.82 kg) load, Figure 13(b), was premature and was caused by this buckling. The initiation of buckling is indicated by the reduction in amplitude of the received Lamb wave signals around 4000 lbs (1814.34 kg) as shown in Figure 13(c).

The second approach is the monitoring of the natural frequency of the region that has a potential for buckling. This condition was simulated in a 4-in.-wide, 49-in.-long, and 0.125-in.-thick glass fiber epoxy composite strip. Finite element analysis of the vibration characteristics of this bar at various stages of buckling indicated that easily recognizable changes in frequency and mode shapes result even in the early stages of buckling. Experimental measurement of the frequency and mode shapes confirmed the feasibility of this approach. In this experiment, the bar was excited at the third natural frequency corresponding to the unbuckled configuration using a five-cycle windowed sine wave. The oscillation in the unbuckled bar lasted longer than 1 s. The duration of oscillation dropped significantly even in the early stages of buckling when the deformation was less than 0.25 in., providing a clear indication of the instability. Figure 14(a) shows the experimental configuration and Figure 14(b) shows the different time responses of the beam for different stages of prebuckling.

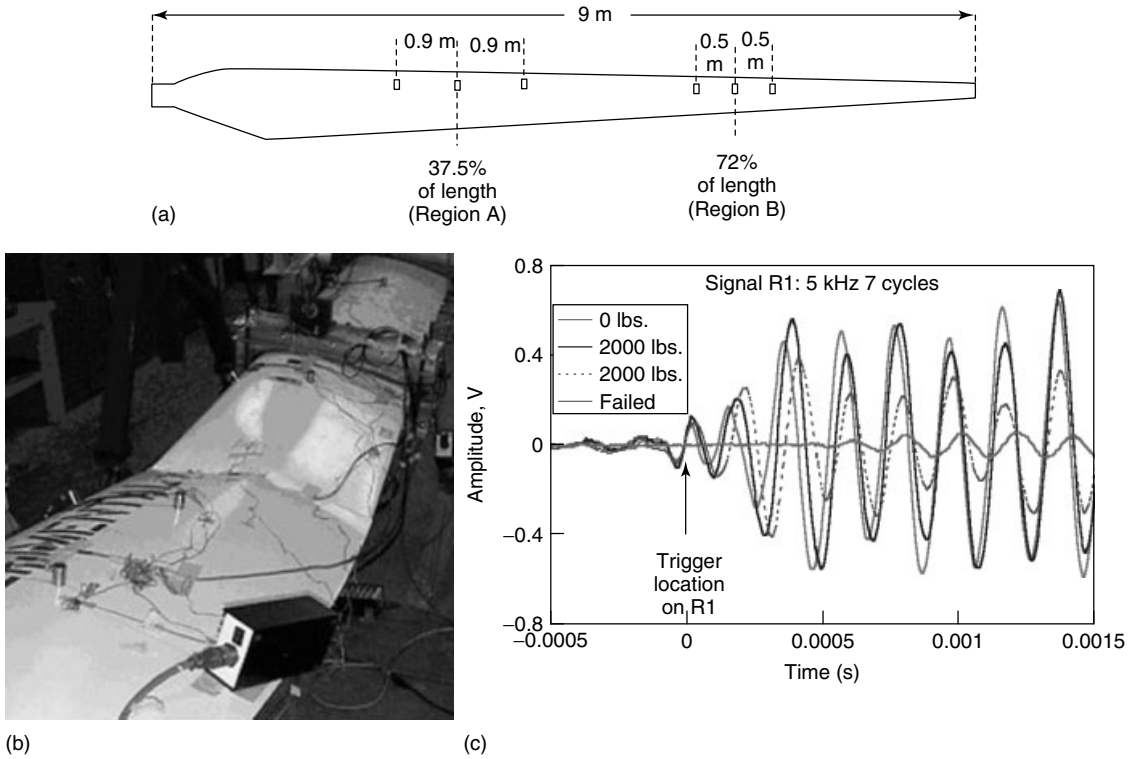


Figure 13. Early detection of local buckling in a WTB using low-frequency Lamb wave signals.

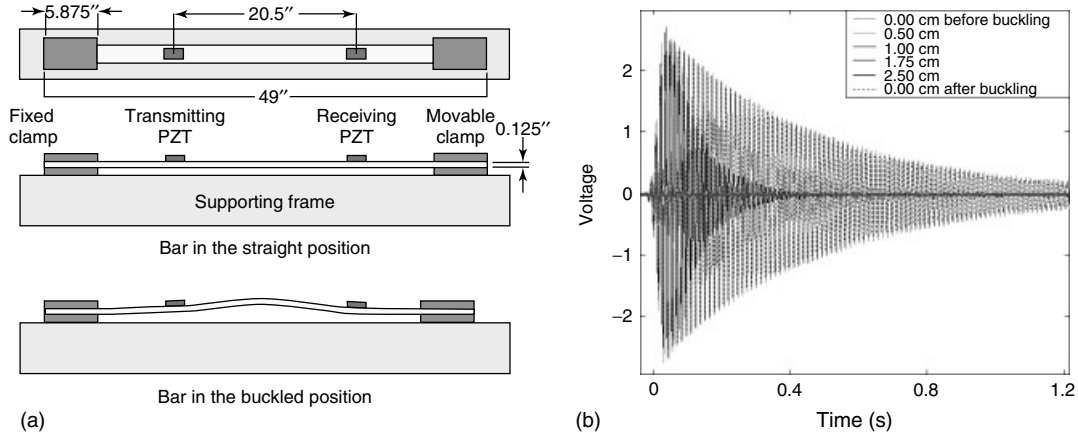


Figure 14. Detection of buckling through changes in natural frequency.

5 MULTISTATE CONTINUOUS SENSORS

This section describes recent advances in continuous sensor technology that allows different types of distributed sensors to be utilized inexpensively, reliably, and with low weight for monitoring complex structures such as WTs. The continuous sensor architecture is for use in an SNS described above. SNS is an electronic sensor network developed using highly

distributed interconnected continuous sensors coupled to a parallel signal-processing system to enable *in situ* real-time monitoring of WTBs. The generic SNS can use almost any type of passive sensor. Examples include piezoelectric ceramic wafers to monitor AEs due to crack propagation, and carbon nanotube (CNT) neurons (threads or film) to monitor strains, detect large cracks, and to detect electrolytes related to early stage corrosion and hygrothermal degradation. In structural applications, an SNS can detect damage such as composite fiber breaking and tell when impact damage is propagating. SNS is low cost and simple because no predamage data, actuation, or multiplexing is needed. Also, it is relatively insensitive to drift in properties of the sensors owing to environmental effects because reference data is not used. From an operational viewpoint, SNS is more than a monitoring system—it is an artificial central nervous system that monitors the condition of the structure without interruption and identifies when anomalous events and early degradation occur. It also defines the operational performance envelope and provides confidence in terms of safety and reliability of the system. In structures, degradation often occurs at complex built-up or joined sections where the load and acoustical paths are complex. SNS is excellent for use in these areas of high feature density where other monitoring techniques are impractical to use. Multistate continuous sensors extend the usefulness of SNS.

This section describes ongoing development of a suite of multistate continuous sensors, including MEMSs continuous accelerometers, CNT crack and corrosion sensors, pressure, temperature, and other types of sensors for use in an SNS. It reduces signals from many sensors to a small amount of essential information that can be equated to degradation of the system. Different types of sensors can be used with an SNS, but the design of the continuous sensor might differ for these different types of sensors. Moreover, the SNSAP parameters change depending on the type of continuous sensor. Details of filtering and thresholding also will, in general, be different for every sensor type. We are presently developing a suite of multistate continuous sensors for use in an SNS. The goal is to achieve integration of multiple sensor types into a single signal-processing architecture.

Dual use application of the SHM techniques for WT monitoring and other applications should be considered to reduce development cost and to take maximum advantage of technology. Besides mechanical applications such as monitoring WTBs, sensor networks are becoming increasingly important to safeguard the environment and homeland, and for medical applications. SNS has potential application in many types of sensor systems to simplify the hardware by reducing the number of wires and A/D convert boards. SNS is designed to sense anomalous values of a particular state variable over a large area using a continuous sensor. Figure 15 shows some of the individual sensors and sensor materials that can be used to form continuous sensors. The design of different multistate sensors for an SNS is discussed in the following text. It is also important to model the transfer function of the continuous sensor and SNS for new types of sensors. In particular, individual sensor nodes must be integrated properly to form continuous sensors—they cannot be arbitrarily connected. The rows overlap the column neurons and are electrically insulated. The size of the cells in the grid can be small to detect small damage. A description of the different sensor types is as follows.

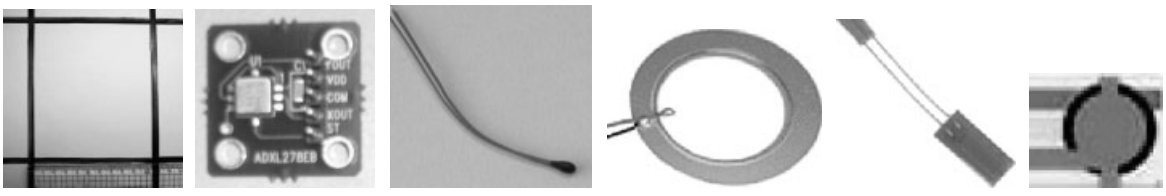


Figure 15. Sensor elements that can be used to form continuous sensors: (a) nanotube impedance film sensors to detect corrosion, cracking, and delamination for an SNS; (b) MEMS accelerometer on a circuit board; (c) thermistor that can from a continuous temperature sensor; (d) PZT commercial wafers with brass backing used to detect acoustic emissions; (e) strain gauge sensor; and (f) pressure mapping sensor.

1. Carbon nanotube continuous sensor

This sensor is to detect cracks, delamination, and corrosion damage based on electrical impedance. A prototype CNT-based sensor has been built and cracking and electrolyte representing corrosion have been detected based on the impedance of the sensor. CNT film sensors, highly distributed on the surface or embedded within a composite, are used to form a continuous sensor for crack and corrosion monitoring. An example of nanotube neurons on a simple panel is explained. A highly distributed network of nanotube continuous sensors, which sense along their entire length and have a biomimetic architecture, allows coverage of large structures. An example of the change in resistance as a crack passes through a single neuron is shown in Figure 16(a). There is a small change in resistance as the crack begins propagating through the neuron. The change becomes larger and approaches infinity as the crack passes through the neuron. An example of the change in capacitance as electrolyte is put on a single neuron is shown in Figure 16(b). Up to a factor of 50 increase in capacitance occurs because of the double layer supercapacitance property of nanotubes. The change in resistance owing to the electrolyte is only 6% as shown in Figure 16(b). Therefore, crack and corrosion sensing are mostly decoupled and can be measured using the same neuron. Signal processing using the nanotube neurons is greatly simplified because the electrical properties of the neurons, rather than structural waves, are used to characterize damage.

A prognostic method for damage modeling can be developed based on the changes in electrical parameters of the CNT neurons. The modeling of the neuron is based on the Randal Warburg circuit. The CNT is a strain sensor (piezoresistive effect) and a highly sensitive corrosion sensor (electrochemical impedance spectroscopy (EIS) effect). A model should be developed to predict the crack length based on the EIS spectra of the nanotube composite. An electrode configuration can also be used on a nanocomposite plate where the nanotubes are dispersed in the polymer matrix. Low-cost carbon nanofibers can replace high-cost CNT for large applications of an SNS.

An important application of the CNT neuron in WTB is health management of polymer matrix composites (PMCs). The CNT neuron can be developed for service life monitoring of PMCs used WTB structural applications. Two of the primary life-limiting mechanisms in polymer composites are hygrothermal degradation and oxidative degradation. Hygrothermal and oxidative degradation can lead to chemical changes in the resin system causing cracking and embrittlement in the surface layers of the composites. Within the oxidized layer of the composite, the tensile strength, strain to failure, flexural strength, density, and toughness decrease while the modulus increases. Surface cracks provide pathways for the transport of moisture and oxidants to the fiber/matrix interfaces that act as high-diffusion paths thereby increasing the degradation rate. The CNT neuron is a passive continuous sensor that

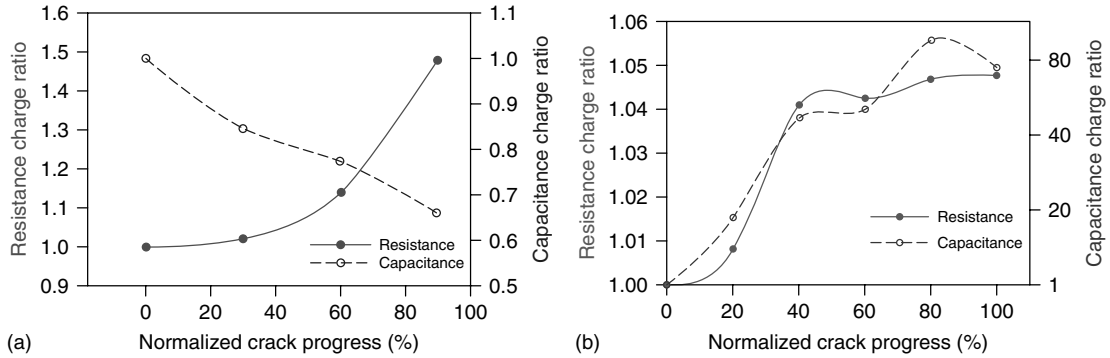


Figure 16. Change in capacitance and resistance of a carbon nanofiber neuron owing to simulated damage: (a) crack propagation damage, the resistance changed from 9.3 kΩ to infinity, the capacitance changed from 44 pF to zero; (b) electrolyte corrosion damage, the resistance changed from 117 to 123 kΩ, which is a 5% increase, the capacitance changed from 19.1 to 1109 pF, a 60 times increase.

can monitor the electrical impedance of composite materials to indicate degradation of PMCs used in WT structures. Future work should investigate the scientific, technical, and commercial feasibility of the CNT neuron for PMC health management including monitoring oxidative degradation for a neat resin aged in low temperature and oxidizing environments.

2. MEMS capacitive accelerometer
Accelerometer evaluation boards from Analog Devices Inc. were used to build a four-element continuous sensor. The accelerometers are two axis each and measure in-plane acceleration. The four accelerometers were used to form a continuous accelerometer with four inputs and one output in the x axis and a second continuous accelerometer with four inputs and one output in the y axis. These accelerometers work based on the capacitance of a MEMS cantilever. The accelerometers required a modified circuit to form the continuous sensor to sum and amplify signals and to prevent low-frequency signal feed-through. Initial testing was successful, opening up the use of MEMS continuous accelerometers for vibration measurement, impact detection, and condition monitoring of bearings, unbalance, and damage.

3. Thermistor continuous temperature sensor
This sensor is to monitor operating temperatures based on resistance of the sensor. Certain types of damage may cause an increase in temperature. The continuous temperature sensor may indicate hot spots in a structure.

4. Piezoelectric Lead Zirconium Titanate (PZT) wafers
These are being tested with a brass backing to sense AEs owing to cracks and delamination in metallic and composite structures. Prototypes were tested in the lab with good results. This sensor is more rugged than PZT wafers without a backing. Studies using the transfer function voltage output of the sensor divided by strain input (V_0/S) show that a large number of piezoelectric sensors will not reduce the quality of the data obtained for AE detection. Transfer functions must be determined for other sensor types in the project.

5. Continuous strain gauge
This sensor can be formed using miniature strain gauges in series or parallel. Equivalent electric circuits can be used to develop a transfer function. The strain gauge continuous sensor will be tested in the project to monitor large strain simultaneously at multiple gauges.

6. Continuous pressure sensor
This sensor can be built using a piezoresistive or capacitive film sensor typically with applications in biomechanics. The piezoresistive and capacitive pressure sensors are easy to incorporate as a continuous sensor. This sensor has many types of commercial applications.

Other sensor types include an inclinometer, humidity sensors, proximity sensor, magnetic sensor, flow sensor, and liquid level sensors. Another class of sensors is nanowire sensors. A nickel nanowire neuron continuous sensor is expected to be useful to detect cracks, delamination, and corrosion damage using electrical impedance or eddy current. The Ni nanowires are being produced at University of Cincinnati. Use of multiple sensor types and fusing data is possible using SNS and is an area of future work.

5.1 Manufacturing an SNS

Features of an SNS include the following: (i) it is a passive system; (ii) detects anomalous events; (iii) continuous sensors combine 10 or more sensor nodes into a one wire sensor; (iv) highly distributed continuous sensors called *neurons* can detect small damage on large structures with complex geometry and high feature density; (v) bioinspired parallel signal processing reduces the number of channels of data acquisition from tens or hundreds to four; (vi) sensor modules can be developed for different types of sensors (piezoelectric, piezoresistive, thermistor, strain, magnetic, etc.); (vii) multistate sensing is possible in one system architecture; (viii) sensing AEs to detect cracking is possible using high-bandwidth data-acquisition boards (up to 1 MHz); (ix) sensing vibration using accelerometers is possible using moderate-bandwidth data-acquisition boards (up to 50 kHz); and (x) sensing strain, pressure, temperature is possible using low-bandwidth data-acquisition boards (up to 1 kHz). Overall, SNS is

simple, low cost, redundant, and reliable. Other types of SHM systems may also have many of the attributes mentioned above. It is suggested that some type of SHM test bed be developed for WTBs to practically evaluate the different techniques for SHM.

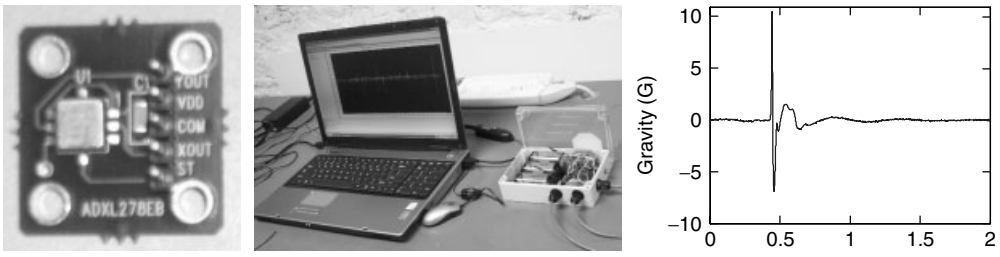
A plan for manufacturing SNS should be developed such that the SNS is available for prototype testing on operating WTs. To a first-order approximation, the cost of an SNS can be quite reasonable (∼$20 000, including the analog electronics, computer, A/D boards, and sensors), and the system can be small and lightweight. Dual use commercialization of SNS should be possible. In military, the technology can be applied to aircraft for structural and engine applications, and for exhaust wash structures to monitor high-temperature PMCs. For commercial applications, the technology can be applied to civilian and military aircraft and nonaerospace applications such as bridges, buildings, rocket casings, and any structure with complex geometry that must be monitored for degradation in real time.

6 WIRELESS MEMS ACCELEROMETERS FOR SHM IN ROTATING SYSTEMS

Existing accelerometers and the data-acquisition systems used for structural analysis and monitoring studies are often cumbersome to use. Therefore, a wireless system using small MEMS accelerometers is being developed. The first stage of this development has been performed successfully at the University of Cincinnati. The accelerometer system was built by integrating commercial MEMS accelerometers (ADXL 278, Analog Devices Inc.) with wireless sensor technology from MicroStrain Inc. The wireless measurement system can be installed on a rotating WTB without being tethered to any external devices or computers. The current prototype system has eight MEMS accelerometers wired to an A/D converter and wireless transmitter (MicroStrain Inc.). The cost is about $3000. This battery-powered system (Figure 17) is in initial testing and continuously streams acceleration data wirelessly to a laptop computer with a bay station. This wireless system cannot store much data locally, but seems suitable for WT use. Battery life is a limitation and power harvesting from blade vibration may be used to charge the batteries. SNS could be adapted to the wireless system.

6.1 Damage detection and fatigue testing of complex joints

Detecting and locating cracks in WT structural components that have high feature densities is a very challenging problem and is not discussed much in the literature. In general, few SHM techniques have been applied to the monitoring of joints and complex structural geometries. However, reliable low-cost assessment of joints is crucial to maintain operational availability and productivity, reduce maintenance cost, and prevent catastrophic failure of WTs. SNS and other SHM systems should be deployed on components or sections of composite structures and tested together to understand the capabilities of the different approaches for SHM. SNS



(a) (b)

Figure 17. Wireless data-acquisition system with eight MEMS accelerometers: (a) MEMS accelerometer/board, battery-powered wireless module, laptop computer with bay station and (b) one channel response due to a shock input. This system was assembled at the University of Cincinnati using commercial components. Continuous accelerometers can be used to allow 32 or more individual MEMS accelerometers to be used with this system.

could become a standard signal-processing architecture for general SHM systems, where reducing the number of data-acquisition channels and hardware is of critical importance.

6.2 Resources for SHM

An increasing number of publishers, companies, and research organizations are developing and providing information, sensors, and measurement systems that may be used for SHM including for WTs. A partial list of resources for SHM components and systems is given in [41–54].

7 SUMMARY AND CONCLUSIONS

Different techniques for monitoring wind turbines and particularly WTBs are discussed. Overall, we can say that monitoring degradation in WT structures is a complex problem that might be solved using multiple types of sensors and different damage detection techniques. New signal-processing architectures are also needed that simplify the hardware and reduce the cost of the overall health monitoring system. An SNS was developed as a simple approach for health monitoring of WTs. The neural system listens for AEs to detect damage in a WTB. The SNS was tested during proof testing of WTB and multiple damage zones were located. Subsequent postfailure sectioning of the failed blade showed close correlation between the predicted and actual damage locations. An SNS with multistate continuous sensors is also being developed to measure other variables such as pressure, acceleration, and electrical impedance to identify different types of damage. Power harvesting and wireless data transmission from the rotating blades to the fixed frame are other areas where research is needed. It is suggested that a test bed for WTBs be developed to practically evaluate the different techniques for SHM.

ACKNOWLEDGMENTS

The work related to the SNS was supported by the NREL under subcontract number XCX-2-31214-01. Mr Alan Laxson was the technical monitor of this project. Much of the help in making the prototype of the SNSAP was provided by engineers from Texas Instruments and Analog Devices. Also Mr Henry Westheider, Instrumentation specialist, and Mr Doug Hurd, Machinist, in the Department of Mechanical Engineering at the University of Cincinnati provided valuable help in building the SNS prototype. The static test of the 3-TEX-100 blade was performed at NREL test facility in Golden, CO. The LABVIEW expertise provided by Dr Vahan Gevorgian of NREL during the on-site testing of the SNSAP on the WTB is gratefully appreciated. Mr Derek Berry and Dr Mansour Mohamed are the principal technical leads from TPI Composites and 3 TEX, Inc., respectively. Mr Scott Hughes and Mr Jeroen van Dam are the NREL engineers who assisted during the WTB testing.

REFERENCES

[1] Migliore P, van Dam J, Huskey A. *Acoustic Tests of Small Wind Turbines*, AIAA-2004-1185, 1983.

[2] NREL Publications, National Renewable Energy Laboratory, Wind Technology Center, Golden, CO, http://www.nrel.gov/wind/ (accessed Apr 2008).

[3] Bently Nevada, *Product Overview of Wind Turbine Condition Monitoring*, http://www.proximityprobes.com/applications/wind.htm (accessed Jun 2006).

[4] Wells R, Hamstad MA, Mukherjee AK. On the origin of the first peak of acoustic emission in 7075 aluminum alloy. *Journal of Materials Science* 1983 **18**:1015–1020.

[5] Sundaresan MJ, Schulz MJ, Ghoshal A. *Structural Health Monitoring Static Test of a Wind Turbine Blade*, NREL/SR-500-28719, Subcontractor report, March 2002.

[6] Dutton AG, *et al.* Acoustic emission monitoring from wind turbine blades undergoing static and fatigue testing. *Proceedings of the 15th World Conference on Non-Destructive Testing*. Roma, 15–21 October 2000, http://www.ndt.net/article/wcndt00/papers/idn553/idn553.htm.

[7] Joosse PA, Blanch MJ, Dutton AG, Kouroussis DA, Philippidis TP, Vionis PS. Acoustic emission monitoring of small wind turbine blades. *Journal of Solar Energy Engineering, Transactions of the ASME* 2002 **124**:446–454.

[8] Berry D. *Wind Turbine Blades Manufacturing Improvements and Issues*, Feb 2004, Available online at http://www.sandia.gov/wind/2004BladeWorkshopPDFs/DerekBerry.pdf.

[9] Sutherland HJ. *On the Fatigue Analysis of Wind Turbines*, Report SAND99-0089. Sandia National Laboratory, 1999.

[10] Sutherland HJ, Mandell JF. Effect of mean stress on the damage of wind turbine blades. *Journal of Solar Energy Engineering* 2004 **126**(4):1041–1049.

[11] Sandia National Laboratory, Wind Energy Group, *Annual Workshop Proceedings*, http://www.sandia.gov/wind/topical.htm (accessed Apr 2008).

[12] Chang FK (ed). Structural health monitoring. *Proceedings of the Workshop on Structural Health Monitoring*, Stanford University, Stanford, CA. DEStech Publications, 1997, 1999, 2001, 2003, 2005.

[13] Smart Materials Nanotechnology Lab, http://www.min.uc.edu/∼mschulz/smartlab/smartlab.html (accessed Apr 2008).

[14] *Structural Health Monitoring: An International Journal* 2002.

[15] Miller RK, Hill EVK (eds). Acoustic emission testing. *Nondestructive Testing Handbook*, *Third Edition*. American Society for Nondestructive Testing, 2005; Vol. 6.

[16] Larwood S, Musial W. Comprehensive testing of Nedwind 12-Meter wind turbine blades at NREL. *19th American Society of Mechanical Engineers (ASME) Wind Energy Symposium*, 2000.

[17] Sorensen BF, *et al. Fundamentals for Remote Structural Health Monitoring of Wind Turbine Blades—A Preproject*, Report Riso-R-1336(EN). Riso National Laboratory: Roskilde, May 2002.

[18] Musial W. *Energy in the Wind Presentation*. NREL, Kidwind Teachers' Workshop, May 2005.

[19] Harvey RL. *Neural Network Principles*. Prentice Hall, 1994.

[20] Zigmond MJ, Bloom FE (eds). *Fundamental Neuroscience*. Academic Press: San Diego, CA, 1999.

[21] Kirikera GR. *An Artificial Neural System for Structural Health Monitoring*, MS thesis, University of Cincinnati, August 2003.

[22] Kirikera GR. *A Structural Neural System for Health Monitoring of Structures*, Ph.D., University of Cincinnati, August 2006.

[23] Kirikera GR, *et al.* Mimicking the biological neural system using active fiber continuous sensors and electronic logic circuits. *SPIE 11th International Symposium, Smart Sensor Technology and Measurement Systems*. San Diego, CA, 14–18 March 2004; pp. 148–157.

[24] Kirikera GR, Shinde V, Schulz MJ, Ghoshal A, Allemang R, Sundaresan MJ. Damage localization in composite and metallic structures using a structural neural system and simulated acoustic emissions. *Mechanical Systems and Signal Processing* 2007 **21**(1):280–297.

[25] Kirikera GR, Shinde V, Schulz MJ, Ghoshal A, Sundaresan MJ, Allemang RJ, Lee JW. A structural neural system for real time health monitoring of composite materials. *Structural Health Monitoring—An International Journal* 2008 **7**(1):65–83.

[26] Kirikera GR, Shinde V, Schulz MJ, Sundaresan MJ, Hughes S, van Dam J, Nkrumah F, Grandhi G, Ghoshal A. Monitoring multi-site damage growth during quasi-static testing of a wind turbine blade using a structural neural system. *Structural Health Monitoring—An International Journal* 2007 (Accepted, In Press).

[27] Ghoshal A, Martin WN, Schulz MJ, Chattopadhyay A, Prosser WH. Simulation of asymmetric Lamb waves for smart sensing and actuation systems in plates. *Shock and Vibration Journal* 2005 **12**(4):243–271.

[28] Lee JW, Kirikera GR, Kang I, Schulz MJ, Shanov V. Structural health monitoring using continuous sensors and neural network analysis. *Smart Materials and Structures* 2006 **15**:1266–1274.

[29] van Dam J, Hughes S. *NWTC Structural Testing Test Report—Static Test of the 3-TEX 9meter Blade*, NWTC-ST-3TEX-STA-01-1205, Internal NREL report, April 2006.

[30] Frank Pai P. *SHM and Use of the Nonlinear Finite Element Code GESA*, http://web.missouri.edu/∼umcengrmaeweb/faculty/pai/pai.html (accessed Apr 2008).

[31] Ghoshal A, Sundaresan MJ, Schulz MJ, Frank Pai P. Structural health monitoring techniques for wind turbine blades. *Journal of Wind Engineering and Industrial Aerodynamics* 2000 **85**(3, 24):309–324.

[32] Insensys, *Structural Health Monitoring Using Fiber optics*, http://www.insensys.com/wind_downloads.asp (accessed Apr 2008).

[33] Lading L, McGugan M, Sendrup P, Rheinlander J, Rusborg J. *Fundamentals for Remote Structural Health Monitoring of Wind Turbine Blades—A Preproject, Annex B–Sensors and Non-Destructive Testing Methods for Damage Detection in Wind Turbine Blades*, Riso-R-1341(EN), Riso National Laboratory, Roskilde, May 2002.

[34] Pitchford CW, Grisso BL, Inman DJ. Impedance-based structural health monitoring of wind turbine

blades. In *Proceedings of the SPIE*. SPIE, April 2007; Vol. 6532, pp. 65321I.

[35] Drewry MA, Georgiou GA. A review of NDT techniques for wind turbines. *Insight* 2007 **49**(3): 137–141.

[36] *A Review of NDT Techniques for Wind Turbines*, http://en.wikipedia.org/wiki/Wind_turbine (accessed Mar 2007).

[37] McMillan D, Ault G. Towards Quantification of Condition Monitoring Benefit for Wind Turbine Generators. European Wind Energy Conference and Exhibition, EWEC 2007, http://ewec2007proceedings.info/index2.php?page=searchresult&tr=3 (accessed Apr 2008).

[38] Risoe National Laboratory, http://www.risoe.dk/ (accessed Apr 2008).

[39] The American Wind Energy Association (AWEA) 2007. http://www.awea.org/.

[40] Caithness Windfarm Information Forum. http://www.caithnesswindfarms.co.uk/ 2007.

[41] *Smart Materials and Structures Journal*, http://www.iop.org/EJ/journal/SMS (accessed Apr 2008).

[42] *Journal of Intelligent Material Systems and Structures*, http://www.sagepub.com/journalsProdDesc.nav?prodId=Journal201582 (accessed Apr 2008).

[43] Acellent Technologies. http://www.acellent.com/ 2007.

[44] Physical Acoustics Corporation, http://www.pacndt.com/index.aspx?go=research&focus=/capabilities/structural_health_monitoring.htm (accessed Apr 2008).

[45] Advitam. http://www.advitam-group.com/ 2007.

[46] Sawyer T, Aragon G. *Structural Health Monitoring is Sensitive Subject*, August 2007, http://enr.construction.com/features/transportation/archives/070808c.asp.

[47] ASMOTE. http://www.asmote.com/ 2007.

[48] Motion and Control Group, University of Sheffield, http://www.dynamics.group.shef.ac.uk/health/health.html (accessed Apr 2008).

[49] Ultra Electronics Limited, http://www.ultracontrols.aero/Products/pdf/15_crackdetection.pdf (accessed Apr 2008).

[50] Smart Fibers, http://www.smartfibres.com/SHM.htm (accessed Apr 2008).

[51] CrackFirst, http://www.twi.co.uk/j32k/unprotected/band_1/c1352.html (accessed Apr 2008).

[52] Los Alamos National Laboratory, *Structural Health Monitoring*, http://www.lanl.gov/damage_id/ (accessed Apr 2008).

[53] Ben Franklin Center of Excellence in Structural Health Monitoring, http://www.esm.psu.edu/wiki/research:cjl9:structural_health_monitoring (accessed Apr 2008).

[54] Center for Intelligent Material Systems and Structures, http://www.cimss.vt.edu/ (accessed Apr 2008).

[55] Mohamed MH, Wetzel KK. 3D woven carbon/glass hybrid spar cap for wind turbine rotor blade, *Transactions of the ASME, Journal of Solar Energy Engineering* 2006 **128**:562–573.

[56] Mohamed MH. *Wind Blade Spar Cap and Method of Making*, USP No. 7,377,752, May 27, 2008.

Chapter 148
Large Rotating Machines

Tomasz Gałka
Institute of Power Engineering, Warsaw, Poland

1 INTRODUCTION

The term *large rotating machines* refers mainly to turbine-generator units and high-capacity pumps, fans and compressors employed in power industry, process industry, mining, etc. "Large" means high power or, more precisely, high intensity of energy transformation processes. From the point of view of technical condition monitoring and assessment, they have four important features:

- complexity, which implies that such machines are characterized by a large number of technical condition parameters;
- individuality, which means that two machines of the same design often differ substantially from the point of view of condition assessment;
- critical nature, which means high cost, high reliability demand, and high risk involved in a failure (life hazard, damage, production loss, and environmental impact);
- long service life, usually with a number of major overhauls.

These features have a major influence on methods used for structural health monitoring, as well as on relevant measuring equipment and data processing techniques.

In general, monitoring large rotating machines has two principal aims:

- detection, identification, and assessment of failures (to maintain operational parameters, provide operational safety, and optimize overhauls);
- tracing condition degradation processes (to determine life consumption and/or residual life).

Both these aims are accomplished by using the same technical condition symptoms, but they differ in procedures and algorithms employed for this purpose.

Historically, the beginnings of structural health monitoring in large rotating machines can be related to the introduction of measurement of process parameters, which were later supplemented by vibration measurements. Apart from performance monitoring, they were initially employed for maintaining

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

operational safety. In the 1960s, the first generation of systems for continuous diagnostic monitoring was introduced. These systems were based on analog techniques and employed relatively simple data processing algorithms. The second-generation systems, which appeared in the 1970s, featured more advanced processing techniques, like fast Fourier transform (FFT). Third generation (expert systems), which saw widespread introduction in the 1990s, was based entirely on digital techniques and featured advanced software for data processing and presentation, as well as certain elements of diagnostic reasoning [1]. It seems justified to state that measurement techniques have reached a stage of maturity and certain patterns have been established. On the other hand, development in the field of data processing is very intensive; in particular, Artificial Intelligence (AI) methods are introduced to a growing extent.

2 CONDITION SYMPTOMS

Because of their complexity, large rotating machines are characterized by a number of technical condition parameters, which can be described by a vector $\mathbf{X}(\theta) = \{X_1(\theta), X_2(\theta), \ldots, X_m(\theta)\}$, where θ denotes machine lifetime.[a] Usually these parameters cannot be measured directly and the technical condition is determined on the basis of symptoms, i.e., measurable quantities covariable with the machine condition [2]. For a given machine, we can define a vector of symptoms $\mathbf{S} = \{S_1(\theta), S_2(\theta), \ldots, S_n(\theta)\}$. A fundamental relation is thus given by

$$\mathbf{S}(\theta) = \Phi[\mathbf{X}(\theta)] \tag{1}$$

where Φ denotes an operator. In some cases, for a symptom $S_i(\theta) \in \mathbf{S}$, we can assume a lack of cross relations, which means that

$$\vee_j \frac{\partial S_i}{\partial X_k} \approx 0 \text{ if } k \neq j \tag{2}$$

which implies that equation (1) can be reduced to the simplest possible form of a diagnostic relation

$$S_i(\theta) \approx \Phi[X_j(\theta)] \tag{3}$$

Such an approach sometimes yields analytical expressions, which can be directly employed in condition assessment procedures [3]. Their suitability is, however, limited to particular machine types or even individual examples. In practical applications, measurable symptoms depend not only on technical condition, but also on control parameters and interference, so that equation (1) has to be replaced by

$$\mathbf{S}(\theta) = \Phi[\mathbf{X}(\theta),\ \mathbf{R}(\theta),\ \mathbf{Z}(\theta)] \tag{4}$$

where $\mathbf{R}$ and $\mathbf{Z}$ denote control and random interference vectors, respectively.

Let us assume, for the time being, that influences of control and random interference can be either neglected or accounted for in some way (this issue shall be discussed later). Then, given long service life and overhauls performed at certain intervals, the time history of a symptom $S_i(\theta)$ can be considered as a combination of two functions. The first represents life consumption processes, inevitably resulting from normal operation. These processes can be described by monotonically increasing functions [4] with time constants of the order of machine service life. The second one represents the influence of overhauls and failures. Overhauls are represented by step changes. Failures manifest themselves as either very fast (almost step) changes or continuous evolution with a comparatively short time constant. This corresponds to the two principal aims of condition monitoring mentioned above.

2.1 Vibration-based symptoms

Because of high information content, vibration patterns are the most important source of diagnostic symptoms for condition monitoring of rotating machines. Various vibration characteristics provide information on machine condition [5, 6]; the most useful and commonly employed characteristics are

- absolute vibration spectra (bearings, casings, and foundations),
- relative vibration trajectories (orbits),
- vibration time histories or trends (evolution).

Large rotating machines produce broadband vibration spectra with many individual components. These components can be grouped in two categories. The first includes those resulting directly from rotary

motion, i.e., harmonics and subharmonics of the fundamental frequency f_0 that corresponds to the rotational speed. They are generally referred to as *harmonic components*. The second includes those generated by medium flow interaction with the fluid-flow system, sometimes called *blade components*. An example, for a steam turbine rotating at 3000 rpm, is given in Figure 1.

Subharmonic components are related to malfunctions like fluid-induced rotor instability (oil whirl and whip) and partial rub in journal bearings [7, 8]. Instability produces $n \times f_0$ components ($n < 1$), which often take the form of a broadband "hump" in the subharmonic range [7, 9]. Rub usually produces $(n/m) \times f_0$ components ($n/m < 1$, both n and m small integer values), which are clearly visible as peaks in narrowband spectra [7]. Shaft trajectory (orbit) analysis is usually conclusive.

Harmonic components are related to common malfunctions like unbalance, rotor bow, coupling misalignment, shaft crack, loose part, or debris accumulation. All these faults can result in the $1 \times f_0$ component increase, which usually provides the first warning. The $2 \times f_0$ component is usually related to misalignment [7, 10], but can also indicate a rotor crack. Higher harmonic components are more difficult to interpret; they sometimes result from faulty couplings, but can also be produced by fluid-flow system components [11]. In general, diagnostic relations for these symptoms are complex and cannot be reduced to simple and most convenient forms of the type given by equation (3). Diagnostic reasoning has to be augmented by other symptoms. Much can be deduced from how these components change with operational parameters and rotational speed (which can be determined during machine run-down) [7]. Correlation analysis (see below) can also yield conclusive results.

Blade components result from interaction between fluid-flow system components and medium flow. Frequencies of these components can be expressed by

$$f_w = l \cdot u \tag{5}$$

$$f_k = z \cdot u \tag{6}$$

$$f_{(k+w)/2} = (z + l) \cdot \frac{u}{2} \tag{7}$$

$$f_{(k-w)/2} = (z - l) \cdot \frac{u}{2} \tag{8}$$

where l and z denote number of blades in rotor stages and bladed diaphragms, respectively. Components with frequencies given by equations (7) and (8) result

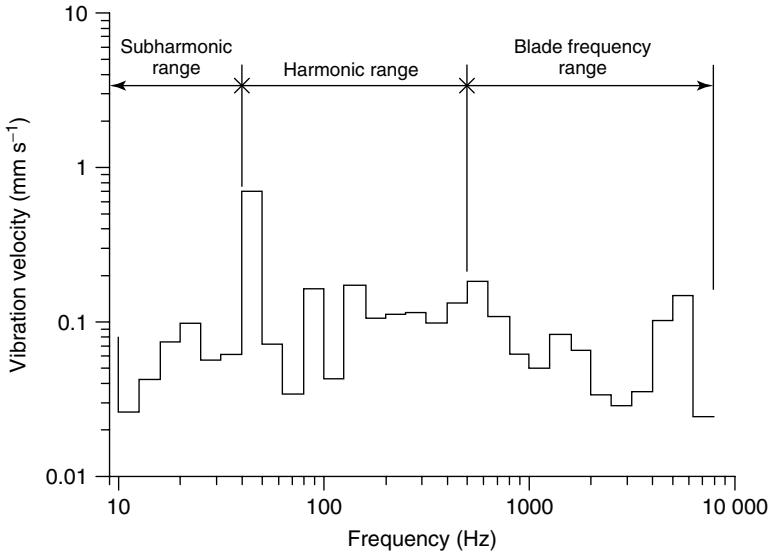


Figure 1. Example of a rotating machine with absolute vibration at the third octave (23% constant percentage bandwidth (CPB)) spectrum (200-MW steam turbine). A logarithmic horizontal scale is convenient because of the broad frequency range. Spectral analysis is dealt with in detail in **Statistical Time Series Methods for SHM**.

from interaction between adjacent stages and are in a way analogous to the treatment of nonlinear vibration without damping, with two harmonic components of exciting force [12]. Amplitudes of these components depend on the condition of fluid-flow system elements [10, 13] and can be employed as diagnostic symptoms.

Vibration trajectory or orbit represents a path of the shaft centerline in a plane perpendicular to the rotor [7]. In large rotating machines, relative vibrations of shafts in bearings are usually measured in two perpendicular directions (results are sometimes represented as vectors). For safety reasons, relative vibration amplitudes have to be kept within acceptable limits, determined by the machine manufacturer, in all operating conditions (including passing critical rotational speed or speeds during start-up). Vibration orbits are, however, much more useful as a diagnostic tool, as information is contained not only in amplitudes, but also orbit shape and direction of precession, as well as shaft centerline position as a function of operational parameters (rotational speed, load, temperature field, etc.). Normally, an orbit is elliptical with a single phase marker, which indicates domination of the f_0 component [7, 8].

Analysis of relative vibration orbits can yield conclusive results for malfunctions like rotor instability, rotor rub, excessive radial load, and misalignment. This is especially valuable for distinguishing malfunctions that produce similar absolute vibration patterns. In practice, analysis of absolute vibration is more convenient for determining amplitude–frequency characteristics. On the other hand, distorted orbits (banana-shaped, flattened, bent, or showing multiple loops) are indicative of certain malfunctions.

Information on machine condition is contained not only in absolute and relative vibration patterns, but also in vibration time histories (evolution). As mentioned above, slow monotonic increase is typical for normal life consumption in machines designed for long service life. Malfunctions result in "abnormal" evolution (Table 1) [5, 14]. Both evolution type and time constant (duration) provide information on malfunction type. It should be noted that in large rotating machines, time constants can vary from seconds to months or even years (Figure 2).

Trend analysis aimed at tracing life consumption processes can be performed only if available data cover a sufficiently long period. Such an approach can, however, be very useful if control and interference cannot be neglected (equation 4). In most cases, we can reasonably assume that components of **R** and **Z** vectors have no long-time trend; so for ith component we have

$$\frac{\Delta R_i}{\Delta\theta} = \frac{[R_i(\theta_0 + \Delta\theta) - R_i(\theta_0)]}{\Delta\theta} \to 0 \text{ if } \Delta\theta \to \infty \tag{9}$$

Table 1. Symptom assessment: example for absolute vibration evolution in a machine lifetime

No	Type of evolution	Duration (time constant)	Damage or malfunction
1	Simple, monotonic	1 day or more	Deformation of machine casing
		Several minutes to several hours	Shaft deformation, thermal unbalance
		A few seconds	Dry friction
2	Complex, approximately monotonic	Days to months	Deformations of casings and foundations
		Hours to days	Changes of natural frequencies
		Several minutes to hours	Thermal unbalance
3	Rapid, discontinuous	Below 2 s	Blade loss or damage, cracks of rotor elements
4	Exponential	Hours to weeks	Fatigue cracks
		10–20 min	Dry friction
5	Cyclic or diverging	Varying	Sliding friction (e.g., generator seals)
		A few seconds	Malfunction of gears
6	Cyclic, complex	Cyclic	Malfunctions of governing system
7	Rapid, random	Varying	Bearing instability, unstable medium flow

After [14], reproduced by permission of Queens University, Kingston.

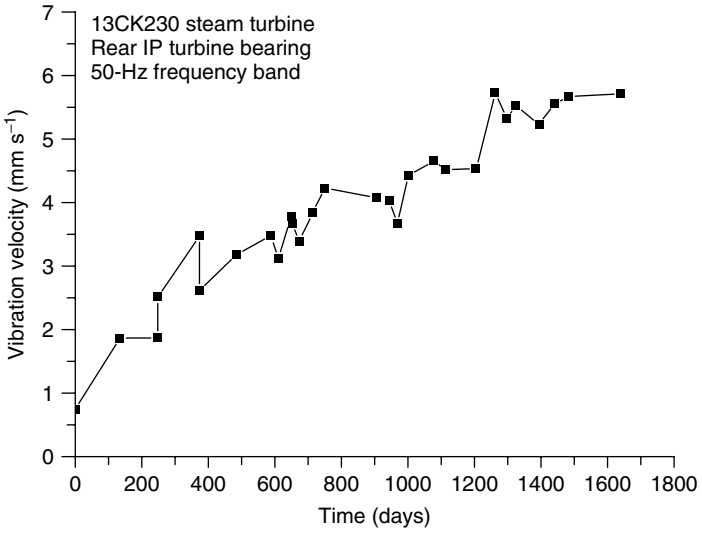


Figure 2. Example of a vibration trend analysis; slow increase of the $1 \times f_0$ component of absolute vertical vibration of a turbine bearing, caused by a slowly developing rotor bow.

$$\frac{\Delta Z_i}{\Delta\theta} = \frac{[Z_i(\theta_0 + \Delta\theta) - Z_i(\theta_0)]}{\Delta\theta} \to 0 \text{ if } \Delta\theta \to \infty \tag{10}$$

Thus, if we analyze an entire time history to extract a trend, we may neglect dependence of **S** on **R** and **Z**, provided that this time history covers a period ($\Delta\theta$) that is sufficiently long.

2.2 Other condition symptoms

Symptoms other than vibration-based ones are often provided by physical quantities that are monitored during machine operation, for either safety or performance control reasons. They are often useful in augmenting diagnostic reasoning. Their interpretation is in many cases quite straightforward.

2.2.1 Acoustic symptoms

Large rotating machines are usually operated in an environment characterized by high acoustic background level. This implies that interference (equation 4) markedly influences measured symptom values. Correlation between vibration-based and acoustic symptoms (sound intensity levels in individual frequency bands) in the harmonic frequency range is in many cases good [15], but the former are much more convenient from the point of view of measurement and interpretation. Acoustic symptoms have certain potential for detecting faults of control valves and labyrinth seals, but their application is in practice often limited to organoleptic noise level assessment, employing no measuring devices and based on individual experience.

2.2.2 Temperature

In thermal rotating machines (steam or gas turbines), the temperature at certain points is monitored continuously. This refers to bearings and casings. Increase of bearing temperature (and outlet oil temperature) is a symptom of excessive bearing load, usually caused by misalignment. Such malfunction results from faulty assembly, bearing pedestal displacement during start-up, or foundation distortion. Symptoms of this type can be useful if interpretation of vibration-based misalignment symptoms is unclear.

Monitoring casing temperatures is important, especially in steam turbines. High-pressure and intermediate-pressure casings are thick-walled elements and temperature gradients during transients (especially start-ups) must be kept within acceptable limits, to avoid excessive stresses. This limitation is imposed by both safety and life consumption reasons. Time histories of temperature in individual points and

temperature differences, especially between upper and lower casing parts, allow to detect water ingress into turbine, which results in intensive cooling of the lower part, large deformation, and usually, rotor jam. Interpretation of such symptoms is usually straightforward.

2.2.3 Displacement

During transient states (especially start-ups), axial displacement is monitored to keep clearances within acceptable limits. This is especially important in large steam turbines operating at high steam temperature, as thick-walled casings heat up much slower than rotors. Normally, displacements measured at individual points result directly from thermal expansion and/or axial forces imposed by medium flow. Substantial discrepancy is a symptom of expansion hindrance (e.g., due to slide block jam). Such symptoms are usually quite easy to interpret.

Radial (vertical and/or horizontal) displacement of machine elements may result in shaft line distortion that sometimes significantly influences dynamic behavior. This is particularly important in large machines with long shaft lines (in large steam turbines, even about 70 m). Excessive distortion usually leads to misalignment and problems with load distribution between bearings. This can be caused, e.g., by improper temperature field in foundation and support structure. Misalignment and load bearing problems can usually be identified with vibration-based symptoms [9]; however, to perform necessary readjustment of bearing positions, displacement measurements may be needed [16]. They are usually performed during start-ups and should cover the entire period from preliminary heating to steady-state operation, which in the largest machines can take several days. Figure 3 shows difference in the vertical displacements of two adjacent bearings in a turbine-generator unit, plotted against time [10]; in this case, the difference is rather large, but stabilization is achieved in a comparatively short time. Such measurements are sometimes time-consuming and tedious (especially if laser methods cannot be employed), but the interpretation of results is straightforward.

2.2.4 Electric parameters

Many large rotating machines include either electric motors (pumps, fans, and compressors) or generators (steam- or gas-turbine driven power generating units). Electric parameters, mainly voltage and current, are measured for both performance monitoring and operational safety reasons. They also provide input data for torsional vibration monitoring

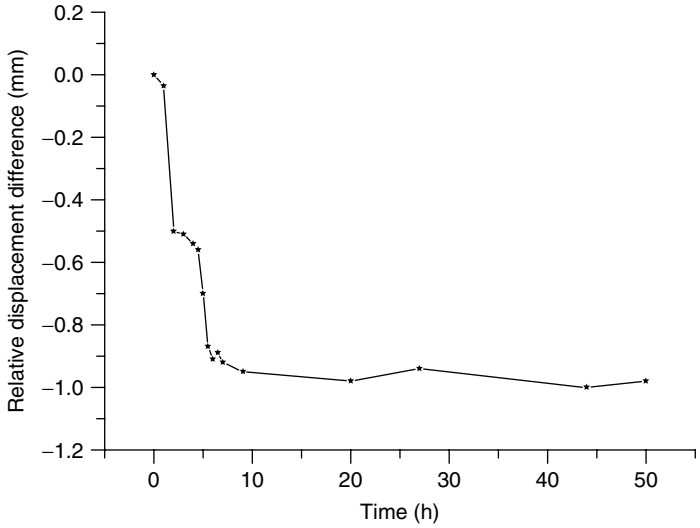


Figure 3. Relative difference in vertical displacements between rear turbine and front generator bearings plotted against time; developing misalignment is clearly visible [after [10], reproduced by permission of Polish Academy of Sciences].

systems in turbogenerators [17] and can be useful in identifying causes of failures resulting from accelerated fatigue.

In electric motors, failures like short-circuits or broken rotor bars produce uneven temperature distribution, which can lead to vibration patterns resembling unbalance [7]. In such cases, measurements of electric parameters may yield conclusive results.

2.2.5 Others

In steam and gas turbines, endoscopy is widely used for determining the condition of rotor blades [18]; state-of-the-art methods employ advanced image acquisition and processing techniques.

Large steam turbines are fitted with on-line or off-line systems for monitoring temperature, pressure, and steam flow. These quantities are necessary for energy conversion diagnostics, i.e., determining operational parameters like specific heat consumption and thermal efficiency [19]. They can also be used to monitor degradation of turbine fluid-flow system.

In compressors and pumps, pressure measurements are fundamental for performance monitoring. They are also important for operational safety (avoiding unstable operation conditions—stall or surge), which is particularly important with high flow rates and high-pressure ratios [20]. These quantities are of supplementary nature in condition assessment based on other symptoms.

Methods employing acoustic emission, widely employed in nondestructive testing, have not found widespread use in the monitoring of large rotating machines, despite their considerable potential [21].

3 DIAGNOSTIC RELATIONS

Development of structural health monitoring in large rotating machines has so far included four stages, namely,

1. symptom acquisition (measurement of physical quantities and data processing);
2. qualitative diagnosis (fault detection and identification);
3. quantitative diagnosis (fault extent or, more generally, life consumption determination);
4. prognosis (residual life determination).

Diagnostic relations can thus be either qualitative or quantitative.

In general, qualitative relations are of binary type: object condition is either "good" or "faulty"; symptom is either present or not present. For complex objects, like large rotating machines, such an approach is adequate only in few cases. Quantitative relations are provided by symptom value versus failure advance or life consumption, both members being continuous functions.

Diagnostic experiments of active type, involving large rotating machines are seldom possible; determination of diagnostic relations has therefore to be based on models. At first, these models were developed directly from an analysis of processes that are sources of diagnostic symptoms. Such an approach is exemplified by vibrodiagnostic models [3, 10, 22] that, in general, relate vibration components in individual frequency bands to elementary vibration sources. These models were initially of qualitative type. With the introduction of criteria values (see below), they were supplemented with a tool for quantitative condition assessment.

Because of complexity, interpretation of symptoms is often equivocal, i.e., the same symptom may result from different malfunctions or faults. $1 \times f_0$ component of absolute vibration is perhaps the most representative example (see Section 2.1). Thus, condition assessment on the basis of such models inevitably involves a measure of correct diagnosis probability. Only in some specific cases (e.g., detection of fluid-induced instability on the basis of subharmonic spectral components) can this probability be very high. Substantial improvement can be achieved if two or more symptoms appear simultaneously; in more general terms, this means employing correlation. A good example is provided by misalignment detection (three symptoms—increase of the $2 \times f_0$ component in absolute vertical vibration spectra, banana-shaped shaft orbit, and oil temperature change at the bearing outlet). Correlation of vibration time histories in various machine points can also be a useful symptom [11].

Much effort has been spent on modeling dynamic behavior of rotors and bearings [7, 8, 23, 24]. Finite elements method (FEM) is used to model the line of rotors. Various methods have been developed for modeling bearings and supporting structures (bearing pedestals, foundations, etc.). Even cursory description

of these complex issues is certainly beyond the scope of this article; from the point of view of condition assessment, this approach can be summed up as follows:

- the model is developed for a particular machine (rotor or line of rotors, bearings, supporting structure);
- certain imperfections that represent "real" malfunctions (like unbalance, misalignment, rotor or blade crack, bearing displacement, or skew) are introduced in the model;
- dynamic response of the model to these imperfection is determined (vibration spectra in individual measuring points, shaft orbits in bearings);
- results are validated in laboratory conditions.

Validation on full-scale rotors is seldom possible and scaled-down laboratory systems are used for this purpose [8].

Modal analysis provides a valuable tool for detecting failures that modify natural frequencies, in particular shaft cracks, which are among the most dangerous ones. It has been shown that transverse shaft cracks modify natural frequencies, but resulting changes are rather small, typically below 1% [25]. Experimental investigations [26] have pointed at a need for advanced measurement and signal-processing techniques to reveal corresponding spectral components. These methods have not yet found widespread applications, but they have a great potential; much work is currently under way. Modal vibration methods are dealt with in detail in **Modal–Vibration-based Damage Identification**.

3.1 Normalization of symptom values

In the above considerations, it has been assumed, more or less explicitly, that relation between symptoms and condition parameters given by equation (1) is valid. Usually, however, equation (4) is far more appropriate, which means that influences of control and interference have to be taken into account. In fact, this is the principal cause of difficulties in direct application of model-based simulation or laboratory-scale experiment results to real machines, operating in an industrial plant environment.

Normalization of the influence of interference vector components is, in general, not possible, as some of them are not measurable. Proper measurement procedures can alleviate the problem, but not solve it completely. Normalization of control parameters is usually a complex issue. Experimental investigations have shown that dependence of diagnostic symptom values on overall parameters, like machine load, cannot be neglected [27]; moreover, this dependence is strongly nonlinear. The problem is, in fact, even more complicated. Consider a given moment $\theta = \theta_0$. We have

$$P(\theta_0) = F_1[\mathbf{R}(\theta_0)] \tag{11}$$

where P is a parameter that describes "intensity" of operation (load, discharge, mass flow, etc.). At the same time, for a given symptom S_i, we have

$$S_i(\theta_0) = F_2[\mathbf{R}(\theta_0)] \tag{12}$$

Usually it is convenient to speak in terms of $S_i(P)$ relations, which are comparatively simple to determine experimentally [25]. However, the function given by equation (11) is usually unique, but not single, as various $\mathbf{R}$ vectors can yield the same value of P. Any normalizing function of the $S_i(P)$ type will thus be only an approximation. This complicates modeling, as many additional factors have to be accounted for.

Many faults produce significant changes of the symptom vector components, so qualitative diagnosis can be provided even despite these influences. Correlation analysis can also be very useful. In tracking long-time trends, related either to normal life consumption or slowly developing faults (Table 1), equations (9) and (10) are essential; usually there are considerable fluctuations, but "net" increase can be observed if a sufficiently long period is covered (Figure 4).

3.2 Symptom limit values

In practice, a decision has often to be made whether a machine can be further operated or not, and diagnostic procedures are expected to provide arguments. This is particularly important for critical machines. In a sense, this can be viewed as a return to the "binary approach" and thus some threshold symptom value has to be determined. This can be based on the analysis of energy transformation and dissipation processes, which leads to the energy processor

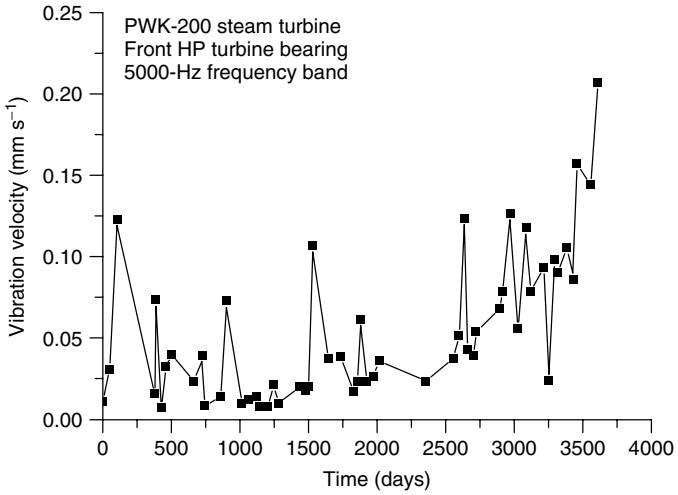


Figure 4. Example of a vibration trend; despite fluctuations, an increasing trend is clearly visible.

(EP) model [4, 28]. Application of the statistic decision theory and symptom reliability concept [4, 29] leads to so-called symptom limit value. Limit value (in some monitoring systems referred to as *alert value*) is defined as one corresponding to condition deterioration to a point beyond which the object cannot fulfill all requirements and should be considered faulty. This can be expressed as

$$S_i(\theta) < S_l \Rightarrow \text{GOOD condition}$$
$$S_i(\theta) \geq S_l \Rightarrow \text{FAULTY condition} \qquad (13)$$

where S_l denotes the limit value. At this point, the decision has to be made whether to operate the machine at reduced parameters (e.g., lower load or flow rate), reschedule overhauls, or perform a repair. Such an approach can be schematically illustrated by Figure 5.

The basic value represents a new object with no faults or malfunctions and, in theory, is the lowest possible value of the symptom under consideration. The admissible value is determined from safety considerations (usually by the machine manufacturer)

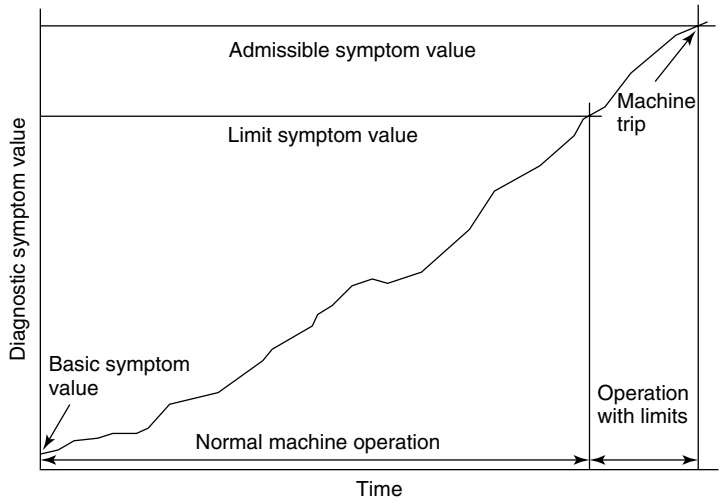


Figure 5. Machine condition assessment on the basis of criteria symptom values.

or imposed by relevant standards and basically is not related to condition assessment. This value should not be exceeded during steady-state operation; if this happens, the machine has to be tripped. In some monitoring systems, this value is referred to as a *danger* level.

In general, a limit value can be attributed to any diagnostic symptom, but in monitoring systems, such values are determined mainly for vibration-based symptoms [30]. For many large rotating machines, basic and limit values have to be determined individually, even for identical objects; steam turbines provide a good example [30].

4 MONITORING PHILOSOPHY

Typically, a large rotating machine will have a system for measuring parameters important for process control and operational safety. Some of these parameters also provide information important for technical condition assessment, but this task will usually be accomplished by a separate monitoring/diagnosing system. In general, two approaches can be distinguished: continuous (on-line) and periodic (off-line) monitoring. The choice will always be a compromise between demands and costs.

Since introduction in late 1960s, on-line monitoring of large rotating machines has evolved from comparatively simple systems for data acquisition, alarms generation, and protection to complex expert systems featuring leading-edge data processing techniques and AI methods. Several off-the-shelf systems from various manufacturers are available at the market, which facilitate various degrees of versatility and can be usually tailored to a specific application. Despite their trade names, in most cases, they feature only data acquisition, processing, and storage, with few diagnostic capabilities, if any. Exemplary layout of one such system is shown in Figure 6.

In many applications (e.g., in mining industry), several machines at various, often distant, locations have to be monitored simultaneously, which imposes high demands on data transmission. Owing

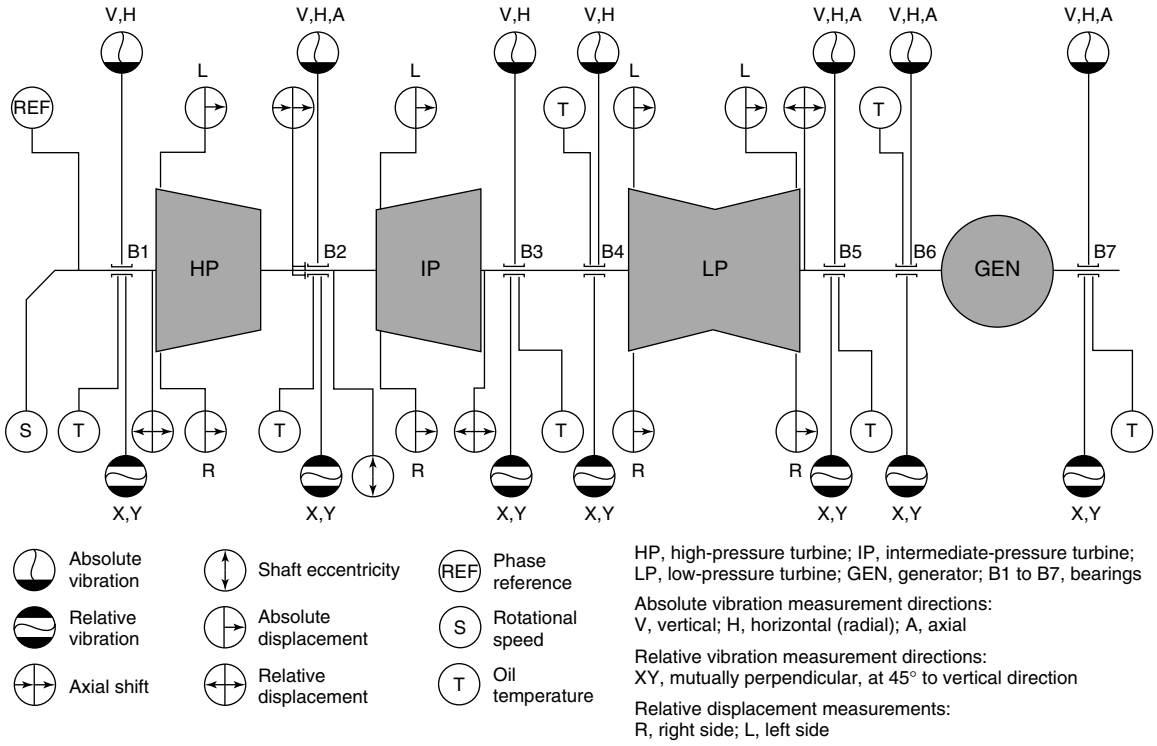


Figure 6. Example of a monitoring system layout for a 200-MW steam turbine.

to the large amount of data acquired from individual sensors, data compression can be of great importance. An example of such a distributed system is shown in Figure 7.

Advanced diagnostic systems, which incorporate expert systems, knowledge base, and database from both modeling and operational experience and employ AI methods are developed individually for particular machine types. The first systems of this kind were developed in the 1980s for large steam turbines in nuclear power plants. These systems are very costly, so their application is justified only for large machines with high reliability demand and extremely severe consequences of potential failure.

Much cheaper off-line systems are based on measurements performed at certain time intervals. They obviously cannot provide instantaneous reaction for a failure characterized by a very short "time constant" (Table 1), but have a number of advantages:

- array of measuring points and measurement settings can be adjusted to current machine status;
- there is some choice of transducers with characteristics usually superior to that of permanently installed ones (for example, vibration transducers permanently installed on rotating machines seldom exceed frequency range of 1 kHz, which excludes blade frequency range);
- measuring equipment and software can be easily upgraded as new versions become available.

Off-line monitoring/diagnosing systems use commercial (usually portable) equipment for data acquisition and various software packages for post-processing and supporting diagnostic reasoning. Such

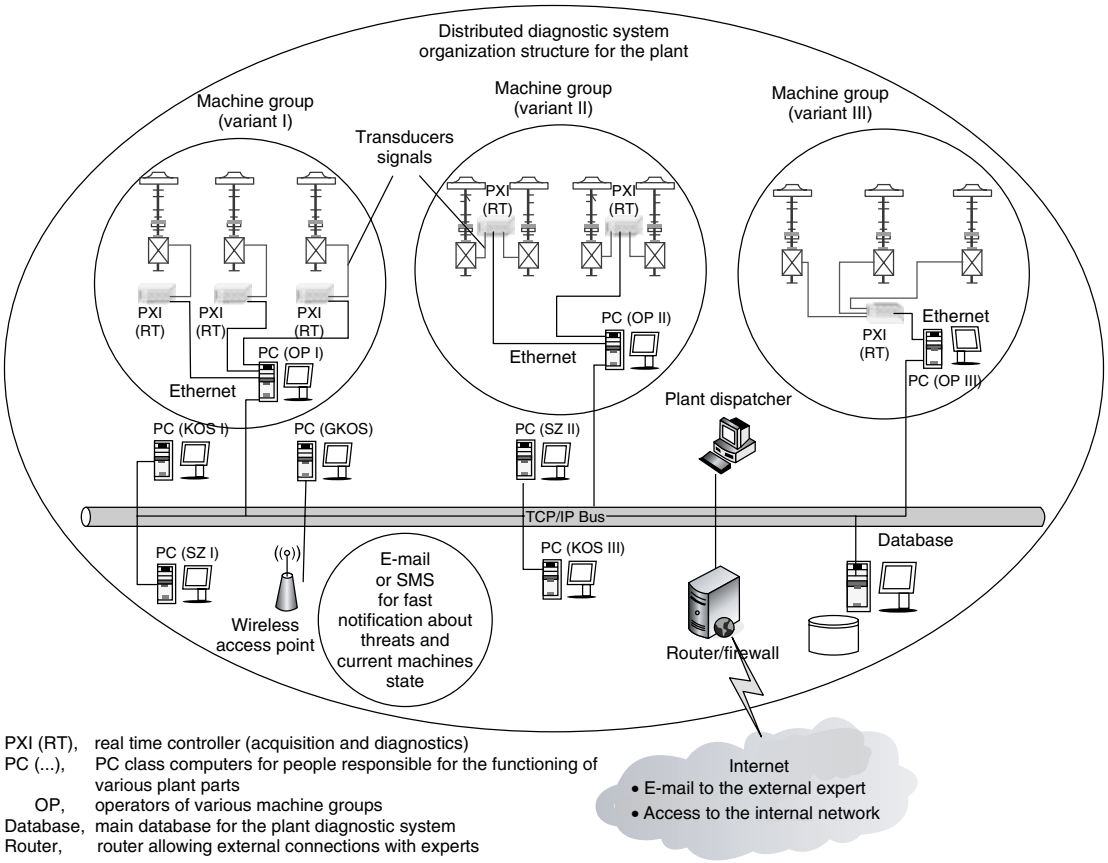


Figure 7. Exemplary structure of a plant distributed diagnostic system [after [31], reproduced by permission of Instituto de Telecomunicações, Coimbra].

an approach, combined with proper on-line operation monitoring, can be a good and cost-effective alternative to expensive on-line systems.

5 FUTURE DIRECTIONS

Structural health monitoring should be seen in the widest context of machine—or plant—maintenance improvement [32]. Certainly, the ultimate and ideal aim is to eliminate all unplanned maintenance [33]. In reality, there will always be a compromise between some acceptable level of failure risk on one hand and some acceptable percentage of unnecessary maintenance on the other, with all associated costs taken into account. This is exemplified by the Neyman–Pearson rule of risk assessment [4], which itself forms the basis for symptom limit value determination [30]. Improvement in this field calls for a system that is able to give a reliable prognosis or, more specifically, reliable residual life estimation. This means taking the next step in the development of condition monitoring (see Section 3).

State-of-the-art systems for structural health monitoring have reached an advanced stage of development as far as damage identification and quantitative assessment are concerned. On the other hand, still relatively few systems feature forecasting of technical condition evolution. Owing to the complexity of symptom–condition relations (see equation (4)), simple prediction methods are not adequate and much development is necessary. Given the demands imposed by the need for overall maintenance improvement, this will be an important, and probably the most important, trend in this field in the near future.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to Professor Czesław Cempel for encouragement and discussion, which has helped me much in giving this text its final form.

END NOTES

[a.] Lifetime is here and in the following denoted by θ rather than t to avoid confusion with so-called dynamic time, which is used in equations of motion (*see* **Free and Forced Vibration Models**).

REFERENCES

[1] Kanki H, Yasuda C, Umemura R, Itoh C, Miyamoto C, Kawaguchi T. Vibration diagnostic expert system for steam turbines. *Proceedings of the CISM/IFToMM Symposium 'Diagnostics of Rotating Machines in Power Plants'*. Springer: Wien, New York, 1994; pp. 25–36.

[2] Cempel C. Multidimensional condition monitoring of mechanical systems in operation. *Mechanical Systems and Signal Processing* 2003 **17**(6): 1291–1303.

[3] Orłowski Z. A model for operational diagnostics of steam turbines. *Mechanical Systems and Signal Processing* 1995 **9**(2):215–222.

[4] Natke HG, Cempel C. *Model-Aided Diagnosis of Mechanical Systems*. Springer: Berlin, Heidelberg, New York, 1997.

[5] Morel J. *Vibration des Machines et Diagnostic de leur État Mécanique*. Eyrolles: Paris, 1992.

[6] Südmersen U, Pietsch O, Liu Y, Reimche W, Stegemann D. Advanced condition monitoring of power plants by vibration analysis. *Proceedings of the 3rd International Conference on Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques*. CETIM: Senlis, 1998; pp. 97–107.

[7] Bently DE, Hatch CT. *Fundamentals of Rotating Machinery Diagnostics*. Bently Pressurized Bearing Press: Minden, 2002.

[8] Kiciński J. *Rotor Dynamics*. Institute of Fluid-Flow Machinery, Polish Academy of Sciences: Gdańsk, 2006.

[9] Orłowski Z, Gałka T. Cavitation et vibrations autoexcitées dans une turbine á contre-pression. *Proceedings of the S.F.M. Congress*. Société Française de Mécaniciens: Courbevoie, 1995; pp. 513–522.

[10] Orłowski Z. Diagnostics of steam turbines. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 2001 **49**(2):233–247.

[11] Gałka T. Higher harmonic components in steam turbine vibration velocity spectra: a case study. *Proceedings of the 20th International Congress COMADEM'07*, Faro. Instituto de Telecomunicações: Coimbra, 2007; pp. 351–359.

[12] Magnus K. *Schwingungen: Eine Einführung in die Teoretische Behandlung von Schwingungproblemen*. B.G.Teubner: Stuttgart, 1976.

[13] Gałka T. Assessment of the turbine fluid-flow system condition on the basis of vibration-related symptoms. *Proceedings of the 16th International Congress COMADEM 2003*. Växjö University Press: Växjö, 2003; pp. 165–173.

[14] Orłowski Z, Gałka T. Assessment of vibrodiagnostic symptoms: an example for a 360 MW steam turbine. *Proceedings of COMADEM 95 International*. Queens University: Kingston, 1995; pp. 673–680.

[15] Orłowski Z, Gałka T. Application of noise analysis in power plant turbine diagnostics. *Proceedings of INTER-NOISE'99 Conference*, Fort Lauderdale, FL. INCE: Poughkeepsie, 1999; pp. 1075–1078.

[16] Lapini GL, Rossini T, Gadda E, Benanti A. *In-Service Measurements of Turbo-Generator Bearing Vertical Misalignment*. CISE Report, Milano, 1984.

[17] Chow JH, Javid SH, Sanchez-Gasca JJ, Bowler CEJ, Edmonds JS. Torsional model identification for turbine generators. *IEEE Transactions on Energy Conversion* 1986 **EC-1**(4):83–91.

[18] Charchalis A. Applications of diagnosing of naval gas turbines. *Proceedings of the 14th International Congress COMADEM 2001*. Elsevier: Oxford, 2001; pp. 489–494.

[19] Głuch J. The advantages of off-line thermal and flow diagnostics in power industry. *Proceedings of the 7th European Conference on Turbomachinery*. ETC: Athens, 2007; pp. 543–552.

[20] Botros KK, Henderson JF. Developments in centrifugal compressor surge control—a technology assessment. *ASME Journal of Turbomachinery* 1994 **116**:240–249.

[21] Holroyd TJ, Randall N. Field application of acoustic emission to machinery condition monitoring. *Proceedings of the 5th International Congress COMADEM 1993*. University of the West of England: Bristol, 1993; pp. 217–222.

[22] Orłowski Z, Gałka T. Modeling of the steam turbine fluid-flow system for technical condition assessment purposes. *Applied Mechanics in the Americas, Vol.9: Proceedings of the 7th Pan American Congress of Applied Mechanics*. AAM/Universidad de la Frontera, 2002, pp. 557–560.

[23] Bachschmid N, Pennacchi P. Model based malfunction identification from bearing measurements. *Proceedings of the 7th International Conference on Vibrations in Rotating Machinery*. Nottingham, 2000; pp. 571–580.

[24] Markert R, Platz R, Siedler M. Model based fault identification in rotor systems by least square fitting. *Proceedings of the ISROMAC-8 Confenerce*. ISROMAC: Honolulu, 2000; pp. 901–915.

[25] Gałka T. Normalization of vibration measurements: unnecessary complication or important prerequisite? *Proceedings of the Second International Symposium on Stability Control of Rotating Machinery ISCORMA-2*. Bently Pressurized Bearing Press: Gdańsk, 2003; pp. 722–731.

[26] Sol JC. *Vibration d'une Structure Comportant une Fissure*, EdF Report P34–185. Electricité de France, 1980.

[27] Orłowski Z, Gałka T. Application of modal analysis for steam turbine diagnostics. *Proceedings of COMADEM'96*. Sheffield University Press: Sheffield, 1996; pp. 889–898.

[28] Cempel C, Natke HG. The modeling of energy transforming and energy recycling systems. *Journal of Systems Engineering* 1996 **6**:79–88.

[29] Cempel C. Limit value in practice of vibration diagnosis. *Mechanical Systems and Signal Processing* 1990 **4**(6):483–493.

[30] Gałka T. Application of energy processor model for diagnostic symptom limit value determination in steam turbines. *Mechanical Systems and Signal Processing* 1999 **13**(5):757–764.

[31] Mäczak J. Structure of distributed diagnostic systems as a function of particular diagnostic task. *Proceedings of the 20th International Congress COMADEM'07*, Faro. Instituto de Telecomunicações: Coimbra, 2007; pp. 617–624.

[32] Tsang AHC. A strategic approach to managing maintenance performance. *Journal of Quality in Maintenance Engineering* 1998 **4**(2):87–94.

[33] Crocker J. Prognostics in aero-engines. *Proceedings of the 16th International Congress COMADEM 2003*. Växjö University Press: Växjö, 2003; pp. 145–154.

Chapter 150

Prognostics and Health Management of Electronics

Michael Pecht

Center for Advanced Life Cycle Engineering (CALCE), University of Maryland, College Park, MD, USA

1 INTRODUCTION TO THE PREDICTION OF RELIABILITY

The traditional reliability prediction methods for electronic products include Mil-HDBK-217 [1], 217-PLUS, Telcordia [2, 3], PRISM [2], and FIDES [4]. All these methods assume the components of the system have constant failure rates with "pi-factor" modifiers to account for various quality, operating, and environmental conditions. There are numerous problems with these types of modeling approaches, and these have been mentioned in hundreds of papers [5, 6]. The general consensus has been that these methods should never be used, because they are inaccurate for predicting actual field failure events, and they are highly misleading, and can result in poor designs [3].

In the Mil-HDBK-217A documentation published in 1965, there was only a single point failure rate for all monolithic integrated circuits, regardless of the stresses, the materials, or the architecture. Mil-HDBK-217B was published in 1973, with the RCA/Boeing models simplified by the Air Force to follow a exponential distribution. In 1979, Mil-HDBK-217C was published to "band-aid" the problems. To keep pace with the accelerating and ever changing technology base, Mil-HDBK-217C was updated to Mil-HDBK-217D in 1982 and to Mil-HDBK-217E in 1986. In 1991, Mil-HDBK-217F became a prescribed US military reliability prediction document [7]. In the meantime, IEEE 1413 standard provided guidance on the appropriate elements of a reliability prediction [8]. The IEEE 1413.1 guidebook provides a summary of the evaluation of the common methods of reliability prediction described in this document. That information should be utilized

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

for determining which reliability prediction method is appropriate in a particular application [9].

The physics-of-failure (PoF) approach and design-for-reliability (DfR) methods have been developed by the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland with the support of industry, government, and other universities. PoF is an approach that utilizes knowledge of a product's life-cycle loading and failure mechanisms to perform reliability modeling, design, and assessment. The approach is based on the identification of potential failure modes, failure mechanisms, and failure sites for the product at a particular life-cycle loading condition. The stress at each failure site is obtained as a function of both the loading conditions and the product geometry and material properties. The use of PoF modeling approaches for electronic components and devices, like those used for mechanical systems, are a powerful tool in support of electronic prognostic capabilities. This is because the root cause of almost all electronic devices or component failures is often mechanical—something physically breaks at a subcomponent, solder joint, connection, layer, delamination, etc., level. Solder fatigue models are already under development and show promise [10, 11].

Prognostics and health management (PHM) is a method that permits the assessment of the reliability of a system under its actual application conditions. When combined with PoF models, it is thus possible to make continuously updated predictions based on the actual environmental and operational condition monitoring of each individual product.

Assessing the extent of deviation or degradation from an expected normal operating condition (i.e., health) for electronics provides data that can be used to meet several critical goals, which include (i) providing advance warning of failures; (ii) minimizing unscheduled maintenance, extending maintenance cycles, and maintaining effectiveness through timely repair actions; (iii) reducing the life-cycle cost of equipment by decreasing inspection costs, downtime, and inventory; and (iv) improving qualification and assisting in the design and logistical support of fielded and future systems [12, 13]. The importance of PHM has been explicitly stated in the US Department of Defense 5000.2 policy document on defense acquisition, which states that "program managers shall optimize operational readiness through affordable, integrated, embedded diagnostics and prognostics, embedded training and testing, serialized item management, automatic identification technology, and iterative technology refreshment" [14]. Thus, a prognostics capability has become a requirement for any system sold to the Department of Defense.

2 MODELING OF STRESS AND DAMAGE UTILIZING LIFE-CYCLE LOADS

Life-cycle loads of a product can arise from manufacturing, shipment, storage, handling, operating, and nonoperating conditions. The life-cycle loads (thermal, mechanical, chemical, electrical, and so on), either individually or in various combinations, may lead to performance or physical degradation of the product and reduce its service life. In the stress–damage prognostics approach, the extent and rate of product degradation depends upon the magnitude and duration of exposure to loads (usage rate, frequency, and severity). The life-cycle loads are monitored *in situ*, and used in conjunction with PoF-based damage models to assess the degradation due to cumulative load exposures.

In studies made by Ramakrishnan and Pecht [15] and Mishra *et al.* [16], the test vehicle consisted of an electronic component-board assembly placed under the hood of an automobile and subjected to normal driving conditions in the Washington, DC, area. The test board incorporated eight surface-mount leadless inductors soldered onto an FR-4 substrate using eutectic tin–lead solder. Solder joint fatigue was identified as the dominant failure mechanism. Temperature and vibrations were measured *in situ* on the board in the application environment. Using the monitored environmental data, stress and damage models were developed and used to estimate consumed life.

Shetty *et al.* [17] applied the PHM methodology for conducting a prognostic remaining-life assessment of the end effector electronics unit (EEEU) inside the robotic arm of the space shuttle remote manipulator system (SRMS). A life-cycle loading profile for thermal and vibration loads was developed for the EEEU boards. Damage assessment was conducted using physics-based mechanical and thermomechanical damage models. A prognostic estimate using

a combination of damage models, inspection, and accelerated testing showed that there was little degradation in the electronics and they could be expected to last another 20 years.

Mathew *et al.* [18, 19] applied the PHM methodology in conducting a prognostic remaining-life assessment of circuit cards inside a space shuttle solid rocket booster (SRB). Vibration time history recorded on the SRB from the prelaunch stage to splashdown was used in conjunction with physics-based models to assess the damage due to vibration and shock loads. Using the entire life-cycle loading profile of the SRBs, the remaining life of the components and structures on the circuit cards was predicted. It was determined that an electrical failure was not expected within another 40 missions.

Gu *et al.* [20] developed a methodology for monitoring, recording, and analyzing the life-cycle vibration loads for remaining-life prognostics of electronics. The responses of printed circuit boards (PCBs) to vibration loading in terms of bending curvature were monitored using strain gauges. The interconnect strain values were then calculated from the measured PCB response and used in a vibration failure fatigue model for damage assessment. Damage estimates were accumulated using Miner's rule after every mission and then used to predict the life consumed and remaining life. The methodology was demonstrated for remaining-life prognostics of a PCB. The result was also verified by the real time to failure of the components by checking the components' resistance data.

Simons and Shockey [21] performed a PoF-based prognostics methodology for failure of a gull-wing lead power supply chip on a DC/DC voltage converter PCB assembly. First, three-dimensional finite element analyses (FEA) were performed to determine strains in the solder joint due to thermal or mechanical cycling of the component. The strains could be due to lead bending resulting from the thermal mismatch of the board and chip and those resulting from local thermal mismatch between the lead and the solder, as well as between the board and the solder. Then the strains were used to set boundary conditions for an explicit model that could simulate initiation and growth of cracks in the microstructure of the solder joint. Finally, on the basis of the growth rate of the cracks in the solder joint, estimates were made of the cycles to failure for the electronic component.

Nasser and Curtin [22] applied PHM methodology to predict failure of the power supply. They subdivided the power supply into component elements based on specific material characteristics. Predicted degradation within any single or combination of component elements could be rolled up into an overall reliability prediction for the entire power supply system. Their PHM technique consisted of five steps: (i) acquiring the temperature profile using sensors; (ii) conducting FEA to perform stress analysis; (iii) conducting fatigue prediction of each solder joint; and (iv) predicting the probability of failure of the power supply system.

Searls *et al.* [23] undertook *in situ* environment loading, such as temperature measurements, in both notebook and desktop computers used in different parts of the world. In terms of the commercial applications of this approach, IBM has installed temperature sensors on hard drives (Drive–TIP) [24] to mitigate risks due to severe temperature conditions, such as thermal tilt of the disk stack and actuator arm, offtrack writing, data corruptions on adjacent cylinders, and outgassing of lubricants on the spindle motor.

Vichare *et al.* [25, 26] also conducted *in situ* health monitoring of notebook computers. The authors monitored and statistically analyzed the temperatures inside a notebook computer, including those experienced during usage, storage, and transportation, and discussed the need to collect such data both to improve the thermal design of the product and to monitor prognostic health. After the data was collected, it could be used to estimate the distributions of the load parameters. The usage history was used for damage accumulation and remaining-life prediction.

The European Union funded a project from September 2001 through February 2005 named *Environmental Life-cycle Information Management and Acquisition* (ELIMA) for consumer products, which aimed to develop ways of better managing the life cycles of products. It used technology to collect vital information during a product's life to lead to better and more sustainable products [27, 28]. The objective of this work was to provide a basic model for predicting the remaining lifetime of parts removed from products, based on the dynamic data collected by the ELIMA system. The ELIMA technology included sensors and memory built into the

product to record dynamic data such as operation time, temperature, and power consumption. This was added to static data about materials and manufacturing. As a case study, the member companies monitored the application conditions of a game console and a household refrigerator. The work concluded that for the remaining-lifetime prediction, in general, it was essential that the environments associated with all life intervals of the equipment be considered. These included not only the operational and maintenance environments but also the preoperational environments, when stresses imposed on the parts during manufacturing, assembly, inspection, testing, shipping, and installation might have a significant impact on the eventual reliability of the equipment. Stresses imposed during the preoperational phase were often overlooked.

Tuchband and Pecht [29] presented the use of prognostics for military line replaceable units (LRUs) based on their life-cycle loads. The study was part of an effort funded by the Office of the Secretary of Defense to develop an interactive supply chain system for the US military. The objective was to integrate prognostics, wireless communication, and databases through a web portal to enable cost-effective maintenance and replacement of electronics. This study showed that prognostics-based maintenance scheduling could be implemented into military electronic systems. The approach involves an integration of embedded sensors on the LRU, wireless communication for data transmission, a PoF-based algorithm for data simplification and damage estimation, and a method for uploading this information to the Internet. Finally, the use of prognostics for electronic military systems enabled failure avoidance, high availability, and reduction of life-cycle costs.

3 CANARY DEVICES APPROACH

Canary devices mounted on the actual product have been used to provide advance warning of failure due to specific wearout failure mechanisms. The word "canary" is derived from one of coal mining's earliest systems for warning of the presence of hazardous gas using the canary bird. Because the canary is more sensitive to hazardous gases than humans, the death or sickening of the canary was an indication to the miners to get out of the shaft. The same approach, using canaries, has been employed in PHM. Canary devices were integrated into a specific component, device, or system design and incorporated failure mechanisms that occur first in the embedded device. These embedded canary devices (also called *prognostics cell*) were noncritical elements of the overall design providing early incipient failure warnings before actual system or component failure [12].

Mishra and Pecht [30] studied the applicability of semiconductor level health monitors by using precalibrated cells (circuits) located on the same chip with the actual circuitry. The prognostics cell approach was commercialized by Ridgetop Group to provide an early warning sentinel for upcoming device failures [31]. The prognostic cells were available for 0.35, 0.25, and 0.18 μm CMOS processes. The time to failure of these prognostic cells could be precalibrated with respect to the time to failure of the actual product. The stresses that contributed to degradation of the circuit included voltage, current, temperature, humidity, and radiation. Since the operational stresses were the same, the damage rate was expected to be the same for both the circuits. However, the prognostic cell was designed to fail earlier owing to increased stress on the cell structure by means of scaling. For example, scaling could be achieved by controlled increase of the current density inside the cells. With the same amount of current passing through both circuits, if the cross-sectional area of the current-carrying paths in the cells was decreased, a higher current density was achieved. Both the structure and the loading could be scaled. Further control in current density could be achieved by increasing the voltage level applied to the cells. Higher current density led to higher internal heating, causing greater stress on the cells. When a current of higher density passed through the cells, they were expected to fail faster than the actual circuit [30]. Currently, prognostic cells are available for semiconductor failure mechanisms such as electrostatic discharge (ESD), hot carrier, metal migration, dielectric breakdown, and radiation effects.

The extension of this approach to board-level failures was proposed by Anderson and Wilcoxon [32], who created canary components (located on the same PCB) that include the same mechanisms that lead to failure in actual components. Anderson *et al.*, identified two prospective failure mechanisms: (i) low

cycle fatigue of solder joints, assessed by monitoring solder joints on and within the canary package; and (ii) corrosion monitoring using circuits that are susceptible to corrosion. The environmental degradation of these canaries was assessed using accelerated testing, and degradation levels were calibrated and correlated to actual failure levels of the main system.

Goodman *et al.* [33] used a prognostic cell to monitor time-dependent dielectric breakdown (TDDB) of the metal-oxide semiconductor field-effect transistor (MOSFET) on the integrated circuits. The prognostic cell was accelerated to failure under certain environmental conditions. Acceleration of the breakdown of an oxide could be achieved by applying a voltage higher than the supply voltage, to increase the electric field across the oxide. When the prognostics cell failed, a certain fraction of the circuit lifetime was used up. The fraction of consumed circuit life was dependent on the amount of overvoltage applied and could be estimated from the known distribution of failure times.

4 PoF-BASED PHM IMPLEMENTATION APPROACH

The general PHM methodology is summarized in CALCE and shown in Figure 1. It includes two main parts: virtual life assessment and real prognostics assessment. Design data, expected life cycle, failure modes, mechanisms, and effects analysis (FMMEA), and PoF models can be the input for the virtual life assessment to get a better reliability assessment based on the historical and current data. The next step is to get the *in situ* sensor data, bus monitor data, diagnostics data, and maintenance records for prognostics assessment in real product life cycles. Three methodologies for *in situ* prognostics assessment include (i) the use of expendable devices, such as "canaries" and fuses that fail earlier than the host product to provide advance warning of failure; (ii) the monitoring of parameters and the extraction of features that are precursors to impending failure [34]; and (iii) the use of stress and damage models that employ life-cycle loads (e.g., usage conditions, temperature, vibration, radiation). Approaches 1 and 3 are PoF related [16], so in this article, the focus is on detailed PoF-based prognostics, which is discussed in a later paragraph.

The PoF methodology is founded on the premise that failures result from fundamental mechanical, chemical, electrical, thermal, and radiation processes. The objective of the PoF methodology in the PHM process is to calculate the cumulative damage accumulation due to various failure mechanisms for a product in a given environment. The approach to implement PoF into PHM can be based on the FMMEA, which is shown in Figure 2. This approach

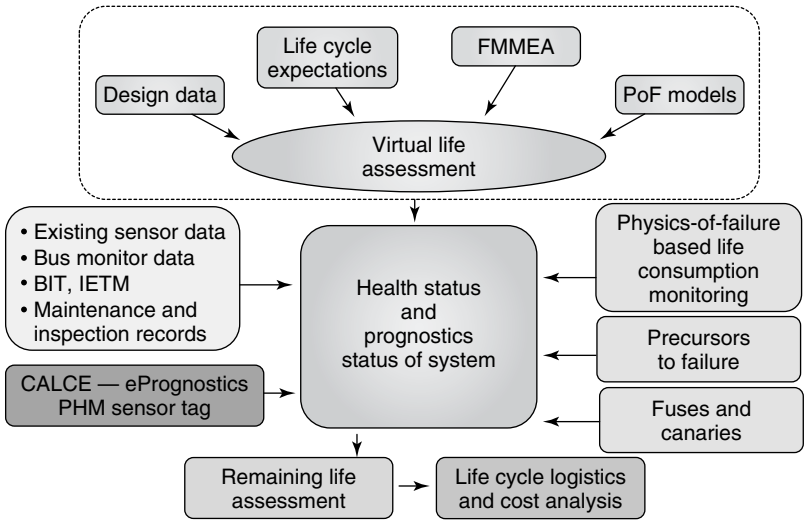


Figure 1. CALCE PHM methodology.

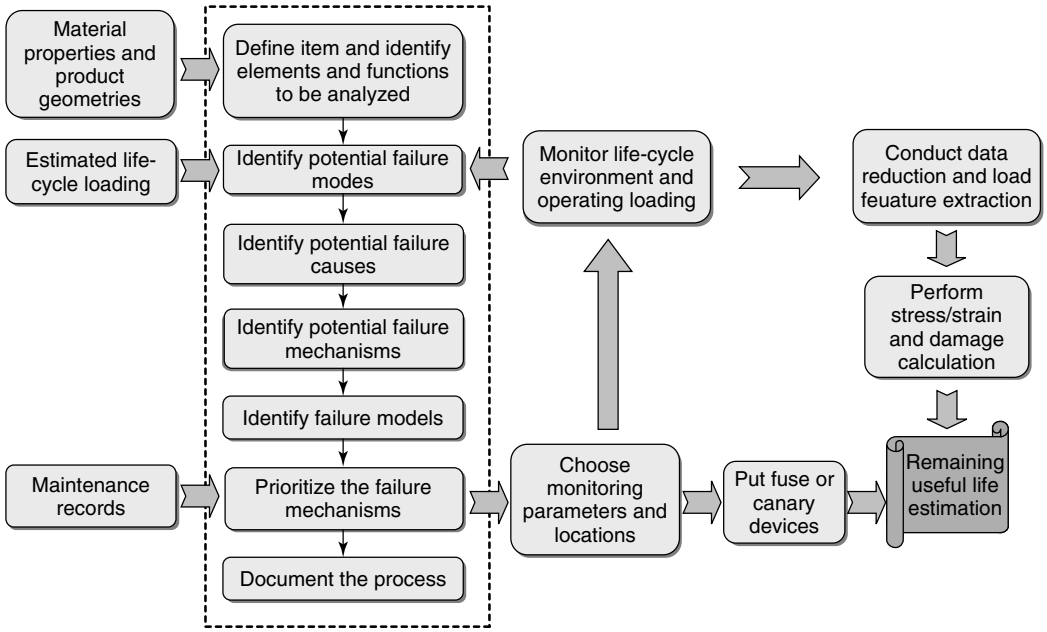


Figure 2. FMMEA-based PHM approach.

consists of design capture, identification of potential failure, and reliability assessment [15].

Design capture is the process of collecting structural (dimensional) and material information about a product to generate a model. This step involves characterizing the product at all levels, i.e., parts, systems, as well as physical interfaces. The potential failure identification step involves using the geometry and material properties of the product, together with the measured life-cycle loads acting on the product, to identify the potential failure modes, mechanisms, and failure sites in the product. This task is best performed through virtual qualification, which is a simulation-based methodology used to identify and rank the potential failure mechanisms. The reliability assessment step involves identification of appropriate PoF models for the identified failure mechanisms. A load–stress analysis is conducted using material properties, product geometry, and the life-cycle loads. With the computed stresses and the failure models, an analysis is conducted to determine the cycles to failure and then the accumulated damage is estimated using a damage model. Actually, the PoF methodology can provide a systematic approach to reliability assessment early in the design process.

4.1 Failure mode, mechanism, and effect analysis

A failure mode is the effect by which a failure is observed to occur [35]. It can also be defined as the way in which a component, subsystem, or system could fail to meet or deliver the intended function. All possible failure modes for each identified element should be listed. Potential failure modes may be identified using numerical stress analysis, accelerated tests to failure (e.g., HALT), past experience, and engineering judgment. The failure mode needs to be observable directly by methods such as visual inspection, electrical measurement, or other tests and measurements.

A failure cause is defined as the specific process, design, and/or environmental condition that initiated the failure, whose removal will eliminate the failure. Knowledge of potential failure causes can help identify the failure mechanisms driving the failure modes for a given element. One method of looking for causes is to review the life-cycle loads item by item to evaluate whether any of the loads there can cause the failure.

Failure mechanisms are the physical, chemical, thermodynamic, or other processes that result in

Table 1. Failure mechanisms, relevant loads, and models in electronics

Failure mechanisms	Failure sites	Relevant loads	Failure models
Fatigue	Die attach, wirebond/TAB, solder leads, bond pads, traces, vias/PTHs, interfaces	ΔT, T_{mean}, dT/dt, dwell time, ΔH, ΔV	Nonlinear power law (Coffin–Manson)
Corrosion	Metallizations	M, ΔV, T	Eyring (Howard)
Electromigration	Metallization	T, J	Eyring (Black)
Conductive filament formation	Between metallization	M, ∇V	Power law (Rudra)
Stress driven diffusion voiding	Metal traces	S, T	Eyring (Okabayashi)
Time-dependent dielectric breakdown	Dielectric layers	V, T	Arrhenius (Fowler–Nordheim)

Δ, cyclic range; V, voltage; T, temperature; S, stress; ∇, gradient; M, moisture; J, current density; H, humidity.

failure. Failure mechanisms are categorized as either overstress or wearout mechanisms. Overstress failure arises because of a single load (stress) condition, which exceeds a fundamental material strength. Wearout failure arises because of cumulative damage due to loads (stresses) applied over an extended time or number of cycles. Within current technology, PHM can only be applied in the wearout failure mechanisms. Typical wearout failure mechanisms for electronics have been summarized in Table 1 [36].

Failure models help quantify the failure through evaluation of time to failure or likelihood of a failure for a given geometry, material construction, environmental, and operational condition. For wearout mechanisms, failure models use both stress and damage analysis to quantify the damage accumulated in the product.

When using the canary devices PHM approach, the geometries or material properties of the prognostics cell can be scaled to accelerate the failure under user conditions, on the basis of potential failure mechanisms. When using the modeling of stress and damage approach, environmental and usage load profiles are captured using sensors. Sensor data is then converted into a format that can be used in the failure models.

In the life cycle of a product, several failure mechanisms may be activated by different environmental and operational parameters acting at various stress levels, but, in general, only a few operational and environmental parameters, and failure mechanisms, are responsible for most failures. High priority mechanisms are those with high combinations of occurrence and severity. Prioritization of the failure mechanisms provides an opportunity for effective utilization of resources.

4.2 Life-cycle loading monitoring

The life-cycle environment of a product consists of manufacturing, shipment, storage, handling, operating, and nonoperating conditions. The life-cycle loads (thermal, mechanical, chemical, electrical, and so on), either individually or in various combinations, may lead to performance or physical degradation of the product and may reduce its service life [1]. The extent and rate of product degradation depends on the magnitude and duration of exposure (usage rate, frequency, and severity) of such loads. If one can measure these loads *in situ*, the load profiles can be used in conjunction with damage models to assess the degradation due to cumulative load exposures. The typical life-cycle loads have been summarized in Table 2 [12].

4.3 Data reduction and load feature extraction

Experience has shown that even the simplest data collection systems can accumulate vast amounts of data quickly, requiring either a frequent download procedure or a large-capacity storage device [37].

Table 2. Life-cycle loads

Load	Load conditions
Thermal	Steady-state temperature, temperature ranges, temperature cycles, temperature gradients, ramp rates, heat dissipation
Mechanical	Pressure magnitude, pressure gradient, vibration, shock load, acoustic level, strain, stress
Chemical	Aggressive versus inert environment, humidity level, contamination, ozone, pollution, fuel spills
Physical	Radiation, electromagnetic interference, altitude
Electrical	Current, voltage, power

The main reasons for using data reduction in life-consumption monitoring are reduction of storage space; reduction in data-logger CPU load; and alignment with life-prediction models. The efficiency measures of data reduction methods should consider gains in computing speed and testing time; the ability to condense load histories without sacrificing important damage characteristics; and estimation of the error introduced by omitting data points.

The CALCE group has studied the accuracy associated with a number of data reduction methods such as ordered overall range (OOR), rainflow cycle counting, range-pair counting, peak counting, level crossing counting, fatigue meter counting and range counting.

Embedding the data reduction and load parameter extraction algorithms in to the sensor modules as suggested by Vichare *et al.* [25] can lead to a reduction in on-board storage space, lower power consumption, and uninterrupted data collection over longer durations. As shown in Figure 3, a time–load signal can be monitored *in situ* using sensors, and further processed to extract (in this case) cyclic range (Δs), cyclic mean load (S_{mean}), rate of change of load ($\mathrm{d}s/\mathrm{d}t$), and dwell time (t_{D}) using embedded load extraction algorithms. The extracted load parameters can be stored in appropriately binned histograms to achieve further data reduction. After the binned data is downloaded, it can be used to estimate the distributions of the load parameters. This type of output can be readily input into fatigue damage accumulation models.

In Vichare's study [26, 38], the temperature data was processed using two algorithms: (i) OOR to convert an irregular time–temperature history into peaks and valleys and also to remove noise due to small cycles and sensor variations and (ii) a three-parameter rainflow algorithm to process the OOR results to extract full and half cycles with cyclic range, mean, and ramp rates. The approach also involved optimally binning data in a manner that provides the best estimate of the underlying probability density

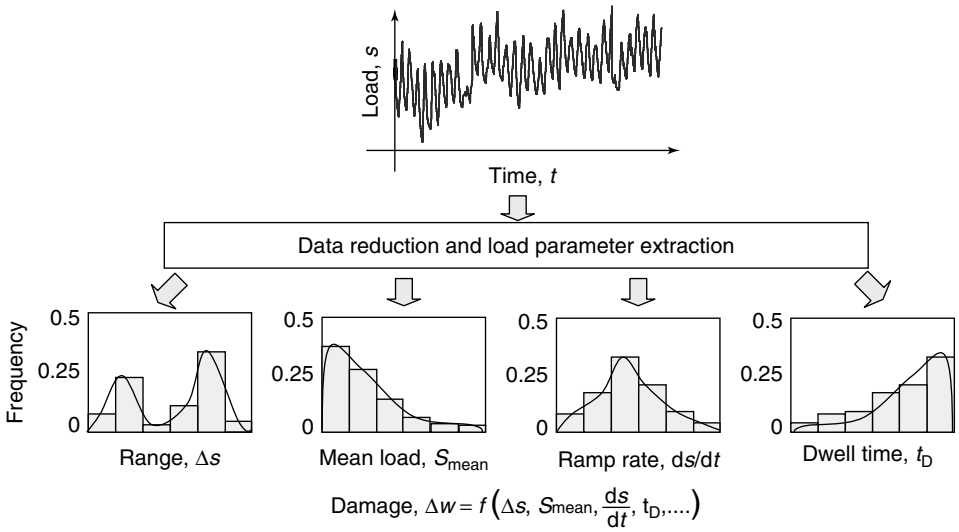


Figure 3. Load feature extraction.

function of the load parameter. The load distributions were developed using nonparametric histogram and kernel density estimation methods. The use of the proposed binning and density estimation techniques with a prognostic methodology was demonstrated on an electronic assembly.

4.4 Damage assessment and remaining-life calculation

Temperature and vibration are the most common causes of electronics failure [38]. The PoF models used to calculate the damage caused by temperature and vibration loadings are summarized in Figure 4. Damage caused by temperature can be calculated in time domain using Coffin–Manson's model. This approach has been demonstrated by Mishra *et al.* [16], Vichare *et al.* [39], and Cluff *et al.* [40]. Damage caused by vibration can be calculated in both time and frequency domains. The time domain, which has been demonstrated by Gu *et al.* [20], can use Basquin's model. The frequency domain, which has been demonstrated by Mishra *et al.* [16], can use first-order Steinberg's model [41].

4.5 Uncertainty implementation and assessment

The PoF model can be used to calculate the remaining useful life. However, it may still not be possible to make logistics decisions with certainty. Hence, it is necessary to identify the uncertainties in the prognostic approach and assess the impact of these uncertainties on the remaining-life distribution to make risk-informed decisions. Uncertainty analysis for prognostics implementations gives the prediction more meaning. With uncertainty analysis, a prediction can be expressed as a distribution rather than a single point. The prediction can be expressed as a failure probability.

Gu *et al.* [42] implemented the uncertainty analysis of prognostics for electronics under vibration loading. Gu identified the uncertainty sources and categorized them into four different types: measurement uncertainty, parameter uncertainty, failure criteria uncertainty, and future usage uncertainty. Then, the approach for implementing the uncertainty analysis was presented and shown in Figure 5 [42]. It utilized a sensitivity analysis to identify the dominant input variables that influence the model output. With information of input parameter variable distributions, a Monte Carlo simulation was used to provide a distribution of the accumulated damage. From the accumulated damage distributions, the remaining life was then predicted with confidence intervals. A case study was also presented, which used an experiment with an electronic board under vibration loading and a step by step demonstration of the uncertainty analysis implementation. The results showed that the experimentally measured failure time was within the bounds of the uncertainty analysis prediction.

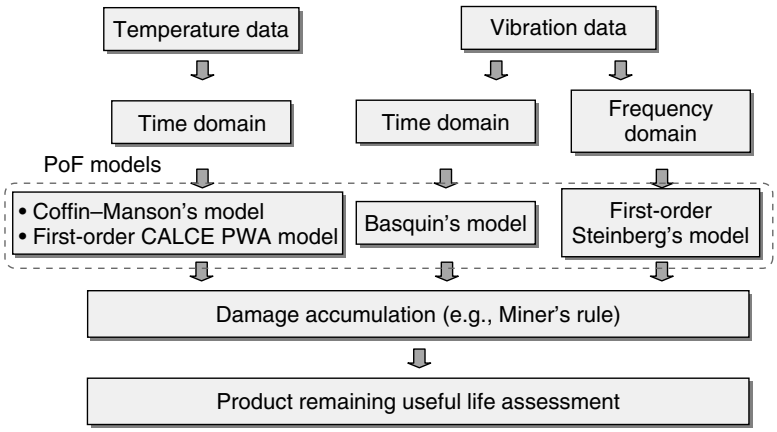


Figure 4. Damage calculation approach for temperature and vibration data.

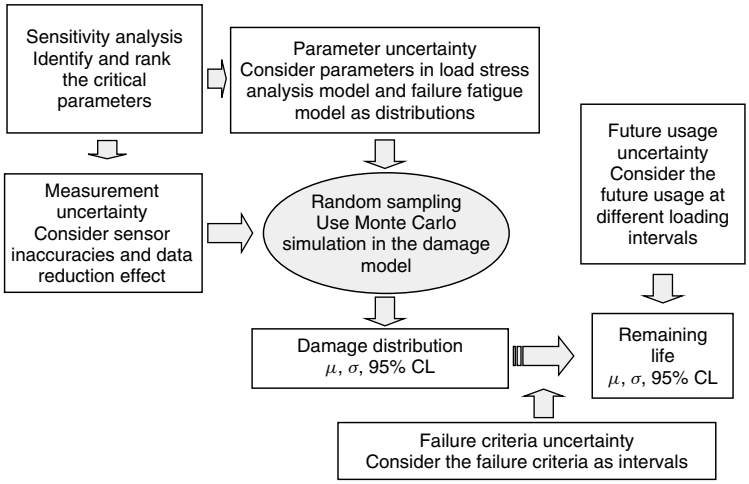


Figure 5. Uncertainty implementation for prognostics.

5 APPLICATION OF PoF IMPLEMENTATION FOR PHM

Prognostics based on the PoF model can be used in different areas, such as new products, legacy systems, and storage reliability prediction. Examples are discussed in the following paragraphs, along with a comparison of PoF-based PHM and data-driven-based PHM.

If the new product has not been manufactured, it is impossible to use the data-driven method since there will be no data available for training the algorithm. The PoF method normally only needs to change the material properties or geometries to model the new products, since most new products are not brand new and they have several previous versions or similar products can be referenced. Then it can save the time for performing prognostics for new products. On the other hand, the PoF approach makes it possible to give guidance for new product designs. The PoF approach incorporates reliability into the design process by establishing a scientific basis for evaluating new materials, structures, and electronics technologies. In addition, it can give feedback (such as failure site and failure mode) from life-cycle monitoring. Therefore PoF-based PHM is suitable for new products, and it reduces the design margin.

The legacy system continues to be used because of the prohibitive time and/or cost of replacing or redesigning it, though it is often less competitive and less compatible with modern equivalents [43]. Legacy systems can be found in both military and commercial sectors. PoF-based PHM to legacy systems can provide significant benefits. In the meantime, the data-driven approach lacks sufficient data to train the algorithm. Few legacy systems are in existence, and thus there are no extras for training purpose. The PoF-based PHM approach is based on an understanding of the life-cycle conditions of the legacy system and its failure modes and mechanisms. The first step is to use all available information (such as previous loading conditions, maintenance records, and so on) to assess the health status of the legacy system. The second step is to calibrate the health status using individual unit data so that an assessment of individual legacy systems' health can be derived. The third step involves the use of sensors and prognostic algorithms to update the health status on a continual basis to provide the most up-to-date prognosis of the system [44].

Improper storage is one factor that can precipitate failures in electronic products, especially in a hostile environment [45]. The temperature and humidity of storage areas are typically prime environmental factors. The extent and rate of product degradation depends upon the magnitude and duration of exposure (usage rate, frequency, and severity) to such loads. If one can measure these loads *in situ*, the load profiles can be used in conjunction with damage models to assess the degradation due to cumulative load exposures. Depending on the quality of the storage spaces,

environmental factors such as vibration, shock, fungi, sand and dust, and radiation might also come into the picture. Hence, the PoF-based PHM method is ready to be implemented. In the meantime, the limitation for the data-driven approach is that it can only detect failure when near the failure point. It is difficult to assess the remaining life from the beginning or middle of storage. The PoF approach will be more suitable since, for storage conditions, the loading will not change frequently, and dwell loading will become the dominant effect. On the basis of the PoF model, it is possible to assess the product reliability after a period of being stored, to indicate the remaining life, and to determine whether it can survive the next mission.

6 CONCLUSIONS

Traditional reliability predictions based on handbook methods are generally inaccurate and misleading. In this article, we have shown that PoF-based PHM is more suitable for reliability assessment. PHM can provide previously unknown information on life-cycle environmental and operational conditions, and previously unknown information on failure modes and mechanisms. It can also help prevent premature failures, and provide information on remaining life. In the future, owing to the increasing amount of electronics in the world and the competitive drive toward more reliable products, PoF-based PHM is being looked upon as a cost-effective solution to improve the reliability of electronic products and systems. Currently, more research should be focused on building physics-based damage models for electronics, obtaining the life-cycle data of product, and assessing uncertainty in remaining useful life prediction to make the PHM more realistic. Along with that, advance sensor technologies, communication technologies, decision making methods, and return of investment methods also need to be investigated.

REFERENCES

[1] Bowles JB. A survey of reliability prediction procedures for microelectronic devices. *IEEE Transactions on Reliability* 1992 **41**(1):2–12.

[2] Sinnadurai N, Shukla AA, Pecht M. A critique of the reliability analysis center failure rate model for plastic encapsulated microcircuits. *IEEE Transactions on Reliability* 1998 **47**(2):110–113.

[3] Telcordia SR-332, *Reliability Prediction Procedure (RPP) for Electronic Equipment*, 2001.

[4] FIDES Group, *FIDES Guide Issue A: Reliability Methodology for Electronic Systems*, 2004.

[5] Wong KL. What is wrong with the existing reliability prediction methods? *Quality and Reliability Engineering International* 1990 **6**:251–258.

[6] Pecht M, Kang WC. A critique of MIL-Hdbk-217E reliability prediction methods. *IEEE Transactions on Reliability* 1988 **37**(5):453–457.

[7] Pecht M, Nash F. Predicting the reliability of electronic equipment. *Proceedings of the IEEE* 1994 **82**(7):992–1004.

[8] IEEE Standard 1413–1998, *IEEE Standard Methodology for Reliability Prediction and Assessment for Electronic Systems and Equipment*. IEEE, December 1998.

[9] IEEE Standard 1413.1–2002, *IEEE Guide for Selecting and Using Reliability Predictions Based on IEEE 1413*. IEEE, February 2003.

[10] Hess A, Calvello G, Frith P, Engel SJ, Hoitsma D. Challenges, issues, and lessons learned chasing the "Big P": real predictive prognostics part 2. *Aerospace Conference*. IEEE, March 2006, pp. 1–19.

[11] Gu J, Vichare N, Tracy T, Pecht M. Prognostics implementation methods for electronics. *53rd Annual Reliability and Maintainability Symposium (RAMS)*. Orlando, FL, 2007.

[12] Vichare N, Pecht M. Prognostics and health management of electronics. *IEEE Transactions on Components and Packaging Technologies* 2006 **29**(1): 222–229.

[13] Leonard CT, Pecht M. Improved techniques for cost effective electronics. *Proceedings of the Annual Reliability and Maintainability Symposium*. Orlando, FL, 1991; pp. 174–182.

[14] DoD 5000.2 Policy Document, *Defense Acquisition Guidebook, Chapter 5.3—Performance Based Logistics*, December 2004.

[15] Ramakrishnan A, Pecht M. Life consumption monitoring methodology for electronic systems. *IEEE Transactions on Components and Packaging Technologies* 2003 **26**(3):625–634.

[16] Mishra S, Pecht M, Smith T, McNee I, Harris R. Remaining life prediction of electronic products using life consumption monitoring approach.

Proceedings of the European Microelectronics Packaging and Interconnection Symposium. Cracow, 16–18 June 2002; pp. 136–142.

[17] Shetty V, Das D, Pecht M, Hiemstra D, Martin S. Remaining life assessment of shuttle remote manipulator system end effector. *Proceedings of the 22nd Space Simulation Conference*. Ellicott City, MD, 21–23 October 2002.

[18] Mathew S, Das D, Osterman M, Pecht M, Ferebee R. Prognostic assessment of aluminum support structure on a printed circuit board. *International Journal of Performability Engineering* 2006 **2**(4):383–395.

[19] Mathew S, Das D, Osterman M, Pecht M, Ferebee R, Clayton J. Virtual remaining life assessment of electronic hardware subjected to shock and random vibration life cycle loads. *Journal of the IEST* 2007 **50**(1):86–97.

[20] Gu J, Barker D, Pecht M. Prognostics implementation of electronics under vibration loading. *Microelectronics Reliability Journal* 2007 **47**(12): 1849–1856.

[21] Simons JW, Shockey DA. Prognostics modeling of solder joints in electronic components. *Aerospace Conference*. IEEE, March 2006.

[22] Nasser L, Curtin M. Electronics reliability prognosis through material modeling and simulation. *Aerospace Conference*. IEEE, March 2006.

[23] Searls D, Dishongh T, Dujari P. A strategy for enabling data driven product decisions through a comprehensive understanding of the usage environment. *Proceedings of IPACK'01*. Kauai, HI, 8–13 July 2001.

[24] Herbst G. *IBM's Drive Temperature Indicator Processor (Drive-TIP) Helps Ensure High Drive Reliability*, IBM White Paper, October 1997, http://www.hc.kz/pdf/drivetemp.pdf (accessed Sep 2005).

[25] Vichare N, Rodger P, Eveloy V, Pecht M. Environment and usage monitoring of electronic products for health assessment and product design. *International Journal of Quality Technology and Quantitative Management* 2007 **4**(2):235–250.

[26] Vichare N, Rodgers P, Eveloy V, Pecht MG. In-situ temperature measurement of a notebook computer—a case study in health and usage monitoring of electronics. *IEEE Transactions on Device and Materials Reliability* 2004 **4**(4):658–663.

[27] Bodenhoefer K. Environmental life cycle information management and acquisition—first experiences and results from field trials. *Proceedings of Electronics Goes Green 2004+*. Berlin, 5–8 September 2004; pp. 541–546.

[28] ELIMA Report, *D-19 Final Report on ELIMA Prospects and Wider Potential for Exploitation*, April 2005, http://www.ELIMA.org (accessed Dec 2005).

[29] Tuchband B, Pecht M. The use of prognostics in military electronic systems. *Proceedings of the 32nd GOMACTech Conference*. Lake Buena Vista, FL, 19–22 March 2007; pp. 157–160.

[30] Mishra S, Pecht M. In-situ sensors for product reliability monitoring. *Proceedings of the SPIE* 2002 **4755**:10–19.

[31] Ridgetop Semiconductor-Sentinel Silicon™ Library, *Hot Carrier (HC) Prognostic Cell*, August 2004.

[32] Anderson N, Wilcoxon R. Framework for prognostics of electronic systems. *Proceedings of International Military and Aerospace/Avionics COTS Conference*. Seattle, WA, 3–5 August 2004.

[33] Goodman D, Vermeire B, Ralston-Good J, Graves R. A board-level prognostic monitor for MOSFET TDDB. *IEEE Aerospace Conference*. Big Sky, MT, 2006.

[34] Zhang G, Kwan C, Xu R, Vichare N, Pecht M. An enhanced prognostic model for intermittent failures in digital electronics. *IEEE Aerospace Conference*. Big Sky, MT, March 2007.

[35] Pecht M. *Product Reliability, Maintainability, and Supportability Handbook*. CRC Press: New York, 1995.

[36] Vichare N, Rodgers P, Eveloy V, Pecht M. Environment and usage monitoring of electronic products for health (reliability) assessment and product design. *IEEE Workshop on Accelerated Stress Testing and Reliability*. Austin, TX, October 2005.

[37] Harris R, McNee I. Physics of failure (PoF) approach to life consumption monitoring (LCM) for military vehicles, part 2. *Third International Conference on Health and Usage Monitoring—HUMS*. Melbourne, 2003; pp. 55–64.

[38] Vichare N, Rodgers P, Pecht M. Methods for binning and density estimation of load parameters for prognostics and health management. *International Journal of Performability Engineering* 2006 **2**(2):149–161.

[39] Vichare N, Pecht M. Enabling electronic prognostics using thermal data. *Proceedings of the 12th International Workshop on Thermal Investigation of ICs and Systems*. Nice, Côte d'Azur, 27–29 September 2006.

[40] Cluff K, Barker D, Robbins D, Edwards T. Characterizing the commercial avionics thermal environment for field reliability assessment. *Proceedings-Institute of Environmental Sciences* 1996: 50–57.

[41] Steinberg D. *Vibration Analysis for Electronic Equipment*, *Third Edition*. John Wiley & Sons, 2000.

[42] Gu J, Barker D, Pecht M. Uncertainty assessment of prognostics implementation of electronics under vibration loading. *AAAI Symposium on Artificial Intelligence for Prognostics*. Arlington, VA, 2007.

[43] Madisetti VK, Jung Y-K, Khan MH, Kim J, Finnessy T. Reengineering legacy embedded systems. *IEEE Design and Test of Computers* 1999 **16**(2): 38–47.

[44] Tuchband B, Vichare N, Pecht M. A method for implementing prognostics to legacy systems. *Proceedings of IMAPS Military, Aerospace, Space and Homeland Security: Packaging Issues and Applications (MASH)*. Washington, DC, 6–8 June 2006.

[45] Zhang Y, Pecht M, Lantz L. A case study of IC storage failures in Taipei trains. *Microelectronics Reliability* 1998 **38**:1811–1816.

FURTHER READING

Dasgupta A, Barker D, Pecht M. Reliability prediction of electronic packages. *Proceedings of the Annual Reliability and Maintainability Symposium*. IEEE, 1990, pp. 323–330.

Ramakrishnan A, Pecht M. Load characterization during transportation. *Microelectronics Reliability* 2004 **44**(2): 333–338.

Chapter 151
Multiwire Strands

Francesco Lanza di Scalea[1], Ivan Bartoli[1], Piervincenzo Rizzo[2], Alessandro Marzani[3], Elisa Sorrivi[3] and Erasmo Viola[3]

[1] *Department of Structural Engineering, University of California, San Diego, CA, USA*
[2] *Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA, USA*
[3] *DISTART, University of Bologna, Bologna, Italy*

1 INTRODUCTION

High-strength, multiwire steel strands are widely used in civil engineering such as prestressed concrete structures and cable-stayed or suspension bridges. Material degradation of the strands, usually consisting of indentations, corrosion, or even fractured wires, may result in a reduced load-carrying capacity of the structure that can lead to collapse. A survey involving the study of more than 100 stay-cable bridges [1] pessimistically reported that most of them were in danger mainly because of cable defects. Strand failures that caused bridge collapses are well documented [2–4]. Hence, it is of interest to develop structural health monitoring systems for strands that can detect defects as well as monitor applied loads to alert on any prestress loss.

Techniques proposed to detect defects (corrosion or wire fractures) in cables and strands include magnetic flux leakage [5], time-domain reflectometry [6, 7], and methods based on elastic waves, namely, acoustic emission testing [8–13] and ultrasonic testing [8, 14–22]. For the purpose of monitoring stress levels in the strands, methods of testing include fiber-optic strain sensors [23, 24], modal analysis based on the vibrating cord theory [25–27], magnetic permeability measurements [28–31], and wave-based ultrasonic testing [32–38].

The ultrasonic testing technique for strands is normally applied as guided-wave testing because it exploits the strand's waveguide geometry. The advantages of this technique over the others mentioned above include (i) the possibility of using

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

transducers permanently attached to the strand for continuous structural monitoring, (ii) the potential for providing simultaneous defect detection and stress monitoring capabilities for the strands with the same sensing system, and (iii) the possibility for detecting both active defects and preexisting defects toggling between the modes of "passive" acoustic emission testing and "active" ultrasonic testing within the same sensing system.

This article presents recent advances in the area of multiwire strand monitoring by guided ultrasonic waves (GUW), focusing on the "active" mode involving external generation and detection of waves. Wave propagation models are first presented to predict the complicated dispersive solutions of the strand waveguide in both free and embedded configurations. Experimental results are then presented with application to defect detection and stress monitoring in free strands. For defect detection purposes, the proposed system uses magnetostrictive transducers, wavelet-based signal processing, and statistical pattern recognition. For stress monitoring purposes, the proposed approach uses piezoelectric transducers, which monitor the ultrasonic "cross talk" between individual wires comprising the strand. The potential and difficulties of the field implementation of the technique are discussed in Section 5.

2 WAVE PROPAGATION MODELS FOR HELICAL AND AXIS-SYMMETRIC WAVEGUIDES

Proper modeling of guided waves propagating in the strands is important to identify modes, enhance sensitivity to detects, and select low-loss mode–frequency combinations.

The semianalytical finite element (SAFE) method is an effective tool to model waveguides of arbitrary cross section [39–41]. Here, the SAFE method is used to model the multimode and dispersive behavior of guided waves in a free, seven-wire strand (a pretwisted waveguide) and in a rod embedded in grout and concrete (an axis-symmetric multilayer waveguide).

The results presented here pertain to modal solutions of the defect-free structure. Therefore, they are of some use in the design of a guided-wave monitoring system. An improvement can consist of studying the interaction of waves with defects, which requires a different modeling strategy (e.g., a local–global approach).

2.1 Free strand (pretwisted waveguide)

The single strand can be represented as a pretwisted waveguide, for which the formulation proposed in [42] can be used. The equations of motion for this structural component can be written using a body coordinate system (ξ, η, ζ) that is related to the fixed Cartesian framework (x, y, z) through the relations (Figure 1):

$$\begin{aligned} \xi &= x\cos(\alpha z) + y\sin(\alpha z) \\ \eta &= -x\sin(\alpha z) + y\cos(\alpha z) \\ \zeta &= z \end{aligned} \tag{1}$$

where α is the uniform rate of pretwist in the axial direction. The displacement vector ($\mathbf{u}$), strain vector ($\boldsymbol{\varepsilon}$), and stress field ($\boldsymbol{\sigma}$) at a point with respect to the body coordinate system are as follows:

$$\begin{aligned} \mathbf{u} &= \begin{bmatrix} u_\xi & u_\eta & u_\zeta \end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{\varepsilon} &= \begin{bmatrix} \varepsilon_{\xi\xi} & \varepsilon_{\eta\eta} & \varepsilon_{\zeta\zeta} & \varepsilon_{\eta\zeta} & \varepsilon_{\xi\zeta} & \varepsilon_{\xi\eta} \end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{\sigma} &= \begin{bmatrix} \sigma_{\xi\xi} & \sigma_{\eta\eta} & \sigma_{\zeta\zeta} & \sigma_{\eta\zeta} & \sigma_{\xi\zeta} & \sigma_{\xi\eta} \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{2}$$

Linear strain compatibility relations and constitutive equations can be summarized as

$$\boldsymbol{\varepsilon} = \mathbf{L}_{\xi\eta}\mathbf{u} + \mathbf{L}_{\zeta}\mathbf{u} \quad \boldsymbol{\sigma} = \tilde{\mathbf{C}}\boldsymbol{\varepsilon} \tag{3}$$

where $\tilde{\mathbf{C}}$ is the complex constitutive matrix that describes the viscoelastic properties of the material [41] while $\mathbf{L}_{\xi\eta}$ and $\mathbf{L}_{\zeta}$ are differential operators as defined in [42]. The weak form of the governing balance equation for the given structure can be obtained via Hamilton's variational principle:

$$\begin{aligned} \delta \int_{t_1}^{t_2} \Bigg(\frac{1}{2} \int_{\zeta} \Bigg\{ \iint_{\Omega} \rho \dot{\mathbf{u}}^{\mathrm{T}} \dot{\mathbf{u}} \, \mathrm{d}\xi \, \mathrm{d}\eta \\ - \iint_{\Omega} \boldsymbol{\varepsilon}^{\mathrm{T}} \tilde{\mathbf{C}} \boldsymbol{\varepsilon} \, \mathrm{d}\xi \, \mathrm{d}\eta \Bigg\} \, \mathrm{d}\zeta \Bigg) \, \mathrm{d}t = 0 \end{aligned} \tag{4}$$

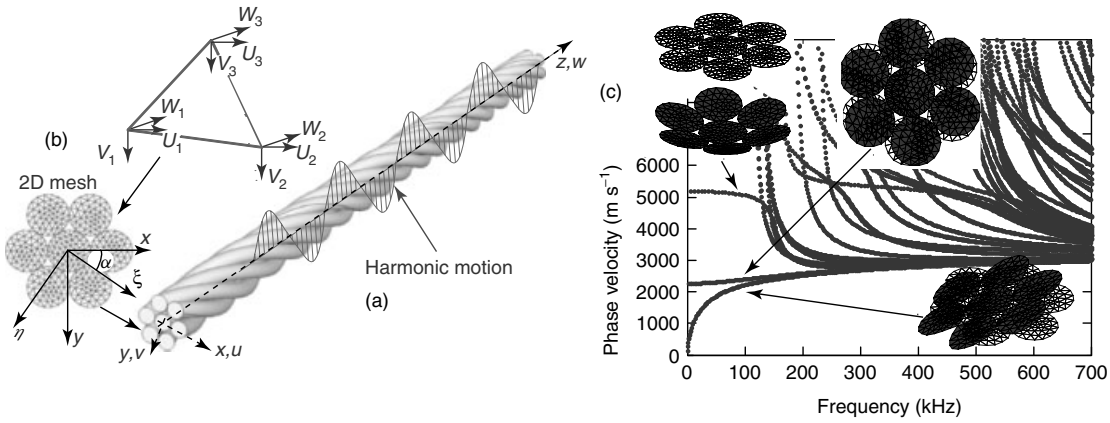


Figure 1. SAFE method. (a) Pretwisted waveguide representing a seven-wire strand, (b) generic element of the surface Ω, and (c) phase velocities and mode shapes for the strand at 100 kHz.

where a dot indicates derivative with respect to the time variable t, ρ is the mass density, and Ω is the area of the waveguide cross section. Assuming a harmonic motion along the propagation direction, $z \equiv \zeta$, and interpolating the displacement components over the cross section by standard finite element method, the approximated displacement field in the eth element takes the form

$$\mathbf{u}^{(e)} = \begin{bmatrix} \sum_{n=1}^{N} N_n(\xi,\eta) U_{\xi n} \\ \sum_{n=1}^{N} N_n(\xi,\eta) U_{\eta n} \\ \sum_{n=1}^{N} N_n(\xi,\eta) U_{\zeta n} \end{bmatrix}^{(e)} e^{\mathrm{i}(k\zeta - \omega t)} = \mathbf{N}(\xi,\eta)\mathbf{q}^{(e)} e^{\mathrm{i}(k\zeta-\omega t)} \tag{5}$$

where k is the guided-wave wavenumber and ω is the circular frequency.

Equation (5) is the core of the SAFE method where i represents the imaginary unit, $\mathbf{q}^{(e)}$ is the displacement vector of the eth element with nodal displacement components $U_n = U_{\xi n}$, $V_n = U_{\eta n}$, $W_n = U_{\zeta n}$, and $\mathbf{N}$ is the shape function matrix. Substitution of the last expression in the Hamilton's principle and subsequent standard finite element assembling result in the following wave equation:

$$\left[k^2\mathbf{K}_1 + \mathrm{i}k\mathbf{K}_2 + \mathbf{K}_3 - \omega^2\mathbf{M}\right]_M \mathbf{Q} = \mathbf{0} \tag{6}$$

with M as the number of total degrees of freedom of the cross-sectional mesh. For the sake of brevity, expressions of the stiffness matrices $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{K}_3$, and mass matrix $\mathbf{M}$ are not shown here but can be found, for example, in [42]. The eigenvalue problem in equation (6) is a two-parameter $(k-\omega)$ eigensystem. If material damping is neglected and only propagative modes are of interest, the wavenumber can be assigned as a real number and ω is adopted as eigenvalue parameter. Thus, for each wavenumber k in input, M eigenfrequencies ω_m and M eigenvectors $\mathbf{Q}_m$ are obtained, corresponding to propagating waves. Once ω_m is known, the dispersion curves can be easily computed. The phase velocity for the mth mode can be evaluated by the expression $c_{\mathrm{ph}} = \omega_m / k$. $\mathbf{Q}_m$ describe the cross-sectional mode shapes of the mth mode.

This formulation was adopted to study the dispersion properties of the structural component shown in Figure 1(a). The waveguide is a 15.24-mm-diameter (0.6-in.), high-strength steel grade 270, seven-wire strand. The pitch of the helical wire is equal to $p = 17 \times$ diameter and the material has the following nominal properties: bulk longitudinal velocity $c_{\mathrm{L}} = 5960\,\mathrm{m\,s^{-1}}$, bulk shear velocity $c_{\mathrm{T}} = 3260\,\mathrm{m\,s^{-1}}$, and density $\rho = 7700\,\mathrm{kg\,m^{-3}}$. The

mesh, shown in Figure 1(b) and generated by Matlab's "pdetool", consists of 323 nodes and 512 triangular elements with linear interpolation displacement functions. The phase velocity dispersion curves are shown in Figure 1(c) up to a frequency of 700 kHz. Notice the complexity of the modes. Mode shapes computed by SAFE for the three fundamental modes (longitudinal, torsional, and flexural) are shown in the same plot. These results are important to select suitable modes in experimental tests, as well as to identify the modes being measured from their arrival times.

2.2 Rod embedded in grout and concrete (axis-symmetric waveguide)

The SAFE method can be used to study the case of a strand embedded in grout and concrete. In the case of embedded member, it is important to model the ultrasonic energy leakage into the surrounding medium. The objective is to identify mode–frequency combinations, which propagate within the strand with minimum attenuation losses. The strand is approximated here by a straight rod, since the focus is on the interlayer energy leakage. The attenuation dispersion curves can be obtained using complex stiffness matrices in the SAFE formulation as described in [41].

The embedded rod can be treated as an axis-symmetric multilayer waveguide. For axis-symmetric geometries, it is convenient to develop the wave equations by using a cylindrical reference system, with the cross section lying in the $r-\theta$ plane and the z axis being parallel to the longitudinal direction of the waveguide (Figure 2a). The displacement, strain, and stress vectors variables at a point of coordinates (r, θ, z) and time t are, in this case, defined as follows:

$$\begin{aligned} \mathbf{u}(r,\theta,z,t) &= \left[u_r u_\theta u_z\right]^{\mathrm{T}} \\ \boldsymbol{\varepsilon} &= \left[\varepsilon_{rr}\ \varepsilon_{\theta\theta}\ \varepsilon_{zz}\ \varepsilon_{\theta z}\ \varepsilon_{zr}\ \varepsilon_{r\theta}\right]^{\mathrm{T}} \\ \boldsymbol{\sigma} &= \left[\sigma_{rr}\ \sigma_{\theta\theta}\ \sigma_{zz}\ \sigma_{\theta z}\ \sigma_{zr}\ \sigma_{r\theta}\right]^{\mathrm{T}} \end{aligned} \tag{7}$$

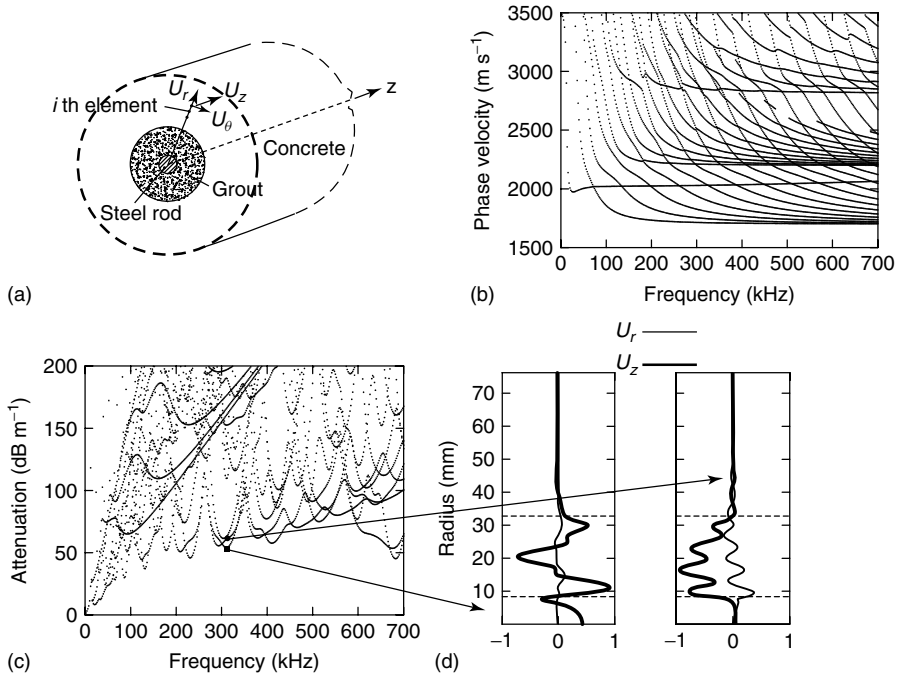


Figure 2. SAFE solutions for a 15.24-mm steel rod embedded in a 63.5-mm grout layer and a 152.4-mm concrete layer (axis-symmetric modes). (a) System modeled, (b) phase velocity curves, (c) attenuation curves, and (d) displacement mode shapes of two modes at low-loss points.

The mechanical variables in this equation are related by the compatibility and constitutive equations written in the cylindrical coordinate system

$$\boldsymbol{\varepsilon} = \mathbf{L}\mathbf{u} \quad \boldsymbol{\sigma} = \tilde{\mathbf{C}}\boldsymbol{\varepsilon} \tag{8}$$

where $\mathbf{L}$ is the compatibility matrix.

Subdividing the cross section into finite elements, the approximate displacement field in the element is

$$\mathbf{u}^{(e)} = \begin{bmatrix} \sum_{n=1}^{N} N_n(r) U_{rn} \\ \sum_{n=1}^{N} N_n(r) U_{\theta n} \\ \sum_{n=1}^{N} N_n(r) U_{zn} \end{bmatrix} e^{\mathrm{i}(n\theta + kz - \omega t)} = \mathbf{N}(r)\mathbf{q}^{(e)} e^{\mathrm{i}(n\theta + kz - \omega t)} \tag{9}$$

where n is the circumferential order of the mode, $\mathbf{N}(r)$ is the matrix of the shape functions, and $\mathbf{q}^{(e)}$ is the element's nodal displacement vector. The discretization is performed here only along the radial direction (r) using monodimensional elements owing to symmetry. Hamilton's principle can be used to generate the governing equations of motion:

$$\int_{t_1}^{t_2} \left\{ \int_z \int_0^{2\pi} \int_{R_\mathrm{i}}^{R_\mathrm{o}} \left[\delta\boldsymbol{\varepsilon}^{\mathrm{T}} \tilde{\mathbf{C}} \boldsymbol{\varepsilon} + \delta \mathbf{u}^{\mathrm{T}} (\rho \ddot{\mathbf{u}}) \right] r \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}z \right\} \times \mathrm{d}t = 0 \tag{10}$$

where R_i and R_o are the waveguide inner and outer radius, respectively. The finite element procedure reduces equation (10) to a problem similar to the one in equation (6) [43]:

$$\left[k^2 \mathbf{K}_1 + kn\mathbf{K}_2 + ik\mathbf{K}_3 + n^2 \mathbf{K}_4 + in\mathbf{K}_5 + \mathbf{K}_6 - \omega^2 \mathbf{M} \right]_M \times \mathbf{Q} = \mathbf{0} \tag{11}$$

Nontrivial solutions of equation (11) can be obtained by solving a three-parameter ($n - k - \omega$) eigensystem. Herein, n is assigned to obtain the nth order axial symmetric modes and k is adopted as the eigenvalue parameter for a given frequency ω. The M-dimensional second-order eigenproblem for k can be solved by recasting equation (11) into a $2M$ first-order form as follows:

$$[\mathbf{A}(n, \omega) - k\mathbf{B}(n, \omega)]_{2M} \quad \mathbf{U} = \mathbf{0} \tag{12}$$

For details on the complex matrices $\mathbf{A}$ and $\mathbf{B}$ see, for example, [41]. For each frequency ω, $2M$ complex eigenvalues k_m and $2M$ complex eigenvectors $\mathbf{Q}_m$ are obtained, corresponding to right-propagating and left-propagating waves. The first M components of $\mathbf{Q}_m$ describe the cross-sectional mode shapes of the mth mode. In this case, the phase velocity for the mth mode can be evaluated as $c_{\mathrm{ph}} = \omega/\mathrm{Real}(k_m)$, where $\mathrm{Real}(k_m)$ is the real part of the wavenumber. The imaginary part of the wavenumber is the attenuation, $att = \mathrm{Imag}(k)$, in nepers per meter (8.686 dB).

Figure 2 shows the SAFE results for a 15.24-mm-diameter (0.6-in.) steel rod embedded in a 63.5-mm (2.5-in.) outer diameter layer of grout and a 152.4-mm (6-in.) outer diameter layer of concrete. By simply discretizing a radius of the multilayer waveguide, the SAFE routine efficiently computed phase velocity (Figure 2b), attenuation (Figure 2c), and cross-sectional mode shapes (Figure 2d). Material properties assumed in the simulation are summarized in Table 1. Only axis-symmetric modes ($n = 0$) are shown. Twenty-five quadratic elements were used to discretize each of the three layers. The particular displacement mode shapes plotted correspond to small attenuation losses of two modes at 310 kHz. It can be seen that no energy is present in the outer concrete layer at both of these low-loss points.

Table 1. Material properties used in the SAFE model of the steel rod embedded in grout and concrete

	Longitude velocity, c_L ($\mathrm{m\,s^{-1}}$)	Shear velocity, c_T ($\mathrm{m\,s^{-1}}$)	Density, ρ ($\mathrm{kg\,m^{-3}}$)	Longitude attenuation, κ_L (Np/wavelength)	Shear attenuation, κ_T (Np/wavelength)
Steel rod	5960	3260	7700	0.003	0.008
Grout layer	2810	1700	1600	0.043	0.100
Concrete layer	3758	2090	2152	0.186	0.229

However, one of the two modes generates substantial displacements within the steel rod. This kind of analysis can help in designing a structural monitoring system that concentrates the ultrasonic energy within the rod and maximizes the inspection range.

3 DEFECT DETECTION IN STRANDS

3.1 Experimental setup and procedure

Defect detection results will be presented for the grade 270, seven-wire twisted strand with a diameter of 15.24 mm (0.6 in.). A notch was machined, perpendicular to the strand axis, in one of the six peripheral wires by saw-cutting with depths increasing by 0.5-mm steps to a maximum depth of 3 mm (Figure 3). A final cut resulted in the complete fracture of the helical wire (broken wire (bw)), which was the largest defect examined corresponding to a depth of 5 mm. The smallest notch depth of 0.5 mm corresponded to a 0.7% reduction in the strand's cross-sectional area. The largest notch depth of broken wire corresponded to a 15.6% reduction in the strand's cross-sectional area.

The strand was subjected to a 120-kN tensile load, corresponding to 45% of the material's ultimate tensile strength (UTS). Magnetostrictive transducers resonant at 320 kHz were used to excite and detect longitudinal guided waves in the strand (Figure 3). This frequency was chosen since it is known to propagate with little losses in loaded, free strands [36]. The distance between the transmitting and the receiving transducers, d_1 in Figure 3, was fixed at 203 mm (8 in.) in all tests. By sliding the transmitter/receiver pair along the strand, tests were conducted at the five different notch-receiver distances, d in Figure 3, of 203 mm (8 in.), 406 mm (16 in.), 812 mm (32 in.), 1016 mm (40 in.), and 1118 mm (44 in.). The latter was the largest distance allowed by the rigid frame of the hydraulic loading. Five-cycle tonebursts centered at 320 kHz, modulated with a triangular window, were used as generation signals. Signals were acquired at a sampling rate equal to 33 MHz and stored after different number of digital averages, namely, 500, 50, 10, 5, 2, and 1 (single generation).

Two time windows were selected for the direct signal and the defect reflection measured by the receiver. The gated waveforms were then processed through the discrete wavelet transform (DWT) using the Daubechies of order 40 (db40) mother wavelet. For a 33-MHz sampling frequency, the 320-kHz frequency of interest was contained in the sixth level of DWT decomposition, according to

$$f_j = \Delta \times \frac{F}{2^j} \tag{13}$$

relating the reconstructed frequency f_j at level j to the center frequency F of the mother wavelet, the scale 2^j, and the signal sampling frequency Δ. Thus, the sixth level was the only one considered in the further analysis (pruning). Representative results of the pruning process are shown in Figure 4, presenting the signals reconstructed from the first six DWT detail decomposition levels ($D_1 - D_6$). The original signal was taken without any averages. The D_6 reconstruction correctly identifies (at around 140 μs) the reflection from a 2.5-mm-deep notch in the helical wire. Since levels 1–5 will merely reconstruct noise, they were eliminated in the DWT analysis process.

Subsequent to pruning, the sixth decomposition level was subjected to the thresholding process. The

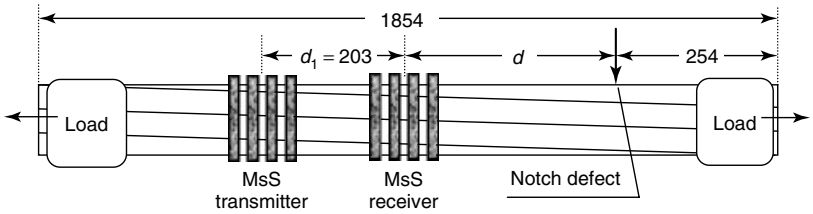


Figure 3. Experimental setup for defect detection in a strand using reflections of guided waves excited and detected by magnetostrictive transducers (dimensions in millimeter).

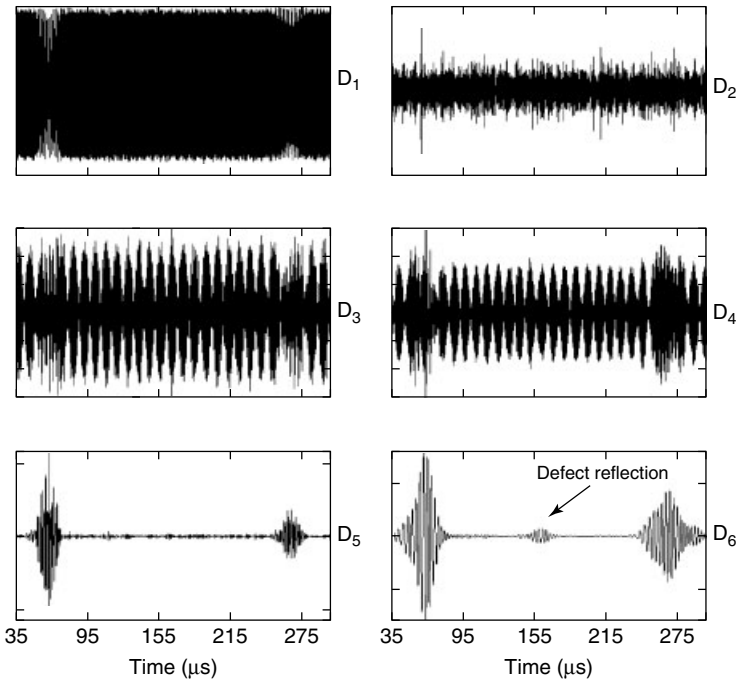


Figure 4. Signals reconstructed after pruning the DWT coefficients at the first six decomposition levels.

threshold chosen to select the relevant wavelet coefficients is an important variable that affects the sensitivity of the defect sizing. An optimum threshold combination for the direct signal and the defect reflection was searched on the basis of obtaining the largest sensitivity to defect size through a variance-based reflection coefficient. It was found that the larger sensitivities were obtained when setting more severe thresholds on the defect-reflected signals, with little effect of the thresholds imposed on the direct signal. On the basis of the findings in [20], optimum thresholds were fixed at 20% of the maximum wavelet coefficient amplitude for the direct signal, and at 70% of the same quantity for the defect reflection.

A "reflection" damage index (**DI**) vector was constructed from the ratios between certain features of the reflected signal, $F_{\text{reflection}}$, and the same features of the direct signal, F_{direct}:

$$\mathbf{DI} = \left[\frac{F_{\text{reflection},\, i}}{F_{\text{direct},\, i}} \right] \tag{14}$$

After parametric studies, the following four features were used to compute a four-dimensional **DI**: variance, root mean square, peak amplitude, and peak-to-peak amplitude of the thresholded wavelet coefficients at level six. All **DI** components showed a quite linear dependence in a semilogarithmic scale on the notch depth, and a relatively negligible dependence on the defect position for notches between 1.5 and 3 mm in depth. The experimental data for two of these components are shown in Figure 5. The results for very small notches, below 1 mm in depth, were less stable against varying distances owing to the poorer SNRs of the defect reflections. The results for the broken wire case (5-mm-deep notch) also showed an increased dependence on the notch-receiver distance, with **DI** components generally increasing for defects located further away from the receiver. This trend is opposite to what would be expected considering wave attenuation effects, and its origin is probably associated with the interference of multiple propagating modes that is distance dependent. It was also found that the **DI** component based on the variance of the wavelet coefficient vector (Figure 5) had the largest sensitivity to notch depth compared to all other components.

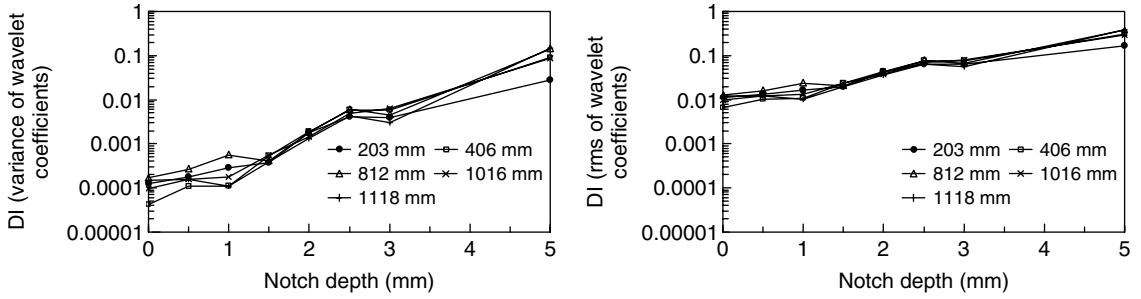


Figure 5. Components of the damage index vector measured from the variance and from the root mean square of the thresholded wavelet coefficients at the sixth decomposition level.

3.2 Statistical defect classification

A multivariate statistical analysis was performed to discriminate the defect indications from random noise, which may be present in the measurements. The Mahalanobis squared distance (MSD), D_ζ, was used as the discordancy test according to [44]:

$$D_\zeta = (\mathbf{x}_\zeta - \bar{\mathbf{x}})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{x}_\zeta - \bar{\mathbf{x}}) \qquad (15)$$

where $\mathbf{x}_\zeta$ is the potential outlier vector, $\bar{\mathbf{x}}$ is the mean vector of the baseline, $\mathbf{K}$ is the covariance matrix of the baseline, and T represents a transpose matrix. In the present study, since the potential outliers were always known *a priori*, D_ζ was calculated exclusively without contaminating the statistics of the baseline data.

The baseline distribution was obtained from the ultrasonic signals stored after averaging over 10 acquisitions and corrupted by two different levels of white Gaussian noise. The noise signals were created by the MATLAB *randn* function. The random noise increased the sample population and simulated possible variations in SNR of the measurements that can be originated, in practice, by a number of factors including changing sensor/structure ultrasonic transduction efficiency, and changing environmental temperature affecting ultrasonic damping losses. The *randn* function generates arrays of random numbers whose elements are normally distributed with zero mean and standard deviation equal to 1. The function was premultiplied by a factor that determines the noise level. Factors equal to 0.01 and 0.1 were considered as "low noise" and "high noise", respectively. For each noise level, 300 baseline samples were created.

The same approach was taken to generate a large number of data for the damaged conditions. Six of the seven total notch sizes discussed in the previous section were considered. The 10 average signals acquired for each of the six defects were corrupted by the low noise level and the high noise level, generating a total of 300 samples for each damage size. These samples represented the testing data of the algorithm. A total of 2100 samples data were thus collected for each noise level. The added noise can be quantified in terms of SNR by the following expression:

$$SNR(\mathrm{dB}) = 10\ \mathrm{Log} \left[\frac{\sum_{i=1}^{N} s_i^2 \Big/ N}{\sum_{i=1}^{N} u_i^2 \Big/ N} \right] \qquad (16)$$

where s_i and u_i are the amplitudes of the ultrasonic signal and the noise signal, respectively, and N is the number of points. The SNR between the direct signal and the two 0.01 and 0.1 noise levels was about 43 and 23 dB, respectively. The SNR between the reflection from the 3-mm-deep notch and the two 0.01 and 0.1 noise levels was about 32 and 12 dB, respectively. Clearly, the latter two values decreased with decreasing notch depth.

3.3 Defect detection results—"low" noise

The MSD computed from the four-dimensional **DI** of all samples, including the baseline data and the damage data, calculated for the low noise level of

0.01 are summarized in Figure 6(a). The mean vector and the covariance matrix were determined from the 300 **DI** vectors associated with the undamaged condition of the strand. The horizontal line in this figure represents the 99.73% confidence threshold value of 21.579. Eight baseline samples are outliers, and thus false positive indications. Clear steps can be seen for increasing levels of damage. All damaged conditions were properly classified as outliers and thus there were no false negative indications. The MSD values showed good discrimination between all defect sizes, including the smallest notch depths, confirming that it is advantageous to combine multiple GUW features to provide a large sensitivity to the defects. Nevertheless, compared to previous multivariate outlier analyses in structural monitoring applications, the dimension of the **DI** was still kept at a very low value by selecting only four features of the DWT-processed wave signals.

3.4 Defect detection results—"high" noise

The MSD results of the **DI** corrupted with the high noise level of 0.1 are shown in Figure 6(b). The 99.73% confidence threshold was now computed as 18.137. Compared to the low noise results of Figure 6(a), it is clear that the heavier noise corruption compromises the ability to detect the notch depths below 2.0 mm, corresponding to a 5% reduction in strand's cross-sectional area. The ratios of correctly classified outliers below 5% area reduction were only 12/300, 7/300, and 1/300 for notch depths of 0.5, 1.0, and 1.5 mm, respectively. Above the 5% area reduction, the sensitivity to defect detection was also degraded with the increasing noise level; for example, the MSD values for the 2-mm-notch depth in Figure 6(b) are 4 orders of magnitude smaller than the corresponding values in Figure 6(a). The reduced

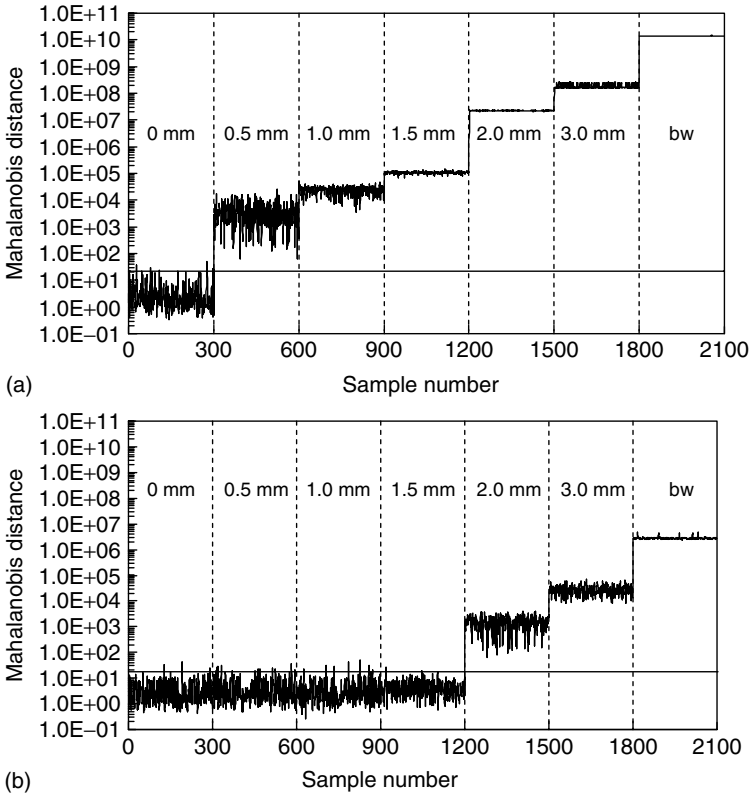


Figure 6. Mahalanobis squared distance for the baseline (undamaged) and damaged strand data corrupted with the low-level noise (a) and the high-level noise (b).

Table 2. Results of defect detection from outlier analysis: number of outliers $n/300$ for the various damage sizes and two levels of noise

Noise level	Damage size (notch depth, mm)						
	0	1.5	1	1.5	2	3	5.0 (bw)
0.01	8/300	300/300	300/300	300/300	300/300	300/300	300/300
0.1	3/300	12/300	7/300	1/300	300/300	300/300	300/300

number of false positive indications (three against eight) is the only improvement over the low noise level.

Table 2 summarizes the number of outliers detected in the multivariate analyses for both levels of noise considered; the outliers are false positive indications for the baseline data (damage size 0) and, instead, correct indications of anomalies for the defect data.

4 STRESS MONITORING IN STRANDS

It is known that the application of an axial load to a multiwire strand generates radial forces between the individual wires comprising the strand [45]. After extensive studies of wave propagation in loaded seven-wire strands, it was determined that the most effective strategy for stress measurement is one which monitors the ultrasonic "cross talk" between the peripheral and the central wires as a function of applied axial load. For this purpose, ultrasonic transmitters and receivers must be small enough to probe the individual wires. Two ultrasonic features, which are presented in the following sections, were found suitable for monitoring applied stress in free, seven-wire strands. These are the interwire energy leakage and the interwire frequency shift.

4.1 Experimental setup and procedure

Tests were performed at UCSD's Powell Labs on a SATEC M600XWHVL, 600 kips capacity, pneumatic test apparatus configured for tensile loading. A single specimen, again the grade 270 seven-wire strand ($UTS = 1.86$ GPa), was tested at a length of 1.82 m (72 in.). Figure 7(a) shows a picture of the strand on the testing machine.

A variety of ultrasonic transducers were installed on the specimen as shown in Figure 7(b). The focus here is on the piezoelectric (PZT) transmitter installed on a peripheral wire, PZT 3 in Figure 7(b), and on the two piezoelectric sensors installed at the strand's bottom end, PICO(C) and PICO(P) in Figure 7(b). Pictures of these transducers are also shown in Figure 7(c). PZT 3 was a 10 mm $\times$ 3 mm, rectangular piezoelectric plate, which excited waves in the peripheral wire. The waves were detected at the strand's end by both sensor PICO(C) (Physical Acoustics Corporation), installed on the central wire, and sensor PICO(P) installed on the same peripheral wire as PZT 3.

Load–unload cycles were performed with 11 load steps in each cycle. The steps were based on 70% of ultimate load or 182.4 kN (41 kips), consisting of a 0%, 20% (8.2 kips), 40% (16.4 kips), 60% (24.6 kips), 80% (32.8 kips), 100% (41.0 kips), and down to 80, 60, 40, 20, and 0%.

At each load step, the excitation frequency to PZT 3 was swept in the ranges 50–700 kHz and 700 kHz–2 MHz with a ~300-V peak-to-peak, 3-cycle Hanning-modulated toneburst. The signals measured by PICO(C) and PICO(P) were acquired and analyzed using a National Instruments data-acquisition unit.

4.2 Stress monitoring results—interwire energy leakage feature

The amount of ultrasonic energy leaking from the peripheral wire to the central wire was expected to increase with increasing load applied to the strand as a result of the increasing interwire contact. The "interwire energy leakage" feature was calculated as the ratio between the root mean square (rms) of the central wire signal from PICO(C) and the root mean square of the peripheral wire signal from

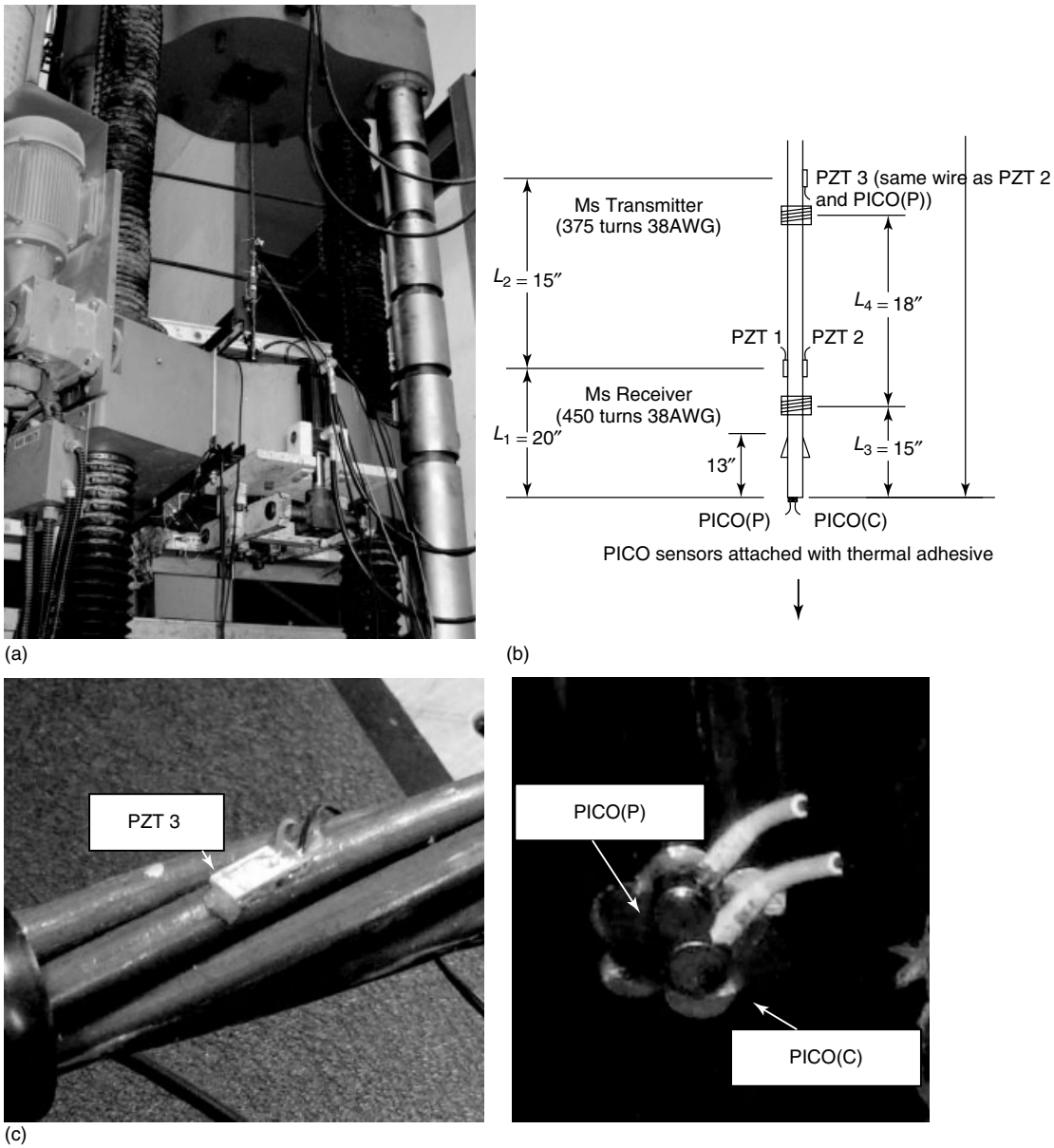


Figure 7. (a) The 1.82-m, seven-wire strand installed in the SATEC testing machine for stress monitoring tests; (b) ultrasonic sensor layout; and (c) pictures of the piezoelectric transmitter (PZT 3) on the peripheral wire and the two piezoelectric receivers on the strand's bottom end probing the central wire, PICO(C), and the peripheral wire, PICO(P).

PICO(P). This feature was calculated on the first arrival. The normalization makes the feature independent of generation power and thus eliminates the need for a baseline measurement, once the distance of the transmitter from the strand's end is fixed.

Figure 8 shows the normalized energy leakage measured in the low-frequency range of 300–700 kHz in a load–unload cycle. It can be seen that the interwire leakage increases with increasing load as expected. The behavior is consistent between load

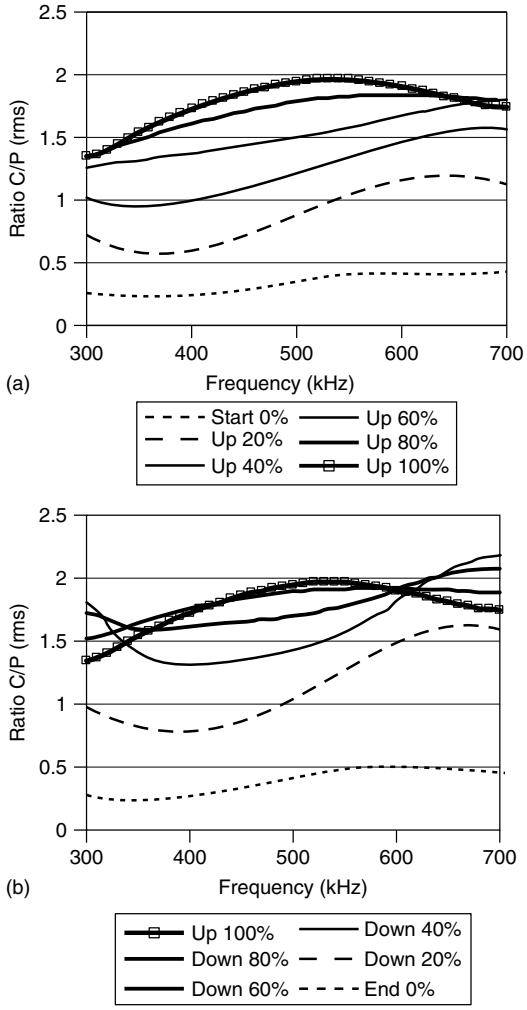


Figure 8. Normalized energy leakage between peripheral wire and central wire as a function of load applied to the strand, in the range 300–700 kHz. PZT 3 transmitting, PICOs (C) and (P) receiving. (a) Load ramp and (b) unload ramp. 100% load = 70% ultimate load (U.T.L.) or 182.4 kN.

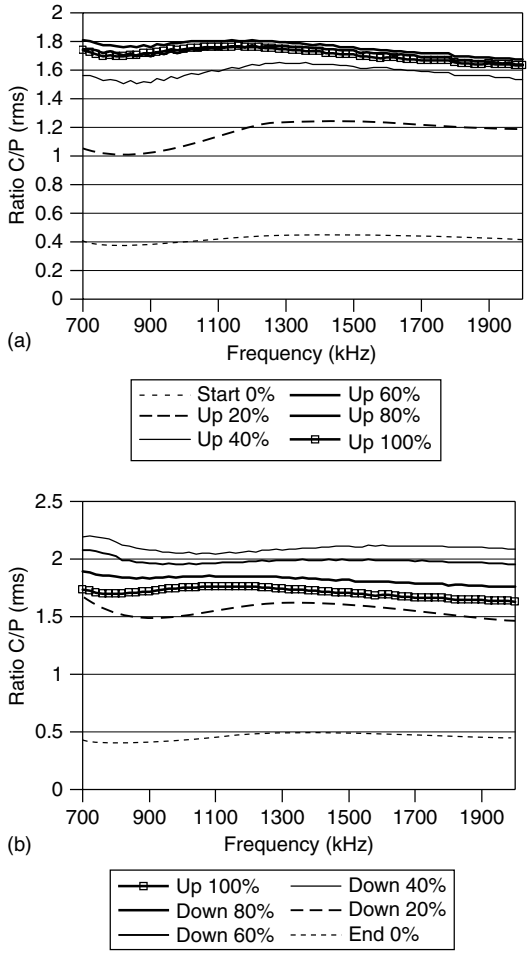


Figure 9. Normalized energy leakage between peripheral wire and central wire as a function of load applied to the strand, in the range 700 kHz–2 MHz. PZT 3 transmitting, PICOs (C) and (P) receiving. (a) Load ramp and (b) unload ramp. 100% load = 70% U.T.L. or 182.4 kN.

and unloading ramps. Also, the trend is such that individual load levels can be well discriminated, in addition to allowing a clear distinction between loaded and fully unloaded strand. Maximum sensitivity to load levels is observed between 400 and 500 kHz.

The same energy leakage feature was monitored in the high-frequency range of 700 kHz to 2 MHz. The results are shown in Figure 9 for a load–unload cycle. The results confirm the trend of increasing interwire leakage with increasing load level. However, compared to the low-frequency range of Figure 8, there is less discrimination among load levels, although the ability to distinguish between loaded and fully unloaded strand is maintained. Also, discrepancies exist in Figure 9 between load and unload ramps. These discrepancies appear owing to interwire contact stresses, which are not released at once during unloading until the load is completely removed. This apparent hysteretic behavior of the strand is only seen in the higher frequency range where the waves are more sensitive to

localized stress/strain fields. For these reasons, the low-frequency range of Figure 8 remains better suited than the high-frequency range for load level monitoring.

4.3 Stress monitoring results—interwire frequency shift feature

The frequency shift feature is also attractive because of its robustness against transducer structure contact and baseline-free characteristics.

The frequency feature is better discussed in light of Figure 10, which shows the peripheral-to-central wire transmissibility in terms of rms of the PICO(C) signal under the PZT 3 excitation, again calculated on the first arrival. Two distinct peaks of maximum transmission are seen in the ranges of 50–130 kHz and 130–250 kHz, respectively. The two peaks have opposite trends with increasing load, with the first one shifting toward higher frequencies and the second one shifting toward lower frequencies, and hence the need for examining the two peaks independently. Comparing load and unload ramps, it can also be seen that the hysteretic behavior of the strand only appears in the higher frequency range where the waves are sensitive to the residual, localized interwire stresses.

The shift in peak frequency relative to the unloaded peak frequency is shown in Figure 11 for the low-frequency range of 50–130 kHz during a load–unload cycle. The plot shows the expected increase in peak transmission frequency with increasing load level, which is consistent with the vibrating cord theory. This behavior is expected since the low frequencies capture the vibrational behavior of the strand as a whole. The shift is as high as 50% between fully loaded and fully unloaded strand. It is also possible to discriminate among different load levels. The trend is consistent between load and unload ramps. These results demonstrate that frequency shift in the range of 50–130 kHz is another promising feature for load level monitoring in the strand.

Figure 12 shows the relative frequency shift for the high-frequency range of 130–250 kHz in the same load–unload cycle. It can be seen that the frequency decreases with increasing load. This result is due to the fact that higher frequencies capture the vibrational behavior of the individual wires rather than that of the entire strand. The difference between load and unload ramps also confirms the sensitivity to residual interwire stresses. The high-frequency range of Figure 12 does not offer the same load discriminating capability as the low-frequency range of Figure 11. However, it still allows to distinguish between a loaded and a fully unloaded strand.

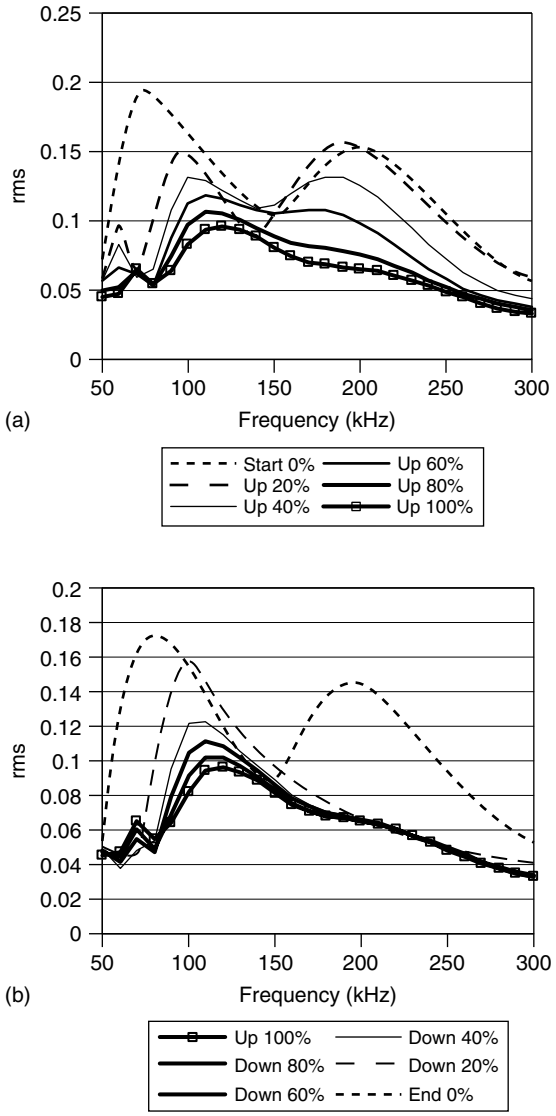


Figure 10. Energy leakage between peripheral wire and central wire as a function of load applied to the strand, in the range 50–300 kHz. PZT 3 transmitting, PICO(C) receiving. (a) Load ramp and (b) unload ramp. 100% load = 70% U.T.L or 182.4 kN.

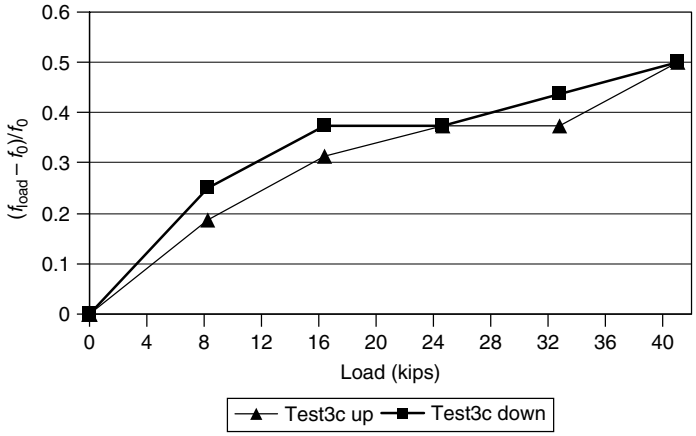


Figure 11. Shift in peak transmission frequency between peripheral and central wire as a function of load applied to the strand, in the range 50–130 kHz. PZT 3 transmitting and PICO(C) receiving. Maximum load 41 kips = 70% U.T.L.

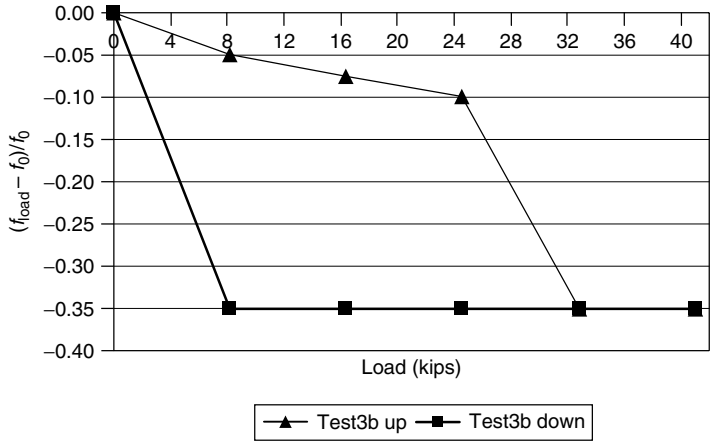


Figure 12. Shift in peak transmission frequency between peripheral and central wire as a function of load applied to the strand, in the range 130–250 kHz. PZT 3 transmitting and PICO(C) receiving. Maximum load 41 kips = 70% U.T.L.

5 DISCUSSION AND CONCLUSIONS

This article presents a technique based on ultrasonic guided waves to monitor the structural condition of multiwire strands used in prestressed concrete structures and cable-stayed or suspension bridges. The ultrasonic technique has the potential for providing both defect detection and stress monitoring in the strands.

The SAFE method was used to model the multimode and dispersive behavior in a free, seven-wire strand (a pretwisted waveguide) and in a rod embedded in grout and concrete (an axis-symmetric multilayer waveguide). In the second case, ultrasonic leakage from the rod into the surrounding media was modeled to identify mode-frequency combinations, which propagate within the rod with minimum attenuation losses for long-range monitoring.

For defect detection purposes, an array of magnetostrictive transducers probing the strand as a whole is proposed. Proof-of-principle results for the detection of notch-like defects were shown on the basis of a reflection damage index vector containing four components. A multivariate statistical analysis based on an outlier analysis showed the ability to distinguish the notches, which were located as far away as

1100 mm from the transducers, from simulated digital noise. The algorithm was able to properly flag notches as small as 0.5 mm (0.7% strand's area reduction) for SNRs on the order of 32 dB. For higher noise level, corresponding to SNRs on the order of 12 dB, the properly flagged notches were as small as 2 mm (5% strand's area reduction).

For stress monitoring purposes, piezoelectric transducers probing the individual wires near the strand's end are proposed. The best transducer layout was found when ultrasound excitation is performed on a peripheral wire and ultrasound detection is performed on the central wire and the peripheral wire at the strand's end. Two features proved suitable for stress monitoring. The first feature, the interwire leakage between the peripheral and the central wire, does not require a baseline once normalized, and in the range of 400–500 kHz it appears effective not only to detect a complete loss of stress but also to quantify the level of applied stress. The second feature was the shift in peak frequency of the peripheral-to-central wire transmissibility spectrum. The frequency shift proved most sensitive in the range of 50–130 kHz; it does not require a baseline and it can detect both a complete loss of stress and the level of applied stress.

The experimental results presented were obtained in the case of free strands. For embedded strands, the sensitive frequency ranges may change as a result of the different waveguide problem. However, the general trends of ultrasonic features as a function of defects and applied stress are expected to be similar to those found for the free waveguide case.

In an actual posttensioned structure, the stress monitoring technique simply requires access to a peripheral wire of the strand (at a location ∼1 m into the embedded portion) to install the PZT transmitter, and access to the strand's free end to install the piezoelectric sensors. If these access points are granted, existing structures can be instrumented. For new structures, the strands would clearly be instrumented before prestressing. Because of their geometry, the installation of the magnetostrictive transducers for defect detection requires access to the entire strand surface; hence, the defect detection method is more applicable to new structures where the strands are instrumented before installation. Clearly, the technique remains to be proven in the field.

ACKNOWLEDGMENTS

The strand monitoring project was funded at UCSD by the US National Science Foundation under grant # 0221707 (Dr S-C. Liu, Program Manager), and by the California Department of Transportation under contract # 59A0538 (Dr C. Sikorsky, Program Manager). Funding was also provided by the Italian Ministry for University and Scientific & Technological Research MURI (40%). The topic of research is part of the research thrusts of the Center of Study and Research for the Identification of Materials and Structures (CIMEST) at the University of Bologna.

REFERENCES

[1] Watson SC, Stafford D. Cables in trouble. *Civil Engineering* 1988 **58**:38–41.

[2] Woodward RJ. Collapse of Ynys-y-Gwas bridge, West Glamorgan. *Proceedings of the Institution of Civil Engineers, Part 1* 1988 **84**:635–669.

[3] Parker D. Pacific bridge collapse throws up doubt on repair method. *New Civil Engineer* 1996: 3–4.

[4] Chase SB. Smarter bridges, why and how? *Smart Materials Bulletin* 2001 **2**:9–13.

[5] Scheel H, Hillemeier B. Location of prestressing steel fractures in concrete. *ASCE Journal of Materials in Civil Engineering* 2003 **15**(3):228–234.

[6] Liu W, Hunsperger RG, Folliard K, Chajes MJ, Barot J, Jhaveri D, Kunz E. Detection and characterization of corrosion of bridge cables by time domain reflectometry. *Nondestructive Evaluation of Bridges and Highways III*. SPIE, 1998; Vol. 3587, pp. 28–39.

[7] Chajes M, Hunsperger R, Liu W, Li J, Kunz E. Void detection in grouted post-tensioned bridges using time domain reflectometry. *Proceedings of the Transportation Research Board*. Washington, DC, 2003; Vol. 3853.

[8] Kwun H, Teller CM. *Nondestructive Evaluation of Steel Cables and Ropes Using Magnetostrictively Induced Ultrasonic Waves and Magnetostrictively Detected Acoustic Emissions*, US Patent 5,456,113, 1995.

[9] Casey NF, Laura PAA. A review of the acoustic-emission monitoring of wire ropes. *Ocean Engineering* 1997 **24**:935–947.

[10] Rizzo P, Lanza di Scalea F. Acoustic emission monitoring of carbon-fiber-reinforced-polymer bridge stay

cables in large-scale testing. *Experimental Mechanics* 2001 **41**(3):282–290.

[11] Cullington DW, MacNeil D, Paulson P, Elliott J. Continuous acoustic monitoring of grouted post-tensioned concrete bridges. *NDT and E International* 2001 **34**:95–105.

[12] Fricker S, Vogel T. Site installation and testing of a continuous acoustic monitoring. *Construction and Building Materials* 2007 **21**(3):501–510.

[13] Ansari F. Fiber optic health monitoring of civil structures using long gage and acoustic sensors. *Smart Materials and Structures* 2005 **14**:S1–S7.

[14] Kwun H, Teller CM. Detection of fractured wires in steel cables using magnetostrictive sensors. *Materials Evaluation* 1994 **52**:503–507.

[15] Pavlakovic BN, Lowe MJS, Cawley P. The inspection of tendons in post-tensioned concrete using guided ultrasonic waves. *Insight—NDT and Condition Monitoring* 1999 **41**:446–452.

[16] Pavlakovic BN, Lowe MJS, Cawley P. High-frequency low-loss ultrasonic modes in imbedded bars. *Journal of Applied Mechanics* 2001 **68**:67–75.

[17] Beard MD, Lowe MJS, Cawley P. Ultrasonic guided waves for inspection of grouted tendons and bolts. *ASCE Journal of Materials in Civil Engineering* 2003 **15**(3):212–218.

[18] Rizzo P, Lanza di Scalea F. Load measurement and health monitoring in cable stays via guided wave magnetostrictive ultrasonics. *Materials Evaluation* 2004 **62**(10):1057–1065.

[19] Rizzo P, Lanza di Scalea F. Ultrasonic inspection of multi-wire steel strands with the aid of the wavelet transform. *Smart Materials and Structures* 2005 **14**(4):685–695.

[20] Rizzo P, Lanza di Scalea F. Feature extraction for defect detection in strands by guided ultrasonic waves. *Journal of Structural Health Monitoring* 2006 **5**(3):297–308.

[21] Reis H, Ervin BL, Kuchma DA, Bernhard JT. Estimation of corrosion damage in steel reinforced mortar using guided waves. *ASME Journal of Pressure Vessel Technology* 2005 **127**:255–261.

[22] Ervin BL, Bernhard JT, Kuchma DA, Reis H. Estimation of general corrosion damage to steel reinforced mortar using frequency sweeps of guided mechanical waves. *Insight—NDT and Condition Monitoring* 2006 **48**(11):682–692.

[23] Bronnimann R, Nellen PhM, Sennhauser U. Reliability monitoring of CFRP structural elements in bridges with fiber optic Bragg grating sensors. *Journal of Intelligent Material Systems and Structures* 1999 **10**:322–329.

[24] Zhang W, Gao J, Shi B, Cui H, Zhu H. Health monitoring of rehabilitated concrete bridges using distributed optical fiber sensing. *Computer-Aided Civil and Infrastructure Engineering* 2006 **21**: 411–424.

[25] Casas JR. A combined method for measuring cable forces: the cable-stayed Alamillo bridge, Spain. *Structural Engineering International* 1994 **4**(4):235–240.

[26] Tabatabai H, Mehrabi AB, Yen WP. Bridge stay cable condition assessment using vibration measurement techniques. *Structural Materials Technology III*. SPIE, 1998; Vol. 3400, pp. 194–204.

[27] Cunha A, Caetano E, Delgado R. Dynamic tests on large cable-stayed bridge. *Journal of Bridge Engineering* 2001 **6**(1):54–62.

[28] Bouchilloux P, Lhermet N, Claeyssen F. Electromagnetic stress sensor for bridge cables and prestressed concrete structures. *Journal of Intelligent Material Systems and Structures* 1999 **10**:397–401.

[29] Wang M, Lloyd GM, Hovorka O. Development of a remote coil magnetoelastic stress sensor for steel cables. *Health Monitoring and Management of Civil Infrastructure Systems*. SPIE, 2001; Vol. 4337, pp. 122–128.

[30] Wang G, Wang ML, Zhao Y, Chen Y, Sun B. Application of EM stress sensors in large steel cables. *Smart Structures and Materials*. SPIE, 2005; Vol. 5765, pp. 395–406.

[31] Sumitro S, Kurokawa S, Shimano K, Wang ML. Monitoring based maintenance utilizing actual stress sensory technology. *Smart Materials and Structures* 2005 **14**(3):S68–S78.

[32] Kwun H, Bartels KA, Hanley JJ. Effect of tensile loading on the properties of elastic-wave in a strand. *Journal of the Acoustical Society of America* 1998 **103**(6):3370–3375.

[33] Chen H-L, Wissawapaisal K. Application of Wigner-Ville transform to evaluate tensile forces in seven-wire prestressing strands. *ASCE Journal of Engineering Mechanics* 2002 **128**:1206–1214.

[34] Washer G, Green RE, Pond RB. Velocity constants for ultrasonic stress measurements in prestressing tendons. *Research in Nondestructive Evaluation* 2002 **14**(3):81–94.

[35] Lanza di Scalea F, Rizzo P, Seible F. Stress measurement and defect detection in steel strands by

guided stress waves. *ASCE Journal of Materials in Civil Engineering* 2003 **15**(3):219–227.

[36] Rizzo P, Lanza di Scalea F. Wave propagation in multi-wire strands by wavelet-based laser ultrasound. *Experimental Mechanics* 2004 **44**(4):407–415.

[37] Rizzo P. Ultrasonic wave propagation in progressively loaded multi-wire strands. *Experimental Mechanics* 2006 **46**:297–306.

[38] Bartoli I, Rizzo P, Lanza di Scalea F, Marzani A, Sorrivi E, Viola E. Structural health monitoring of strands in P/C structures by embedded sensors and ultrasonic guided waves. *Fracture Mechanics of Concrete and Concrete Structures Conference*. Catania, 2007; pp. 1029–1036.

[39] Huang KH, Dong SB. Propagating waves and edge vibrations in anisotropic composite cylinders. *Journal of Sound and Vibration* 1984 **96**:363–379.

[40] Hayashi T, Song WJ, Rose JL. Guided wave dispersion curves for a bar with an arbitrary cross section, a rod, and rail example. *Ultrasonics* 2003 **41**: 175–183.

[41] Bartoli I, Marzani A, Lanza di Scalea F, Viola E. Modeling wave propagation in damped waveguides of arbitrary cross section. *Journal of Sound and Vibration* 2006 **295**:685–707.

[42] Onipede O, Dong SB. Propagating waves and end modes in pretwisted beams. *Journal of Sound and Vibration* 1996 **195**:313–330.

[43] Nelson RB, Dong SB, Kalra RD. Vibrations and waves in laminated orthotropic circular cylinders. *Journal of Sound and Vibration* 1971 **18**: 429–444.

[44] Worden K, Manson G, Fieller NRJ. Damage detection using outlier analysis. *Journal of Sound and Vibration* 2000 **229**:647–667.

[45] Machida S, Durelli AJ. Response of a strand to axial and torsional displacement. *Journal of Mechanical Engineering Science* 1973 **15**(4):241–251.

Chapter 154
Landfills

Kai Münnich, Jan Bauer and Klaus Fricke
Leichtweiss-Institute, Department of Waste and Resource Management, Technical University of Braunschweig, Braunschweig, Germany

1 INTRODUCTION

Sanitary landfills are still most commonly used in the global scheme of things to dispose of municipal solid waste (MSW). The disposal of waste is always the origin of unwanted emissions, which might have a negative impact on the environment at large and the human health in particular. As a result of the industrial and economical improvement of the last decades, which is often very closely related to an uninhibited growth of the cities especially in the developing and low–income countries, problems with uncontrolled waste disposal have become more and more serious. One reason is that landfills, which were built in former times at the periphery of the cities, are now very often directly surrounded by residential quarters, so the residents are often directly affected by the waste disposal. From the multitude of different impacts [1] due to the imminent danger and long-term hazards, the focus is set here on

- fluid emissions and
- mechanical failure.

These different kinds of risks are interlaced; only a separate consideration of specific failure mechanism allows insight into a part of the *in situ* situation. Owing to the mechanical failure of the landfill, for example, caused by nonuniform settlements, changes occur in the microbial and chemical conditions inside the landfill body. Rainfall and oxygen permeate into the disposed waste and cause renewed aerobic conversion of biodegradable organic substances. The gas thus produced leads to direct emissions and, because of the rise of gaseous pressure, to interferences with the mechanical stability of the landfill body. The same effect on stability occurs when only rainfall infiltrates the waste and the percolation of the water is hindered by impermeable layers (e.g., plastic foils or daily soil cover).

The large landfill catastrophes of the last years (Table 1) with the collapse of large parts of the landfill body were among others caused by a combination

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Table 1. Landfill catastrophes of the last years

Year	Location	Cause of failure	Volume displaced
1997	Bogota, Colombia	Pore pressure caused by leachate recirculation	800×10^3 m^3
1997	Durban, South Africa	Pore pressure caused by codisposal of liquid waste	160×10^3 m^3
2000	Manila, Philippines	Shear failure following heavy rainfall	$13–16 \times 10^3$ m^3
2005	Bandung, Indonesia	Mechanical failure caused by fire and heavy rainfall	2700×10^3 m^3

This list is not exhaustive. The table was drawn from a number of sources, which appear in the reference list [2–6].

(a) (b)

Figure 1. (a) Payatas landfill slope failure [Reproduced from Ref. 5. © Kavazanjian *et al.*, 2005] and (b) Leuwigajah valley after the landfill failure—downhill view. [Reproduced from Ref. 6. © Koelsch *et al.*, 2005.]

of inappropriate waste disposal (poor compaction) and high water saturation of the waste (heavy rainfall, no drainage system at all or not functioning). In developing countries residential developments are very often located close to landfills, which results in accidents with many fatalities and destruction of these structures (Figure 1).

Over the last few years, the knowledge of the potential long-term risk of landfills has resulted in the intensification of a couple of regulations concerning the construction and controlling of landfill bodies. In all countries having directives concerning the waste disposal on landfills, it is therefore regulated by law that landfills must be built as a capsulate construction with a base and surface sealing. These very complex engineering structures are—unlike most other structures such as bridges or roads—intended to last forever or at least as long as the wastes represent a potential danger. Another important aspect of landfills is on one hand their dimension (e.g., the Fresh Kills Landfill of New York City has an area of 1.5 km^2 and a volume of about 750 million m^3) and on the other hand the inhomogeneity of the waste, whose chemical and physical properties are strongly changing with time.

It is, therefore, necessary to monitor the functioning of the installed equipment in and around the landfill as well as the mechanical behavior of the landfill body in all phases of the lifespan of the landfill. The different phases (operation, closure, and aftercare) may last for centuries depending on the volume of the landfill and the disposed waste [7, 8]. The main focus in the legal requirements regarding the monitoring of landfills is set on the measurement of settlement of the landfill surface, the quantity and quality of polluted water leaving the landfill, and the groundwater quality in the surroundings of the landfill.

However, it remains unclear if these parameters are sufficient and significant to make a reliable statement about the status of the landfill body and its hazard potential. As the monitoring program might become

very expensive, it is necessary to reduce the measurements to the required minimum, because the costs for the monitoring, even in the closure and aftercare phase, have to be received during operation. So, from the point of view of the landfill operator, monitoring is often only an imposition, which, on the one hand, has to be done because of the requirements made by the supervisory authority and on the other hand when data registration is necessary for technical processes (i.e., leachate or gas treatment).

2 MONITORING OF LEACHATE EMISSIONS IN LANDFILL DRAINAGE SYSTEMS

In the phase of waste disposal, precipitation freely infiltrates the waste and slowly percolates downward to the landfill base. During this transport process the liquid reacts with the waste material to produce leachate, which is thereby often characterized by high concentrations of contaminants. A generation of leachate can be observed at nearly each landfill, even under arid conditions, in the case of the disposal of high volumes of organic material or sludges. The quantity of leachate depends on the local climatic conditions, the water content in the waste, and the daily amount of disposed waste and its compaction, among others. To avoid a backwatering of leachate in the waste and the affiliated problems of stability, a drainage system has to be installed at the base of the landfill. The functioning of this system, made of a gravel filter in combination with drainage pipes (concrete or high density polyethylene (HDPE)-material), has to be controlled by establishing a water balance for the site [9]. If possible, a camera inspection of the drainage pipes has to be made at regular intervals to identify local damage spots and settlement measurements have to be made in view of controlling the minimum slope of the pipe for a free outflow of leachate [10]. For water balance or the modeling of the long-term leachate discharge and its quality, the specific discharge per unit area is often assumed to be constant for the landfill. Although this assumption is made because of the inhomogeneity of the disposed waste, the distribution of the biological activity, problems with incrustation processes in the drainage layers, and the partial accumulation of leachate, a nonuniform discharge might be expected to be more realistic.

At present, very often the emission of leachate from landfills can be classified according to quantity and quality only at the input of the purification plant, which means an integration of the whole landfill with its different phases of activity. When measuring at the outlet of a single drainage pipe, the catchment area is reduced from several hectares for the whole landfill to a maximum of about 12 000 m^2 for a single pipe (according to German regulations, maximum length is 400 m and maximum distance between two pipes is 30 m). This catchment area might even be too large, because the mass of waste disposed on the area might be very high (e.g., about 500 000 t at a landfill height of 40 m) and besides the MSW shows a large variability in its physical and biochemical properties.

In order to fine-tune the network of discharge measurements, the optimum would be to measure the discharge at any place at the landfill base sealing. For existing landfills, which have priority in this article, a manipulation of the landfill, for example, the installation of discharge-measuring gauges across the base sealing, is not possible. Therefore, the existing landfill facilities have to be used. Hence, a measuring system to determine the discharge in unreachable pipes has been developed, which allows the leachate discharge volume to be assessed at various points within the leachate collection pipes.

The measuring device to be developed had to fulfill various requirements, to achieve the greatest possible measured lengths in order to include the entire landfill surface, where possible. Simultaneously, it must be reliable and easy to operate as well as fulfilling the usual safety requirements for work at landfills, as, for example, the explosion protection requirements. Furthermore, it should be possible to integrate other measuring systems for the classification of leachate quality.

From these requirements, a commercially available camera system, which is commonly used for the inspection of drainage pipes, was used and adapted to the special purposes. A spillway weir was installed on the camera head, which can be moved in all directions (Figure 2). The placement of the weir and the required temporary sealing of the leachate flow in the pipe can be controlled over the movement of the camera head. A supplementary data transfer cable was not necessary because almost the entire

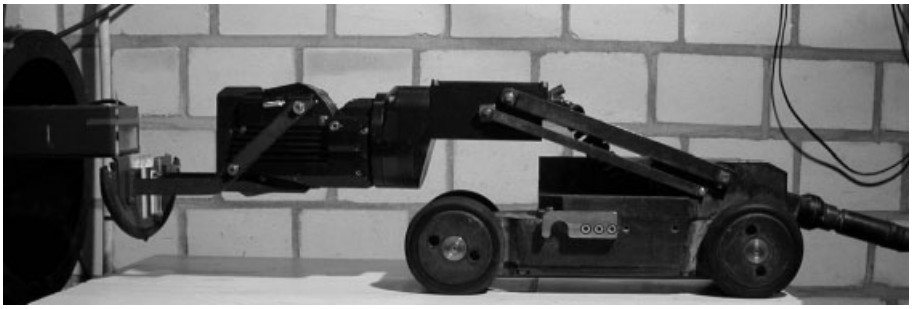


Figure 2. Camera lafette with elevated weir (on the left side).

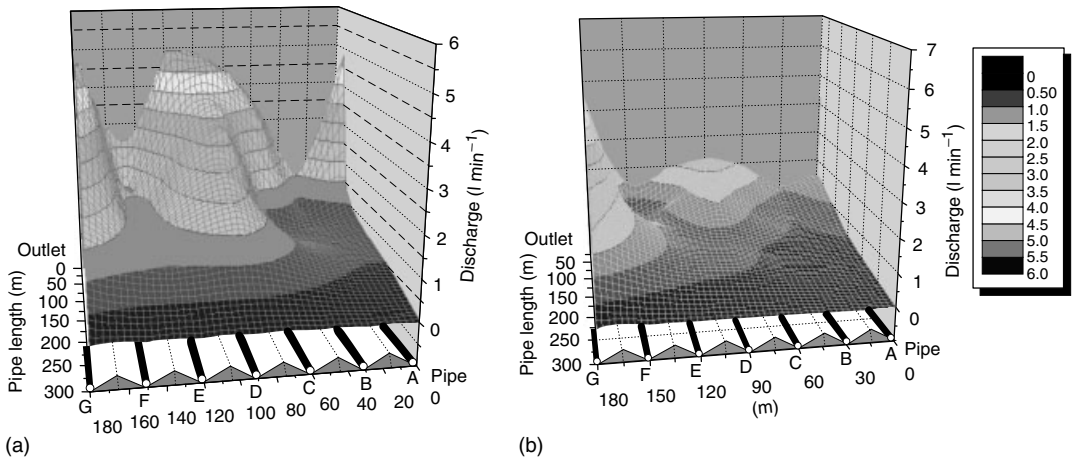


Figure 3. Discharge volume at the base of a landfill in (a) winter and (b) summer.

spillway weir is visible through the camera. The height of the impounded water is measured over a storage tube with millimeter scale, which is readable through the camera. In addition, a sensor for electrical conductivity and temperature is installed on the weir to identify changes in the quality of the leachate. The efficiency of the camera carriage is reduced by the installation of the spillway weir only very minimally.

A typical result of the monitoring of the leachate discharge into drainage pipes is shown in Figure 3. As expected, the discharge at the beginning of the pipe is small and increases with the length of the pipe and reaches the maximum value at the outlet. Not expected were the large differences in the discharge volume in-between the single pipes, although the interspace is only 30 m. The high discharge volume in drains D and E is caused by a recirculation of leachate in the landfill on the surface. Repetitions of the measurements at different dates show that the maximum discharge volume and the pipes, where it can be observed, are changing depending upon climatic and operation conditions.

Starting from the measured results and the known geometry of the landfill base, a surface-specific calculation of discharge can be made. The total length of the drainage pipe can be divided into parts of different lengths by changing the measuring point distances. Thus, the specific collection area can be adapted to the conditions present at the landfill. An example is shown in Figure 4. During the first measurements in April 2000, the distances between the measuring points were 50 m, and these grid distances were reduced for the following measurements. As the landfill segment was already surface covered at the time of the measurements, it was possible to make references to irregularities during waste deposit or surface installment through the measurements. Inspections made in this regard showed a faulty area in the surface

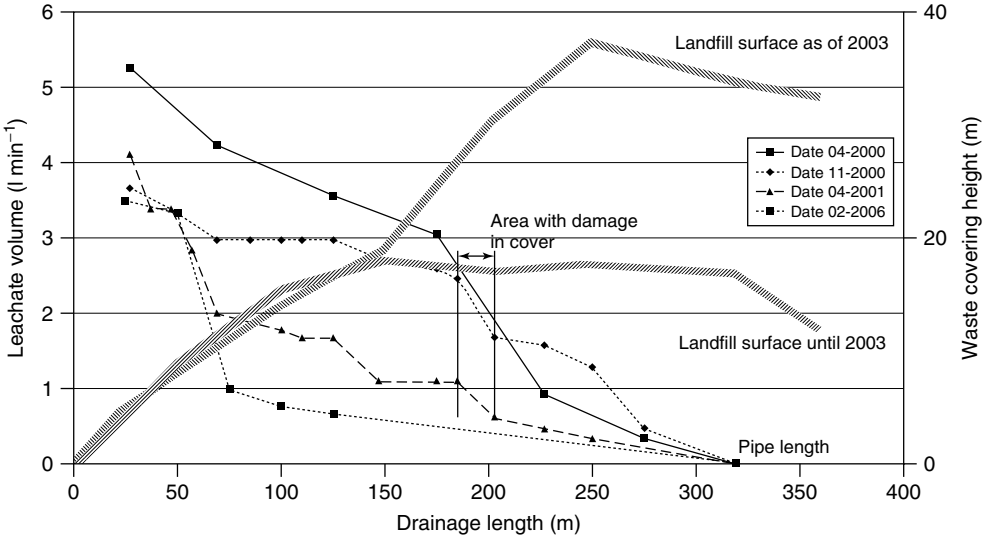


Figure 4. Leachate discharge of one particular drainage pipe.

covering. After the repairing, the volume coming into the collector is reduced significantly.

The distribution over the drainage length also shows that the volume discharged on the first, approximately, 75 m of the pipe increases with time. This might be caused by the partial deformation of the pipes so that the leachate flows beside the pipe in the gravel layer and is collected only at the outlet of the pipe because of the sealing system in the landfill slope. This assumption is supported by the observations made while driving into the pipe. In all pipes where the measurements have been made a backwater in the drainage system, could be observed over a distance of more than 15–20 m starting from the outlet of the pipe. This would mean that a part of the landfill is saturated at the base, which would have an important influence on the slope stability of the landfill in these areas. A supplementary measurement of the slope of the installed drainage pipes shows that, in parts, settlement of the liner system give result to partial backwater of leachate.

Since measurement within the pipes can only be conducted at relatively long time intervals, information regarding the variability of the leachate discharge in time cannot be gathered. For this reason another measuring apparatus was constructed, which can continuously record the outflow volume and the quality of leachate at single outlets of drainage pipes. Here also, the principle of measurement is a spillway weir, but the reading tube was replaced with a pressure sensor connected to a data log. In arbitrary time intervals, the backwater height and the electrical conductivity are recorded.

The results of monitoring over several years show that during waste disposal the discharge fluctuates over short periods, but that this fluctuation approximately balances itself out over a longer period. Furthermore, the discharge increase after a period of rain is recognizable, practically nearly without time delay. Also, the changes in landfill operation, e.g., the removal of the surface cover and the disposal of new waste in certain areas, can be tracked (Figure 5). As a consequence, the discharge increases with a certain time delay. Also, the above-mentioned bad spot in the sealing system was recorded as an increase in the discharge volume of the leachate. After the repair, the discharge decreases very fast to values as before. The placement of the final surface cover leads to a slow decrease in the leachate volume collected. With these data, the storage capacity and the degree of saturation of the waste can be evaluated.

3 MONITORING OF THE DEFORMATION OF LANDFILLS

Waste is a very compressible material, which is piled up in landfills to heights of 50–60 m or even more. As

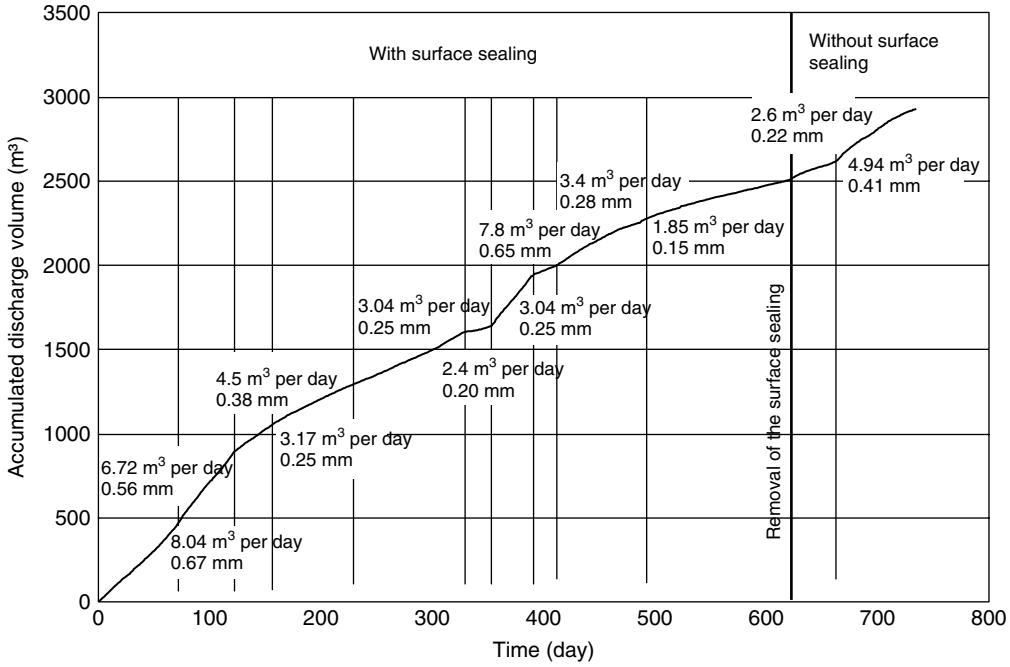


Figure 5. Results of continuous discharge measurement.

a consequence of the self-weight of the waste and the degradation of parts of the waste, nonuniform settlement of the surface occurs. A magnitude of up to 40% of the initial height of the landfill settlements represents a serious threat to the structures installed in the landfill (e.g., top cover system and gas drainage). In general, deformation measurements of landfill bodies are only conducted on the surface and in drainage pipes at the landfill base. The measurements at the landfill base are made to control the geomechanical properties of the underlying liner system and subsoil. In the past, some landfill collapses were caused by the breakdown of the subsurface soil under the high load of the waste. In most of these cases the plane consists of soft soil (e.g., [11]).

The monitoring of the surface is commonly made with conventional geodetic techniques as triangulation and distance measurement, with electronic tacheometers, or with a global positioning system (GPS). In some cases, internal instruments like inclinometers, extensometers, or buried plates/plates are installed [12]. The measured movements of the internal instruments must not be identical to those of the surrounding waste, as the instruments have different physical properties compared to waste (weight, stiffness, etc.).

In Figure 6, the settlement behavior of two different areas of one landfill are shown. Area 1 was filled with waste from 1967 to 1980 and then covered with a soil material. The measurement bolts were installed in 1993. Area 2 was filled from 1980 to 1990 and the bolts were installed in April 1997. The curves make it clear that even 25 years after finishing, the waste disposal settlement occurs. For both cases, it must be taken into account that the measurement started some years after the last disposal of waste, which means that the major settlement has not yet been recorded.

All these measurements are very precise, but only punctual, so the areal resolution to detect stability problems of landfills is not sufficient. A solution to this problem might be the ground-based 3D laser scanning technology, which is in use for more than 10 years in civil engineering for monitoring dams, landslides, etc [13]. Until now, 3D laser measurements were only made within research works on different landfills [14, 15]. The main advantages of this technology, the point density and the ease of implementation, may result in a more frequent

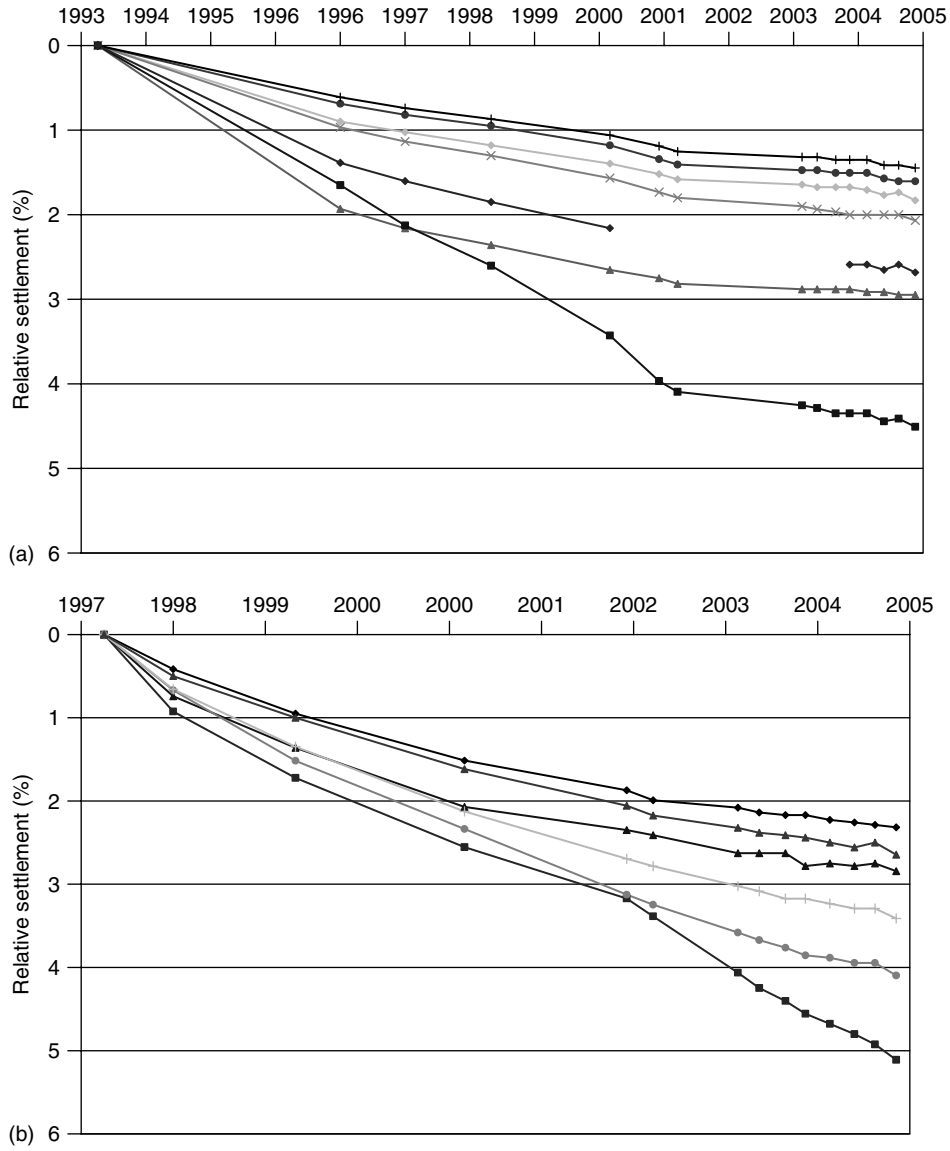


Figure 6. Surface settlement at different points of an older (a) and newer (b) area of a landfill.

measurement so that a more continuous representation of the landfill is enabled. A better understanding of the hydrobiophysical and mechanical interactions between waste and structures is possible.

These surface measurements are not always sufficient to describe a possible failure of the landfill slope, because the horizontal movement of the waste inside the landfill is often larger than that on the surface. The effect can be observed on the surface, where gas or leachate shafts have a strong inclination very often. To get an idea of the three-dimensional movement, an approximately 10-m-high and 10-m-wide ditch was dug for research purposes into the slope of an older landfill. The ditch is equipped with several measuring bolts on the slope surface to detect the three-dimensional deformation by tacheometry and differential global positioning system (DGPS) measurements. Also, eight measuring pipes at four

levels were installed into the slope of the ditch and in a range of up to 5 m into the landfill body. One of their purposes was to measure the vertical deformation inside the landfill body. The open walls were covered with foil, in order to prevent the penetration of oxygen into the landfill body and to avoid an influence on stability during heavy rainfall. During the excavation work waste samples were taken, in order to classify the waste and determine the density. The waste was weighed and the waste volume was calculated by laser scanning. The hydrostatic profile measurement of the measuring pipes confirms the order of magnitude of the settlement of the ditch wall. Regarding the settlement velocity over time, the pipes for the settlement investigation show seasonal settlement behavior with larger settlements after rain events. Furthermore, the results show a settlement that is not constant over the length of the pipe, without a significant deformation of the initial pipe length profile, but the more the pipe reaches inside the waste body, the settlement rate grows. A comparison between the beginning of the pipe on the back wall of the ditch and the end of the pipe inside the landfill body shows a 40% increase in the settlement, on an average, inside the landfill body. The difference between the beginning and end of the pipe increases from the head of the slope to the bottom.

Horizontal deformations occur in a magnitude of up to 14 cm in the three-year survey period. The horizontal movements, on an average, are approximately 75% of the settlement values. There is no significant correlation between horizontal deformations and settlement, and the results of the measurements do not show that they are influenced by each other (Figure 7). In this case, the horizontal deformations depend on the geometry of the slope excavation and the landfill topography. The main horizontal movements occur in the range of the slope foot because of the shear load of the whole landfill body. Even the horizontal movement exceeds the settlement values in the lower level of the back wall. Remarkable is the overall low horizontal movement of the two vertical

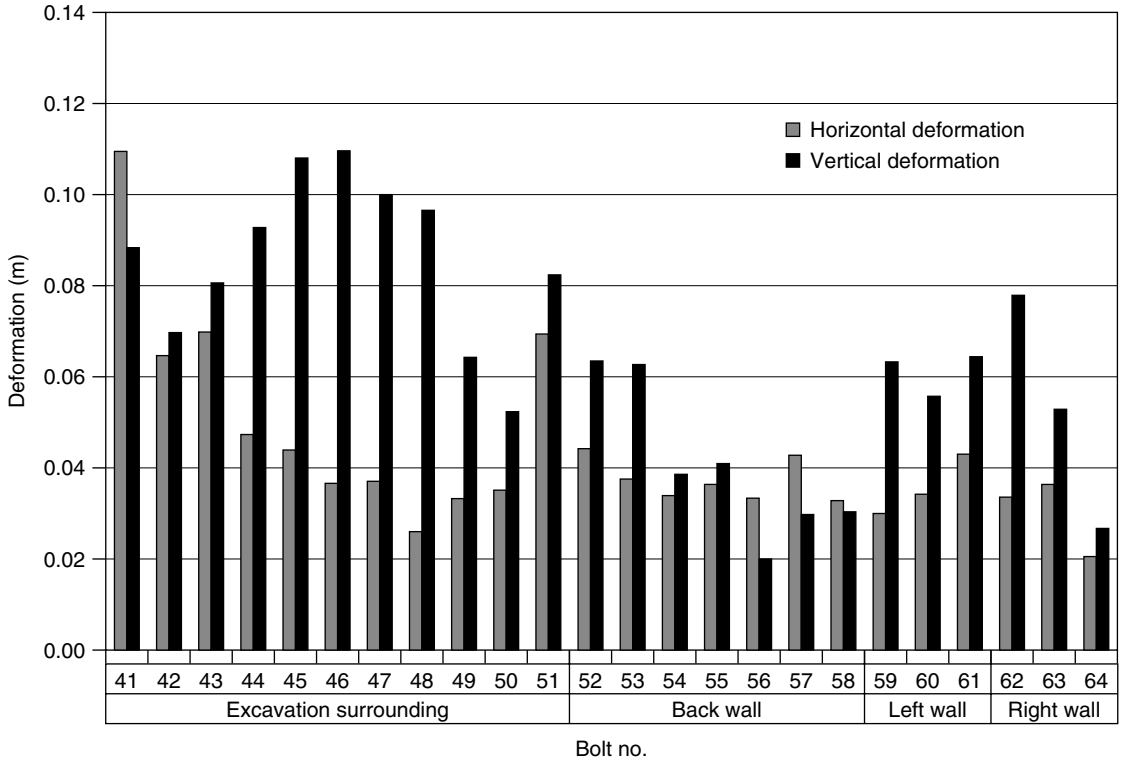


Figure 7. Deformation of surface points in a 2.5-year survey period.

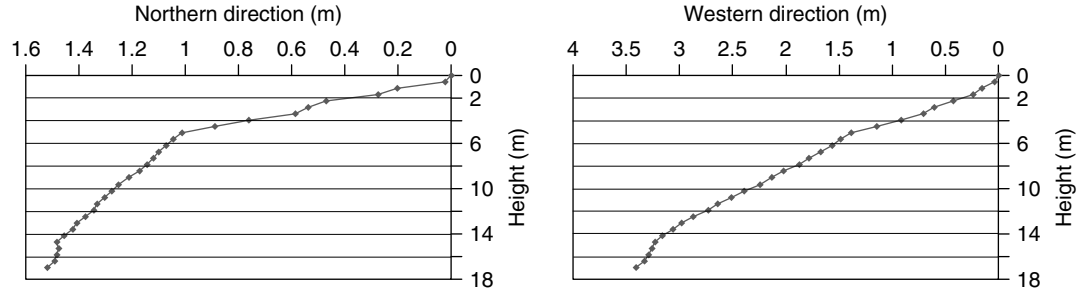


Figure 8. Horizontal deformation of a gas shaft.

side walls and the nearly vertical back wall of the excavation, which was not to be expected and is an indication of the stability properties of MSW.

To determine the horizontal displacement inside landfill bodies, inclinometers are sometimes installed in the existing shafts or special shafts are drilled in the waste. These inclinometers are used since a long time in the survey of buildings, dams, etc. The accuracy of these instruments is very high, but the disadvantage is that because of their construction, strong local deformations cannot be determined as the guiding tubing is ruptured. New inclinometers have been developed, which allows the use of existing gas shafts for the measurement [16]. This allows measurements to be conducted without destroying the surface cap, and it reduces the costs of measurements, since new pipes need not be installed. Furthermore, it is possible to measure pipes that have already been deformed over time, giving more information about landfill movement. For measurements, the probe is lowered into the shaft on a steel cable with the aid of a winch. After initial measurements at the shaft base, the probe is pulled in portions that correspond to the length of the probe. In each portion, the inclination of the pipe, the pivoting of the probe, and the exact time of measurement are recorded. Exact positioning in millimeters of height is achieved through the use of a highly accurate cable length counter. Centering at the beginning of every measurement is performed by the pressure-controlled centering device of the probe. The horizontal turn of the probe is determined by a rotation rate sensor. The electronic gyroscope determines the deviation from a direction independent of the earth's magnetic field so that the piping of the well or other iron items do not influence the measurement.

The first measurements made in an old landfill show that the horizontal deformations are high as expected compared with values measured in geotechnical engineering. A horizontal displacement of about 3.5 m of the bottom of a 17-m-long shaft could be observed (Figure 8).

The horizontal deformations were so high that the serviceability of the vertical gas wells was impaired. In spite of these deformations, no mechanical failure of the waste, which leads to cracks on the landfill surface or on slopes, could be detected.

4 SUMMARY

At present, knowledge about the physical behavior of landfill bodies is still insufficient. More precisely, no information is as of yet available on the connection between the mechanical strength and the hydraulic behavior, which is why precise conclusions concerning the deformation behavior cannot be drawn. Such data are important for the serviceability of the technical equipment installed in landfill sites and to avoid the worst case scenario of slope failures.

Monitoring techniques commonly used in structural and civil engineering, which allow a registration of data with high accuracy, are often not suitable for landfills. The cause of it may be seen in the large values of movements/displacements in waste, which may not result in the failure of the structure. Furthermore, as landfills are dynamic systems, which change their physical behavior with the ongoing process of biodegradation of the waste material, the critical points for monitoring are not always obvious and they are changing with time.

The latest landfill adapted technologies to monitor the leachate emissions and mechanical behavior of landfills are presented. Monitoring with these tools allows a better spatial resolution of important data

such as leachate generation, leachate quality, and horizontal and vertical movement of the waste body. The *in situ* measurements on leachate discharge and vertical and horizontal deformation should be completed by laboratory tests for the physical fundamentals.

The collected data and the data from the literature emphasize the need for monitoring the behavior of landfills at different stages of the lifespan of the landfill. Although these actions are often costly and their need is not always evident at first sight, it is essential to start monitoring during the disposal of waste and not to stop when the landfill capacity has been reached. Only a long-term monitoring of the relevant parameters enables the landfill owner and operator to get a better understanding of landfill behavior as a response on the specific conditions of the site (climate, waste composition, volume of waste, etc.).

ACKNOWLEDGMENTS

We would like to thank the German Research Foundation (DFG) for their financial support of the Collaborative Research Center 477.

REFERENCES

[1] El-Fadel M, Findikakis AN, Leckie JO. Environmental impacts of solid waste landfilling. *Journal of Environmental Management* 1997 **50**:1–25.

[2] Blight GE, Fourie AB. Catastrophe revisited—disastrous flow failures of mine and municipal solid waste. *Geotechnical and Geological Engineering* 2005 **23**:219–248.

[3] Brink D, Day PW, du Preez L. Failure and remediation of bulbul drive landfill: Kwazulu-Natal, South Africa. *Sardinia '99, Seventh International Waste Management and Landfill Symposium*. CISA, 1999, pp. 555–562.

[4] Hendron DM, Fernandez G, Prommer PJ, Giroud JP, Orozco LF. Investigation of the cause of the 27 September 1997 slope failure at the Dona Juana Landfill. *Sardinia '99, Seventh Int. Waste Management and Landfill Symposium*. Sardania, 1999; pp. 545–554.

[5] Kavazanjian Jr E, Merry SM. The 10 July 2000 Payatas landfill failure. *Tenth International Waste Management and Landfill Symposium, SARDINIA 2005*. Sardinia, 2005.

[6] Koelsch F, Fricke K, Mahler C, Damanhuri E. Stability of landfills—the Bandung dumpsite disaster. *Tenth International Waste Management and Landfill Symposium, SARDINIA 2005*. Sardinia, 2005.

[7] Sundqvist JO. System engineering models for waste management. *Proceedings from the International Workshop held in Gothenburg*. AFR Report 229. Swedish Environmental Protection Agency: Stockholm, 1998.

[8] Krümpelbeck I. *Untersuchung zum langfristigen Verhalten von Siedlungsabfalldeponien, Veröffentlichung des Lehrstuhls für Abfall- und Siedlungswasserwirtschaft der Bergischen*. Universität—Gesamthochschule: Wuppertal, 2000, Heft 3.

[9] Münnich K, Collins H-J. Evaluation of the water balance of municipal waste landfills. *International Waste Management and Landfill Symposium Proceedings SARDINIA 2001*. Cagliari, 2001.

[10] Kölsch F, Collins H-J. *Kontinuierliche Höhenvermessung von nicht begehbaren Rohren—Untersuchungen auf Deponien und in Abwasserleitungen*. Der Bauingenieur H.6. Springer Verlag, 1992.

[11] Reynolds RT. Geotechnical field techniques used in monitoring slope stability at a landfill. *3rd International Symposium on Field Measurements in Geomechanics*. Oslo, 1991.

[12] Jang Y-S, Kim Y-I. Behavior of a municipal landfill from field measurement data during a waste-disposal period. *Environmental Geology* 2003 **44**:592–598.

[13] Bitelli G, Dubbini M, Zanutta A. Terrestrial laser scanning and digital photogrammetry techniques to monitor landslide bodies. *Proceedings of the ISPRS Congress*. International Society for Photogrammetry and Remote Sensing: Istanbul, 2004.

[14] Bauer J, Münnich K, Fricke K. Settlement processes of landfill bodies—long-term survey of a slope deformation. *Tenth International Waste Management and Landfill Symposium, SARDINIA 2005*. Cagliari, 2005.

[15] Olivier F, Lhomme D, Gourc JP, Hidra M. The measurement of landfill settlement using terrestrial 3D laser scanner imaging. *Tenth International Waste Management and Landfill Symposium, SARDINIA 2005*. Cagliari, 2005.

[16] Goedecke H, Ziehmann G, Fricke K. Measurement of horizontal deformation of landfill bodies. *International Waste Management and Landfill Symposium Proceedings SARDINIA 2003*. Cagliari, 2003.

Chapter 152

SHM and Lifetime Management of Industrial Piping Systems

Frank Schubert[1], Bernd Frankenstein[1], Thomas Klesse[1], Klaus Kerkhof[2], Xaver Schuler[2], Herbert Friedmann[3], Fritz-Otto Henkel[3] and Helmut Wenzel[4]

[1] *Fraunhofer Institute for Nondestructive Testing (IZFP-D), Dresden, Germany*
[2] *MPA Universität Stuttgart, Stuttgart, Germany*
[3] *Wölfel Beratende Ingenieure, Höchberg, Germany*
[4] *VCE Holding GmbH, Vienna, Austria*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

Significant parts of industrial piping systems are partly or totally inaccessible for traditional nondestructive inspections, e.g., by being insulated or by being located at extreme positions. Therefore, it is not possible to detect damages without acceptable time and effort. For these kinds of problems, reliable monitoring techniques are needed. These techniques have to account for the fact that damages in piping systems manifest themselves on different scales ranging from global vibrational changes due to damaged support elements up to local changes caused by cracklike defects. Therefore, the corresponding monitoring system should have two "observing eyes", a *global* one for the overall condition and a *local* one for crucial error-prone parts of the structure. This dualism can be ideally realized by a structural health monitoring (SHM) system based on guided elastic waves and vibrations. The implementation of such a system including hardware and software development, decision support system (DSS), and subsequent

lifetime management of industrial piping systems is one of the main goals of the ongoing European SAFE PIPES project. Intermediate results of this work are presented in the following.

2 THE SHM "GLOBAL EYE": MODEL-BASED MODAL ANALYSIS

2.1 Preface

The global eye of the SHM system is represented by model-based vibration analysis. Its main objective is the detection of changes of boundary conditions that influence the life cycle of a piping system. The method is based on the fact that changes in system stiffness and boundary conditions are reflected in significant changes of experimentally detectable and numerically computable natural frequencies and related mode shapes. Besides the piping system with rigid supports this means, above all, changes in the stiffness of supporting constructions, e.g., loss of load-bearing capacity of aged spring hangers, increase of hysteresis with spring and constant hangers, load rearrangement due to heating and cooling processes of piping systems with constant hangers and the failure of pipe clamps due to fatigued bolts. In view of the fact that this method detects actual changes in the system, it can be used at any point of the life cycle of a system. The finite element method (FEM) modeling of a partly fatigued pipe, however, is highly problematic. For this, it is desirable to simulate and measure the undamaged zero state of a system. The individual steps in the context of model-based modal analysis are described as follows.

2.1.1 FEM simulation as a basis

In the first step, the design state of the piping system is transferred to a FEM model for static and dynamic analysis. The mechanical system of the pipe has to be modeled with appropriate accuracy, whereas first discrepancies may already occur between the idealized model and the real pipe since structural modifications can sometimes be decided only on site. Possible deviations can be discovered in the zero state, e.g., by walkdowns, and the model then must be updated accordingly. However, if a pipe that is already damaged or aged has to be simulated, it is hardly possible to determine the complete life-cycle consumption and other changes experienced by a pipe during operation and to describe such changes in a FEM model with sufficient accuracy.

2.1.2 Measurements

Kinematic quantities, e.g., accelerations, vibration velocities, and deflections are measured together with strains at appropriate points at anchoring elements and at the pipe itself. Forced vibrations of the system with excitations from plant operation as well as vibrations initiated by artificial or ambient excitation are measured. For operational vibrations and ambient vibrations, the determination of the excitation by measurement is often impossible. For measurement of free vibrations, initiated by artificial excitations, it is necessary to choose an excitation that is suitable to excite an industrial piping system, due to its frequency content and excitation energy, without the danger of damage to the structure. For measurement of modal quantities (experimental modal analysis, EMA) ambient vibrations (wind, soil tremor, and general plant operation) or artificial impacts (impulse hammer, shaker, and snap back) can be used as excitation. The resulting modal parameters serve as basis for updating the FEM models.

2.1.3 Experimental and operational modal analysis

An EMA provides natural frequencies, mode shapes, and modal damping values of the system, however, additionally requiring the determination of the excitation force of the structure. Since it is hardly possible to measure the excitation of operational vibrations and ambient vibrations, the operational modal analysis (OMA) has been currently developed. Within the OMA framework, explicit knowledge or control of the input excitation is no longer necessary.

Related to natural frequencies and mode shapes, the results of EMA or OMA correspond to those of FEM simulations. However, with the last-named simulation method besides modal quantities, displacements and stresses can be simulated and presented for the entire structure.

2.1.4 Model updating

By comparison of measurement and analysis results that can be carried out individually or by software-based model update, the model can be verified. With a close matching of results, the FEM model verified by measurement presents a safe basis for prognostic statements concerning planned modifications of the piping system, e.g., of the anchoring concept. In the case of modifications of the pipe during operation, the comparison of the measuring data with the unchanged simulation model vice versa allows the drawing of conclusions concerning the damage state. For example, a change in stiffness due to corrosion or due to a locked spring hanger may lead to a shift of natural frequency or a change of mode shape [1, 2]. For that purpose, in the FEM model, regions where damage is expected must be identified. These "unsafe regions" allow a *local* adaptation of physical parameters such as stiffness and mass. These parameters are modified in the unsafe regions until the deviations between the identified (measured) natural frequencies and mode shapes and the simulated data are minimized.

As long as measurement and simulation results match, the structure does not show any damage that influences the stiffness. If deviations emerge from the comparison of natural frequencies determined by measurements and those determined by FEM simulation, a damage of the pipe must be assumed. The model update results in a modified simulation model, which again correlates with the modified measuring data and thus iteratively describes the modifications or damage of the pipe, respectively. However, comprehensive research work is still necessary in the updating of the damping.

2.1.5 Reassessment of the pipe

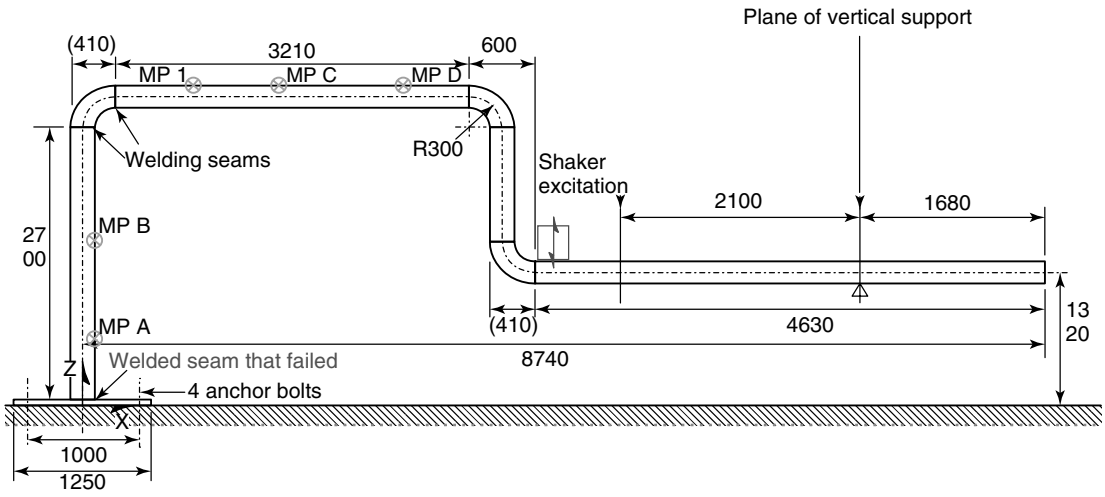
One of the simulation results of the model, subject to modification through model update, consists in the new stress distribution as a consequence of the modified load-bearing behavior. By comparison with relevant codes, the stresses in the piping system can be reassessed and statements concerning the exploitation of the system are possible. The analysis of the updated stress distribution and the load history of the system (heating, cooling, vibration cycles, and singular events) provide information on the life-cycle consumption or the remaining lifetime, respectively.

2.2 Example A: modal analysis based on shaker excitation

The following test carried out at a mock-up of a piping system gives an insight into the procedure of the modal analysis method. A piping system equipped with instruments was excited in its first natural frequency until a failure of the welded seam between base plate and pipe was produced (Figures 1 and 2). This damaging could be traced. The failure of the welded seam could be implemented in the FEM model of the pipe by model updating.

At the bottom of the upward branch, the piping system was welded on a base plate (Figures 1a and 2a); on the right side it was held vertically. Directly following the descending branch, the excitation was applied by means of a shaker (Figure 1c). Within the scope of a FEM analysis, the first vertical natural frequency was determined to be 5.09 Hz. An attempt was made to excite the pipe in this resonance frequency continuously [3]. By this directed excitation, a failure of the welded seam at both sides of the base point (Figure 2) could be produced after approximately 30 000 load cycles. Despite continuing excitation, the amplitudes at all measuring points of the mock-up decreased significantly (Figure 3a). By this damage—representing a considerable loss of stiffness at the clamping—the natural frequencies of the system were shifted downward. For instance, the first natural frequency at 5.09 Hz was shifted to 3.59 Hz. The shift of natural frequencies can clearly be seen in the waterfall diagram shown in Figure 3(b).

This measurement data then served as input for the model update. At the hot spots (high-stressed areas) of the pipe where the damage was expected, unsafe model regions were defined. The restraints as well as the elbows were used for this purpose. If the stiffness of an element changes due to a flaw, this also influences the stiffness and, therefore, the natural frequencies and modal matrix. Thus, the software must be capable to identify the flaw and its location. At the destruction of the clamping, a decrease from 5.09 to 3.59 Hz (30%) of the first vertical natural frequency was observed in that system. After approximately 18 iterations within the model updating program, the natural frequencies of the *measured* data and the changed *analytical* model matched, meaning that the model is successfully updated.



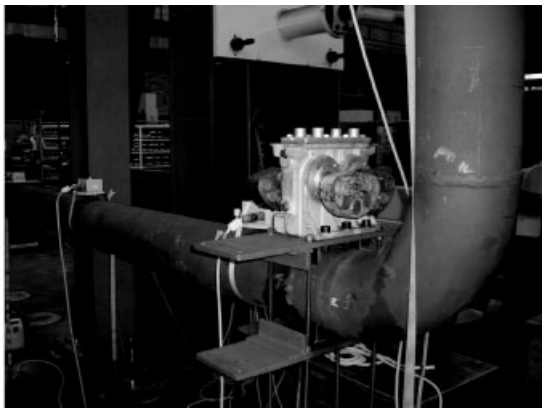


Figure 1. Steel pipe mock-up with sensors (indicated by MP x), and shaker. (a) Sketch of mock-up, (b) photo of mock-up, and (c) photo of the shaker.

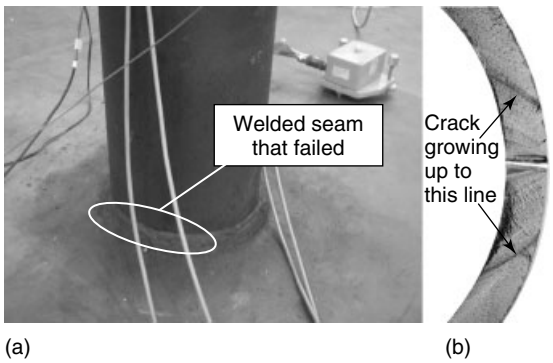


Figure 2. (a) Failure of the welded seam caused by shaker excitation. (b) The remaining load-carrying cross section at the end of the test.

2.3 Example B: modal analysis based on ambient vibrations

In this case, no artificial excitation of the system is necessary, but the stochastic excitation by operational oscillations and oscillations of the environment—so-called ambient vibrations—is sufficient to excite the system even during revision of the facility. For this purpose, highly sensitive seismic velocity sensors are used. This analysis method is safer than conventional methods where artificial excitation, e.g., by shakers or impulses, might damage the investigated structure. In special cases, measurements can be taken directly on

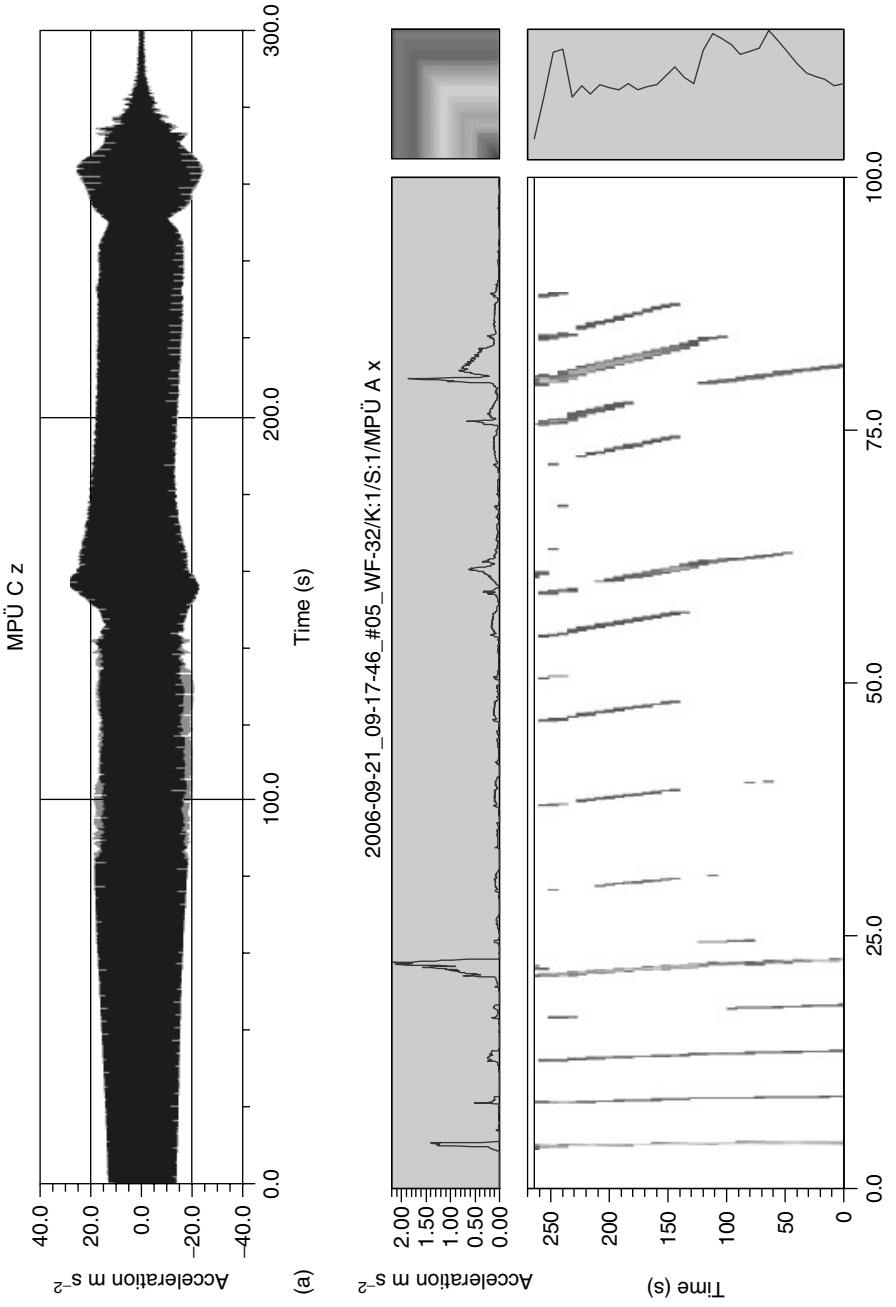


Figure 3. Modal analysis results at the pipe mock-up. (a) Reduction of accelerations at measuring point C. During this process, the natural frequencies decreased with progressing crack propagation. Higher frequencies are more sensitive to changes in the stiffness (b).

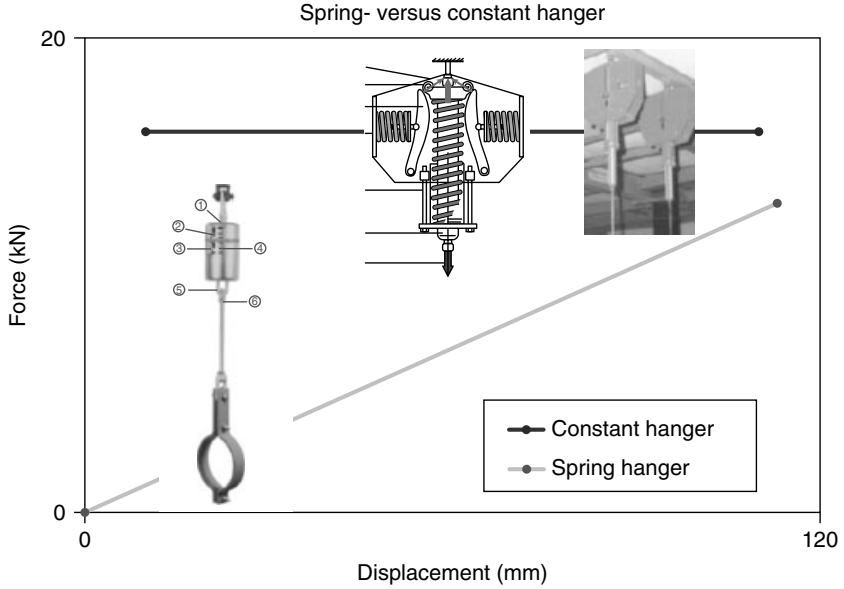


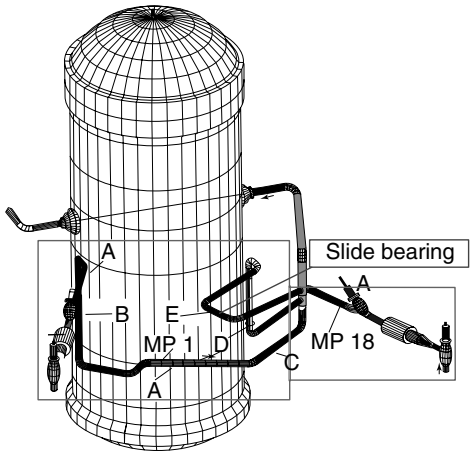
Figure 4. Load-bearing behavior (load-displacement curve) of typical piping supports such as spring hangers and constant hangers.

the insulation of a pipe. As no artificial excitation is necessary, the expenditure is limited. With this excitation mechanism no "force inputs" are necessary, only "outputs" coupled by a reference signal are used [4, 5]. This technique is very well known and is established as "OMA method" (see Section 2.1.3). The number of the usually expensive transducers can also be limited by this method. Groups of measuring points are recorded by only a few transducers, then shifted with respect to time and later on coupled by a reference signal.

Within the scope of a national research project [6], the potential of the method for identifying current conditions of piping systems was demonstrated especially regarding the state of spring hangers and constant hangers. The hangers are supports of the piping to bear the loads (dead loads, loads due to thermal expansion, earthquake loads, etc.). They are situated between piping and building. A spring hanger consists of a mechanical spring with a linear force-displacement behavior. The constant hanger consists of a mechanism, which transfers only constant loads between piping and building within a defined range independent from the displacements between piping and building. This is often meaningful because of large piping movements due to elongations caused by rare temperature differences (Figure 4). The analysis of these tests revealed the following:

- The results of the FFT analysis (fast Fourier transform) yield definite spectral densities of the measured velocities at all measuring points.
- Changes in systems with spring hangers could be recorded with high accuracy. Measurements and calculations coincide.
- The simulated locking of a spring hanger leads to a change in the mode shapes.
- An amplitude dependency of some mechanical hanger-parameters complicates the updating of the calculated model to the experimental modal results.
- The locking of a constant hanger leads only to a small change of the natural frequency of the investigated system; however, the resulting nodal point generates a characteristic mode shape.

Another task during an application to a piping system with nominal diameters DN 350/DN 250 in a nuclear power plant in Germany (Figure 5) was to prove whether the calculated frequencies and related modes are actually present at the real structure, especially the mode shapes in the range of 10 Hz. Precalculations led to the result that at this frequency



Emergency cooling piping
20TH24 and 20TH34
Measured area:
MP 1-T-MP 18
Boundary conditions:
(A) Z-stop:
$c_z = 1.3 \times 10^5$ N mm^{-1},
(B) Parallel springs:
$c_z = 2 \times 66.6 \times 10^5$ N mm^{-1},
(C) Z-stop:
$c_z = 1.3 \times 10^5$ N mm^{-1},
(D) Parallel struts:
$c_z = 2 \times 2.8 \times 10^4$ N mm^{-1},
(E) Slide bearing:
$c_z = 2.66 \times 10^4$ N mm^{-1}

Figure 5. Emergency cooling piping (20TH24 and 20TH34) and measured area from measurement point 1 (MP 1) to measurement point 18 (MP 18). Locations and type of hangers are given on the right.

level the switch-off of the high-pressure pump of the emergency core cooling system induces resonance effects and, therefore, a comparison of measurement and calculation was of great interest.

The area of measuring points is marked in Figure 6(a). The support of the piping system consists of the quasi-fixpoints (FP) at the reactor pressure vessel (RPV) and at the two penetrations through the containment, of a slide bearing at measuring point 14, and of struts and spring hangers. The piping system was partly insulated between the tee and the RPV. By FFT analysis of the single signals the natural frequencies can be obtained. From the transfer functions between these signals and a reference point signal, the mode shapes can be determined. For this purpose, a 3-D-transducer was installed at measuring point 8 as a reference point during all measurements.

The natural frequencies were derived from the spectral densities of the velocity signals. The corresponding experimentally determined first mode shape is shown in Figure 6(b). The list of all measured frequencies is documented in the first row of Table 1. The first comparison between measurement and design calculation did not show a satisfactory agreement (Table 1, columns 1 and 2). A view on the measured mode shapes leads to the conclusion that at the low-excitation level applied, sliding at the single slide bearing did not occur. A modification of the calculation model (Table 1, column 3) also did not lead to the desired coincidence. Some frequencies coincide, but not the first one. Only the mode shapes of the first two frequencies look similar.

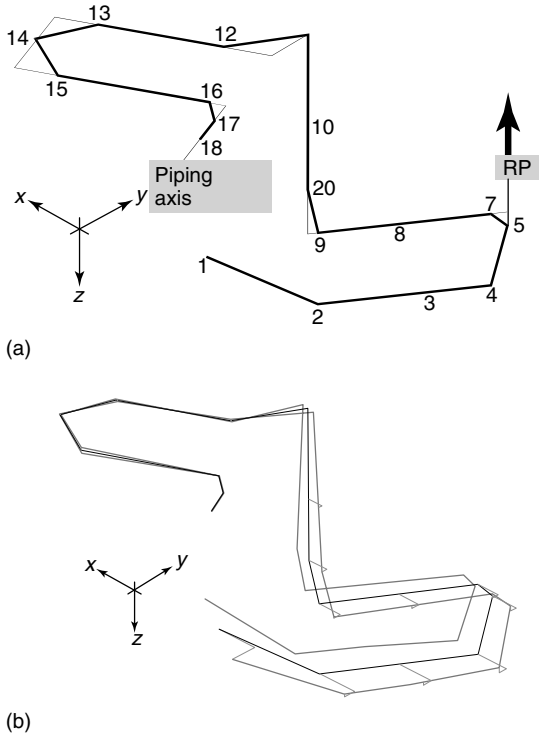


Figure 6. Location of measurement points at the emergency cooling piping (a) and experimentally determined vibration mode at 2.5 Hz (b).

Table 1. Comparison of measured frequencies at a nuclear power plant with those of different FEM calculations

1	2	3	4
Measurement (Hz)	Design calculation (Hz)	Model—modification slide bearing → fix point (Hz)	New FEM calculation (Hz)
2.5	1.43	1.48	2.4
	2.33	2.51	
	2.75		
3.8	3.26	3.19	3.6
	4.18		
5.2	5.35	5.24	6.1
	6.16	6.26	
6.8	6.56	6.44	6.9
8.1	7.71	7.76	7.5/8.3
9.1	9.49	9.51	8.8
10.0	9.96	10.12	9.7/11.9

These large differences are due to the typical construction of a hanger support as shown in Figure 7. From this picture, it becomes clear that the support with two struts with two single-edge-connected hinges has not only vertical direction stiffness values as assumed in the design calculation but also stiffness values in other directions (horizontal translation and rotation around the pipe axis). This implies overall that the pipe is supported stiffer than assumed in the design calculation. For this reason, a parameter study based on a finite element calculation with the code ABAQUS was carried out with a more detailed model in which such constructions were modeled with more suppressed degrees of freedom according to present construction drawings. This led to an increase in the frequency of the first mode compared to the model of the design calculation. One result is shown in Table 1, column 4. Also, the number of the calculated natural frequencies, therefore, declines in the frequency range of 0–10 Hz, and now the mode shapes agree with the measurements quite well. Further experiences with OMA were gained during the European project SafePipes [7].

3 THE SHM "LOCAL EYE": GUIDED ELASTIC WAVE MONITORING

For crucial error-prone parts of a structure, vibration monitoring can be efficiently supplemented by using elastic waves in the kilohertz frequency range. These ultrasonic waves have a shorter range but are more sensitive to smaller defects and thus can serve as an early warning system raising an alarm long before critical damage occurs. If the wavelengths are comparable with or larger than typical dimensions of the structure (e.g., thickness), the waves are called *guided waves*. In this case, geometrical dispersion cannot be neglected in general (see e.g., [8]).

When using elastic waves for SHM purposes, two different approaches are possible, a passive and an active approach. In the passive SHM system, only sensors are needed and "natural" sources like acoustic emissions (AEs) caused by crack generation and growth are detected. In the active SHM

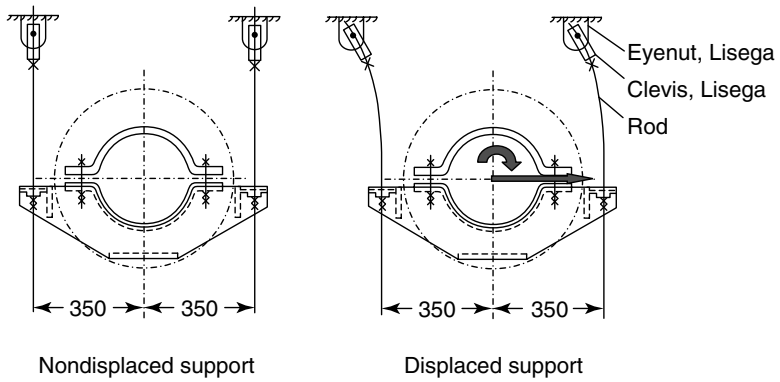


Figure 7. Typical construction of a hanger of the investigated piping system with two struts. The deflection picture on the right shows that additional acting stiffness is present.

system, the transducers are acting as both, sensors and actuators. By using pulse echo or acoustic signature techniques, scattered waves from inside the structure or the changes in acoustic signature response can be detected and used as damage indicator.

In order to implement a monitoring system based on guided waves, the theoretical fundamentals of guided-wave propagation in various materials and structures and their interaction with potential defects have to be investigated first. This can be done via numerical simulation (Figure 8) or by laser detection of elastic wave fields at the surface of the structure.

The simplest case of guided waves can be found in platelike structures where the so-called plate waves or Lamb waves exist. In general, symmetric and antisymmetric wave modes are being distinguished. They are dispersive in general. In most cases, SHM techniques work in the frequency range between 50 and 500 kHz and only the zeroth-order Lamb waves are of particular interest for monitoring applications. In addition to the Lamb waves, horizontally polarized shear waves (SH waves) can also be used. In contrast to the Lamb waves, the zeroth-order SH wave is nondispersive. Numerical and experimental investigations show that each wave mode mentioned earlier shows different sensitivity to specific kinds of damage. The SH0 mode is well suited for crack detection and for any application where a surrounding fluid limits the range of the other modes. The antisymmetric A0 Lamb mode is best suited for determination of delaminations in composites and local changes in wall thickness. The symmetric S0 mode is also suited for crack detection. Since it often represents the fastest wave mode in the structure, it is typically used in cases where first-arrival time picking together with a clear identification of the incoming wave is necessary.

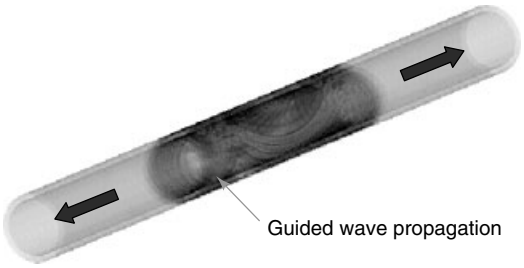


Figure 8. Guided elastic wave propagation in a steel pipe generated by a mechanical radial force point impact on the outer surface of the pipe. The wave front picture was calculated by using the 3-D cylindrical elastodynamic finite integration technique (CEFIT, [9]) and show the helical nature of the elastic wave field. Further interesting numerical and analytical investigations of guided waves in free pipes can be found in [10, 11]. [Reproduced from Ref. 9. © Elsevier, 2004.]

The main difference between a pipe and a plate is the curvature of the pipe producing additional dispersion effects. In guided-wave theory of cylindrical shells, the following naming convention for the different wave modes is commonly used:

1. *Longitudinal* modes are named $L(0, m)$ with $n = 0$ indicating an *axisymmetric* mode and $m = 1, 2, \ldots$ indicating modes of order 1, 2, etc.
2. *Torsional* modes are named $T(0, m)$ with $n = 0$ indicating an *axisymmetric* mode and $m = 1, 2, \ldots$ indicating modes of order 1, 2, etc.
3. *Flexural* modes are named $F(n, m)$ indicating non-axisymmetric modes with $n, m = 1, 2$, etc.

Up to a certain frequency threshold only three wave modes are present, $L(0,1)$, $L(0,2)$, and $T(0,1)$. Their flexural counterparts are $F(1,1)$, $F(1,3)$, and $F(1,2)$. These basic pipe modes can be associated with the fundamental plate modes, i.e., the fast symmetric and weakly dispersive S0 mode ($\cong L(0,2)$), the slow antisymmetric and dispersive A0 mode ($\cong L(0,1)$), and the nondispersive SH wave ($\cong T(0,1)$). Above the frequency threshold mentioned above, higher order modes arise ($T(0,2)$ and $L(0,3)$ in this case), similar to plate diagrams. For the sake of simplicity and clarity, we call these modes "pipe-S0", "pipe-A0", and "pipe-SH" or shorter, P-S0, P-A0, and P-SH in the following, although this procedure is not quite correct formally.

Another peculiarity of wave propagation in a pipe is that waves generated by pointlike sources are propagating along a helical curve around the longitudinal axis. This is shown in Figure 8, where the transient wave field due to a point impact on the outer pipe surface was calculated by using the numerical 3-D CEFIT technique [9]. As a consequence of helical-wave propagation, one and the same wave mode can be detected several times at a certain sensor position since different travel paths from the source to the sensor exist.

In order to demonstrate the existence of different wave modes, various measurements were performed

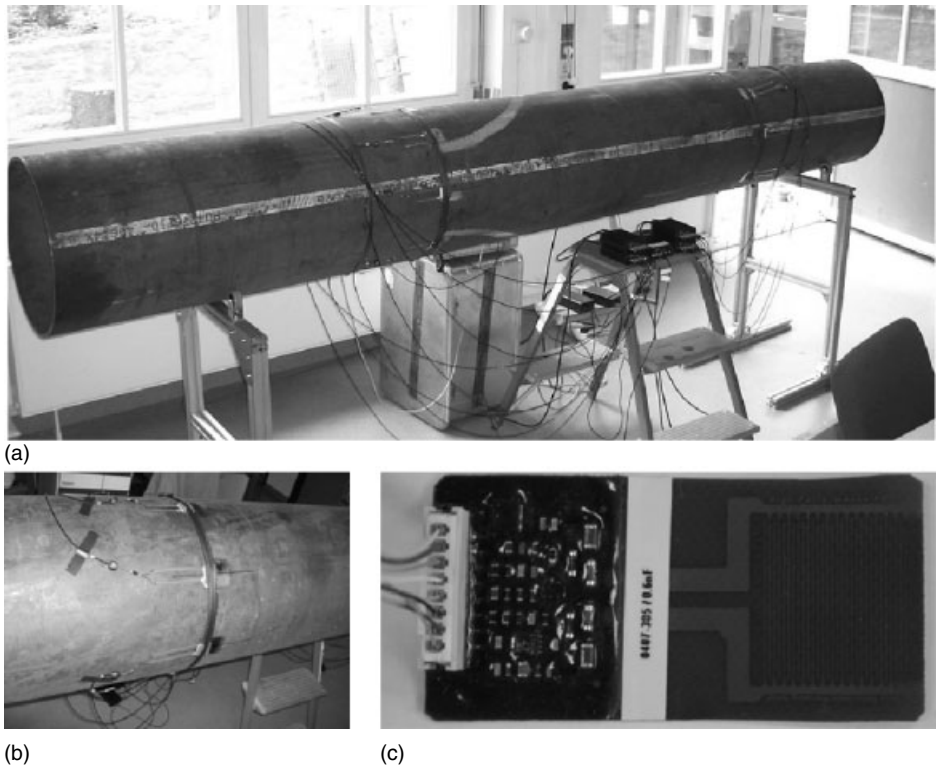


Figure 9. (a) Steel pipe as used for the laboratory measurements (length = 3 m, diameter = 406 mm, wall thickness = 9 mm). (b) In the present case, PZT fiber transducers are glued to the outer pipe surface (b and c). These transducers are characterized by a preferential directivity along the pipe axis and can be used for low-temperature applications up to 100 °C and for frequencies up to 600 kHz. (c) The picture shows a variant where PZT part and preamplifier are combined within one single transducer patch.

using the laboratory setup shown in Figure 9(a). For this low-temperature application, specifically adapted lead–zirconate–titanate (PZT) fiber transducer patches (Figure 9b and c) have been used to excite Ricker wavelets [12] with center frequencies of 70, 126, 240, and 370 kHz, respectively. The detected waveforms at a sensor position lying 150 cm away from the source are given in Figure 10. It should be noted that the time axis is given with a constant offset of +68 μs, i.e., the arrival times have to be corrected for that value before calculating the corresponding wave speeds.

In Figure 10, the first arrival is due to the fastest wave mode present, i.e., the P-S0 mode (or $F(1,3)$). The second elementary wave mode in the first two rows of Figure 10 can be associated with the P-A0 mode (or $F(1,1)$). If the center frequency of the input pulse is increased to 240 kHz, a new wave mode suddenly occurs. It is strongly dispersive and—due to its group velocity—can be identified as the first higher order symmetric mode P-S1 (or $F(1,5)$).

Besides these primary wave modes; secondary wave modes also appear. At 460 μs (−68 μs offset) a wave mode appears at 70 and 126 kHz that is in line with the first-order helical P-S0 mode whose travel distance between the source and receiver is approximately 2 m instead of 1.5 m for the direct wave. The same effect can be observed for the P-A0 mode whose first-order helical counterpart arrives at approximately 700 μs (−68 μs offset). In both cases, the signal shape of direct and helical wave is similar but the amplitude of the helical wave is smaller due to the larger geometrical spreading along the longer propagation path.

Another interesting echo can be found at 665 μs (−68 μs offset). Since the distance between each

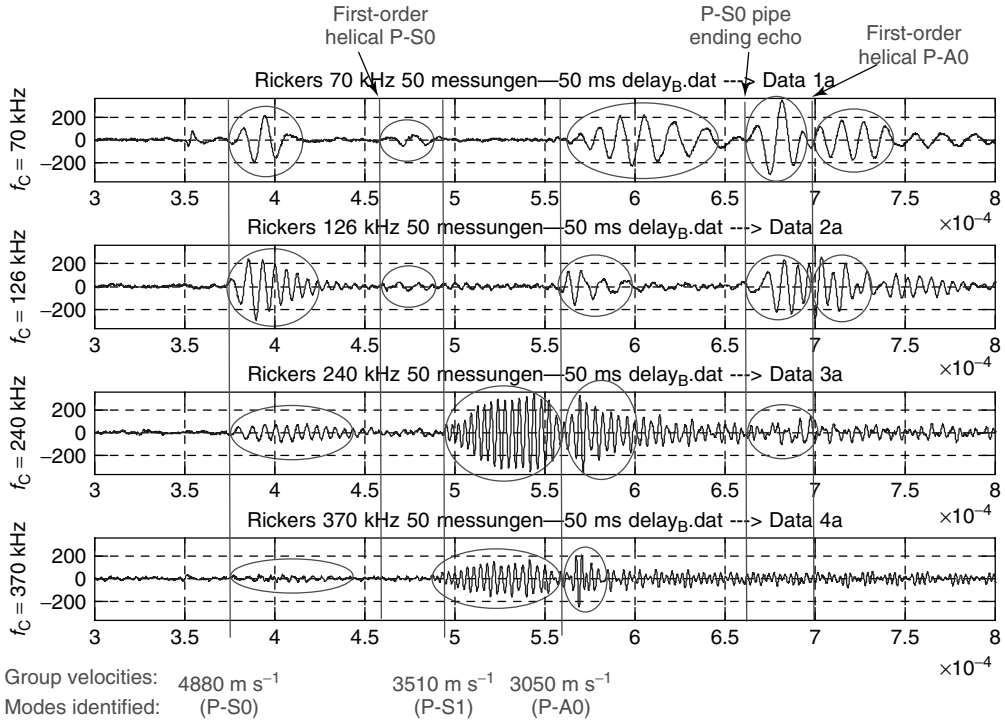


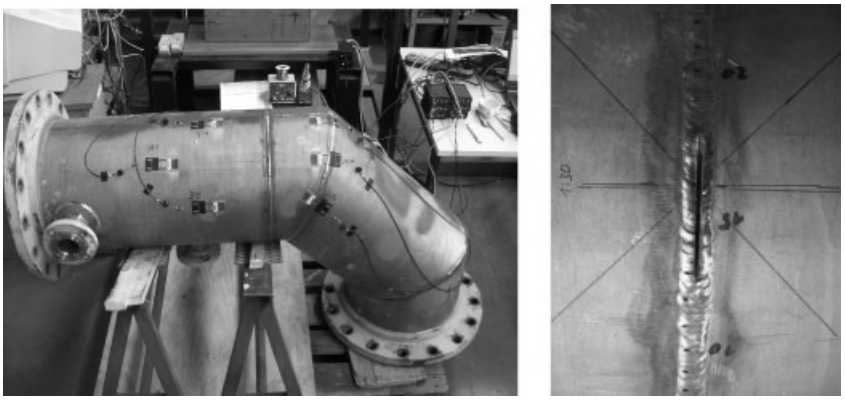
Figure 10. Measured time-domain signals along the steel pipe for four different center frequencies of the input pulse. All elementary wave modes as predicted by theoretical and numerical investigations (including helical modes) could be identified.

of the two transducers to the nearest pipe ending was 75 cm, a wave reflected at the pipe endings reaches the sensor after a propagation path of 3 m. Thus, the wave described above can be identified as the P-S0 echo of the pipe ending. Owing to reflection at the free end and the fact that two different propagation paths with identical length contribute to the signal ("actuator → right pipe ending → sensor" and "actuator → left pipe ending → sensor"), the amplitude of the echo is larger than the amplitude of the primary wave.

3.1 Interaction of guided waves with defects

The interaction of elastic waves with structure-relevant defects represents the most important aspect of guided-wave-based monitoring systems. In general, one can summarise that the higher the frequency the better the spatial and temporal resolution of the monitoring system and the better the sensitivity to small defects. However, for high frequencies, the number of existing wave modes is increased and strong dispersion leads to a complex situation and a severely limited range. The lower the frequency, the smaller the number of wave modes and the larger the obtainable range. However, these advantages are cancelled by a significantly lower sensitivity to small defects. Therefore, to choose a specific frequency for the input pulse always means to make a compromise between flaw sensitivity, on one hand, and obtainable range and dispersion of the corresponding wave modes on the other hand. By using traditional ultrasonic NDE systems working in the megahertz frequency range, very small cracks in the micrometer range can be found but the structural information is usually limited to a small area of the pipe. In order to test the entire pipe, an appropriate scanning device is necessary.

In guided-wave-based SHM, a larger part of the pipe can be examined within one measurement cycle.



(a) (b)

Figure 11. (a) Titanium elbow (wall thickness = 6 mm) (b) with notches artificially introduced into a weld seam. The notch depth was varied between 1 mm (first step) and 5.3 mm (last step). Owing to the sawing, the initial notch length of approximately 12 mm was also increased stepwise.

Owing to the fact that monitoring can be performed in nearly arbitrary time intervals of a few seconds up to a few days, the critical size of defects that have to be found can be increased compared to traditional NDE, which is usually applied in periodic intervals of months or years. Statements from industrial partners revealed that in a typical pipe, cracklike defects, not larger than three times of the wall thickness and not deeper than One-third of the wall thickness, have to be found by an SHM system. For a steel pipe as described earlier (wall thickness = 9 mm), this means that the monitoring system must be able to find cracks not larger than 27 mm and not deeper than 3 mm. As a rough rule of thumb, a defect becomes detectable if its lateral size is at least comparable to the wavelength of the specific wave mode used for the measurements and if its depth is larger than 10–15% of the wall thickness. According to this rule, guided waves in the frequency range between 100 and 200 kHz with wavelengths between 20 and 50 mm as described earlier should be able to meet the requirements of defect sensitivity, on one hand, and sufficiently large range on the other hand.

The main idea of the underlying SHM system is the comparison of the actual state of the pipe with a certain reference state ("baseline approach"). This reference state can either be the pipe without any defects or, alternatively, a state with smaller defect size. For the present investigation at a titanium elbow as shown in Figure 11(a), a reference measurement without any defects was performed first. After that the measurements including the defects were done and compared to the reference state.

In one of the weld seams of the elbow, artificial notches were inserted by sawing (Figure 11b). The depth of the notch was increased from 1 to 5.3 mm in discrete steps. Owing to the sawing, the length of the notch was also increased simultaneously from approximately 12 to 35 mm. For each state, the system response of the elbow due to pulse excitation was determined for different actuator/sensor combinations and was compared with a reference state without defects.

Figure 12(a) shows a typical time-domain signal obtained by using one PZT transducer as actuator and another one as sensor. The center frequency of the input pulse was $f = 150$ kHz in this case. For a 1-mm-deep notch (results not shown here), the difference to the reference measurement without notch is rather small and thus, the linear correlation coefficient between both curves is nearly equal to one. For the 3.5-mm notch as shown in Figure 12 the difference between the curves is significantly larger and, therefore, the correlation coefficient drops to approximately 0.97. For calculation of the correlation coefficients, we used the Hilbert envelope of the signals instead of the signals themselves due to higher robustness.

The measurements were repeated for different notch depths and various center frequencies of the

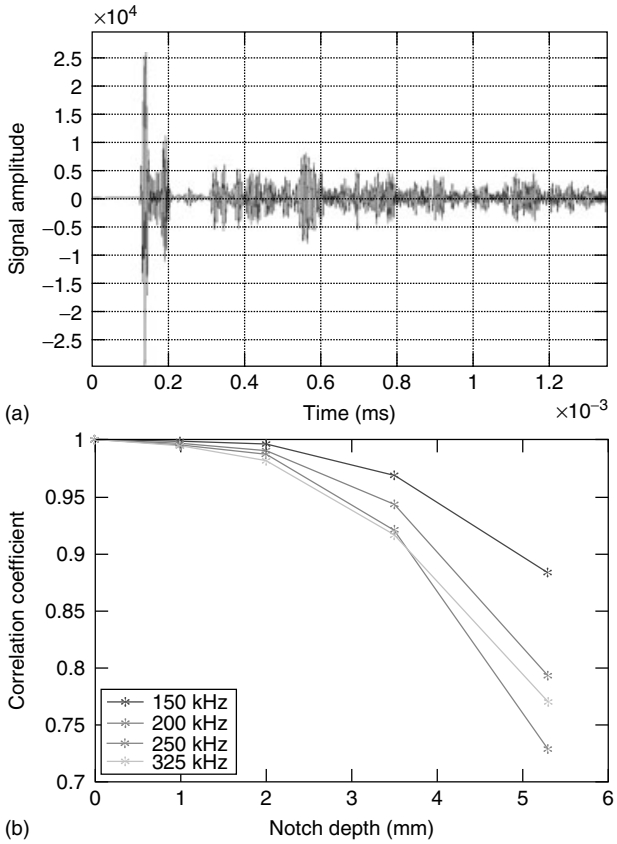


Figure 12. Data evaluation with baseline approach. (a) Typical guided-wave system response of the titanium elbow for a 3.5-mm-deep notch together with the reference signal of the system without notch (the two curves are only discriminable in the colored online version of the Figure). (b) Typical correlation coefficients as a function of notch depth (horizontal axis) and center frequency of the input pulse (different curves), obtained by comparing the particular Hilbert envelopes of the time-domain response with the reference measurements without notch. As a general trend, the correlation coefficient decreases with increasing notch depth and increasing frequency and thus serves as a sensitive damage indicator.

input pulse as well as for different (short and long) propagation paths. In Figure 12(b), the results of the correlation analysis for the short propagation path are displayed for notch depths of 1, 2, 3.5, and 5.3 mm, and for frequencies of 150, 200, 250, and 325 kHz, respectively.

As a general trend, we find that the correlation coefficient decreases with increasing notch depth and also with increasing frequency. These results are physically plausible since in both cases the interaction between guided waves and the defect is enlarged. For the short propagation path, the correlation coefficient drops to values between 0.88 and 0.73. For a longer path (curves not shown here), the correlation coefficient drops to values between 0.931 and 0.937, which is clearly smaller but still significant, even for the lowest frequency.

The results revealed that under laboratory conditions, cracklike defects having a depth of only One/fifth of the wall thickness can be detected by such a monitoring system even in complex geometries with flanges, weld seams, and curvatures, and in cases where the wavelengths are not significantly larger than the lateral size of the defect and where the distance to the source and receiver is rather large. In field tests, a worse signal-to-noise ratio can be expected due to background noise and other disturbances. However, owing to the fact that the coupling

conditions remain constant and the excitation is reproducible, the measurements can be repeated many times in order to increase the signal-to-noise ratio until it reaches an acceptable level.

4 HARDWARE AND GENERAL CONCEPTION OF THE SHM SYSTEM

In order to combine global and local condition monitoring as described in Sections 2 and 3 within a joint SHM system, multichannel network nodes have been designed and developed (Figure 13). Each single node represents a four-channel monitoring system, which is able to manage different kinds of actuators and/or sensors (for low- and high-temperature applications and for low- and high-frequency monitoring). Two different types of nodes are available, one for high-frequency guided-wave monitoring (the local eye) and another type for low-frequency vibration monitoring (the global eye). Each single node can be combined with several others (local or global) resulting in a multichannel monitoring system.

One node consists of four analog input channels per module having a bandwidth of 10 mHz to 10 kHz for the global eye and 20–1000 kHz for the local eye. In the latter case, the sampling is up to 18 Megasamples per second. A 32-bit fix point digital signal processor (DSP), an arbitrary waveform generator, a power amplifier as well as a hardware trigger with synchronization are included. A CAN-Bus interface is normally used but wireless Bluetooth, ZigBee, and nanoNET interfaces are under development. The power supply is 24 V dc. The measured data can be stored on movable memory cards or can be transferred online via USB-Ports.

As explained earlier, the key concept of the new SHM system is the combination of low- and high-frequency monitoring as demonstrated in Figure 14. While the passive low-frequency nodes with acceleration and strain sensors are responsible for global vibration monitoring of the structure (providing information on basic boundary conditions caused by supports and dampers) the active high-frequency guided-wave nodes are used to monitor local crucial and error-prone parts of the piping system. Both levels, i.e., local and global ones, will be combined adaptively dependent on the actual state of the piping system.

The monitoring results from both types of nodes finally merge in a joint decision support system (DSS), which should contribute to the following points in its final version:

1. identification of defects: raise an alarm if a defect is present;
2. localization of defects: if a defect is present, determine its (approximate) position;
3. relevance of defects: determine the kind of defect and state if the defect is relevant for the structural integrity of the pipe; and

Figure 13. Four-channel sensor/actuator nodes for SHM of piping systems. The nodes can be combined to a multichannel measuring system that can be used for both active and passive monitoring. The present nodes are based on a CAN–Bus interface but wireless interfaces are also under development. Details of the hardware implementation can be found in Section 4 and in [13]. [Reproduced with permission from Ref. 13. © SPIE, 2007.]

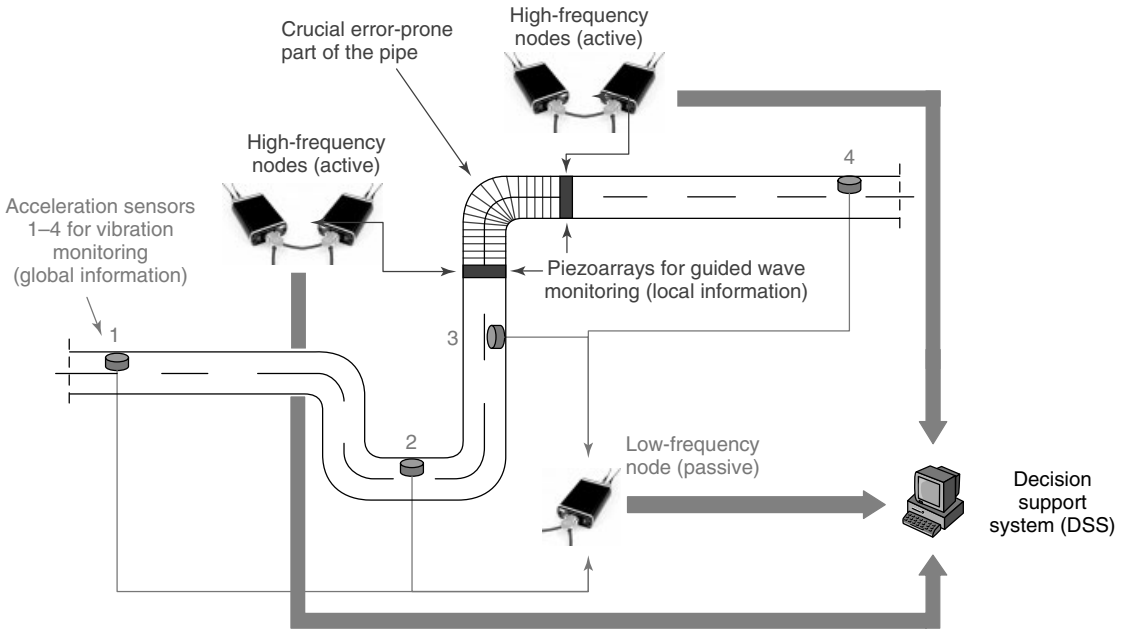


Figure 14. Combination of low- and high-frequency monitoring. While the passive low-frequency nodes with acceleration and strain sensors are responsible for global vibration monitoring of the structure (providing information on basic boundary conditions caused by supports and dampers) the active high-frequency guided-wave nodes are used to monitor local crucial and error-prone parts of the piping system.

4. residual lifetime: try to estimate the remaining lifetime of the structure.

The last two points in the preceding list are embedded in an overall lifetime management of the whole structure and should, therefore, be closely connected with former experience and databases of the plant operator. For this purpose, the DSS will be based on a case-based reasoning approach in order to allow for a close interaction between operator and monitoring system.

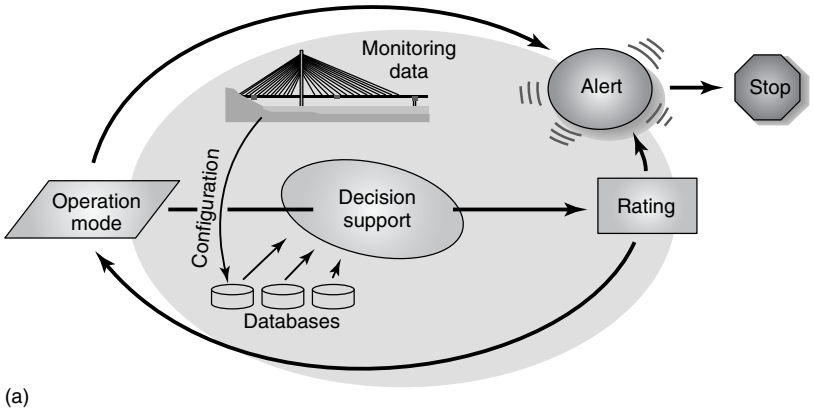
5 DECISION SUPPORT SYSTEM

During the SAFE PIPES project, a DSS called *VCDECIS* has been developed for the assessment of industrial piping systems (Figure 15a). VCDECIS combines previous knowledge with actual measurements (e.g., of ambient vibrations [14]) and offers a variety of options for assessment and damage-detection procedures. It provides all the necessary tools for decision making when monitoring data are available. It allows a combination of various methodologies and provides options for the compensation of environmental influences like temperature, external loading, or radiation from sunshine.

The system consists of five modules which are as follows:

- the operation module (operators tool)
- the databases
- the decision support module
- the alert system
- the output module (takes action if appropriate).

Here, we only focus on the decision support and the alert system. The DSS is defined by rules, which allow an assessment on finding facts and indicators in the available data. Some of the methodologies are complementary and it is intended to conduct parallel processing and comparison of the results. The building of a mean also might make sense in several cases. The rules are the core of the assessment system. They consist of methodologies developed in order to get the maximum information out of the data. The set of rules is permanently updated and enlarged.



(a)

Risk level	Why	What to do
Low	Info	Regular operation
Moderate	Info	Long-term action
Considerable	Show development	Mid-term action
High	Demonstrate risk	Immediate action
Extreme	Show default	Automatic alert, action

(b)

Figure 15. (a) System architecture of the decision support system VCDECIS and (b) the implemented risk levels with cause and consequence. The coloring follows the international practice, i.e risk level Low = green, Moderate = yellow, Considerable = orange, High = red, Extreme = framed blazing red. The colors are only displayed in the coloured online version of the Figure.

The system is started through a configuration file, where the user is asked which methodologies shall be applied and which combinations are desired. A default menu is available that directs the user to a standard evaluation. Each assessment step provides results, which are then stored in the knowledge database for statistical use.

There is an option to go with the process through a neural network that has been trained by the previous applications. Nevertheless, this step is not fully developed, yet owing to insufficient training possibilities for the network and ongoing development work, it is anticipated that support can be provided in the future for the user by this system.

Information on the risk level is also provided to the user. The system has been chosen in connection to alert systems used for natural hazards. Landslide or avalanche alert systems know five distinct risk levels, which are shown in Figure 15(b). When risk level 4 is reached, a consultation with an expert is proposed. The alarm is provided to the operator, who is asked to consult an expert on the observed phenomena. After that the risk level can be up or downgraded. This is the only active interference option.

When risk level 5 is reached, automatic action by the system can be triggered. This could be eventually a red traffic light at the approach to a pipe monitored. The other risk levels provide information for the operator and also ask for his input. As explained in the operation mode, the operator has the opportunity to interfere with the system here and to add his subjective impression to the process. The actions to be performed at each level will have to be defined from case to case. It has to be coordinated with the standard procedure of the plant owners.

6 LIFETIME MANAGEMENT

Lifetime management (Figure 16a) stands for the integration of aging management and economic planning for systems, structures and components (SSC) and, therefore, also for piping systems in power

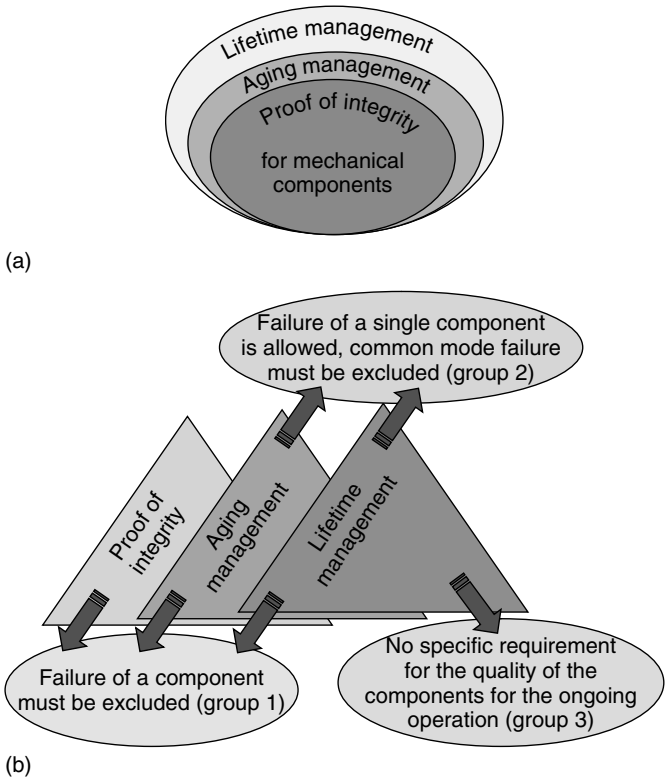


Figure 16. Concept of lifetime management of industrial piping systems. (a) Correlation between lifetime management, aging management, and proof of integrity for systems, structures, and components (SSC); and (b) application of lifetime management, aging management, and proof of integrity for SSC of groups 1, 2, and 3.

plants. The main goals of lifetime management are to optimize the operation, the maintenance and the lifetime of the plants, to maintain an accepted level of safety and performance, and to maximize return on investment over the lifetime of the plant.

Lifetime management has very strong and clear definitions in the field of nuclear power plant, which is shown as follows and reported in [15–19], for example. Various engineering measures are required depending on the safety relevance of the SSC or for reasons of preventive maintenance. Consequently, the SSC have to be divided into three groups (Figure 16b).

The first step within the scope of lifetime management of mechanical components is to select and arrange the SSC and to assign these to group 1, 2, or 3. The classification is according to the requirements of the nuclear codes and standards and if necessary according to plant-specific and safety-related factors. The plant operator is responsible for the classification and an expert has to check it on the basis of the current codes, standards, and the state of the art.

- **Group 1**

Failure of the SSC shall be excluded to avoid subsequent damage, e.g., main coolant lines (MCL). The required quality shall be guaranteed for subsequent operation. The causes of possible in-service damage mechanisms shall be monitored and controlled (proof of integrity). Implementing this "proactive approach" prevents damage.

- **Group 2**

For redundant SSC, the failure of a single part is allowable from a safety relevant point of view. However, common mode failure shall be excluded.

The present quality shall be maintained for subsequent operation. The consequences of possible in-service damage mechanisms shall be monitored (preventive maintenance, time or condition based).

- **Group 3**

There are no defined standards for the quality of the SSC concerning subsequent operation (failure-oriented maintenance).

These guidelines, codes, and standards concern safe operation during the total lifetime (lifetime management), safety against aging phenomena (aging management) as well as proof of integrity (e.g., break exclusion or avoidance of fracture). Within this field, the aging management is a key element. Depending on the safety relevance of the SSC under observation including preventive maintenance, various tasks are required, in particular, to clarify the mechanisms that contribute system-specifically to the damage of the components and systems and to define their controlling parameters, which have to be monitored and checked. Appropriate continuous or discontinuous measures are to be considered in this connection.

7 CONCLUSIONS AND OUTLOOK

In this article, only a short extract of the results obtained within the SAFE PIPES project could be given. The whole results obtained so far revealed that both global vibration monitoring and local guided-wave monitoring of industrial piping systems provide complementary information about the structure under investigation. While vibration monitoring is able to characterize the global condition of the structure due to supports and dampers, the guided-wave module is able to find small cracklike defects and corrosive material degradation in crucial error-prone parts of the components. It can, therefore, be expected that a combination of the "global" and "local eye" will lead to a monitoring system with significantly increased performance and reliability. The development of a joint DSS that evaluates both low- and high-frequency data and that is embedded in an overall lifetime management of industrial piping systems still needs to be completed and is therefore the subject of ongoing work within the SAFE PIPES project.

ACKNOWLEDGMENTS

The present work has been supported by the Commission of the European Communities in the framework of the specific targeted research project SAFE PIPES (Safety Assessment and Lifetime Management of Industrial Piping Systems) under the sixth framework program (NMP2-CT-2005-013898) as well as by the German Bundesministerium für Bildung und Forschung (BMBF) and the Bundesministerium für Wirtschaft und Technologie (BMWI), respectively. This support is gratefully acknowledged. We also thank all our partners in the SAFE PIPES consortium, especially NMW Würzburg for providing us with the PZT fiber transducers and DOW Stade for supply of the titanium elbow.

RELATED ARTICLES

Free and Forced Vibration Models
Fundamentals of Guided Elastic Waves in Solids
Modal–Vibration-based Damage Identification
Ultrasonic Methods
Guided-wave Array Methods
Signal Processing for Damage Detection
Piezoelectricity Principles and Materials
Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors
Sensor Placement Optimization
Ambient Vibration Monitoring
Wind Turbines
Fatigue Monitoring in Nuclear Power Plants
Integrated Sensor Durability and Reliability

REFERENCES

[1] Link M. Updating of analytical models–basic procedures and extensions. In *Modal Analysis and Testing*, *NATO Science Series*, Silva JMM, Maia NMM (eds). Kluwer Academic Publishers, 1999.

[2] Link M, Boettcher T, Zhang L. Computational model updating of structures with non-proportional

damping. *Proc. of the International Modal Analysis Conference IMAC XXIV*, St. Louis, USA, 2006.

[3] Bauer J, Friedmann H, Henkel F-O, Lerzer M, Röhner J. *Model Updating Data and Results*. Part 2, Computational Model-Updating, Deliverable No 4 of the European SAFE PIPES project, FP6-STRP-013898, December 2007.

[4] Mattheis A, Trobitz M, Kussmaul K, Kerkhof K, Bonn R, Beyer K. Diagnostics of piping by ambient vibration analysis. *Nuclear Engineering and Design* 2000 **198**:131–140.

[5] Roos E, Kerkhof K. Zustandsanalyse von Rohrleitungssystemen mittels Modalanalyse bei unterschiedlichen Anregungsintensitäten. *Proceedings of VDI-Schwingungstagung 2000 "Experimentelle und rechnerische Modalanalyse sowie Identifikation dynamischer Systeme"*. VDI Berichte 1550, VDI-Verlag: Kassel, 2000, pp. 307–326.

[6] Kerkhof K. Integrity verification of safety-relevant piping by means of vibration analysis, phase ii: system investigations to determine stiffness of bearings indirectly. *Abschlussbericht zum BMBF/BMWI-Forschungsvorhaben Förderkennzeichen 150 1062*. Staatliche Materialprüfungsanstalt Stuttgart, November 2001.

[7] Kerkhof K, Schwenkkros J. In-situ investigations on the vibration behaviour of plug flow reactors including lifetime-management concepts. *Proceedings of International Marcus Evens Conference 'Process Safety Management and Loss Prevention'*. Geneva, 8–9 November 2007.

[8] Köhler B, Schubert F, Frankenstein B. Numerical and experimental investigation of Lamb wave excitation, propagation, and detection for structural health monitoring. In *Proceedings of 2nd European Workshop on Structural Health Monitoring*, Boller C, Staszewski WJ (eds). DEStech Publications: Lancaster, PA, 2004, pp. 993–1000.

[9] Schubert F. Numerical time-domain modeling of linear and nonlinear ultrasonic wave propagation using finite integration techniques—theory and applications. *Ultrasonics* 2004 **42**:221–229.

[10] Hayashi T, Rose JL. *Guided Wave Simulation and Visualization by a Semianalytical Finite Element Method. Materials Evaluation*. January 2003: 75–79.

[11] Li J, Rose JL. Excitation and propagation of non-axisymmetric guided waves in a hollow cylinder. *Journal of the Acoustical Society of America* 2001 **109**(2):457–464.

[12] Ricker N. The form and laws of propagation of seismic wavelets. *Geophysics* 1953 **18**:10–40.

[13] Frankenstein B, Hentschel D, Schubert F. Monitoring Network for SHM in Aircraft Applications. *Proceedings SPIE Conference on Smart Structures and Materials/NDE*. San Diego, CA, March 2007.

[14] Wenzel H, Pichler D. *Ambient Vibration Monitoring*. John Wiley & Sons: Chichester, 2005.

[15] Roos E, Herter K-H, Schuler X. Lifetime management for mechanical systems, structures and components in nuclear power plants. *International Journal of Pressure Vessels and Piping* 2006 **83**(10):756–766.

[16] Kußmaul K, Blind D. Basis safety—a challenge to nuclear technology. IAEA Spec. Meeting, Madrid, March 5–8, 1979. In *Trends in Reactor Pressure Vessel and Circuit Development*, Nichols RW (ed). Applied Science Publishers Ltd: Barking, 1979.

[17] Kußmaul K. German basis safety concept rules out possibility of catastrophic failure. *Nuclear Engineering International* 1984 **12**:41–46.

[18] IAEA. *AMAT guidelines. Reference Document for the IAEA Ageing Management Assessment Teams (AMATs)*. IAEA Service Series No. 4, March 1999.

[19] IAEA. *Implementation and Review of Nuclear Power Plant Ageing Management Programme*, Safety Report Series No. 15. IAEA: Vienna, 1999.

Chapter 81
Microelectromechanical Systems (MEMS)

Jonas Meyer, Reinhard Bischoff and Glauco Feltrin

Structural Engineering Research Laboratory, Empa, Swiss Federal Laboratories for Materials Testing and Research, Dübendorf, Switzerland

1 INTRODUCTION

Microelectromechanical systems (MEMS) are integrated devices or systems of devices that combine electrical and mechanical components and that have a size, which ranges from the submicrometer to the millimeter level. Miniature mechanical elements such as beams, diaphragms, and springs are fabricated by micromachining techniques from silicon wafers and are combined with microelectronic components and circuits to form microsensors, microactuators, and microengines. The microfabrication technology allows for fabrication of large systems of MEMS devices, which individually perform simple tasks, but in combination can accomplish complicated functions. MEMS allow sensing, controlling, and activating physical and chemical processes on the micro and, by a suitable combination of clusters of MEMS devices, also on the macro scale.

Typical examples of MEMS devices are accelerometers, gyroscopes, strain gauges, pressure and flow sensors, miniature robots, fluid pumps, microvalves, and micromirrors. The term *MEMS* was coined in 1987, long after the first micromachined devices, in particular, microsensors, were commercially available. Equivalent terms for MEMS are microsystems and micromachines.

The term micromachining designates the fabrication of micromechanical parts (such as diaphragms or beams). These parts were fabricated by etching away selected areas of the silicon substrate to obtain the desired micromechanical components. Since the early 1960s, various etching techniques were developed to improve the fabrication of micromechanical components and these techniques form the basis of the so-called bulk micromachining processing techniques. The need for higher design flexibility and better performance, however, gave rise to surface micromachining techniques, in which the so-called sacrificial layers are deposited between structural

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

layers for mechanical separation and isolation. These sacrificial layers are then removed by etching to free the structural layers and to enable mechanical components to move relative to the substrate. Surface micromachining enables the fabrication of complex multicomponent integrated micromechanical structures that would not be possible with traditional bulk micromachining. Details on MEMS and their fabrication technologies can be found in [1–3].

2 MECHANICAL MEMS SENSORS

MEMS devices for sensing mechanical quantities are the most important class of microsensors. The first fabrication of silicon-based MEMS devices started in the late 1950s with the development of pressure microsensors. In 1974, National Semiconductor launched the first high-volume pressure sensor in the market. Silicon pressure sensors are at present the commercially most important microsensor type with a billion-dollar market and large-scale technical applications in different industries like the automotive and aeronautical industry.

The discovery of the piezoresistive effect in silicon and germanium in 1954 enabled the development of silicon-based micromachined strain gauges with a gauge factor 10–20 times greater than those based on metal films. Micromachined piezoresistive strain gauges are now a standard component in accelerometers and pressure sensors.

2.1 Accelerometers

Accelerometers are the commercially second most important type of mechanical microsensors. The design principle of MEMS accelerometers is the same as traditional accelerometers: an inertial mass suspended by a linear elastic mechanical component (micromachined cantilever beam, bridge, or membrane). Accelerations cause inertial forces, which deflect the suspended mass from its zero position. This deflection is converted by a pickup to an electrical signal, which, after a suitable signal conditioning by an internal integrated circuit, appears at the sensor output. The two most prevalent pickup types of MEMS accelerometers are: capacitive and piezoresistive pickup of the seismic mass movement, where capacitive polysilicon surface-micromachined and single-crystal micromachined devices are the most important types.

The amplitude range of capacitive MEMS accelerometers varies between a few g up to $50g$ for applications in air bag systems. Currently, several low-g MEMS accelerometers are commercially available. Table 1 displays a selection of these accelerometers, which have the characteristics to be applicable in structural health monitoring (SHM).

2.2 Application aspects

MEMS sensors have several advantages compared to conventional sensors. They are small, generally low power, highly integrated, and, usually, cheap. These

Table 1. Selection of commercially available MEMS accelerometers

Product	ST microelectronics LIS2L06AL	Analog devices ADXL204	Colybris MS8002.C	Colybris SI-Flex SF1500S	PCB 3711D1FB3G
Number of axes	2	2	1	1	1
Amplitude range (g)	±2.0 (±6.0)	±1.7	±2.0	±3.0	±3.0
Bandwidth (Hz)	0–2000	0–2500	0–200	0–1500	0–100
Sensitivity (mV g^{-1})	660 (220)	595	1000	1200	700
Noise $\mu g/\sqrt{Hz}$	30	170	18	0.5	110
Temperature range (°C)	−40 to +85	−40 to +125	−55 to +125	−40 to +125	−54 to +121
Input voltage (V dc)	2.4–5.5	3–6	2.5–5.5	6–15	5–30
Power consumption (mW)	2.8	1.7	<2	>60	>50
Package size (mm)	5 × 5 × 1.5	5 × 5 × 2	14.2 × 14.2 × 3.8	24.4 × 24.4 × 16.6	21.6 × 21.6 × 11.4[a]
Weight (g)	0.08	<1	1.64	—	77.8[a]

[a] Rugged titanium housing.

qualities enable the deployment of SHM systems with a large number of small sensors, partly integrated into the structure, at affordable costs. When deploying a large number of sensors, the cabling of all sensors with data logging units becomes so labor and cost intensive that it cancels all the advantages of applying MEMS sensors. Therefore, to overcome the limitations of cabling, MEMS sensors are often used in combination with wireless communication technologies. In this area, wireless sensor networks (WSNs) are an emerging technology that heavily bases its sensing capability on small, low-power, and cheap sensors. A WSN is a network of many small intercommunicating computers that are equipped with one or several sensors (*see* **Wireless Sensor Network Platforms**). Since WSNs rely completely on batteries, the application of low-power sensors is a key requirement, and minimizing power consumption is fundamental for extending the operation lifetime of the network. WSNs are being investigated for use in a variety of military, environmental, home, health, and SHM applications. SHM systems based on MEMS devices have been developed and tested in aerospace [4] and automotive [5] engineering. A review of WSNs for SHM is found in [6] and in-depth information about architectures and protocols for WSNs in [7].

3 MEMS-BASED STRUCTURAL HEALTH MONITORING SYSTEM

An application that illustrates very well the potentiality of an MEMS-based WSN is cable tension force monitoring of stay cable bridges. Cable stay forces can be monitored by means of natural frequency estimations based on vibration measurements. By using an appropriate cable model, the relation between the natural frequencies and the tensile cable force can be described.

3.1 Sensor node hardware

A typical network node is composed of one or more sensors, a signal conditioning unit, an analog to digital converter, a data processing unit with memory, a radio transceiver, and a power supply. These components are integrated into an enclosure, which protects the hardware components from mechanical, chemical, and environmental impacts. Figure 1 shows the open rugged and waterproof enclosure, which is designed for applications in harsh conditions. The enclosure contains the hardware components and the power supply, which consists of two 1.5-V batteries with 16 500 mA-h each. External status light emitting diodes (LEDs), switches and connectors allow supervising, interacting, and connecting external sensors, which have to be mounted directly to the structure (e.g., strain gauges).

Many hardware platforms are commercially available that are optimized in terms of power consumption. The prototype network presented in this article is based on the Tmote Sky platform [8]. It features a 6-channel ADC with a resolution of 12 bit, a 16-bit processor with 10-kB RAM and 48-kB program flash memory, and a radio transceiver operating in the 2.4-GHz ISM band with a raw data rate of 250 kbps.

The signal conditioning unit enables to interface various sensing elements. A Sensirion humidity sensor SHT11 [9] is mounted into an opening in the enclosure and allows for temperature and humidity measurements outside of the box. For vibration measurements, the MEMS accelerometer LSI2L06 of ST Microelectronics [10] has been applied because of its good noise performance, low-power consumption, and low costs. The accelerometer and the signal conditioning circuitry, consisting of an amplifier and a low-pass filter, are mounted on a dedicated board (Figure 2).

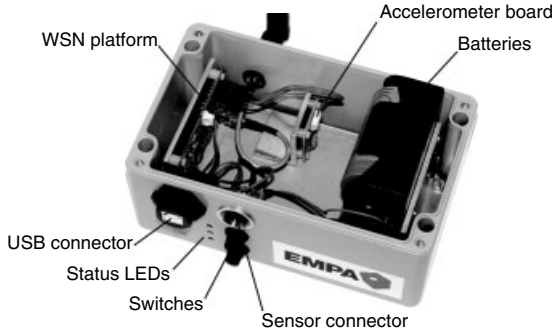


Figure 1. View of a physical sensor node with rugged and waterproof enclosure.

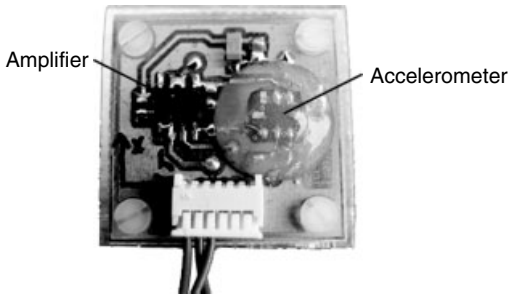


Figure 2. MEMS accelerometer board with amplification and filtering circuitry.

3.2 Sensor node software

The software running on each sensor node sets up the communication links between the sensor nodes, organizes the network topology, synchronizes the nodes, acquires measurements, and performs the data processing. The software is implemented as TinyOS components [11]. TinyOS is a component-based software framework designed for sensor networks and tailored to fit the memory constraints of the sensor nodes. It provides a concurrency model and mechanisms for structuring, naming, and linking software components to form a robust network embedded system.

The basic network functionality is provided by low-level network management components that operate independently from the actual monitoring applications. They are responsible for establishing the wireless links between adjacent nodes, for building the routing tree and for network-wide time synchronization. The monitoring application, which is built on top of these modules, uses this functionality to send and receive data and to have access to global time information. This allows for flexible exchange of the communication and time synchronization components.

A scheduler component forms the core of the actual monitoring application. It manages the data acquisition performed by the sensor node. Its clock is synchronized to the global time. The scheduler configures the measurement and data processing parameters like sampling rates, filter coefficients, thresholds, etc.; it triggers the data acquisition at the scheduled time. In addition to temperature, humidity, and acceleration measurements, information about the internal state of each sensor node (battery voltage) as well as communication parameters of the sensor network (e.g., routing tree) are monitored.

3.3 Data processing

The limited energy resources on each sensor node present the most restricting factor in designing and implementing WSN-based SHM systems. In terms of power consumption, wireless data transmission is much more expensive than data processing. In order to extend the system lifetime, it is therefore preferable to process the raw sensor readings in each sensor node with the aim to significantly reduce the data items that need to be transmitted to the data sink. This strategy is particularly recommended when monitoring vibration-based processes, which produce large amounts of raw data. There are several methods to reduce the size of raw data:

- Data compression encodes the data in a new representation that uses fewer bits than the original, not encoded data. This data reduction is done by using specific encoding schemes, which can either be lossless or lossy [12, 13].
- Data transformation transforms the raw data into a new kind of information that requires less space in terms of bits. Examples of simple data reductions can be maxima, minima, mean values, rms (root mean square), or statistical probability distributions of a physical quantity.
- Data analysis on the sensor node level reduces the amount of data transferred to the network. It differs from the methods described above because the raw data is subjected to an evaluation. The data is analyzed according to given criteria and a decision is taken if the data is relevant or not. Irrelevant data can already be discarded at the sensor node level.

Hence, long-term monitoring with WSN implies decentralized data processing and analysis. However, this is by far not possible for every analysis method. The limited energy resources restrict the complexity of the computational hardware of the sensor node, basically the memory size and computing speed of the hardware, and consequently affect the achievable analysis complexity.

In conventional monitoring systems, natural frequencies can be determined by identifying the

peaks in a frequency spectrum that are computed via averaged spectrogram based on fast Fourier transform (FFT). An efficient in-place FFT computation of 1024 data samples of 16-bit length requires approximately 2 kB. This is a quite large amount of memory usage for low-power microcontrollers, with typical memory sizes of 2–10 kB, since the programs for data processing, task scheduling, time synchronization, and networking must be stored in the very same memory. Therefore, a much less memory demanding method for computing natural frequencies is needed.

Parametric methods of spectral analysis fit this requirement. The natural frequencies are estimated by computing the poles of a spectral model based on a rational function. If the natural frequencies are well separated in the frequency spectrum, a requirement that cable stays usually fulfill quite well, the vibration components associated to a vibration mode can be isolated by filtering the recorded data with a band pass filter. The use of a very simple two-parameter discrete time autoregressive (AR) model is then sufficient to estimate the natural frequency. An algorithm that enables a data reduction by a factor of 500 performs the following steps:

1. The analog signals of the accelerometer are digitalized and stored in a buffer.
2. The offset in the recorded data produced by the earth's gravity is removed by subtracting the average of the recorded data.
3. The recorded data is filtered with a band pass filter to isolate the frequency components close to one of the natural frequencies.
4. A data block is extracted from the filtered data. The size of the data block should contain at least one period of the natural frequency that will be estimated.
5. Using the data block, the parameters of the AR model are fitted. With these parameters, the natural frequency and the damping ratio of the AR model are estimated.
6. The quality of the natural frequency is tested using the estimated damping ratio, since a low damping ratio correlates with a nearly pure harmonic. If the damping ratio is greater than a given threshold, the estimated natural frequency is rejected and step 7 is skipped.
7. The natural frequency estimations that passed the quality test are stored in an array.
8. A new data block is extracted from the filtered data and the steps 5, 6, and 7 are repeated until all data blocks have been processed.
9. The mean value of the natural frequency estimations stored in the array is computed. This represents the estimated natural frequency that is transmitted to the network.

A detailed description of the implemented algorithm can be found in Feltrin *et al.* [14].

4 FIELD TESTS WITH THE MEMS-BASED SHM SYSTEM

The field tests were performed on the Stork Bridge, a two-span cable stayed road bridge with a total length of 124 m. The monitoring was performed on 6 of the 24 cables. Before deploying the monitoring system, a preliminary investigation with standard data-acquisition equipment was performed with the goal to determine the natural frequencies of the six cables and identify the vibration modes with the highest vibration level. This information was used to select the natural frequencies to be tracked by the WSN and to design the band pass filters.

4.1 Overall SHM system

The logical structure of the WSN monitoring system that has been deployed on the bridge is displayed in Figure 3. It is composed of three subsystems. The first subsystem is the WSN that consists of seven sensor nodes: six sensor nodes mounted on the cables, labeled as C21–C26, and the root node, labeled as C0, which is situated under the bridge deck at the abutment (Figure 4). The root node is connected via USB to the base station, which was placed inside the abutment. The base station is powered via the mains supply. The second subsystem is the remote control center that collects all data generated by the WSN and is responsible for the long-term storage of the data. It implements the data visualization and representation tools. Furthermore, this subsystem provides an interface to the operator to observe, control, and configure the WSN remotely. This subsystem was located at the EMPA site in Duebendorf, at a distance of 16 km from the Stork Bridge. The third subsystem forms

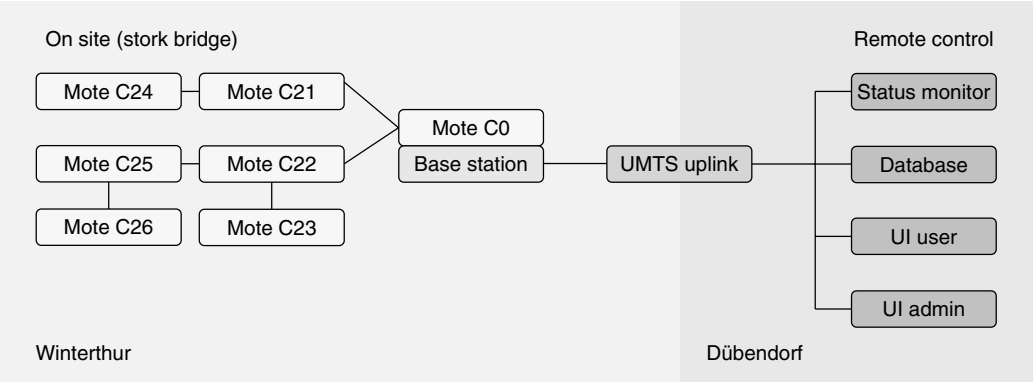


Figure 3. Logical structure of the structural health monitoring system.

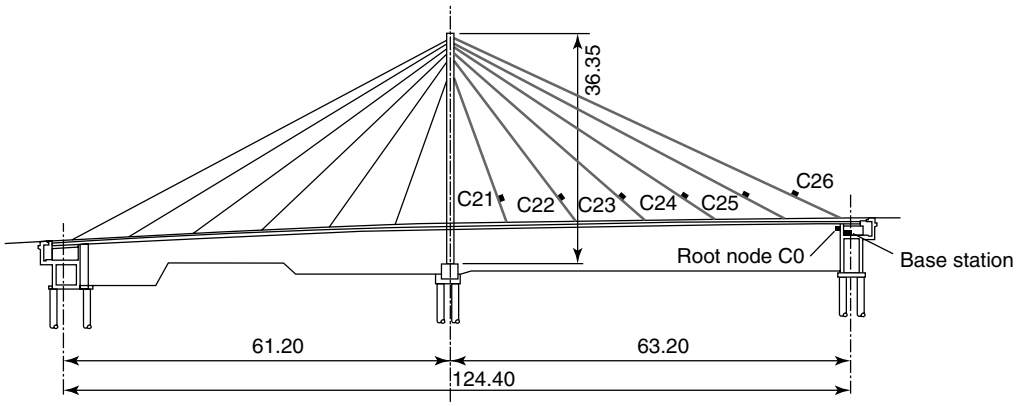


Figure 4. Elevation of the Stork Bridge with structural health monitoring setup.

the communication link between the WSN deployed on the bridge and the remote control center. This link is established via the base station using a standard wireless Universal Mobile Telecommunications System (UMTS) connection. A view to a network node mounted on a cable and to the root node below the bridge is shown in Figure 5.

The most challenging issue was to achieve overall system stability. A major source of instability was data processing. The computation of natural frequencies in one shot occupied the CPU for a long period and spoiled the execution of processes that guarantees the basic network functionality producing frequent system break downs. To overcome this problem, the algorithm was split up into tiny threaded code sections that required limited CPU time and that permitted an execution of basic network processes between two sequential threaded code sections. Furthermore, the integration of data acquisition, data processing, time synchronization, process scheduling, etc. into one software system turned out to be very sensitive to many tiny details regulating their interrelations. A change of duty cycle, for example, could destabilize the system producing system breakdowns within a short time. The modest CPU and RAM resources of the Tmote Sky platform significantly accentuated these problems.

5 RESULTS

Figure 6 displays a typical time history of the accelerations that were captured on the longest cable (C26) with the MEMS accelerometer LSI2L06 at a sampling rate of 100 Hz. The magnitude of the ambient cable vibration is very low. With respect to the 12-bit AD

Figure 5. Views of a sensor node mounted on a cable and the root node in the bridge abutment below the bridge deck.

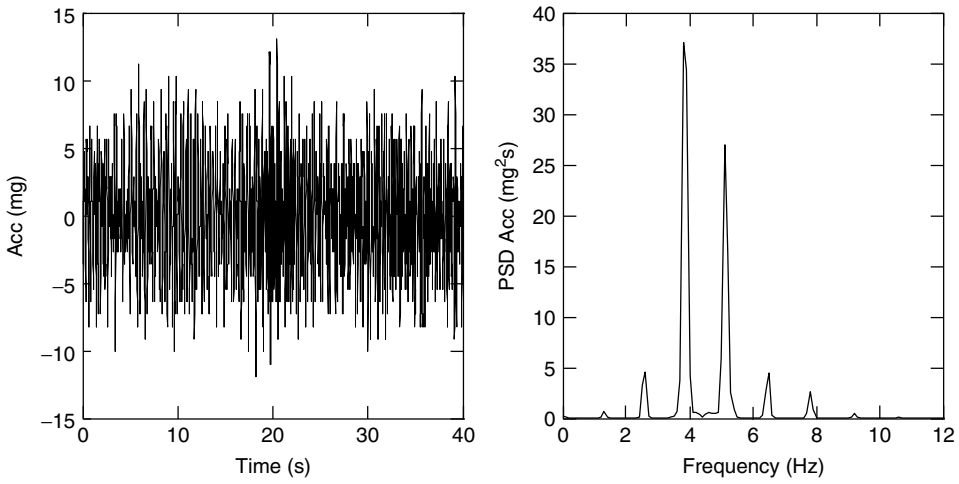


Figure 6. Time history of the accelerations sensed with the MEMS accelerometer LSI2L06 and its power density spectrum.

converter, which maps 1 mg approximately to 1 bit, the maximum of 13 mg is equivalent to 4 bits. Nevertheless, the natural frequencies can still be extracted from the time history as is demonstrated by the power density spectrum shown in Figure 6.

Figure 7 displays the natural frequencies of the cables C24, C25, and C26 of the Stork Bridge during a period of 60 days. The natural frequencies were estimated every minute from ambient vibration data using the algorithm described in this article. The typical rms magnitude of the ambient vibration data was 4–20 mg. The computation of the natural frequency lasts approximately 8 s. The algorithm was implemented in a series of threaded code sections to enable concurrent processes (e.g., time synchronization) to access the CPU. The three bands displayed in Figure 7; demonstrate that the algorithm generates estimations with a significant scattering. The accuracy of individual frequency estimations is within 5–10%, which is a direct consequence of the low level of accelerations and the short data blocks used for estimating the natural frequency (blocks of 50 samples, which correspond to 2–2.5 cycles).

A more accurate estimation of the natural frequency is obtained by using a moving average filter with a span of 200 samples (black curves inside the bands). Relatively small variations of natural frequencies are still detectable. This data processing step was done at the off-site control center with data retrieved from the data base. For monitoring

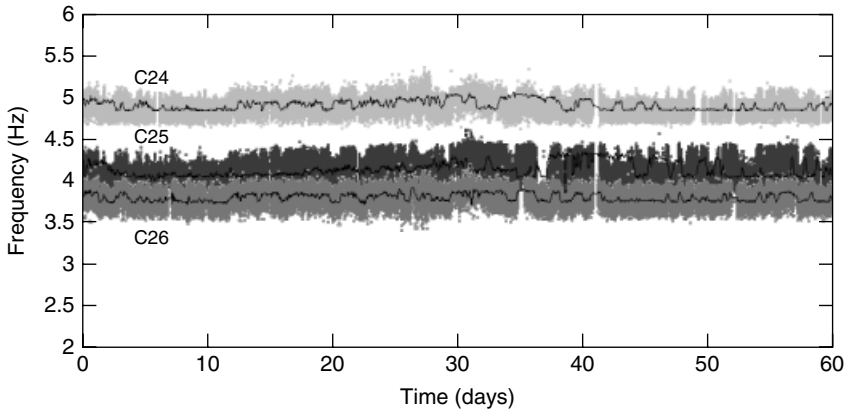


Figure 7. Time history of the natural frequencies of the cables C24, C25, and C26.

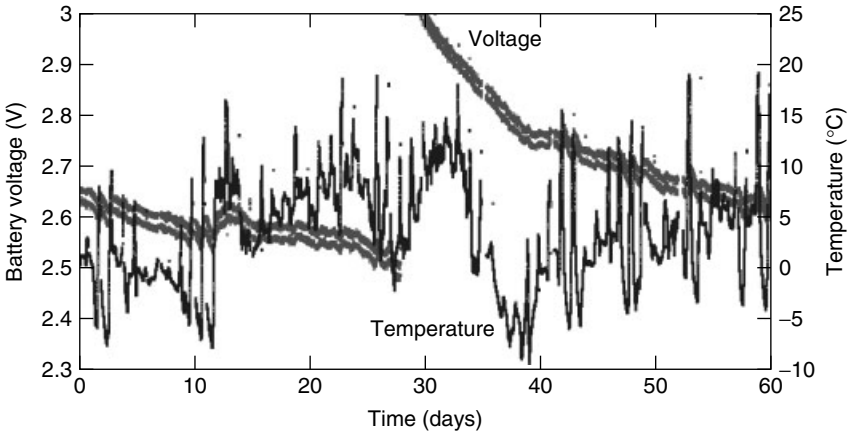


Figure 8. Time history of the battery voltage of the sensor node and the temperature measured on the cable C26.

of cable tension, the accuracy is good enough, since only significant changes are of concern for ensuring structural safety of a bridge.

Figure 8 shows the decay of battery voltage and the temperature on sensor node C26 over a 60-day period. It clearly depicts the dependency of battery capacity on temperature. The voltage graph consists of two lines. This effect is due to the fact that the battery voltage drops about 100 mV when the radio chip is turned on. Since voltage measurements are not synchronized to this switching, some measurements are taken when the radio is on and some when it is off. The voltage drop within 30 days is approximately 0.2 V. The theoretical lifetime of the WSN is approximately three months. This lifetime can be easily extended by a factor of 2 or 3 by extending the time between natural frequency estimations or by decreasing the duty cycle (the ratio of the system on time in a given period of time to the period of time), which was 40% during this test period. The voltage jump at day 29 is due to the replacement of batteries.

The graphs shown in Figures 7 and 8 reveal data losses during some periods of time. The causes of these losses are manifold: data from the sensor nodes is lost during the transport to the base station, stability issues in the communication software on the sensor nodes which lead to communication link breakdown, bugs in the software that render the base station irresponsive and block the reception of the data packets from the sensor nodes, and UMTS link breakdown caused by the telecommunication provider. The tests

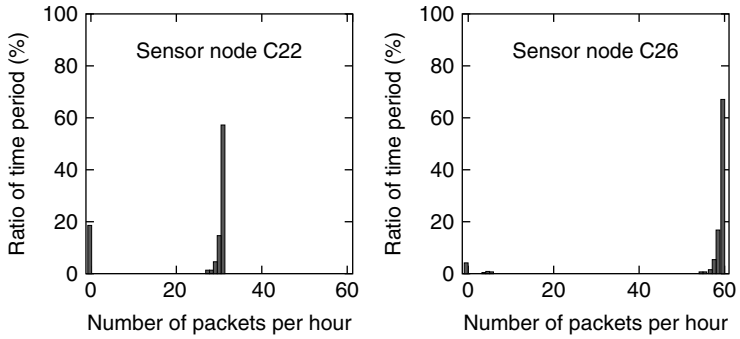


Figure 9. Frequency distribution of the number of valid data packets acquired within an hour of the cables C26 and C22.

demonstrate that data loss is intrinsically linked to the WSN since a lossless communication protocol would be too energy-consuming for field applications. Figure 9 displays the frequency distribution of the number of valid data packets within an hour of the cables C22 and C26 that were recorded at the remote control center during the 60-day period. For cable C26, which is closest to the root node, most of the time (67%), 60 packets out of 60 arrived at the control center. In 5.1% of the time no data was recorded. The same figures of the cable C22, which is farther away from the root node, are 57% with 30 packets out of 30 and 18.7% of the time with no data records. For the nodes C22, the natural frequency estimations occurred every 2 min. The mean data packet arrival rate over the 60 days was 24 out of 30 for C22 and 55 out of 60 for C26. As expected, the data loss increases with increasing distance to the base station. Moreover, the frequency distributions demonstrate that the wireless network is either up with a high reception rate (greater than 90%) or totally down. The case of an intermediate reception rate occurs very infrequently.

6 CONCLUSIONS

The field test on the Stork Bridge demonstrates that long-term monitoring with a low-power WSN is feasible. The feasibility is mainly based on the application of low-cost and low-power MEMS sensing technology and on a significant reduction of raw data that is achieved by decentralized data processing. The test demonstrates that the current technology does not provide the same reliability of mature wired monitoring systems in terms of stability, accuracy, and data loss. This outcome is not surprising since research and application of WSN technology in monitoring is still at a very early stage. The most critical issue is the handling of data processing and basic network functionality on a single CPU with very limited computational and memory resources. However, in the near future, these problems will be less critical since WSN platforms will be available that significantly provide more resources. Furthermore, to avoid conflicts between data processing and basic network functionality, the two tasks can be allocated to two separate low-power CPUs. Energy efficiency of hardware and software components will continue to play a central role. However, the progress in low-power microelectronics, which is driven by the huge market of portable electronic devices, and the application of energy harvesting technologies will mitigate the current limitations regarding power consumption. Nevertheless, the experience on the Stork Bridge demonstrates that with a well-balanced resource distribution between data processing and basic network functionality, a stable system can be achieved with existing low-power WSN platforms that provides useful information with an accuracy that is compliant to the monitoring objectives.

RELATED ARTICLES

On the Way to Autonomy: the Wireless-interrogated and Self-powered "Smart Patch" System

Energy Harvesting using Thermoelectric Materials

REFERENCES

[1] Gardner JW, Varadan VK, Awadelkarim OO. *Microsensors, Mems, and Smart Devices*. John Wiley & Sons: Chichester, 2001.

[2] Franssila S. *Introduction to Microfabrication*. John Wiley & Sons: Chichester, 2004.

[3] Varadan VK, Vinoy KJ, Gopalakrishnan S. *Smart Material Systems and Mems: Design and Development Methodologies*. John Wiley & Sons: Chichester, 2006.

[4] Osiander R, Darrin MAG, Champion JL (eds). *MEMS and Microstructures in Aerospace Applications*. Taylor & Francis: Boca Raton, FL, 2006.

[5] Valldorf J, Gessner W (eds). Advanced microsystems automotive applications. In *International Forum on Advanced Microsystems for Automotive Applications (AMAA)*. Springer: Berlin, 2007.

[6] Lynch JP, Loh K. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest* 2005 **38**(2):91–128.

[7] Karl H, Willig A. *Protocols and Architectures for Wireless Sensor Networks*. John Wiley & Sons: Chichester, 2005.

[8] Polastre J, Szewczyk R, Culler D. Telos: enabling ultra-low power research. *Proceedings of the Information Processing in Sensor Networks/SPOTS*. Berkeley, CA, April 2005.

[9] SHT1x/SHT7x Humidity & Temperature Sensor, http://www.sensirion.com/en/pdf/product_information/Data_Sheet_humidity_sensor_SHT1x_SHT7x_E.pdf, 2007.

[10] ST LIS2L06AL, MEMS inertial sensor, http://www.st.com/stonline/products/literature/ds/11665/lis2l06al.pdf, 2006.

[11] Levis P, *et al.* Tinyos: an operating system for wireless sensor networks. In *Ambient Intelligence*, Weber W, Rabaey JM, Aarts E (eds). Springer: New York, 2005, pp. 115–148.

[12] Lynch JP, Sundararajan A, Law KH, Kiremidjian AS, Carryer E. Power-efficient data management for a wireless structural monitoring system. *Proceedings of the 4th International Workshop on Structural Health Monitoring*. Stanford, CA, 15–17 September 2003; pp. 1177–1184.

[13] Caffrey J, *et al.* Networked sensing for structural health monitoring. *Proceedings of the 4th International Workshop on Structural Control*. New York, 10–11 June 2004; pp. 57–66.

[14] Feltrin G, Meyer J, Bischoff R, Saukh O. A wireless sensor network for force monitoring of cable stays. *Proceedings of the 3rd International Conference on Bridge Maintenance, Safety and Management, IABMAS 06*. Porto, 16–19 July 2006, on CD.

Chapter 155
Reliable Use of Fiber-optic Sensors

Wolfgang R. Habel
Division VIII. 1 Measurement and Testing Technology: Sensors, BAM Federal Institute for Materials Research and Testing, Berlin, Germany

1 INTRODUCTION

When users want to get a measurement task solved, they expect optimum performance from available sensor systems. Preferably, a complete measurement solution consisting of the sensing element, supply and recording device(s), and some accessories is required. Moreover, the users certainly expect that an appropriate application methodology is available. In contrast to this, some users prefer to design the sensor system by themselves; they purchase components, like resistive strain gauges, and, in some cases, rent the measuring equipment. This can become complicated if the components of the measurement system are not fully compatible with the specifications of the measurement task. For a better selection of monitoring components, and to be sure to have created a reliable measurement system, confirmed information about the component's characteristics, reproducibility, long-term stability, expected drifts and creep, parameter limits, and eventually, information about the expected uncertainty of the results measured, is needed.

The design work for a sensor system and the choice of a specific application procedure on site must follow specific demands according to the measurement task. Engineers have to find optimum answers with regard to all components of a sensor system, which also includes the selection of an appropriate sensing method out of a multitude of possibilities. Customers, consulting engineers, and suppliers need a well-founded overview on fiber-optic sensor technologies, available components, and recording systems. They have to analyze the measurement task and the influencing operational and environmental conditions. Depending on these prerequisites, the appropriate sensing system consisting of the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

optical source, optical fiber with a specific coating, sensing element itself (extrinsic or intrinsic type), some components like connectors, beam splitters, and finally a recording device with appropriate specifications can be selected. To find all this information summarized, an encyclopedia like this one or standards and guidelines would be helpful. Unfortunately, there is a lack of clear guidelines that summarize all necessary information for design and operation of specific optical sensor systems. Hence the most appropriate sensor system is sometimes not selected, or the expected measurement uncertainty is not achieved, and the customers are disappointed with the results measured. Guidelines and standards are therefore increasingly requested by different groups such as the following:

- **Customers**

They want basic (not really scientific) information about functional principles and basic specifications of the alternative measurement method to "fiber-optic sensors". They are often interested in quite fairly cheap solutions; however, they are mostly willing to pay an appropriate price if the new technology has advantages over the conventional one. To give them confidence in any discussion with fiber-optic sensor experts or consultants, customers should be able to follow a checklist containing all important aspects.

- **Engineering consultants**

They need, first, an overview about available sensing methods and corresponding components to design fiber-optic sensor systems. Customers would be well advised, if they do not receive one offer only for a measurement system, or an offer from only one company. The engineers should be able to compare different offers to find out the most effective sensor technology (sensor system). High-quality systems e.g., for long-term measurements, require high-performance components and facilities; measurement tasks with low-resolution requirements or for only short-term measurements (e.g., without disconnecting cables) need quality systems with lower system performance. Consulting engineers have to estimate and recommend the system which is the most effective one for the user.

- **Innovative companies, skilled personnel**

They have to deal with clear performance specifications for fiber-optic sensor systems as well as their components. This requires a correct use of the expression of performance determining quantities, static, and dynamic specifications. This must be done in accordance with international standards and must use appropriate vocabulary for describing general terms in the measurement techniques (including the use of SI units). The correct terminology ensures that the interdisciplinary community of e.g., physicists, fiber-optic experts, and mechanical and civil engineers understand each other [1].

- **Manufacturers**

Companies that produce and deliver sensors or sensor components have to provide validated products. This means that they have to confirm by examination and prove that the objectives of the particular requirements for a specific intended use are fulfilled. Validation of products as well as measurement methods can be carried out according to standard ISO/IEC 17025/2000 of the International Standardization Organization (ISO) [2]. Then, for a specific intended use, the sensor or the measurement system works as reliably as has been requested.

Using standards and guidelines, the confidence of the user community in fiber-optic sensor technology will increase. Standards generally facilitate the use of technology. As a competent summary of valuable technical recommendations, they contain all details for the use of a sensor system. There are four major types of standards:

- fundamental standards that define terminology, signs, symbols, and basic conventions;
- test methods as well as data-analysis standards that define measurement characteristics (temperature, pressure, physical, and chemical measurands);
- organization standards that describe company-related procedures such as quality management system, maintenance procedures, product or logistic management; and finally
- specification standards, which in the current fiber-optic sensor development stage is the most relevant standard type, since it comprises the particular characteristics of a product, its performance threshold such as functional parameters (measurement accuracy, stability), interface

and interchangeability, measures enabling cost reduction, environmental protection, and overall health and safety.

Guidelines to help the design and selection of a sensor system must consider all aspects such as the following characteristic features, characteristic features of sensor components, characteristic features of interrogation systems, and application and service aspects. Standards also define validation procedures, which deliver the largest contribution to the user's confidence in fiber-optic sensing systems. Steps necessary for a validation of fiber-optic sensor systems are described below.

2 INTERNATIONAL ACTIVITIES TO DEVELOP FIBER-OPTIC SENSOR GUIDELINES

ISIS Canada Research Network (*I*ntelligent *S*ensing for *I*nnovative *S*tructures) is actively developing guidelines for implementing structural health monitoring (SHM) methods in civil engineering such as fiber-reinforced polymers (FRP) containing integrated fiber-optic sensors. The first manual that gives a brief introduction to select fiber-optic sensor technologies was published in 2001 [3]. It provides a number of specifications and instructions on handling and installation, and considers application-related aspects for fiber-optic sensor systems. However, these recommendations cannot be used as guidelines or technical standards for fiber-optic sensors. Nevertheless, this manual is an important step toward developing guidelines for handling of fiber-optic sensors on site.

Research groups in the US standardization organization National Institute of Standards and Technology (NIST) in Boulder, Colorado, USA, have been involved in research into the behavior of fiber Bragg grating (FBG) sensors and associated devices. The basic metrology considerations developed by them provide valuable input for the definition of system component specifications [4]. Preliminary specifications have been developed for FBG sensors, FBG interrogators, and interferometric sensing systems (*see* **Fiber Bragg Grating Sensors**). These activities were mainly driven by companies that offer devices and systems or those that want to use fiber-optic sensor technology. The US Optoelectronics Industry Development Association (OIDA) provides an effective platform to overcome the barriers to generate a robust fiber-sensor market environment. This objective includes the development of standards and guidelines.

NASA utilizes fiber-optic components for the space flight sensor system, and is therefore obliged to validate the components of the sensor system. They use test procedures for materials validation as well as special test programs developed by the American Society for Testing and Materials (ASTM international) to validate sensor functions under specific space-typical requirements [5]. In this case of a very specific application area, the testing parameters must be adjusted for each component to simulate the environmental conditions or the worst case to be expected. Although there is no document cited comparable to a standard for the validation of fiber-optic sensors, the described test program can be considered as a discussion basis along the way to fiber-optic sensor guidelines.

Under the aegis of the "International Union of Laboratories and Experts in Construction Materials, Systems, and Structures" (RILEM), a new Technical Committee "Fiber-Optic Sensors" (TC-OFS) has been established. This committee specifically aims at developing guidelines and standards for fiber-optic sensor technologies in civil engineering. The general objective of the new RILEM TC-OFS is to promote the proper use of fiber-optic sensors in civil engineering applications such that their advantages can be fully exploited. A state-of-the-art report is under development. On this basis, application guidelines for fiber-optic sensors in civil engineering will provide expertise on the most important questions concerning reliability and stability of such sensor systems. Details can be found on the website of the TC-OFS [6]. The activities of this Technical Committee are—according to the general intentions of the RILEM organization—limited to sensing needs that are of direct relevance to civil engineering. However, it is expected that these guideline activities will serve as a model for other application areas, e.g., composite materials monitoring, monitoring of industrial plants with specific risks, and evaluation of new materials.

Activities related to the development of guidelines and standards have been observed in some

European countries. Investigations concerning reliability and long-term stability of components of fiber-optic sensor systems have been carried out at the Swiss Federal Laboratories for Materials Testing and Research (*EMPA*) in Duebendorf, Switzerland [7]. These investigations were subject to characterization of fiber-optic sensor components under the influence of typical environmental and loading conditions, and for the estimation of their lifetime. These investigations provide fundamental knowledge for guidelines, specifications, and standards.

Specific company-related guidelines have been developed by SMARTEC SA, Switzerland, for validation of SOFO system components [8]. SMARTEC developed their own validation program to check the performance of their sensor products. Such activities reveal the interest of innovative companies to offer validated products, and thus to promote confidence in new technical solutions.

In Germany, activities related to the development of guidelines and standards for fiber-optic sensors have been going on for years in Berlin, at the Federal Institute for Materials Research and Testing (BAM) and in the German Association of Engineers (VDI). VDI provides the platform for the Working Group AK17 on "Fiber Optic Measuring Methods" within the German "Society for Experimental Stress Analysis" (GESA), where a guideline for FBG strain sensors is being developed [9]. VDI guidelines, such as Guideline VDI/VDE/GESA 2635 "Experimental Structure Analysis; metallic bonded resistance strain gauges" [10], are generally regarded as competent recommendations. Such guidelines find an interested community and are considered as generally recognized engineering standards. The VDI core group for development of the FBG strain sensor guidelines consists of experts from companies and research institutes with a number of years of special experience.

"FIDES" (Optical Fibres for New Challenges Facing the Information Society) [11] is another European activity on the development of fiber-optic sensor guidelines established by the European COST Action 299. The Working Group (WG 4) "New Challenges in Fiber Optic Sensors" under this activity addresses the development of guidelines and standards for preferably long gauge length sensors. This working group promotes fruitful discussion and interaction with other groups acting in international standardization groups.

Although few companies have been developing test programs for validation of their products in accordance to their own rules, and research institutions like EMPA or BAM have established validation laboratories, generally binding standards or guidelines for fiber-optic sensors are not yet available.

3 DESIGN OF A FIBER-OPTIC SENSOR SYSTEM—RELIABILITY/STABILITY-RELATED ASPECTS

This section and the ones that follow describe important aspects that have to be considered (in standards and/or guidelines) when a measurement system is designed, sensors are applied onto or in measurement objects, and the measuring system is in service over a number of years. Because FBG strain sensors are very often used for SHM purposes (*see* **Fiber Bragg Grating Sensors**), important aspects of fiber-optic sensor applications are exemplified for FBG sensors. All these aspects can certainly be transferred to other sensor types and systems.

To solve a measurement task by using a fiber-optic sensor, first, the appropriate sensor type with specific performance features has to be selected from the different types that are available (*see* **Fiber-optic Sensor Principles**; **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**; **Fiber Bragg Grating Sensors**). The optimum design of the fiber-optic sensor systems comprises a number of components that have to work reliably together under on-site conditions. Figure 1 shows the general hardware structure of a fiber-optic sensor system.

Behind these physical components, numerous parameters that influence the performance (reliability) of the sensor signal and the stability of the system components are hidden. Most of the commercially available fiber-optic components (connectors, connecting cables, FBG elements, reading devices) are validated according to standard test procedures developed for data-communication purposes, or according to the individual manufacturer's own rules. At least functional tests are carried out. However, the optical measurement equipment is sensitive to ambient temperature and pressure variations. For example, an optical spectrum analyzer (OSA) occasionally used

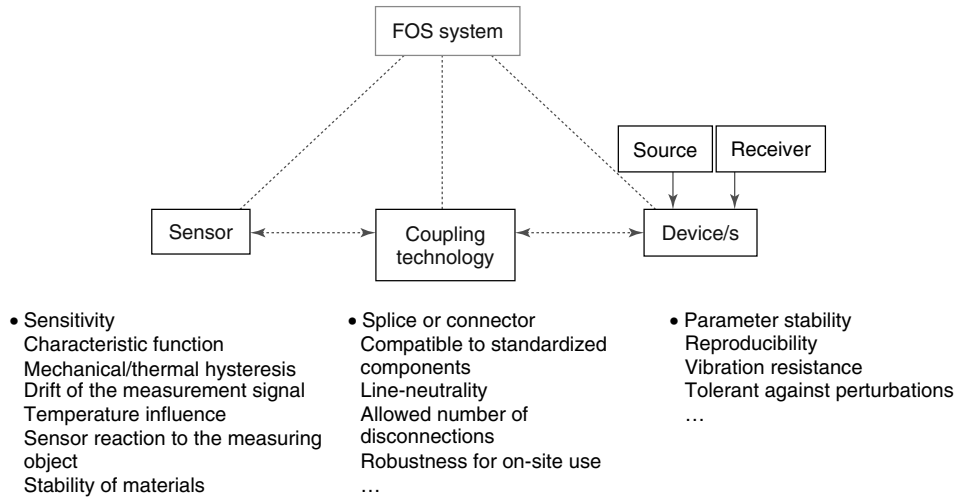


Figure 1. Main components of a fiber-optic sensor system.

for recording the FBG signal can have a temperature sensitivity K_T of about $10\,\mathrm{pm\,K^{-1}}$, which means it corresponds to the same temperature sensitivity of an FBG. If such a read-out device is used to record a signal from an FBG sensor on site, the measurement uncertainty could exceed the acceptable level. On the other hand, the calibration curve for the sensing element and/or the completed sensor sample must be available. If the equipment has to be replaced later by another one with different temperature sensitivity, the measured strain or temperature signal does not represent the previous sensor behavior.

Besides the influence of temperature on all system components, the effect of overall climatic conditions on all the specifications should also be considered. Primarily, the component's behavior under reference climate conditions must be understood. The following list provides an example of some of the important technical features and characteristic quantities that have to be proven for deformation sensors at reference climate:

1. optical characteristics (i.e., attenuation, reflection coefficient, and pulse characteristics, e.g., for FBG);
2. deformation sensitivity;
3. transverse (lateral) sensitivity;
4. ultimate deformability (axial strain, pressure, bending) of fiber sensing area;
5. mechanical hysteresis;
6. fatigue behavior;
7. influence of thermal and mechanical changes along the leading fiber on a sensor signal;
8. ultimate acceptable optical power at reference climate, as well as in general terms;
9. temperature resistance of the whole system;
10. temperature characteristics of sensor material $[\alpha_{\mathrm{Sensor}}(T),\ \alpha_{\mathrm{Sensor}}(T) - \alpha_{\mathrm{Material}}(T)]$;
11. temperature characteristics of optical signal $[\partial n/\partial T]$;
12. influence of local temperature changes on sensor signal;
13. temperature characteristics of sensitivity of deformation $[\partial S_n/\partial T]$; thermal hysteresis.

If sensor systems have to work under very different environmental conditions, performance should also be proven for the limits of intended use. The IEC 61757 standard for fiber-optic sensors [12] and specific sector standards propose few more parameters and test procedures. Some special aspects for the main components of an FBG strain sensor system are described in the following.

3.1 Optical source, detector, modulation, and demodulation

It can be certainly assumed that the optical source delivers a signal with sufficiently stable spectral and intensity characteristics as is needed for the respective

modulation technique. Depending on the sensor signal resolution requirements, devices with different interrogation methods are used. For example, a spectroscopic method using a charged coupled device (CCD) spectrometer provides a simple method with quite a good resolution. In contrast to this, devices that use interrogation methods on the basis of tunable filters, fiber interferometers, or planar integrated optical chips enable very high signal resolution. Stability and reproducibility of the wavelength measurement on FBG sensors decisively determine the uncertainty of the sensor system. This problem can be understood by considering two aspects. The first aspect concerns parameters of the FBG that define the Bragg wavelength. To measure strain or temperature, it is necessary to know the setup wavelength of the grating. The key question is which part of the spectrum should be used for the center wavelength calculation. Figure 2 highlights this problem. Using the definition that the width of a grating peak is defined by the distance between the two minimum points in the peak, this calculated Bragg wavelength λ_B differs from another estimation of the wavelength that is based on the -3-dB level definition. In those cases where the main peak has another side peak, e.g., due to birefringence effects, or has split into asymmetric peaks, the spectral shape has an enormous influence on the λ_B value, and the -3-dB definition must be modified. Unfortunately, there are no recommendations yet in the existing standards or guidelines as to which peak form is recommended or which method for calculating the peak wavelength should be used. To get a good signal-to-noise ratio, a broad spectrum is desired. To achieve a good peak identification, however, a narrow reflection spectrum with a sharp peak is recommended. In any case, peaks with flat or nonuniform tops, asymmetric peaks, and large side lobes must be avoided. Thus, the precision in estimation of the grating wavelength strongly depends on the grating characteristics (mainly strength, uniformity) [13].

The second problem concerns the wavelength stability of the measurement device. Measurement devices for highly precise wavelength reading use standard commercially available etalons similar to gas cells, or at least wavelength-stabilized FBG sensors. It is obvious that only such devices can be recommended for measurements under full climatic and temperature conditions. If customers are only able to fund low-cost devices for a specific task, the relatively high uncertainty in measurement results must be addressed and must be acceptable to the users. Generally, it cannot be expected that customers know or understand all these aspects. Following the allowable uncertainty—and considering the achievable real system-inherent uncertainty—engineering consultants can seriously recommend the measuring method and device that fulfills the customer requirements best.

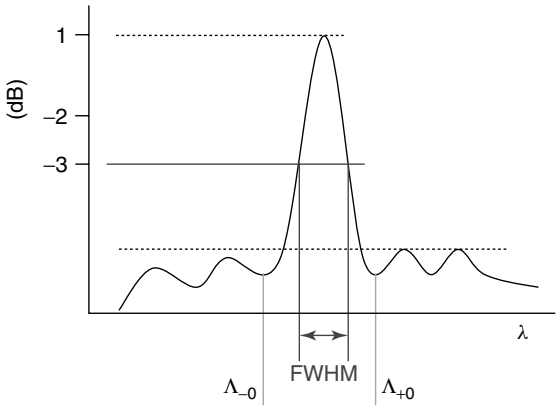


Figure 2. Calculation of the peak wavelength from the spectral characteristics. FWHM (full width at half-maximum) describes the difference between the two independent pulse values (here: wavelength) at which the dependent variable is equal to half (-3 dB) of its maximum.

3.2 Fibers, cables, connectors, and other components

There are an almost unmanageable number of standards for optical fibers, cables, connectors, and other components. All these components are also used in telecommunication and data-communication systems. International standards can therefore be found in the International Standard Classification (ISC), nr. 33.180. To enhance the reliability of components, some investigations were made within COST 270 [14]. For example, the connector group defined a reliability qualification test that resulted in the IEC standard 62005-9-2. In any case, when components developed for telecommunication purposes are used

in sensor systems, their behavior under specific environmental conditions have to be verified.

3.3 Sensing element

From the measurement reliability point of view, the most challenging part of a fiber-optic sensor system is the signal from the sensing element itself. In case of strain sensors, the question is how much of strain is durably transferred from the structure to be evaluated into the sensing element. Depending on the method of sensor application (embedded, clamped, surface-glued, and welded) and type of sensor (sensor fibers, e.g., with gratings inside, sensor patches, or rods), the interaction between the measurand and the sensing element is influenced by numerous factors. From the author's point of view, the appropriate application of FBG strain sensors is very complex and quite difficult. Hence, some sensor aspects related to the reliability of the measurement result will mainly be considered for FBG strain sensors in the following sections.

Almost every expert team uses a different method of attachment or embedment of FBG sensors. Coatings and adhesives with different specifications even for similar measurement tasks are used. This very complex part of the measurement system brings the largest uncertainty—actually, this could even be considered a skills-based uncertainty. In comparison with the large number of commercially offered system components mentioned above, it must be realized that there is still a lack of well-validated methods for the application of FBG sensors. Moreover, only few companies offer application expertise that includes a warranty claim. For example, the embedment of FBG strain sensors in rotor blades of helicopters or wind turbines is, at present, mainly carried out by research institutions or research departments of companies according to their best knowledge because generally accepted technical rules are not yet available. However, the user community interested in fiber-optic sensors promotes and supports more and more research and development (R&D) activities in this area to close the gap in the expert's knowledge and to push the development or completion of corresponding guidelines and standards.

4 VALIDATION—THE METHOD TO ENSURE RELIABLE SENSOR USE

If users want to be sure that they get an appropriate sensor system, they have to ensure that a validated fiber-optic measurement system is used. Users benefit from validation because they get assured information about the performance and limitations of the sensor system. According to the definition in standard ISO/IEC 17025/2000 [2] of the International Standardization Organization (ISO), validation is the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled. It means that validated systems enable consulting engineers, suppliers, and users to evaluate suitability and reliability of the measurement system for the specific use. The key questions of accuracy, long-term stability, and reliable function under environmental influences to be proven by validation depend on the client's needs. Validation is always a balance between cost, risks, and technical possibilities.

Validation should be carried out by institutions that fulfill the requirements for technical competence in the field of fiber-optic sensor technology. The testing institution should preferably be recognized by accreditation bodies. Validation can also be carried out by third-party laboratories (unbiased institutions) that are able to bring in expertise and have the special equipment meet the requirements concerning uncertainty of measurement and traceability to etalons or national standards.

Procedures for validation of the sensor system components comprise

- sensor-related characteristics (measurement range, resolution, sensitivity are the most important);
- measurement uncertainty, repeatability, and reproducibility of data from components of the measurement system;
- proof of stability of the system components characteristics.

Apart from the stability of the sensor system or the reliability of the measurement method, the uncertainty estimation of the measurement results has particular significance. Validation enables getting data that are important for the evaluation of the measurement

uncertainty. Basically, this information can only be achieved under certain conditions:

- all equipment used for tests and or calibrations (including equipment for subsidiary measurements, such as environmental conditions) shall be calibrated before being put into service;
- equipment shall be operated by personnel with sufficient competence in the field of fiber-optic sensors;
- the calibration procedure has to ensure that all measurements are traceable to the International Systems of Unit (SI); the link to SI units may be achieved by reference to measurement standards and high-quality measurement instruments. A calibration laboratory or an unbiased competence center establishes traceability to the SI by means of an unbroken chain of calibrations or comparisons linking the available standards to relevant primary standards of the ISO units of measurement (e.g., national measurement standards). A short remark concerning traceability seems to be necessary (compare it also with EN ISO/IEC 17025 [2, Chapter 5.6] and ISO 10012 [15, Chapter 4.15]): it is clear that the resolution of measurement results achievable with relevant measurement standards (including its repeatability) must approximately be one order better that those of the measurement systems to be evaluated.

However, it is extremely difficult to get validated data for surface-applied or embedded sensors because it is not possible to discriminate between the sensor's behavior and sensor signal influences coming from application (as already pointed out previously). This basic problem leads to a certain amount of uncertainty, which cannot be estimated. In any case, recommendations for application procedures based on fully established expertise will minimize uncertainty.

Not only users but also manufacturers and sales agencies benefit from validation and corresponding guidelines because damage can then be prevented during the production stage. Performance gaps in the measurement system can be displayed and removed by systematic optimization. In case of damage or if measurement results are not accepted by the user, the supplier is able to prove that he has complied with his obligation to exercise due care (product liability).

5 GUIDELINES FOR APPLICATION AND OPERATION OF SENSOR SYSTEMS

Generally, a distinction has to be made between the sensor signal coming from the sensing element (FBG, Fabry–Perot interferometer, distributed fiber-optic sensor) itself and the sensor signal recorded by the applied sensor. The characteristics of the unapplied sensor can more or less differ from the attached one. Moreover, it must be aimed to distinguish between measurement signal-relevant information that comes from the sensor, including surrounding effects, and the contribution from the structure to be evaluated. In almost all cases, it is not possible to discriminate between creep of the structure (measuring object) and creep of the adhesive, e.g., used for bonding of the FBG sensor so that the customer is only able to record the behavior of the structure from the measurement result with a certain sensor-related uncertainty. The largest challenge for fiber-optic sensor experts is to reduce application-related influences.

Whatever type of mechanical sensor is used, enduring reliable strain transfer from the measurement object into the sensing element must be ensured over the complete period of operation. Irrespective of the mode of sensor fixation, the sensor fiber and/or the fiber-optic sensing element has always to be bonded to a support component or to a fixing element, and finally to the measuring object. A number of aspects for consideration arise such as the following:

- characteristic mechanical behavior of fixing components (substrate, adhesives, protective layers/coating), materials combination sensing element/measuring object, and covering must be known and characterized for the total environmental range;
- proof of sensor position accuracy;
- thermally induced stress loading of the sensing element during application;
- sensitivity of all components (sensing element, leading fiber, connectors and splices, optoelectronic components) to mechanical and thermal influences;
- attacks from aggressive media;
- chemical and physical interactions of materials close to the sensor;

- methods to test and evaluate the bonding behavior (creep) of applied sensors over long periods of operation.

Specific aspects have to be considered especially when FBG sensors are used, e.g., possible influences of unwanted mechanical perturbations like transverse pressure or bending of the sensing element. Transverse pressure into the fiber at the location of the grating, or any mechanically induced deformation of the fiber core might change the spectral characteristics of the FBG element. This spectral change can lead to faulty measurement results if the appropriate interrogation system is not used. The possibility of appearance of transverse influence must be estimated, especially when the FBG sensors are embedded in nonhomogeneous materials like reinforced composites or rough materials such as concrete and are subjected to thermally induced material contraction.

Other points concern hydrogen-loaded FBG sensors. A decay of the reflectance can be caused over time by out-diffusion of hydrogen; or applied and permanently strained fibers can be subjected to stress relaxation effects, which lead to a wavelength shift. Finally, decay of the refractive index modulation could occur over time. All these characteristic features have to be considered with regard to long-term measurements. Guidelines will help here to avoid underestimation of these effects.

If strain sensors are to measure not only tensile but also compressive strain, they have to be pretensioned. This procedure is not trivial during application under field conditions because the pre-tension must be held as long as the curing or hardening process is incomplete. Figure 3 shows a methodology on how to introduce strain in FBG arrays during manufacturing of large composite components in a factory. Prefabricated sensor patches could be used alternatively, however, the patches must hold the pre-tension of the sensing element during manufacturing of the patch and later on during potential heat treatment (tempering). In this case, the patch or the adhesive could shrink and the pre-tension of the fixed grating is reduced or lost. The appropriate methodology for application of continuously fixed FBG sensor depends on technological conditions. It must be considered that surface-glued fiber gratings produce an unsymmetrical interface structure, and a reliable and reproducible bonding of the fiber is not easy to reach [16]. Hence, embedment of FBG sensors seems to be less problematic.

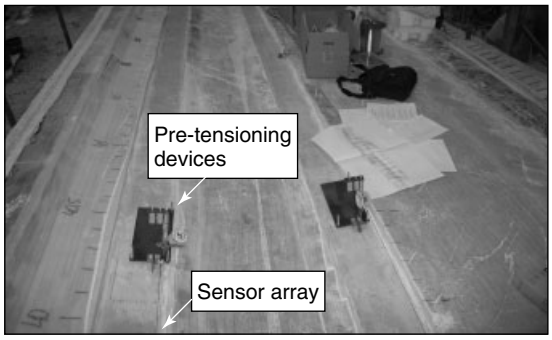


Figure 3. Embedment of FBG arrays into reinforced composite structures using a special device to pre-tension the sensor fibers under defined conditions (here by 0.34%) during the lamination procedure, Photo: BAM.

Embedment of FBG arrays into reinforced composite structures using a special device to pre-tension the sensor fibers under defined conditions (here by 0.34%) during the lamination procedure (Photo: BAM).

Apart from FBG sensor application procedures, arrangement of the wiring of the sensors must not be underestimated. Under harsh mechanical or environmental conditions, splices inside the material to be assessed and ingress/egress points are critical zones. Splices should also be embedded just like the fibers. This is usually possible in huge composite structures with sufficient thickness. However, in steel structures or very narrow tubes, other methods have to be developed. Figure 4 shows a reliable solution to protect the splice joints by storing them in a very small box inside a steel anchor tube. Figure 5 shows the fiber-optic cable at the ingress/egress point, which has to be protected for the process when the anchor is introduced into the bore hole.

On the basis of the manifold experiences of fiber-optic sensor experts in creating ruggedized packaging, guidelines have to list and propose for the user community all technical aspects that are intrinsically tied to application and installation of sensing components on site. This comprises

- preparation of the area where the sensor is to be installed;
- connecting fiber ends and protecting splice areas very close to the measuring object (under field condition);

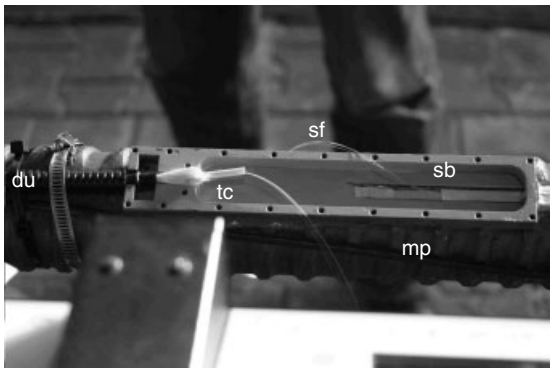


Figure 4. Attachment of FBG arrays onto heavy steel anchors for use in geotechnical engineering (sf, sensor fiber; sb, splice box; mp, micropile/anchor steel; du, duct; tc, transmission cable). Photo: BAM.

Figure 5. Critical points are splice joints (left) and cable egress points (right) in heavy steel anchors (mp, micropile/anchor steel; du, duct; tc, transmission cable). Photo: BAM.

- supervisory procedure to check the appropriateness of the sensor installation;
- protection of sensing area;
- robustness—aging of, e.g., connectors when they are frequently opened;
- mechanical and thermal hysteresis of embedded or applied sensor fibers, e.g., containing FBG sensors;
- identification of zero-point changes when devices have to be disconnected or leading cables have to be cut and exchanged;
- packaging and protection of the ingress/egress areas to make sure that embedded or attached fiber-optic sensors are fully protected over the whole service life of the structure without jeopardizing data integrity;
- sensor reaction to mechanical or thermal impacts (shocks) during operation;
- sensor behavior/aging under repeated vibration;
- sensor behavior (including cabling) under harsh climate conditions;
- durability of sensor-related materials under thermal, chemical, and mechanical conditions.

Additionally, recommendations and advice for repair of applied components of the sensors system should be given in the guidelines. The demand on exchangeability of sensors or components clearly determines the choice of the measurement and application method. Guidelines should also give recommendations for some other critical aspects that can affect reliability and/or stability.

6 TEST METHOD TO CHARACTERIZE MEASUREMENT RELIABILITY

The most critical influence on the reliability of mechanical sensors comes from the stress/strain conditions in the interface area between the sensing element and the material or structure to be measured or monitored. There are at least two boundary layers: (i) the interface between the structural material and the fiber coating and (ii) the interface between the optical fiber and its coating. Because deformation of the structural material must be transferred into the fiber-optic sensor, the mechanical and physicochemical properties of selected and/or existing materials at the interface determine the performance of the sensor. A debonded interface of a continuously attached sensor fiber (Figure 6b) or changes in the interface's behavior due to aging of the involved materials would detrimentally affect the transfer of the measurand to the FBG, and result in erroneous readings. For example, fatigue tests with acrylate-coated FBG embedded in composite samples of wind turbine rotor blades revealed a continuously increasing aging effect in the FBG signal (150 $\mu\varepsilon$ loss after 1 600 000 cycles) for exactly the same deformation values. Some more results are described in [17].

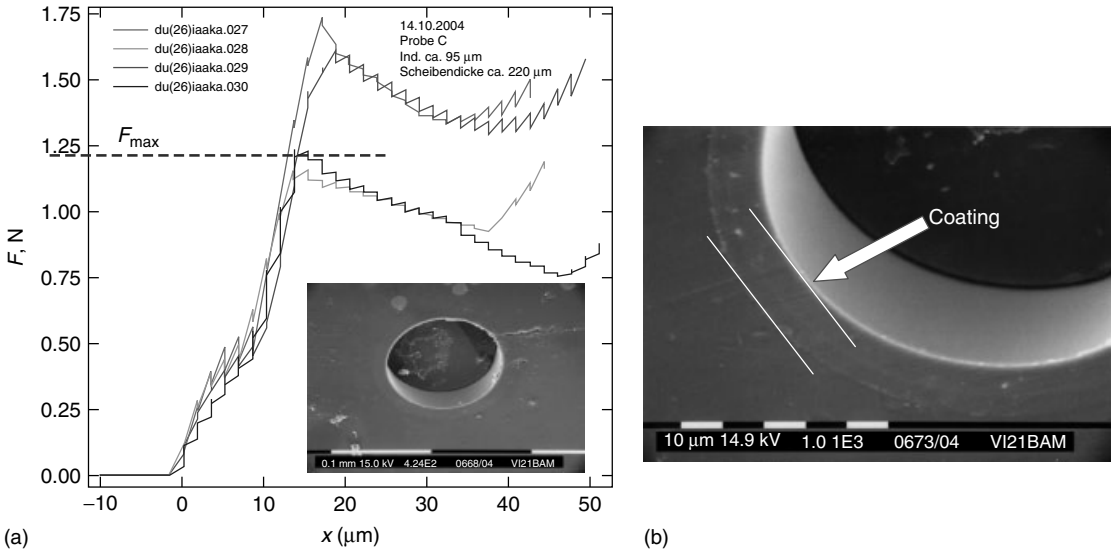


Figure 6. (a) Strain transfer characteristic of PI-coated optical fibers in typical epoxy resin used for composite material (abscissa: displacement of the indenter, ordinate: force at the indenter). (b) Debonding of the coating from the optical fiber beyond the reliable operating range of the sensor [17].

Another systematic influence on the measurement signal of continuously attached fiber-optic sensors comes from the viscoelastic properties of the coating. Coating thickness and Young's modulus have an influence on both the static and dynamic responses of the sensor. In every application (embedment or surface application), the interfacial mechanics has to be well understood. Unfortunately, as of now, there is no complete description of all the correlating effects because test methods to validate the strain transfer behavior have not yet been fully developed. However, standard testing approaches from the field of composite materials research can be used to characterize the interface bonding behavior of embedded or attached fiber-optic deformation sensors. Figure 7 shows such a testing facility to evaluate the bonding behavior of fiber-reinforced materials; this is also used at BAM for characterization of embedded optical sensor fibers [18]. The microindentation testing machine consists of a very stiff beam, which is one-point supported at the middle of the beam. At one end of the beam, a high-precision stepping motor is positioned, which drives the beam in the vertical direction (upward). At its other end, the indenter is fixed at the beam, which introduces the force into the test sample. A microscope is available to position the sample exactly below the indenter needle. The testing procedure is computer controlled; and the force introduced into the fiber as well as the displacement of the end face of the fiber due to the pushing force are automatically measured and recorded. This method is primarily used to determine the appropriate sensor coating for a specific measurement task. Apart from the experimental proof, or at least estimation, of bond strength of the coating at the fiber surface, this test method delivers characteristic behavior of the shear stress development in the border zone coating/structure. Research is needed to correlate these test procedures with other interfacial parameters such as coefficient of friction, thermal residual radial stress, and coating stiffness.

7 OUTLOOK

Significant steps have already been taken toward formulating a set of guidelines and standards for fiber-optic sensors. Few companies and institutions have developed or are developing validation programs to make guideline-like documents available. There is some basic knowledge about materials properties,

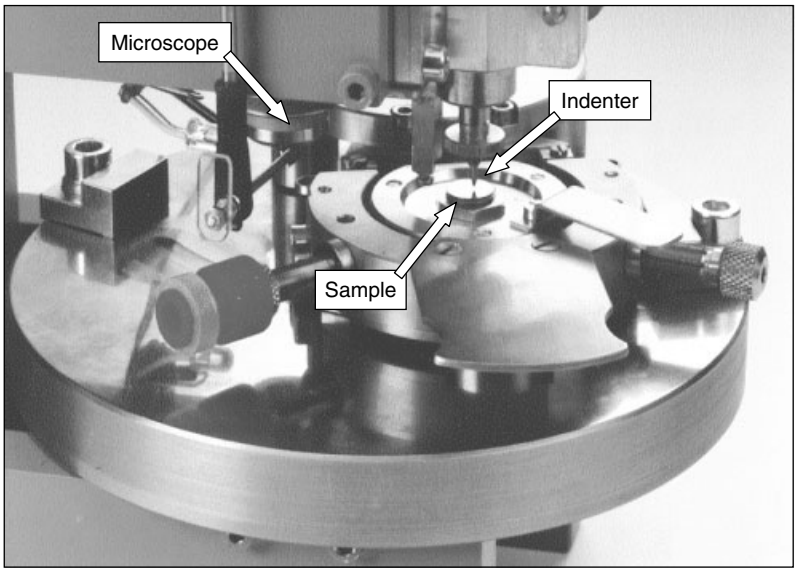


Figure 7. Microindentation test facility for evaluation of the fiber-bonding behavior (Photo: BAM).

interaction of specific sensor materials with environment, and characterization of the sensor behavior under specific operational conditions that are necessary for developing standards and corresponding recommendations.

Few national and also international expert groups, mainly established in the civil engineering and SHM field, have been established and have started work to promote the development of guidelines. There are already some documents that might have the model character for establishing competent international fiber-optic sensor guidelines. With more activities in international expert groups coming to address open questions concerning reliability, reproducibility, stability, and technical rules, confidence in the use of fiber-optic sensors will grow in the user community. Among other fiber-optic sensor technologies, FBG sensor technology is most commonly used in SHM. A set of guidelines will therefore be developed for this sensor type in the near future. Scientists and interested engineers with hands-on expertise are also aiming to establish well-researched guidelines and recommendations on how to handle and install the sensor system outside the laboratory or manufacturing environment, the type of adhesion or bonding material to be chosen, and finally, which influences resulting from application might perturb or possibly damage the sensor signal.

REFERENCES

[1] Habel WR. Fiber optic sensors for deformation measurements: criteria and method to put them to the best possible use. *Proceedings of the SPIE* 2004 **5384**:158–168.

[2] Standard EN ISO/IEC 17025:2000 (trilingual version), *General Requirements for the Competence of Testing and Calibration Laboratories (ISO/IEC 17025:1999)*, International Organization for Standardization, 1999.

[3] Tennyson R (ed). Installation, use and repair of fibre optic sensors. In *Design Manual ISIS-M02-00*, Canada, Spring 2001 and Civionics Specification. Design Manual No. 6 (Chapter 2: Specifications for fibre optic sensors (FOS)). ISIS Canada Research Network, October 2004.

[4] Dyer SD. Key metrology considerations for fiber Bragg grating sensors. *Proceedings of the SPIE* 2004 **5384**:181–189.

[5] Ott MN. Validation of commercial fiber optic components for aerospace environments. *Proceedings of the SPIE* 2005 **5758**:427–439.

[6] RILEM Technical Committee, *Optical Fibre Sensors*, http://www.rilem.net/tcDetails.php?tc=OFS, 2008.

[7] http://www.empa.ch/, 2008.

[8] Inaudi D. Long-term reliability testing of packaged strain sensors. *Proceedings of the SPIE* 2005 **5758**:405–408.

[9] The Association of German Engineers (VDI), *Society of Experimental Stress Analysis (GESA)*. http://www.vdi.de/strukturmonitoring, 2008.

[10] VDI/VDE/GESA-Richtlinie 2635, Part 1 (2007-04), *Experimental Structure Analysis; Metallic Bonded Resistance Strain Gages; Characteristics and Test Conditions* (see also Ref. 9), The Association of German Engineers, 2007.

[11] COST 299, http://www.cost299.org, 2008.

[12] IEC 61757-1 Ed. 1.0 b:1998, *Fibre Optic Sensors—Part 1: Generic Specification*, International Electrotechnical Commission, 1998.

[13] Fernandez AF, Gusarov A, Berghmans F, Kalli K, Polo V, Limberger H, Beukema M, Nellen P. Round-robin for fiber Bragg grating metrology during COST270 action. *Proceedings of the SPIE* 2004 **5465**:210–216.

[14] COST 270, http://www.sckcen.be/cost270/; COST 270 Final report: http://www.cost.esf.org/typo3conf/ext/bzb_securelink/pushFile.php?cuid=253&file=fileadmin/domain_files/TIST/Action_270/final_report/final_report-270.pdf, 2008.

[15] Standard ISO 10012-1, *Quality Assurance Requirement for Measuring Equipment; Part 1: Metrological Confirmation System for Measuring Equipment*, International Organization for Standardization, 1992.

[16] Schlüter V. *Strain Transfer Characterization of Surface Applied Fiber Bragg Grating Sensors*, Young Stress Analyst Competition. British Society for Strain Measurement (BSSM), 2007, pp. 34–38.

[17] Krebber K, Habel WR, Gutmann T, Schram C. Fibre Bragg grating sensors for monitoring of wind turbine blades. *Proceedings of the SPIE* 2005 **5855**:1036–1039.

[18] Habel WR, Schulz E, Kalinka G, Bismarck A. Evaluation of adhesion behaviour of optical fibers for sensors embedded in cementitious materials. In *Fiber Optic Sensors for Construction Materials and Bridges*, Ansari F (ed). Technomic Publishing Co, Inc.: Lancaster-Basel, 1998, pp. 194–206.

Chapter 96

History of SHM for Commercial Transport Aircraft

Roy Ikegami[1] and Christian Boller[2]

[1] *Acellent Technologies, Inc., Sunnyvale, CA, USA*
[2] *Saarland University & Fraunhofer Institute for Non-Destructive Testing, Saarbrücken, Germany (and formerly of The University of Sheffield, Sheffield, UK)*

1 INTRODUCTION

The origins of structural health monitoring (SHM) in commercial aviation are possibly difficult to track. One of the major impacts triggering SHM, which was of course not called SHM at the time, was the comet accidents in the early 1950s. These accidents alerted a variety of parties in aviation, initially more from the military aviation side, to look into loads monitoring for the assessment of structural performance. Associated with this was damage monitoring, where the different methods emerging in nondestructive testing (NDT) were considered. Acoustic emission (AE) methods (*see also* **Applications of Acoustic Emission for SHM: A Review**) were among those first SHM technologies investigated for application to commercial airframe structures. In this method, when a load is applied to a solid structure (e.g., by internal pressure or by external mechanical means), it begins to deform elastically. Associated with this elastic deformation are changes in the structure's stress distribution and storage of elastic strain energy. As the load increases further, some permanent deformation and cracking may occur, which is accompanied by a release of stored energy, partly in the form of propagating elastic waves termed *AEs*. If these emissions are above a certain threshold level, they can be detected and converted to voltage signals by sensitive piezoelectric transducers mounted on the structure's surface. During the 1970s and early 1980s, the Canadian Department of National Defense supported efforts [1–3] to flight test and demonstrate the use of AE methods for detecting the growth of fatigue cracks in airframe structures during flight. The flight testing was performed primarily on military aircraft such as the Canadian CC-130, CF-5, CF-100 and British Tornado. Although these investigations were able to demonstrate the feasibility of utilizing AE methods for the detection of crack growth in aircraft

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

structures during flight, several issues precluded implementation by commercial aircraft operators and original equipment manufacturers (OEMs). Some of the issues are summarized as follows:

- size and weight of equipment required for data acquisition and extensive data storage requirements;
- detailed calibration of the structure was required for the unambiguous detection of crack growth;
- no reliable methods available to distinguish and properly identify AE signals from crack growth versus other high-frequency noise sources such as crack face rubbing, electromagnetic interference (EMI), and airframe structural noise due to in-flight loads;
- reliability of sensors subjected to extreme environments and high loading conditions; and
- lack of quantitative data on crack size, location, and probability of detection (POD).

Recently, owing to the high labor costs for performing detailed structural inspections, commercial aircraft operators and OEMs have developed a renewed interest in SHM methods as a means of conveniently inspecting the locations of airframe structures that are hard to access. In the late 1990s, both Boeing and Airbus made efforts to evaluate several SHM technologies for potential application to their commercial transport aircraft. These efforts are described below.

2 AIRCRAFT OEM AND OPERATOR DEVELOPMENTAL EFFORTS

Although not being implemented today, both the commercial aircraft manufacturers, Airbus and Boeing, have made remarkable progress in the implementation of SHM.

Boeing's ideas of implementing SHM go back to the early concepts proposed by Hickman *et al.* in the early 1990s [4]. After a period of wider exploration, the Boeing Commercial Airplane Company began a collaboration with Delta Airlines to evaluate and flight test several SHM technologies in 1998. The technologies of interest were primarily for the monitoring of corrosion, corrosion by-products, moisture, and strain to determine actual aircraft environments, flight loads, and usage [5, 6]. The Boeing Phantom Works organization had developed an autonomous structural integrity monitoring system (ASIMS), which was a small rugged, battery-powered data-acquisition system that could be placed onboard an aircraft and operated independent of any aircraft system. The ASIMS was capable of interfacing with a variety of environmental and structural sensors and recording both digital and analog sensor data. It was designed to interface with other data-acquisition units and a variety of sensor types including fiber-optic and microelectromechanical system (MEMS) devices, and could be placed in remote, hard-to-access areas of flight vehicles, ships, and ground vehicles to monitor the health of structural components. A ground-based reasoner (GBR) consisting of software running on an off-board computer provided the means for studying, managing, and assessing the data. Delta Airlines provided an in-service Boeing 767-300ER aircraft for the flight testing and data gathering, and the maintenance personnel to install the equipment and periodically download the data from the ASIMS. The data was then sent to Boeing for evaluation.

A variety of sensor systems were placed onto the Delta Airlines 767-300ER aircraft for flight testing and data gathering. The SHM sensor systems included a linear polarization resistance (LPR) corrosion environment sensor and two separate fiber-optic sensor systems. The LPR sensor system developed by Analatom, Inc. is basically an MEMS device that incorporates a galvanic sensor that measures the change in resistance of a host material, which corrodes away. The first fiber-optic system to be included was developed by Blue Road Research, Inc. It utilized sensors consisting of polyimide-coated fiber Bragg gratings (FBGs) to monitor changes in relative humidity (RH). The second system was developed by Luna Innovations, Inc. and used long-period grating (LPG) moisture and humidity sensors along with extrinsic Fabry–Perot interferometric (EFPI) pressure and temperature sensors to monitor environmental conditions conducive to corrosion in aluminum aircraft structures. Figure 1 shows the Boeing ASIMS data-acquisition system (lower left corner) and battery (upper left) integrated with the Luna fiber-optic sensor signal conditioning system. An example of the Analatom LPR/MEMS sensor system is shown in Figure 2. Figure 3 shows the installation of the sensor systems in the forward cargo

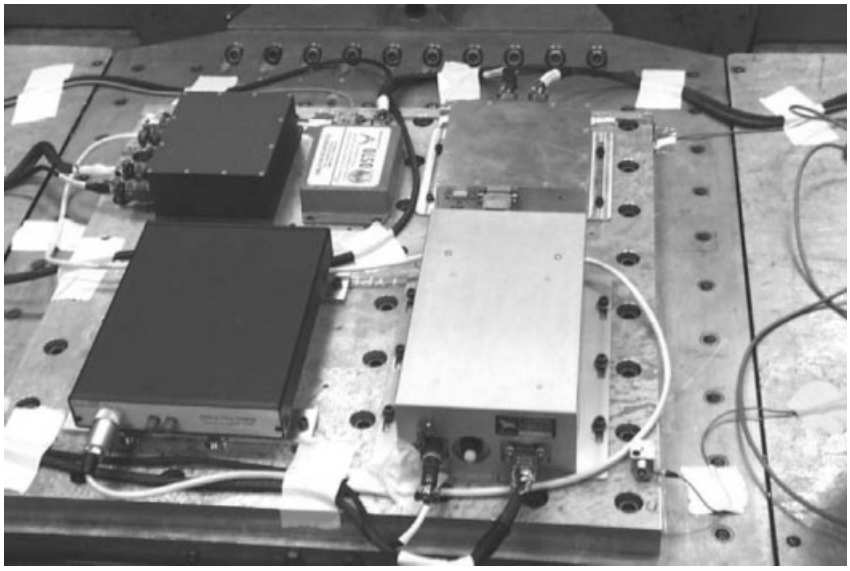


Figure 1. Boeing's ASIMS data-acquisition system and battery integrated with the Luna fiber-optic sensor signal conditioning system.

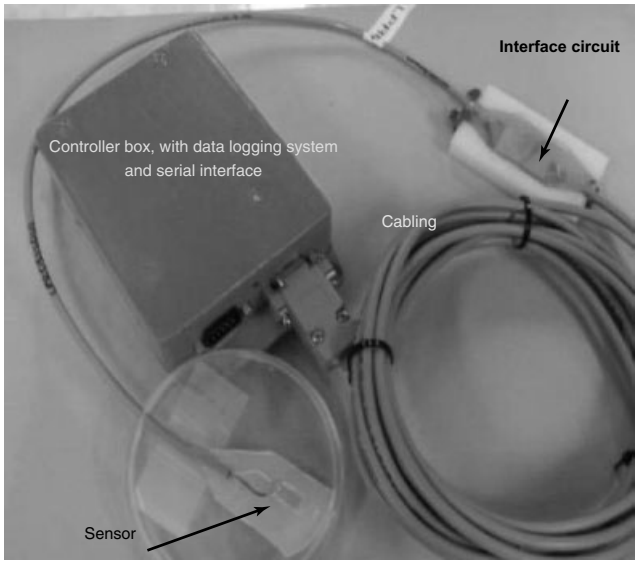


Figure 2. Analatom's microelectromechanical systems (MEMSs) LPR corrosion sensor system.

bay of the 767 flight test aircraft. The sensors were installed beneath the cargo bay flooring in locations that were suspected to be "hot spots" for corrosion. These areas were also chosen for ease of correlating sensor data to actual inspection data that would be provided by Delta maintenance personnel.

Data from the ASIMS and sensors installed in the 767 were downloaded and evaluated for a period of more than two years. Unfortunately, after the airline industry experienced a severe downturn following the terrorist attack of the World Trade Center on September 11, 2001, the effort was discontinued in

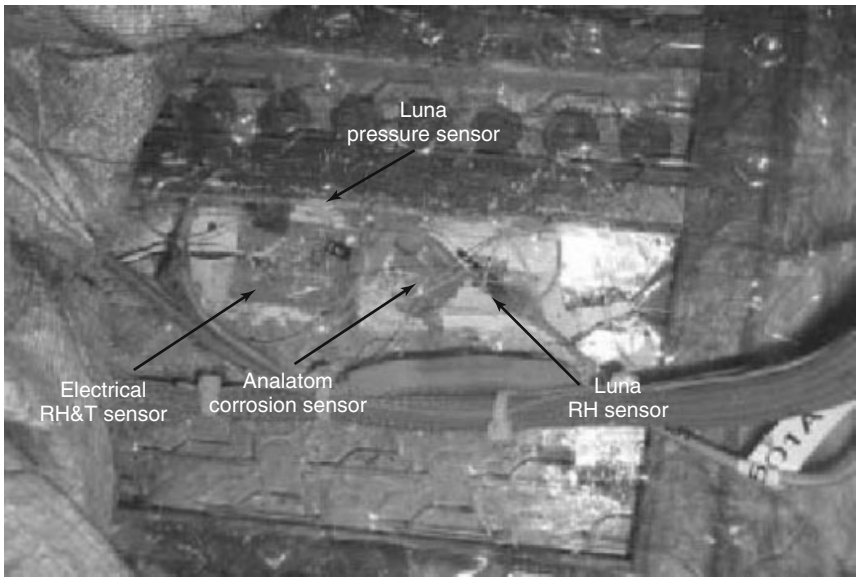


Figure 3. Example ASIMS 767-300ER forward cargo bay installation.

2002 when the airlines began to experience extreme financial difficulties. The effort, however, provided valuable information on the benefits and procedures for installing SHM onto commercial transport aircraft, and the training of airline maintenance personnel for SHM system installation, and data downloading and evaluation.

Another collaborative effort by Boeing, Northwest Airlines, and Structural Monitoring Systems, Ltd (SMS) was started in 2000 to test and evaluate comparative vacuum monitoring (CVM) (*see also* **Comparative Vacuum Monitoring (CVM™)**) systems for monitoring the presence and growth of fatigue cracks in commercial aircraft structures [7]. The basic principle of CVM sensor is that a small volume under a steady-state vacuum is extremely sensitive to leakage, and the resulting change in vacuum level due to air ingress is measurable. The sensors are manufactured from a flexible polymer with channels molded into the applied surface (Figure 4). The sensor is manufactured with a pressure-sensitive adhesive on the applied face and is installed in much the same way as a self-adhesive label. When the sensors have been adhered to the structure under test, these fine channels, and the structure itself, form a manifold of galleries alternately at low vacuum and atmospheric pressure (Figure 5). When a crack develops, it forms a leakage path between the atmospheric and vacuum galleries, producing a measurable change in the vacuum level. The channels should be oriented perpendicular to the expected crack growth direction. The spacing of the channels on the sensor defines the crack size detection limit.

Initial testing was conducted by SMS, with further laboratory validation testing performed at Boeing in Seattle. Testing was conducted on dog-bone-type specimens that had been cut at the centerline. A notch was cut at one of the four bolt holes and a CVM sensor installed on both sides of the plate. Doublers were added and a single line of four bolts along the longitudinal centerline was used to attach the doubler plates to the dog-bone-type specimen. In this way, a high load transfer situation existed between the two halves of the dog-bone specimen and the doubler plates. The CVM sensors were slightly over 0.004″ (0.1 mm) in thickness and were installed directly upon the faying surface of the dog-bone specimen. The standard laboratory equipment offered by SMS was used for crack detection (Figure 6).

After the CVM sensors were tested in a laboratory, they were flight tested on three commercial aircraft beginning in April 2005. A two-year flight test validation program was conducted by the Sandia Corporation in cooperation with the Federal Aviation

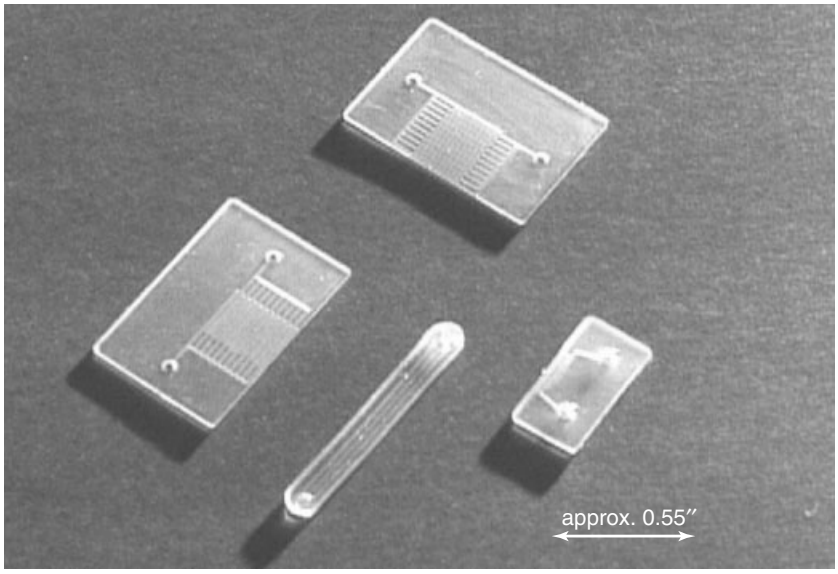


Figure 4. Typical CVM sensors.

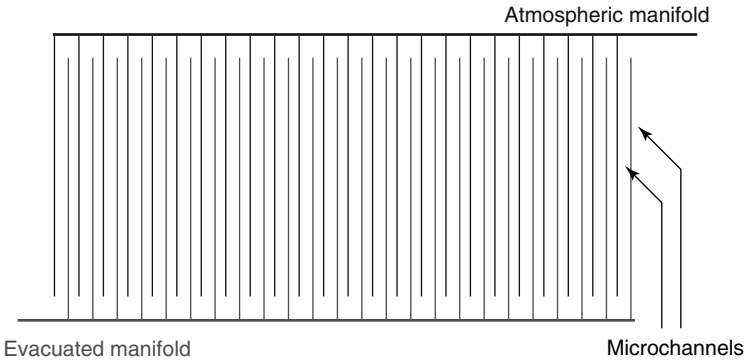


Figure 5. CVM sensor manifolds.

Administration (FAA), Boeing, SMS, a number of US airlines, and the University of Arizona [8]. The program successfully culminated in the FAA certification of the CVM system for a particular nondestructive inspection (NDI) application on a Boeing commercial aircraft, and Boeing's inclusion of CVM technology in its common NDI Manual. This was an aviation industry first for an SHM technology.

Many of the SHM activities of Boeing mentioned above and even more have been summarized in a white paper in 2003 [9]. Here, for the first time, ideas have been presented in a holistic approach including how SHM information would have to be integrated into the aircraft and fleet management process. It is in this paper that possibly for the first time Boeing addressed the concept of a hard landing monitoring system using a neural network-based system.

Although not called SHM at that time, the approach of Airbus toward SHM can be tracked back to the late 1980s when an operational loads monitoring system was developed following the method adopted for military aircraft [10]. The system was configured to be used on the Airbus A320 for monitoring loads specifically in terms of hard landings and limit load exceedances. Although a prototype was realized in hardware, it was hardly flown on a test aircraft

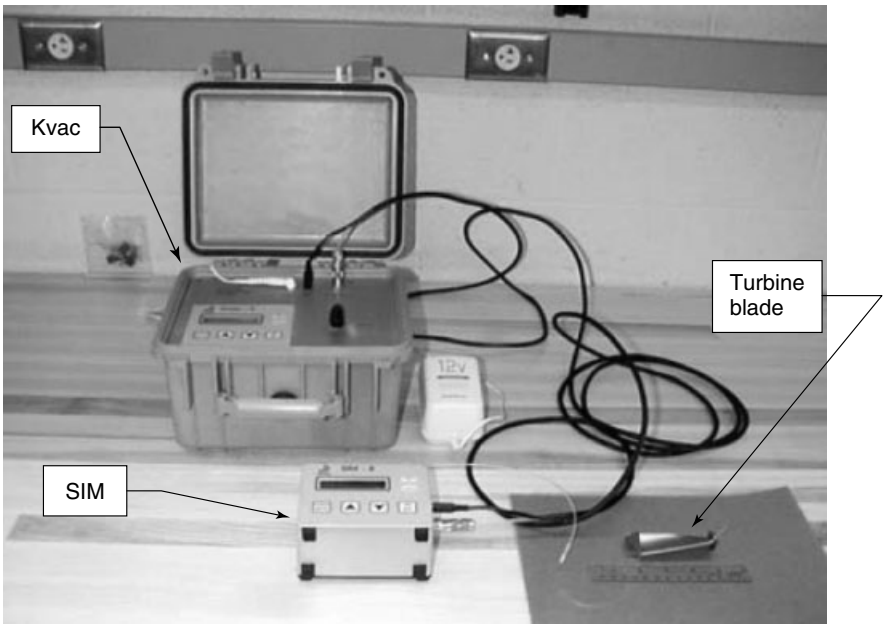


Figure 6. SMS/CVM vacuum-based crack monitoring system (details in **Comparative Vacuum Monitoring (CVM™)**).

officially because economic benefits could not be demonstrated. However, the loads recorded need to be sufficiently precise such that a follow-on residual structural life assessment can be made. This has been a continuous critical point of discussion since the past and becomes even more important once loads monitoring and in-service fatigue assessment become an integral part of an aircraft and fleet management system. Solutions for this are more laborious than anticipated at that time and have resulted in major European research initiatives such as within the EU-funded integrated project *TATEM* that started in 2004 [11].

The initiatives of Airbus toward damage monitoring and SHM, in general, were triggered by the different initiatives in the different countries that were part of the Airbus consortium. A major step in that direction was possibly the MONITOR project, the results of which have been summarized in [12]. It was then that British Aerospace (now BAE Systems) brought in the idea of using acoustic emission and optical FBG sensors for loads monitoring followed by the German side, where Daimler–Benz's corporate research centers, which were also in charge of aerospace technology at that time, brought in the idea of acousto-ultrasonics and optical FBG sensors as well. However, one of the most pioneering publications came from Hofer at MBB Lemwerder (now part of EADS) in 1986 [13] who proposed a transmission-based fiber-optic nervous system (FONS) to be integrated into a composite material, which he had successfully demonstrated in some laboratory experiments. About a decade later a joint test had been done on a multi-riveted lap joint using acousto-ultrasonics [14], which set on further initiatives from Airbus Germany. A remarkable effort was also done by integrating FBG sensors into the rear pressure bulkhead of an Airbus A340 as described in further detail in [15]. The FBG sensors are kept in the aircraft and are subject to further testing with regard to endurance of the sensor system. Activities devoted to CVM were initiated out of Airbus Germany directly as well with initial work starting in the late 1990s, with the first results being reported in [16]. In Spain, work on SHM was initiated at the beginning of the millennium and mainly included monitoring of composite structures either along the Resin Transfer Moulding (RTM) process or within components such as frames, fan cowls or horizontal tail planes [17]. All these activities were finally merged in the EU-funded project called *SMIST*. Details of this as well as the strategic approach of Airbus have been expressed in [18].

A further contribution from civil aviation and Airbus specifically, which set another true milestone, has been H.-J. Schmidt's idea of using SHM for the enhancement of the damage tolerance principle and hence reduced structural weight [19]. This idea has been further pursued and analyzed. It is currently seen that SHM allows for weight reductions of around 20% in the respective components being considered (*see* **Design Benefits in Aeronautics Resulting from SHM**).

According to [18], Airbus has targeted a potential entry into service (EIS) of several competitive SHM technologies for the year 2008. Unlike the early Boeing work, which looked at the development of SHM technologies for enhanced NDI applications for existing aircraft, Airbus envisioned the use of SHM systems permitting new approaches both in the design and the maintenance of new aircraft airframe structures, with the following specific advantages:

- SHM will contribute to reduced structural weight by changing design principles;
- SHM will benefit future designs for composites and metals, maintenance costs will be reduced, and aircraft availability can be increased; and
- SHM will enable new maintenance concepts.

In the late 1990s, Airbus began an evaluation of several SHM technologies [18]. In particular, the following SHM technologies were investigated (i.e. **Acoustic Emission**; **Applications of Acoustic Emission for SHM: A Review**; **Ultrasonic Methods**; **Guided-wave Array Methods**; **Eddy-current Methods**; **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**; **Comparative Vacuum Monitoring (CVM™)**, etc.):

- CVM—see above for a detailed description of the SMS CVM system.
- FBG—see above for a detailed description of the system.
- Acousto-ultrasonics—acoustic waves are sent through the material and received by specific transducers. A change in the local behavior of the material (and hence any structural damage) can be picked up and localized by an array of such transducers.
- Microwave antenna—microwaves are sent and received in a pitch–catch mode inside the material and provide a picture of the water content, which can be used to detect the ingress of water to structures.
- Acoustic emission (see above).
- Eddy-current foil sensors—Eddy currents are generated in the structure. Their pattern and frequency distribution vary according to the presence of cracks or other damages.

Extensive evaluation and testing of the CVM system developed by SMS (*see* **Comparative Vacuum Monitoring (CVM™)**) and an acousto-ultrasonics system developed by Acellent Technologies, Inc (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**) were performed by Airbus. The evaluation of the CVM system by Airbus concluded that it offered a quick and easy way to monitor "hot-spot" areas and thus improve the operational efficiency of an aircraft.

The Acellent acousto-ultrasonics system is based on the use of a network of distributed piezoelectric transducers (sensors/actuators) embedded on a thin dielectric carrier film called the *SMART Layer*, to query, monitor, and evaluate the condition of a structure (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**). Acellent's SHM system includes not only the SMART Layers but also all the hardware and software needed to do the signal transmission, data acquisition, signal processing, and diagnostics. Figure 7 shows an application of this principle to meet the requirements of an aircraft manufacturer. Further applications have been considered such as for monitoring crack growth under composite-bonded repair patches. Figure 8 shows Acellent's SMART Layer applied to a repair patch coupon and the test results displayed as crack growth images and crack length as a function of the number of loading cycles.

The evaluation performed by Airbus included testing of multiriveted aluminum panels under uniaxial constant-amplitude fatigue loading to detect crack initiation at the different rivets as well as the ensuing crack growth [14]. The results of the testing and evaluation performed by Airbus concluded that with a technology like the Acellent SMART Layer, a product was on the way that could be easily bonded on the surface of a structure at relatively low cost. It could be used to automate a troublesome

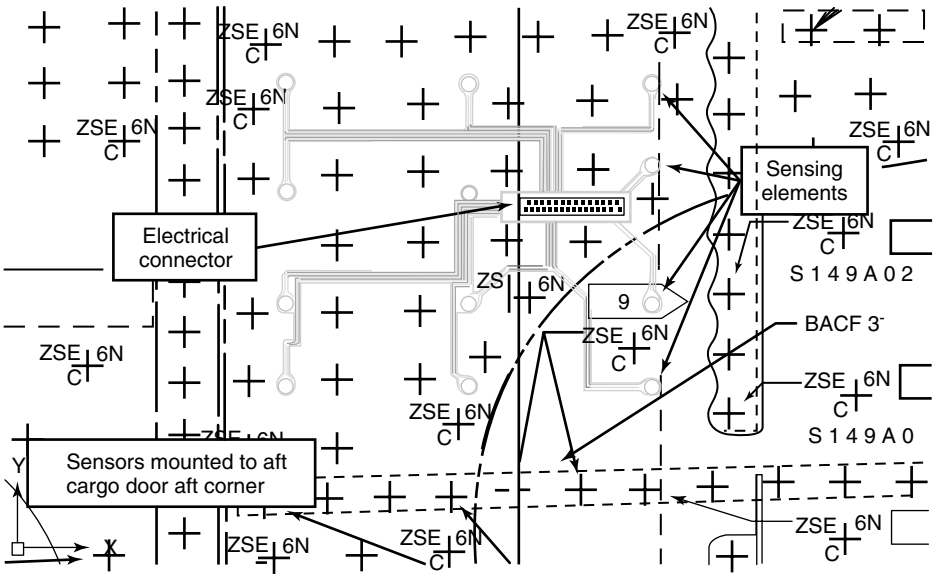


Figure 7. Example of an acousto-ultrasonic monitoring layer installed in an aircraft structure. [Reproduced with permission from Ref. 9. © The Boeing Company, 2003.]

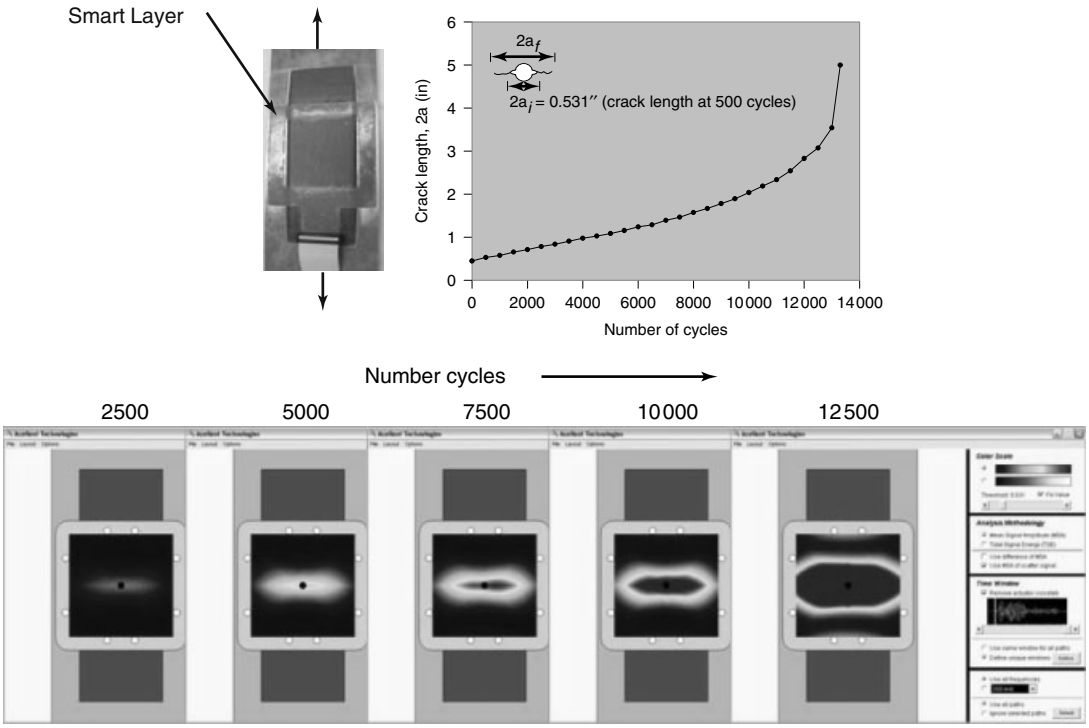


Figure 8. Acellent's SHM system applied to composite-bonded repair patches to monitor crack growth and debonding.

inspection process by avoiding dismantling of complex structural assemblies on an aircraft, and could help make a damage-tolerant design philosophy even more attractive than it is already today.

The meandering winding magnetometer (MWM) sensing principle proposed by Jentek Sensors Inc. is similar to the SMART Layer (*see* **Eddy-current *in situ* Sensors for SHM**). These sensors are shaped field sensors designed as conductive metallic windings, which are placed on a carrier such as Kaptone using microfabrication techniques. A magnetic field is generated from one of these windings, which is then recorded by the other windings. This allows crack propagation in metallic structures to be monitored. The system can also be extended to a high-resolution eddy-current imaging system by introducing a single spatial wavelength or periodic, square-wave inductive drive winding with a linear array of inductive sensing elements. Further to this, the wavelength of the magnetic wave can be varied, which allows to detect cracks, inclusions, and corrosion even in thicker metallic components. The system has been proven to work on cracked aluminum panels, around rivet holes of an aging Boeing 727, and on a C-130 flight deck chine plate. In the 1990s and later, several programs were run by Jentek in cooperation with OEMs and under US government funding including several full-scale tests (such as the Lockheed Martin P-3 test [20]) and numerous coupon tests. For Example, Figure 9(a) provides recent coupon test results performed by the Israeli Air Force (IAF) with support from JENTEK, Figure 9(b) shows a landing

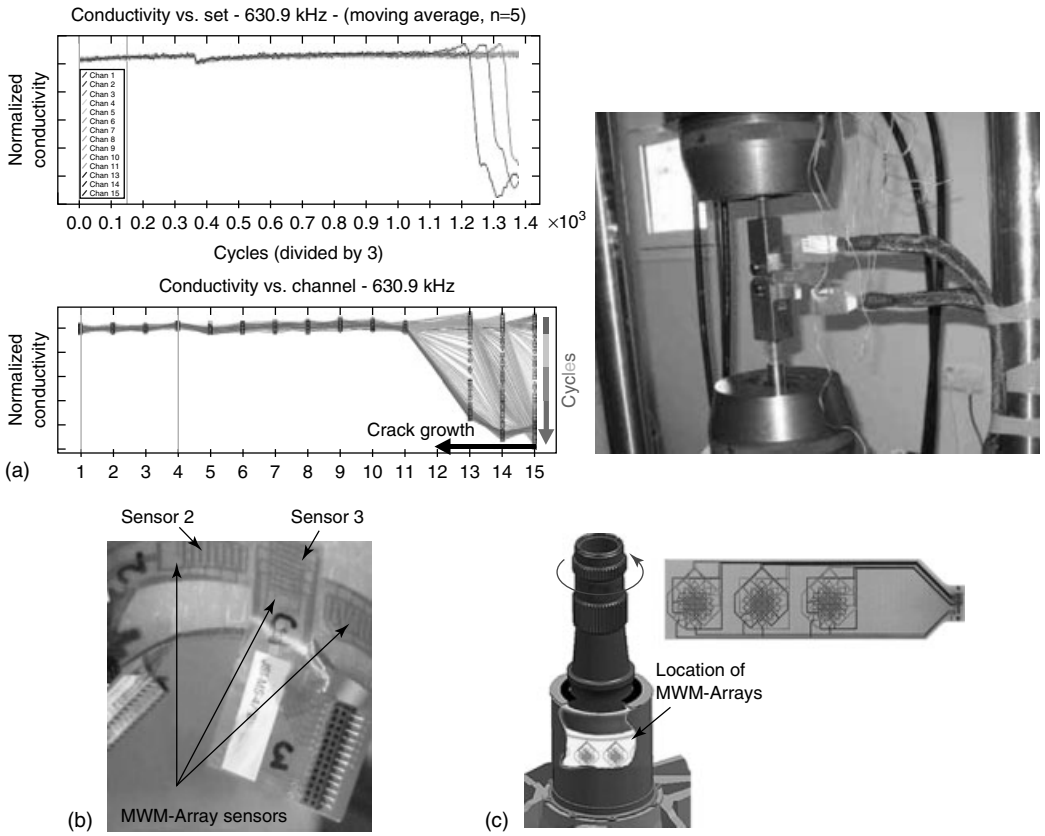


Figure 9. (a) Recent coupon test results performed by the Israeli Air Force with support from JENTEK Sensors; (b) photograph of landing gear component with MWM-Arrays mounted on a critical surface as part of an OEM test; and (c) illustration of an MWM multidirectional stress test recently performed at Boeing Philadelphia. [Figures courtesy of Jentek Sensors Inc.]

gear overload test illustrating MWM stress measurement [21], and Figure 9(c) illustrates an MWM multidirectional stress test recently performed at Boeing Philadelphia [22, 23]. Flight testing of these sensors is now beginning with the IAF and US Air Force. The MWM-Array technology is also used for conventional NDT applications [24].

The Brazilian aircraft manufacturer Embraer is said to place increased emphasis on SHM although not too much has been reported so far [25]. Manufacturers in Japan may be gradually involved into SHM, as can be seen from work being done in large SHM aircraft demonstrators looking at the integration of sophisticated FBG sensors into aircraft composite fuselage structures [26].

3 ROTORCRAFT SHM CERTIFICATION EFFORTS

Over the past three years, commercial aviation regulatory agencies such as the FAA have been increasingly interested in the development of procedures for the certification of SHM technologies for specific aircraft applications. It was previously noted that the FAA was an active participant in the certification of the CVM system for application as an alternative NDI method for a Boeing commercial aircraft. In 2005, the FAA Rotorcraft Directorate initiated a comprehensive program to develop and validate procedures for the certification of health and usage monitoring systems (HUMSs) for commercial rotorcraft [27] and **Experience with Health and Usage Monitoring Systems in Helicopters**. A part of that effort was the development of procedures for the certification of an SHM system for commercial rotorcraft structures [28]. The SHM system would be an integral component of an end-to-end rotorcraft HUMS. As a part of this effort, Acellent Technologies, Inc. is developing a smart patch system (SPS) for rotorcraft structures that will be able to (i) detect fatigue cracks and damages before critical threshold is exceeded or incipient failure occurs and (ii) quantitatively characterize fatigue cracks and damages. This is a five-year project due to be completed in 2010. A prototype

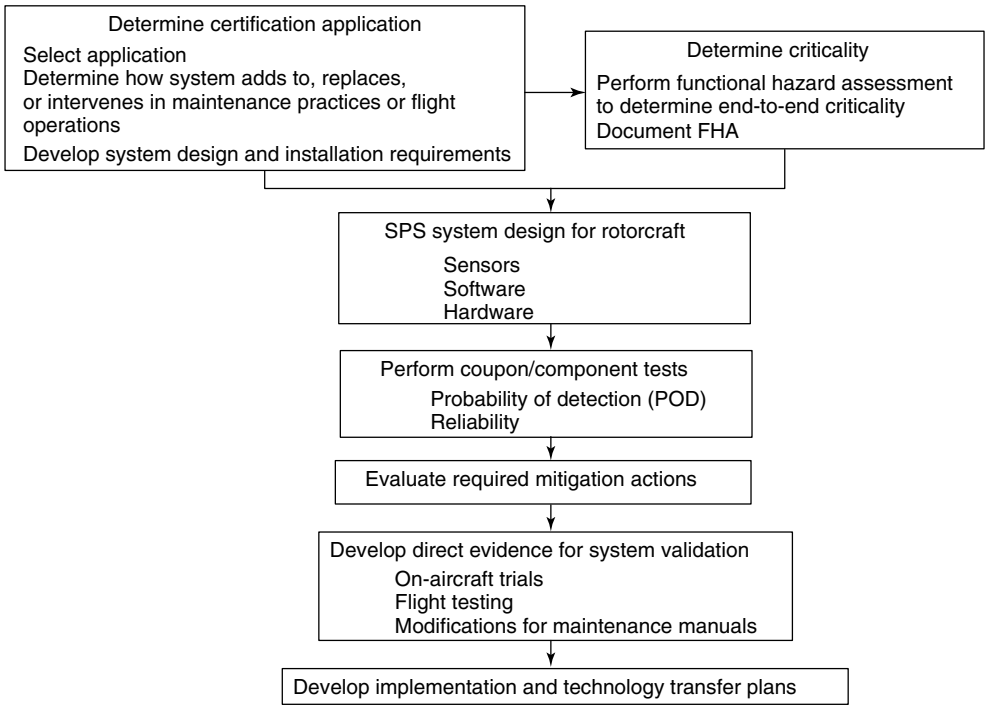


Figure 10. Procedure for the certification of SHM system for commercial rotorcraft applications.

SPS with advanced software, algorithms, and methodologies is being developed, validated, and demonstrated on rotorcraft airframe structures and dynamic components. The basic certification procedure developed to comply with the FAA HUMS certification Advisory Circular (AC) is shown in Figure 10. It addresses the three basic aspects for HUMS certification as described in the HUMS AC, i.e., installation, credit validation, and instructions for continued airworthiness.

4 SUMMARY

Past efforts by commercial aircraft OEMs, airlines, and SHM equipment manufacturers have resulted in the development, demonstration, validation, and certification of a number of SHM technologies that are ready for implementation on commercial transport aircraft. Even though SHM is a relatively new concept to the commercial aircraft industry, the benefits for reducing airframe inspection and repair costs have captured the interest of commercial aircraft operators and the regulatory agencies. Additionally, the commercial aircraft OEMs are now looking at ways to integrate SHM into airframe structural design practices and maintenance concepts for their new aircraft. Although mainly initiated by Airbus and Boeing in commercial aviation, there is hardly any commercial aircraft OEM these days that is not considering SHM to be implemented into its aircraft.

REFERENCES

[1] McBride SL, Maclachlan JW. Acoustic emission monitoring of aircraft structures. *Journal of Acoustic Emission* 1985 **V4**:151–154.

[2] McBride SL, Maclachlan JW. Acoustic emissions due to crack growth, crack face rubbing and structural noise in the CC-130 hercules aircraft. *Journal of Acoustic Emission* 1984 **V3**:1–10.

[3] McBride SL, Maclachlan JW. In-flight acoustic emission monitoring of a wing attachment component. *Journal of Acoustic Emission* 1982 **V1**:223–228.

[4] Hickman GA, Gerardi JJ, Feng Y. Application of smart structures to aircraft health monitoring. *Journal of Intelligent Material Systems and Structures* 1991 **2**:411–430.

[5] Elster J, Trego A, Catterall C, Averett J, Evans M, Jones M, Fielder R. Flight demonstration of fiber optic sensors. Presented at *SPIE Smart Structures and Materials Conference, Smart Sensor Technology and Measurement Systems*. San Diego, CA, March 2003.

[6] Trego A, Clark GJ. Structural health monitoring system: from collection to analysis. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford, CA, September 2005; pp. 1785–1792.

[7] Wheatley G, Kollgaard JR. Automated detection of cracks on the faying surface within high-load transfer bolted specimens. *Proceedings of the SPIE, Smart Nondestructive Evaluation and Health Monitoring of Biological Systems II*. San Diego, CA, B5047, July 2003; pp. 161–168.

[8] German J. *Sensors May Monitor Aircraft for Defects Continuously*, Sandia Corporation News Release, 18 July 2007.

[9] Akdeniz A. *Structural Health Management Technology Implementation on Commercial Airplanes*, Boeing White Paper, 2003.

[10] Ladda V, Meyer H-J. *The Operational Loads Monitoring System OLMS*, NATO AGARD-CP-506; Paper 15, 1991.

[11] www.tatemproject.com, October 5, 2008.

[12] Staszewski WJ, Boller C, Tomlinson GR (eds). *Health Monitoring of Aircraft Structures*. John Wiley & Sons: West Sussex, 2003.

[13] Hofer B. Fibre optic damage detection in composite structures. *Proceedings of the 15th Congress of ICAS*. London, ICAS-86-4.1.2, 1986; pp. 135–143.

[14] Boller C, Ihn J.-B, Straszewski WJ, Speckmann H. Design principles and inspection techniques for long life endurance of aircraft structures. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Stanford, CA, September 2001; pp. 275–283.

[15] Betz D, *et al.* Fibre optic smart sensing of aviation structures. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Munich, Germany, 2001; pp. 306–315.

[16] Stehmeier H, Speckmann H. Comparative Vacuum Monitoring (CVM): monitoring of fatigue cracking in aircraft structures. *2nd European Workshop on Structural Health Monitoring*, DEStech Publishing: Cachan, France, 2004; pp. 367–373.

[17] Esquer PM, Lence FR, Menendez JM. Fibre optic sensors in airbus espana. *1st European Workshop on*

Structural Health Monitoring, DEStech Publishing: Cachan, France, 2002; pp. 1134–1141.

[18] Speckmann H, Henrich R. Structural Health Monitoring (SHM)—overview of technologies under development. Presented at *World Conference on NDT*. Montreal: Stanford, CA, August 30–September 3, 2004.

[19] Schmidt H-J, Schmidt-Brandecker B. Structure design and maintenance benefits from health monitoring systems. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Stanford, CA, 2001; pp. 80–101.

[20] Guadamuz M, Pettit DE, VanOtterloo D. Application of the JENTEK "MWM" sensor to full scale structural testing: a case history, *6th Joint FAA/DoD/NASA Aging Aircraft Conference*, San Francisco, CA, September, 2002.

[21] Goldfine N, Grundy D, Washabaugh A, Craven C, Weiss V, Zilberstein V. Fatigue and stress monitoring with magnetic sensor arrays, *Annual Society for Experimental Mechanics (SEM) Conference, St. Louis, Missouri*, June 2006.

[22] Goldfine N, Sheiretov Y, Dunford T, Denenberg S, Grundy D, Schlicker D, Zilberstein V, Robuck M, Parker C. Magnetic stress gages for torque and load monitoring in rotorcraft. *American Helicopter Society (AHS) 64th Annual Forum*, Montreal, Canada, April 29–May 1, 2008.

[23] Goldfine N, Sheiretov Y, Dunford T, Denenberg S, Grundy D, Zilberstein V. Multi-directional magnetic stress gages. *54th International Instrumentation Symposium; Propulsion Instrumentation Working Group (PIWG)*, Pensacola, FL, May 5–8, 2008.

[24] Goldfine N, Windoloski M, Zilberstein V, Contag G, Phan N, Davis R. Mapping & tracking of damage in titanium components for adaptive life management. *10th Joint NASA/DoD/FAA Conference on Aging Aircraft*, Atlanta, Georgia; April 16–20, 2007.

[25] Chang F-K. An international effort for SHM implementation. *Keynote Lecture held at 4th European Workshop on Structural Health Monitoring*. Krakow, 2008.

[26] Kishi T, Takeda N. Special issue on Japanese smart materials demonstrator program and structures system project. *Advanced Composite Materials*, 2004; Vol. 13(1), pp. 80.

[27] Le D. FAA Health and Usage Monitoring System Research and Development, *FAA HUMS R&D Meeting*, FAA William J. Hughes Technical Center: Atlantic, NJ, 15 February 2005.

[28] Kumar A, Ikegami R, Beard S, Ouyang S, Yu P. Smart patch system for condition based maintenance of rotorcraft. Presented at *6th International Workshop on Structural Health Monitoring*. Stanford University, September 2007.

Chapter 97

Fatigue Monitoring in Military Fixed-wing Aircraft

Matthias Buderath

Product Support, EADS Military Air Systems, Ottobrunn, Germany

1 INTRODUCTION

The general aim of a fatigue monitoring system is the prediction of the remaining fatigue life of each individual aircraft of a fleet. Since military aircraft especially are exposed to a wide range of loads, structural health is one of the substantially limiting factors for their in-service time. Generally, different levels of effort can be applied to realize fatigue monitoring. Low effort on fatigue monitoring results in low costs regarding the monitoring effort but in a high loss of fatigue life. In this case, high safety factors have to be applied or high in-service loads have to be assumed so that structural failure can be excluded.

Increasing investments in the development and production of a new weapon system such as a military aircraft have promoted the tendency to extend the in-service time of existing weapon systems. The initially planned in-service time is ensured by the results of a fatigue test, which is initiated in the design phase. When extending the in-service time, the experience of operational usage, which might differ from the assumptions made in the design phase, becomes more and more important. Changing boundary conditions, such as modified parameters of operational usage and modifications of the weapon system itself, require high accuracy in evaluating individual operational loads.

2 PRINCIPLES OF FATIGUE MONITORING SYSTEMS

Fatigue monitoring has become an essential consideration in regard to the longevity of military fighter and transport airframe structures. Airborne fatigue monitoring systems have been widely discussed in recent decades. Following [1] these systems typically collect

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

operational data for the calculation of the safe life or the inspection interval of the airframe.

The fatigue management of an aircraft is configured in the design process as a result of either the safe life or damage tolerant design principle applied, the load spectrum an aircraft structure is exposed to, the material being used and the shape of the structure itself. All of this results in an aircraft structure's life estimation. This estimate is then certified through a structural fatigue test, following which the aircraft operator collects service load data and puts together a management policy. The process of collecting service load data is termed *fatigue monitoring*, and airworthiness regulations like MIL-A-87221 and Def Stan 00–970 require all military aircraft to be fitted with an onboard fatigue monitoring system. Figure 1 provides an approach as to where results achieved from a fatigue test such as the major airframe fatigue test (MAFT) are taken to identify fatigue critical areas (FCAs) that will then require specific care over the operational life. Another tracking is ongoing for in-service aircraft (and thus in flight), which will allow potential FCAs occurring only in service to be identified and treated accordingly. This latter procedure is specifically important with regard to modifications to be made on an aircraft, which are unavoidable over an aircraft fleet's operational life. All data being generated and collected have a feedback loop into other different blocks of fatigue data generation, which can easily result in a substantial database for fatigue life monitoring [2–5].

2.1 Purposes of structural health and usage monitoring systems

As described more extensively in [1] the only means of managing the fleet in the early days of fatigue management of aircraft was through documenting the number of flight hours or landing cycles. Once the certified number of flight cycles had been reached the aircraft was retired irrespective of its true degree of damage. Later cycle counting methods were developed which allowed some further differentiation in terms of the flight spectra being applied. Subsequently this was further enhanced by the introduction of a peak count method which resulted in so-called fatigue meters (g-counting and fatigue meter formula) and later on, with increased digital storing and processing capability, the recording and utilization of flight parameters and/or strain gauges became a fatigue monitoring standard level. However, the purposes of fatigue monitoring systems remain unchanged. As

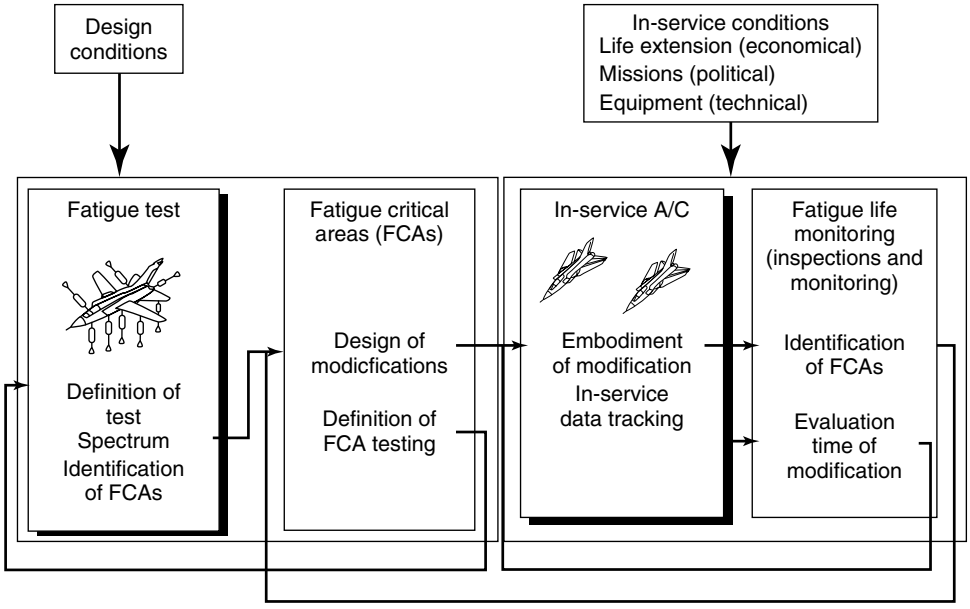


Figure 1. Structural integrity data generation and assessment process during design and operation of military aircraft.

mainly also stated in [1] and **Agile Military Aircraft**, the main purposes are to:

- Fulfill airworthiness requirements to ensure that aircraft are not operated beyond an acceptable level of risk.
- Determine the fatigue life status of a fleet of aircraft throughout its life based on an operational spectrum.
- Determine the actual service load history (many operators in the military aviation sector, and here specifically with fighters, have found that operational usage of an aircraft is significantly more severe than the design spectrum) and to ensure that aircraft are not operated beyond the fatigue damage accumulation threshold for various components as demonstrated through the MAFT (full-scale testing).
- Provide life data that allow the justification and modification of the structural inspection program to be made.
- Improve/optimize the structural integrity management of the fleet (condition-based maintenance (CBM), when done in conjunction with program-based tracking of each aircraft in the fleet). The assertion here is that the utilization of each aircraft is different and that using an average value of aircraft usage is inaccurate when monitoring the whole fleet.
- Detect occurrences (events and load exceedances) of structural overloads in a timely fashion, thus enhancing the fleet safety.
- Assist in the definition of a flight load spectrum for new aircraft of the same type.

Figure 2 summarizes experience gained [6–10] and provides an overview as to when different types of maintenance can only be applied. It becomes obvious that the type of maintenance to be applied is defined during the design stage and that existing aircraft have to stay with basic monitoring system principles in case their type of maintenance cannot be redesigned.

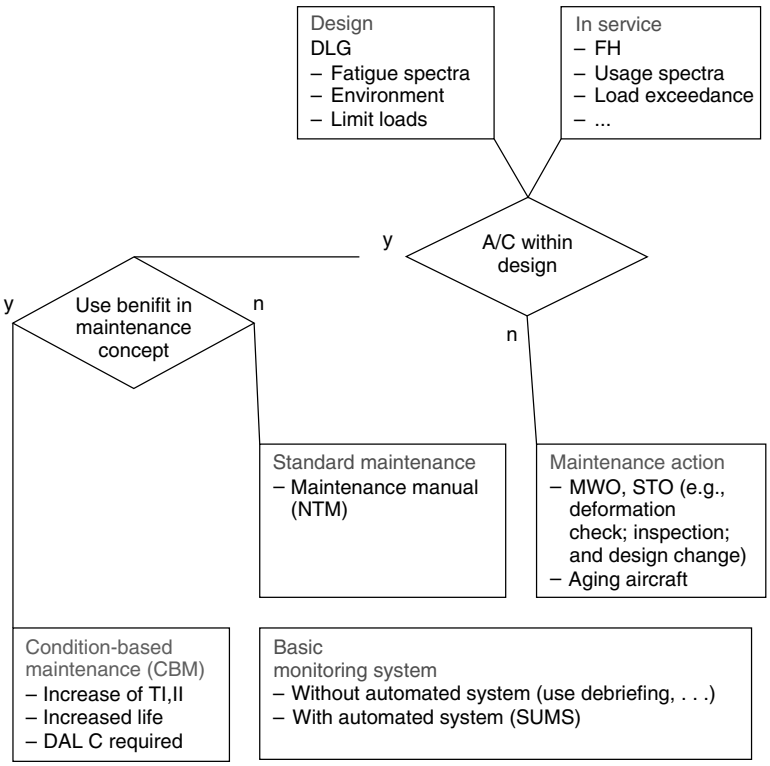


Figure 2. Schematic overview of principles of a basic monitoring system and CBM.

$$UI = \text{usage indicy} \quad \frac{\text{utilization as is}}{\text{utilization as qualified}} \tag{1}$$

$$FI = \text{fatigue indicy} \quad \frac{\text{stress spectrum as is}}{\text{stress spectrum as qualified}} \tag{2}$$

2.2 Functionalities of a fatigue monitoring system

As mentioned above, the process of collecting service load data and combining those with fatigue damage estimation and resulting residual fatigue life estimation is termed *fatigue monitoring*. These service load data generally consist of flight parameters taken from sensors already inherent to the aircraft, which are then converted to the respective operational loads. This approach has the advantage of not requiring any sensors to be added to the aircraft. However, there may be customer requirements or structural necessities that will need the strain (and thus load) sequence to be measured directly in which case a strain gauge (or other sensor) will have to be attached to well-selected locations. Some monitoring systems collect only raw data (time domain data of the respective sensor), while others process the data onboard the aircraft. Besides the question of where the flight data should be processed effectively, the main functionality of an onboard fatigue monitoring system is the recording of all in-service/operational data necessary regarding the fatigue usage of an individual aircraft.

However, the accurate recording of raw data (regarding number of flight parameters/strain gauges, appropriate sample rate, etc.) is only one side of the coin. The other real function of an aircraft fatigue monitoring system lies in the detailed offboard postprocessing and extraction of fatigue-relevant information.

However—and this is the difficulty which is specifically stated in [1] and **Agile Military Aircraft**—military aircraft operators do not follow one standard method of fatigue management as no detailed specifications exist. Design philosophies that feed into fatigue management programs are varied, fatigue test results are interpreted in different ways, and different scatter factors are applied to the fatigue test spectra and fatigue test result. Nevertheless, the necessary growth potential of the basic and enhanced functionalities to cover possible national customer requirements can be summarized for the different categories of onboard and offboard data processing as follows.

2.2.1 Onboard processing

- Recording of an appropriate number of flight-parameter/sensor time histories under consideration of individual sample rates.
- Storage of data on a flight-by-flight basis.
- Upload of user-definable monitoring configuration data without changing the onboard software.
- Load exceedance monitoring including time flag of structural event.
- User-definable allowable loads envelope for load exceedance monitoring.
- Recording span of time around load exceedances.
- Growth potential to add additional parameters and/or sensors.
- Data checking routines; system failure analysis.

2.2.2 On-ground processing

- Download and check of the recorded data regarding completeness and integrity.
- Load exceedance warning, tools for evaluation, and maintenance actions.
- Data processing to calculate/generate fatigue spectra.
- Calculation of usage indices (UIs) and fatigue indices (FIs) based on design assumptions, results of full-scale fatigue tests, etc.
- Growth potential to generate life metrics to support CBM and maintenance optimization.
- Growth potential to generate life metrics for fleet management.
- Growth potential for special technical investigations, e.g., for unexpected in-service problems.
- Database for reevaluation of fatigue-sensitive components.
- Database for unmonitored flight substitution (Sortie Code versus Fatigue Index, etc.).
- Modification of operations to stabilize the rate of fatigue life consumption.

Figure 3 shows the data information flow of onboard data throughout the on-ground fatigue monitoring system. Owing to the complexity of information being presented in terms of decision support and

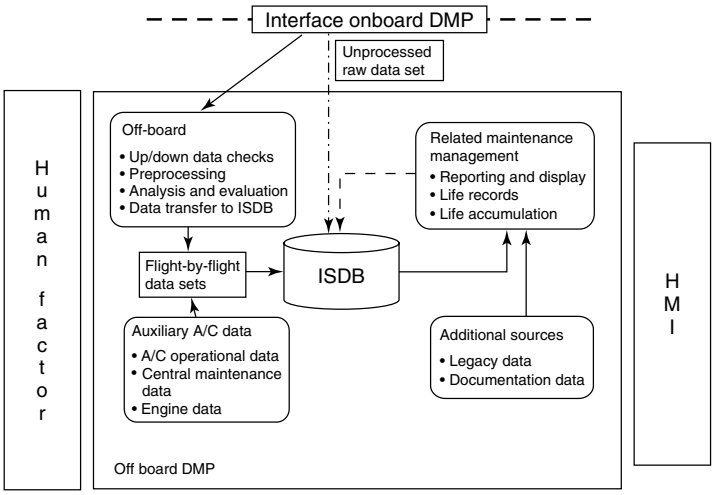


Figure 3. Schematic overview of how to create and how to deal with life consumption.

advisory generation, human factor issues have to be considered as well.

3 FATIGUE MONITORING METHODS

3.1 Flight parameter–based monitoring

The early concepts of aircraft usage monitoring systems were based on flight parameters such as flight hours, mission types, configuration at takeoff, landing weight, and latterly, control surface positions, speed and height among others. With the development of sophisticated acquisition systems, it is now possible to record all necessary parameters as time histories for major load carrying members and subsequently to calculate loads using regression techniques.

These loads can be converted to stresses and strains at fatigue critical locations using transfer functions either derived analytically or by numeric approaches such as finite element modeling. Loading conditions may be separated in terms of tensile, compressive and possibly even multiaxial loads, symmetric or asymmetric flight conditions, different stores configurations, or sub- and supersonic flight.

In summary, it can be stated that flight parameters alone can be used to estimate the dominant load affecting fatigue-sensitive components and, coupled with the use of transfer functions, can be related to stresses at critical locations. Although reasonable levels of accuracy can be achieved, "parameter-only" based approaches for fighter aircraft (using standard flight parameters) can neither account for abrupt maneuvers and gust loads nor directly measure buffet loading. which, for certain components, may contribute significant fatigue damage. For heavy military transport aircraft, these shortfalls may not be relevant.

3.2 Strain gauge–based load monitoring

The sensitive location of the strain gauges can account for aircraft weight and store effects (such as different weapon systems attached to the aircraft), and the variation of principal loads such as the wing root bending moment. Furthermore, strain gauges are sensitive enough to measure abrupt maneuvers and dynamic loading like gusts and buffeting. Thus, the strain gauges must be placed at locations where the fatigue critical loads are representative so that the damge being accumulated can be determined accurately. To achieve this benefit, the location of the strain gauges must be carefully chosen. Following [1], **Agile Military Aircraft**, and others [11–13] care must be taken in particular to ensure that the location of the sensor:

- can be calibrated to the damage-inducing load;
- is dominated by the principal load and insensitive to other loading actions;

- is in an area with a low stress gradient;
- can be directly related to the stress at fatigue-relevant locations;
- is accessible for sensor replacement;
- is replicated on the fatigue test article so that direct comparison can be made; and
- is not dependent on load path variations due to multiple load paths.

3.3 Hot spot monitoring

As stated in reference [14], an alternative philosophy to the strain gauge–based load monitoring is intended to place strain gauges such that they directly monitor the strain/stress at the damage critical locations ("hot spots"). However, there are several problems that can arise:

- The sensor may not be dominated by the principal damage-inducing load and, in particular, it may be difficult to calibrate the sensor response.
- The hot spot may have a high stress gradient.
- There is no guarantee that the maximum strain is monitored.
- If a new hot spot arises, and the gauge does not respond predominantly to the load affecting this new location, then there will be no data available for assessment.
- The hot spot may not be readily accessible.
- Hot spots are not known before finishing the full-scale fatigue test.

3.4 Parameter- and strain gauge–based monitoring

A look at the various fatigue monitoring systems for modern military aircraft being used worldwide reveals that the most popular method is the combination of both flight-parameter and strain gauge–based systems within a fleet. Within this approach, the monitoring methods complement each other. A set of typical parameters being recorded and generated for the different systems is summarized below:

3.4.1 Flight parameter–based system

- Generation of UIs for global overview.
- Parameter/load transfer functions for the generation of load fatigue spectra.
- Parameter/strain transfer function for the generation of "virtual strain gauges" fatigue spectra (based on flight measurements).
- Calculation of FIs based on load/stress fatigue spectra.
- In-flight calibration of strain gauges.

3.4.2 Strain gauge–based system

In addition to the above, this system includes the following:

- load path and hot spot measurements where flight parameters are insufficient and
- support to virtual strain gauge generation (only on dedicated aircraft).

3.5 Processing of collected data

As mentioned before, it makes no sense to record time flight parameters and/or strain gauge responses without subsequent analysis and calculation of UIs, FIs, etc. The purpose of any fatigue monitoring system is therefore the determination of the fatigue life status of an aircraft based on its operational usage.

Independent from the type of in-flight recorder, the amount of collected flight parameters and, if used, the position of the strain gauges, there is still much processing to be conducted before the in-flight data can be used for an assessment of the fatigue usage of an aircraft.

As stated in reference [14], generally, a code is required to format an aircraft's unique data so that it can be processed. It should provide the capability to identify the aircraft (tail number), flight number, aircraft configuration, and mission type. Processing of data recorded is organized in a set of modules in the different actions mentioned below are included.

3.5.1 Data checking and preprocessing module

- Validation of recorded data regarding data integrity (completeness, corrupt data, etc.).
- Determine failed sensors (including sensor information from the flight control system (FCS)).
- Extract number of load exceedances including all information necessary for evaluation and maintenance activities.

- Extract recorded data to time histories.
- Extract number of cycles, e.g., undercarriage cycles and cargo door cycles.
- Extract number and type of landings, e.g., touch and go (TAG), roller, and full stop landings including landing mass, sink rate, bank angle, and drift angle at the moment of touch down.
- Data preprocessing via transfer functions, cross correlations, etc. to get, e.g., aircraft center of gravity (CG), aircraft weight, major loads, and strains/stresses.

3.5.2 Sequence counting module

This module may include a rainflow cycle counting analysis required to generate fatigue spectra for the subsequent calculation of fatigue and thus UIs.

3.5.3 Fatigue module

After generation of the in-service fatigue spectra, fatigue calculations are calibrated against the appropriate fatigue tests having been performed in terms of:

- Global view of the aircraft
  - G spectrum
  - product of N_z *weight spectrum
  - number of, e.g., landings and cargo door cycles
  - product of landing mass × sink rate
  - differential pressure spectrum, etc.
- Major component view
  - wing root bending moment
  - forward and rear fuselage bending
  - fin root bending, etc.
- Local view
  - damage critical areas (hot spots) within the aircraft.

This module generates the life metrics for major components (or even all serialized components) necessary for an advanced maintenance and fleet management concept.

3.5.4 Postprocessing module

This module, whose logic is shown in Figure 4, is the brain of the whole fatigue monitoring system. It is configured as the data management platform (DMP) and updates the database file (in-service database (ISDB)) for each aircraft and, if the accumulated damages equal target values, a warning is given. Furthermore, based on the ISDB maintenance optimization, mission optimization, trend analysis, and fleet management tasks can be performed. The human factor plays an important role on one side, which is the way in which the human being influences the system as such, and the human machine interface (HMI) on the other side, which is the way the communication between the human and the monitoring system is established.

4 AIRFRAME FATIGUE LIFE MONITORING CONCEPTS BASED ON TORNADO AIRCRAFT

The Tornado weapon system was introduced in 1980. Right from the beginning of its in-service time, Tornado was equipped with fatigue monitoring devices to control the remaining fatigue life of critical areas detected during the fatigue test.

Since that time, the concept of fatigue monitoring has been modified, going along with in-service experience and technical innovations. Considering the fact that, for the Tornado weapon system, an extension of the originally scheduled maximum flight hours in the German Air Force is also under discussion, the importance of a reliable fatigue monitoring concept becomes evident. The fatigue monitoring concept applied consists of three major tasks. These are

- in-service measurement;
- mathematical modeling/evaluation of damage calculation; and
- maintenance planning.

The technical innovations led to the fatigue monitoring concepts with increasing accuracy in predicting the remaining fatigue life. As a result of decreasing costs and weight of data-storage capacity, the amount of measured data to represent in-service flight conditions could be enhanced. The corresponding data-evaluation algorithms have gained increased accuracy. This has resulted in a

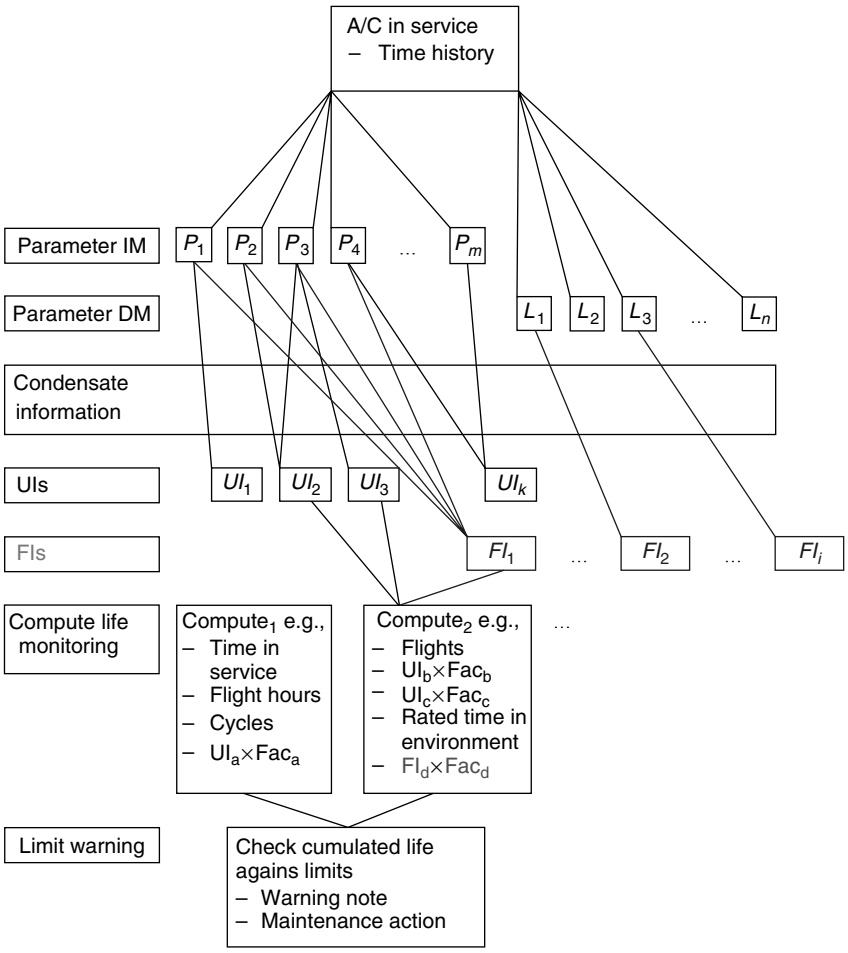


Figure 4. Generic architecture of an on-ground fatigue monitoring system.

subsequent improvement of maintenance planning or, in other words, in maintenance planning that goes along with the actually occurring stresses of individual FCAs [15].

4.1 In-service fatigue monitoring

The "in-service measurement" task for the Tornado weapon system shown in Figure 5 can be described as an integrated concept, consisting of three interrelated components. These are

- individual aircraft tracking (IAT);
- temporary aircraft tracking (TAT); and
- selected aircraft tracking (SAT).

All three components are essentially based on the acquisition of flight parameters, representative of external loads. The components differ with respect to the extent of the measured data.

4.1.1 Individual aircraft tracking (IAT)

The IAT covers a measurement concept, which is applied to all aircraft. A so-called pilot parameter set (PPS) is measured continuously and stored during operational usage. The PPS consists of the parameters N_z, weight, wing sweep angle, and stores. The IAT contributes the data-basis for the calculation of the individual damage index of the FCA.

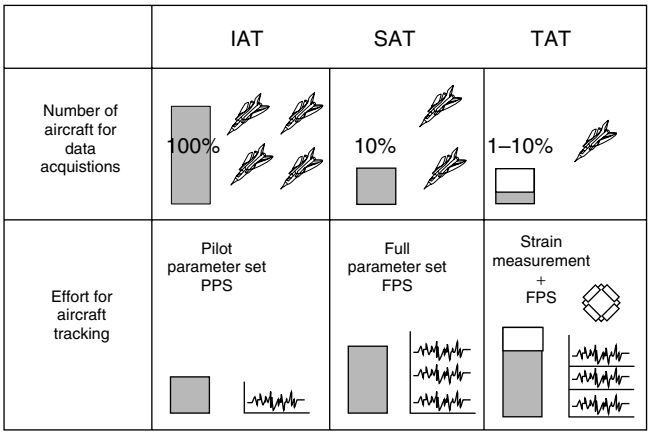


Figure 5. Integrated concept of fatigue monitoring.

4.1.2 Temporary aircraft tracking (TAT)

The TAT is applied to selected aircraft. The measured data set is the most comprehensive one compared with the data sets of the other components. It covers the measurement of the so-called full parameter set (FPS) and the pilot parameter set, which is an inherent part of the FPS. In addition to the flight-parameter sets FPS/PPS, in the scope of TAT, strain is also measured. Strain gauges can be placed at various FCAs. The flight parameters and the simultaneously measured strain allow the adjustment of the coefficients of a flight-parameter-dependent mathematical setup by matching with the actually measured strain.

4.1.3 Selected aircraft tracking (SAT)

The SAT is also applied to selected aircraft. It covers the measurement of FPS and PPS. The SAT is the connecting link between IAT and TAT. The PPS is measured for all aircraft. Starting from a relatively small parameter set, the calculated strain is comparatively inaccurate. The FPS supplies more accurate results regarding the strain, but it is measured only for selected aircraft. To transfer the benefit of a higher accuracy of FPS-based calculated strain, a so-called transfer function is created. It arises from the comparison of the FPS-based calculated strain with the PPS-based calculated strain, which is carried out with the data obtained from the SAT.

Each aircraft of the fleet is thus at least subjected to IAT. Selected aircraft are additionally subjected to TAT or SAT. The general idea behind the concept in data acquisition is that the aircraft that belong to the restricted data-acquisition component (IAT) benefit from the higher accuracy of the limited number of SAT and TAT components. This is an approach that is cost-efficient and simultaneously achieves an overall high accuracy. The flight recorders used with SAT and TAT are distributed on a statistically representative basis throughout the individual squadrons. The simultaneously performed strain/flight-parameter measurements of TAT are cyclically repeated for the same FCA on several aircraft.

4.2 Mathematical modeling

As shown in Figure 6, the precedence of mathematical modeling is coordinated with the concept of in-service data tracking. Results obtained from TAT- and SAT-derived data sets drop into the evaluation of the restricted parameter set of IAT. Because all aircraft are at least subjected to the basic data tracking, this is the basis for the fleet-embracing calculation of the damage factors and, therefore, also the basis for the maintenance planning.

The mathematical model in general should supply the occurring stresses and strains of each aircraft at each FCA by evaluating the incoming flight parameters. The concept of the mathematical modeling has to be adapted to the "physical reality" by considering the system's inherent characteristics and boundary conditions.

Since the comprehensive recording, which is more accurate with respect to the "physical reality", is

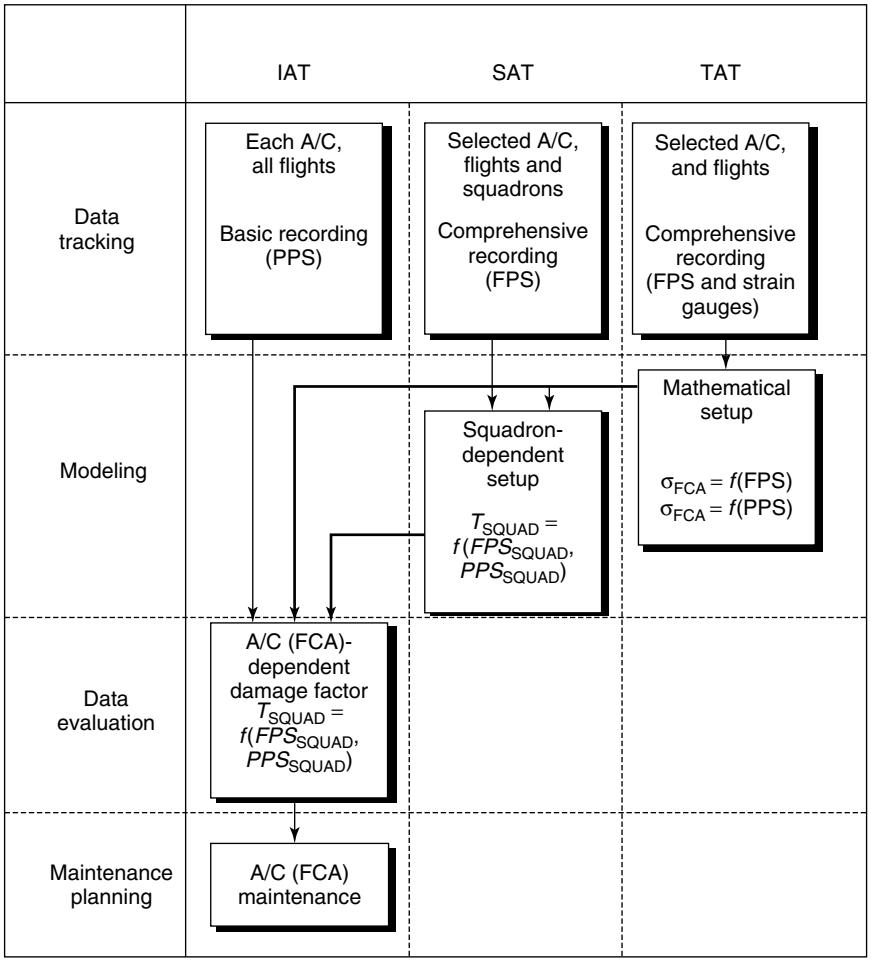


Figure 6. Link of mathematical model with fatigue monitoring.

restricted only to selected aircraft, the concept of modeling also incorporates statistical evaluation methods.

The data obtained from TAT serve for the derivation of the functional relation between the measured flight parameters and the strain at individual FCAs. To choose an appropriate functional setup, first the "physical reality" has to be considered. The strains' respective stresses at particular FCAs depend on the deformation of the aircraft structure. The deformation has to be discussed considering the following two main aspects.

- The external forces and forces resulting from the inertia of masses are the origin of the occurring deformations. The forces can have various points of application depending on the particular maneuver the aircraft is performing. Besides, there are forces that have a deterministic origin as well as forces that occur stochastically, e.g., forces resulting from TAG are induced by pilot action and are therefore of a deterministic origin, whereas forces resulting from gusts have a stochastic origin. Different "classes" of forces might have various influences on functional relations of the strain at a particular FCA.
- The second aspect is that the aircraft itself is not a fixed mechanical setup. Especially, Tornado is a weapon system, which can have various

mechanical setups. The most conspicuous fact in this situation is that the Tornado weapon system makes use of wing sweep technology. Another aspect regarding the mechanical setup of Tornado is the variable mass distribution. Tornado can operate carrying various external loads at different stations. Thus, Tornado might carry the MW1 weapon, weighing more than 2040 kg, under its fuselage or—during another mission—external tanks, weighing 1530 kg maximum each, at the wings. It is obvious that various mechanical setups will have an influence on the strain occurring at particular FCAs.

When setting up a mathematical model to describe strain as a function of flight parameters (Figure 6), the question arises as to how to implement the existence of different aircraft mechanical setups or different classes of external loads resulting from clearly distinguishable sequences of operational aircraft usage. In the concept of the mathematical modeling, this is done by classification. This means that various classes of operational usage are introduced and for each class a set of functions

$$\sigma_{\mathrm{FCA}} = f(FPS) \tag{3}$$

$$\sigma_{\mathrm{FCA}} = f(PPS) \tag{4}$$

has to be derived. In the ideal case both functions should be the same if the PPS parameters are the ones characterizing the loads and thus stresses and strains. However, if this is not the case, then preference will have to be given to the data generated on the basis of FPS.

Figure 7 shows a selection of how classes regarding the mechanical setup of the Tornado weapon system are defined and shows the significant variations a fighter airplane can have in terms of its different configurations.

The next step in formulating the TAT-related mathematical model is to find a mathematical relationship to adjust coefficients by performing a regression analysis. The accuracy of the relationship can be proven by deriving the correlation coefficients between the measured and calculated strain values.

The dependency of strain on a flight parameter can be evaluated by performing correlation analysis. The vertical acceleration N_z is the most influencing flight parameter regarding the strain of an A/C structure. This is the reason why N_z in combination with the

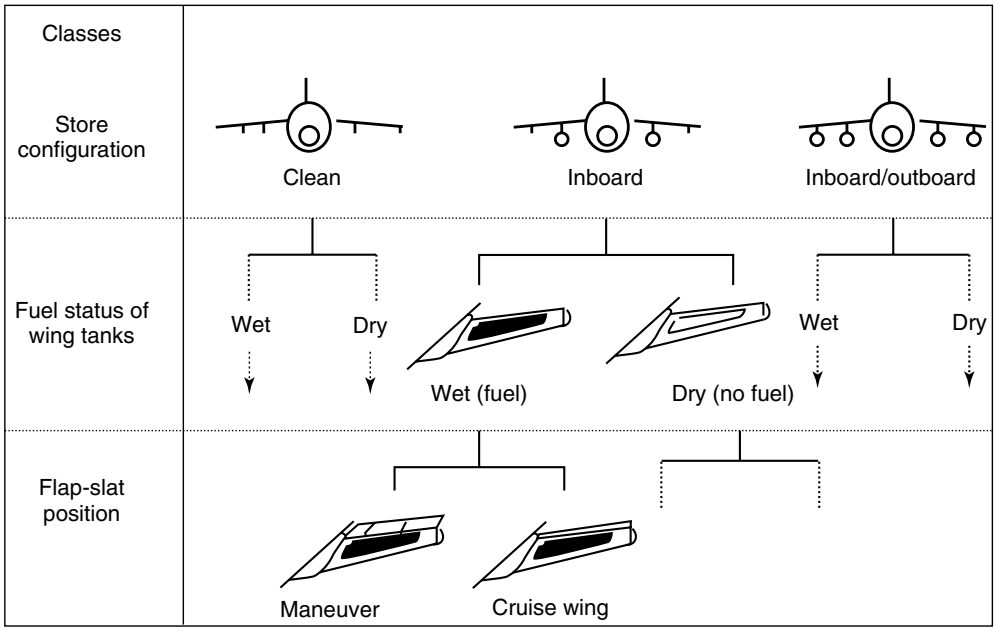


Figure 7. Defined classes.

aircraft weight is the essential flight parameter of the PPS.

Other flight parameters also correlate with strain. The extent of influence depends on the particular flight parameter as well as the location of the FCA. A mathematical setup, which considers various flight parameters, is based on a selection of flight parameters of the FPS obtained with TAT.

4.2.1 SAT-based mathematical modeling

The mathematical modeling in the range of SAT is based on the functional setups obtained from TAT and explained above. These functional setups that rely on the FPS are of a relatively high accuracy, whereas the other setups based on PPS do not fully comprise the influences of maneuvers that are or are not substantially N_z $^*W$-determined. These maneuvers are not considered in the results of the PPS-based function but they are considered in the FPS-based function. Therefore, the PPS-based results always differ from the FPS-based results depending on the occurrence of non-N_z $^*W$-determined.

The structural health monitoring system (SHMS) of Tornado assumes that particular maneuvers are statistically linked to specific tasks of different squadrons. Therefore, the deflection of the FPS-based results to the PPS-based results should be typical of a particular squadron. To benefit from the higher accuracy of a FPS-based result, a squadron-specific transfer function is derived so that less accurate results of a PPS data set can be transformed to a higher accuracy considering the characteristics of squadron-specific operation usage. The target of SAT is to derive squadron-specific transfer functions that determine the relation of FPS-based results to PPS-based results. SAT- and TAT-based mathematical modeling supplies functions that allow the derivation of strain at particular FCAs proceeding from the PPS data. TAT contributes the general functional setup, whereas SAT contributes the squadron-specific influences of the non-N_z $^*W$-determined maneuvers.

4.2.2 Maintenance planning

First experiences regarding FCAs were obtained from the MAFT. For the design of this test, a common load spectrum was defined, which covered the most stringent requirements of the nations participating in the Tornado program. The results of the fatigue test provided the basis for fatigue life assessment in service. The test program revealed the maximum number of safe-life test hours of various FCAs.

Design improvements to modify the FCAs had been derived from the fatigue test results and—as far as feasible—implemented during the manufacturing of the aircraft. Owing to the overlap of fatigue testing and aircraft manufacturing, this was not always possible. Besides, it had to be understood that not all FCAs could be identified by ground testing. Consequently, some retrofit modifications had to be embodied during in-service time for the extension of the fatigue life.

The time of implementing a structural modification is identified by fatigue life monitoring. In case of the FCAs, identified by the fatigue test, the fatigue consumption depends on the usage spectrum of the individual aircraft. Test hours of the fatigue test do not follow a simple correlation with the flight hours of the fatigue test. Therefore, it is necessary to convert the fatigue damage into individual flight hours for the derivation of the point of embodiment for the modification.

Figure 8 schematically shows the major structural modification packages applicable to the German Air Force Tornado fleet and the standardized time of embodiment related to the design life limit. The design life limit (flight hours) represents the number of admissible test hours of the fatigue test.

The respective flight hours are represented by the bars of the chart, e.g., the modification of the center fuselage has to be embodied, when its fatigue life, which corresponds with 100% of the fatigue life according to the design life limit, is consumed. As it is seen, a higher number of flight hours compared with the test hours can be obtained before modification becomes necessary. This means the consumption of fatigue life during operational usage is not as high as the fatigue consumption during testing. The hatched area of the bar indicates the difference between the aircraft with the hardest and the aircraft with the softest operational spectrum of the fleet.

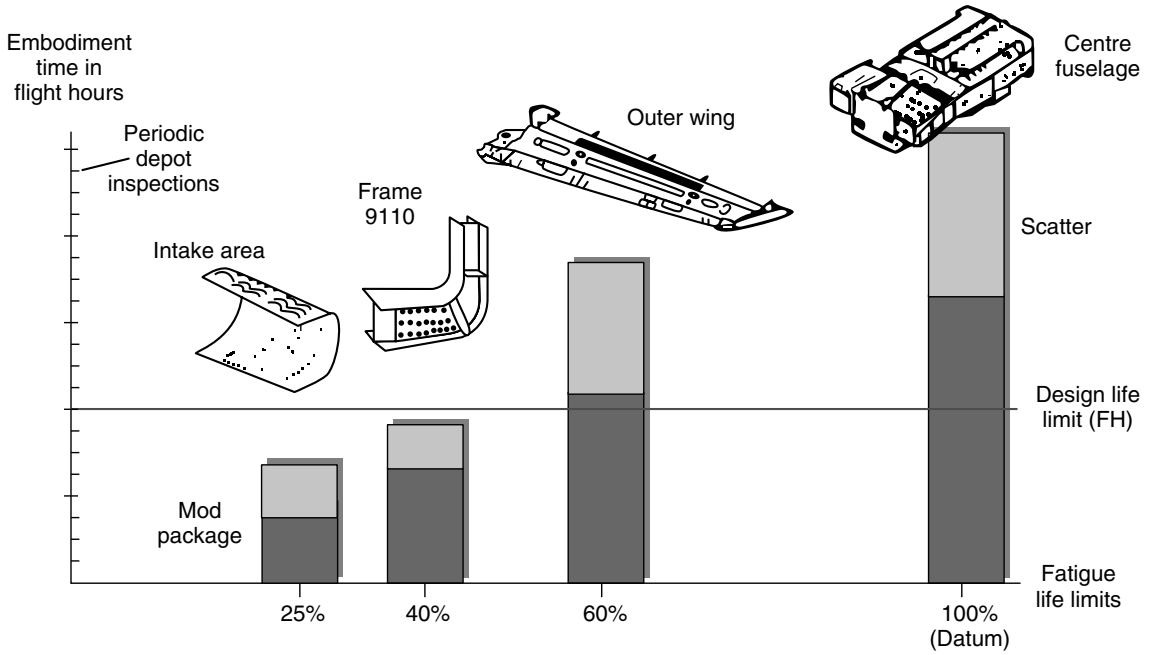


Figure 8. Usage-based maintenance planning.

5 INNOVATION IN FATIGUE LIFE MONITORING

The dynamism in sensor development, loads' modeling and fatigue monitoring techniques these days, which, among others, can be observed in terms of miniaturization, performance, and price, combined with the remarkable progress achieved in sensor signal processing through mushrooming computational power and advanced algorithms, has brought in a new wave of structural technology development that can be entitled "structural health monitoring" (SHM). New and further sensors will allow the monitoring of operational loads at various locations on the aircraft in much more detail, which will further allow the calculation of consumed operational life to be done much more in accordance to the real usage. This will be supported by high detailed loads and finite element models, which allow an accurate calculation of loads based on standard flight parameters. Further to this, there are now more and more sensors emerging that allow the monitoring of damage linking to the traditional usage monitoring of the airframe structure. It can also be observed that the fatigue monitoring, and in the future hopefully also damage monitoring, is becoming an integral part of an aircraft maintenance information system. This article provides a brief overview of the innovation in the field of fatigue monitoring and uses examples from the Tornado Eurofighter and A400M aircraft.

5.1 TORNADO aircraft

The early German Air Force Tornados were equipped with the so-called fatigue meter, which records the cumulative N_z classes within three different wing sweep ranges. As outlined in the previous section, this "basic" classification meets the requirement of a subdivided evaluation for various "mechanical setups" of the Tornado.

Since the Tornado is a weapon system with a wide spectrum of operational usage, the refined concept of fatigue monitoring also considered additional parameters concerning the mass distribution among the aircraft structure as well as various aerodynamic configurations. Thus, the concept using the fatigue meter was replaced by the so-called OLMOS

(onboard life monitoring system) concept. OLMOS and its attached technical devices are still involved in the present evaluations. The data for IAT are stored in the data-acquisition unit (DAU). The device stores the cumulative $N_z{}^*W$ classes depending on the different aircraft setups. These "setup classes" consider the wing sweep angle, the flap and slat position, the external store configuration, and the fuel level in the wing tanks.

Within the scope of TAT, a comprehensive recording of flight parameters is performed. Data are stored by a special device, the maintenance recorder (MR). The data-storage capacity is sufficient to store time sequences of various flight parameters.

The most recent development in data evaluation aims at a comprehensive data recording for each aircraft. The new data-storage device, the flight data recorder (FDR), was introduced to the first aircraft in the German Air Force in 1999 and substitutes the MR. Since the FDR will be installed in each aircraft, this opens new prospects for further development of the data-evaluation algorithms and, hence, for managing the life cycle of the aircraft more efficiently. Figure 9 summarizes the evolution of the different operational load monitoring systems on the German Air Force fleet of Tornados.

5.2 Eurofighter aircraft

Two SHM versions called the *Baseline Fit* and *National Fit* respectively have been developed and are described in further detail below. These reflect the customer requirements for either a parametric-based or strain gauge–based fatigue monitoring system. Both versions are validated by comparison with flight test data correlated to fatigue test results. Figure 10 shows a general overview of the SHM system. The SHM system calculates fatigue life consumed at specified locations on the airframe. The software fitted to the aircraft provides the operator with the option to define the facility via data uploaded by support personnel, whether the location is monitored either flight parameter- or strain gauge–based system.

The SHM system performs real-time fatigue calculations and determines the fatigue life consumed by

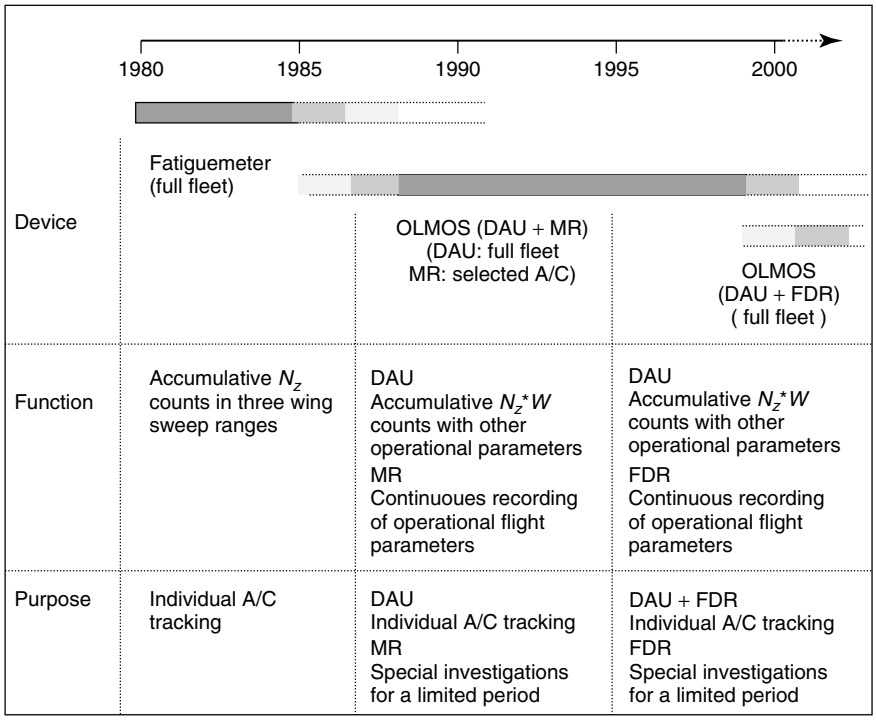


Figure 9. New generation of the Tornado fatigue monitoring system.

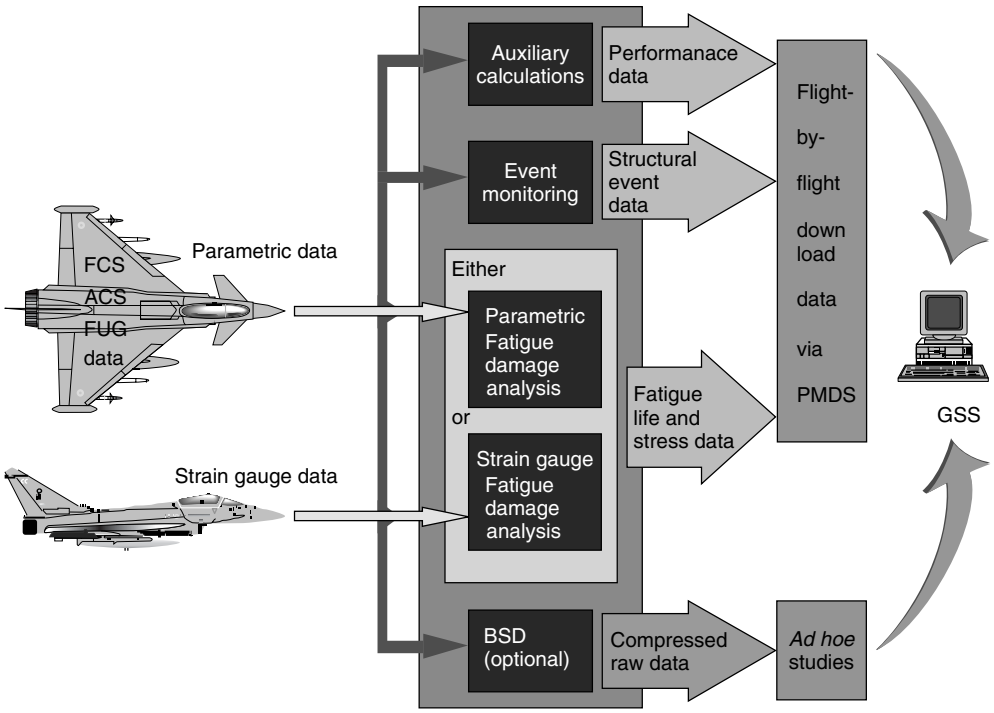


Figure 10. General view of the Eurofighter SHM system.

the airframe with significant structural events and flight performance parameters being also monitored. A facility exists to record parameter and strain gauge time histories, if requested by the operator, for *ad hoc* studies.

5.2.1 Parametric system—Baseline

An overview of the main fatigue calculation processes for the Baseline Fit is shown in Figure 10. Real-time data are captured from the FCS, aircraft control system (ACS), and the fuel gauging system (FUG). The FCS provides aircraft altitude, velocities, and accelerations, the ACS provides information on the aircraft weapons configuration, and the FUG provides fuel mass information. These data are fed into the on-aircraft stress functions, which calculate the stress and hence the damage at each monitored location by comparison with approximately 18 500 templates held in internal memory. Each template, derived from finite element analysis and the results of ground-based airframe fatigue tests, corresponds to a particular aircraft configuration and set of flight parameters. The above process is iterated to generate a history of stress for each location. The stress history is subjected to a real-time range-mean-pairs cycle counting analysis to calculate stress spectra and fatigue damage. The baseline SHM system is based on monitoring 10 locations, with growth potential for monitoring a further 10 locations accommodated by a software change. The system is designed to start operating once the engines have been started and the FCS is active, and it ceases to operate at engine shut down.

5.2.2 National Fit system—strain gauge inputs

An overview of the main fatigue calculation processes for the National Fit system is shown in Figure 11. The National Fit system can record up to 20 analog channels of sensor data. These require sampling and digitizing at rates that are dependent upon the particular monitoring location. A digitizer is required to convert the 20 analog channels to digitized format with a combined sampling rate of at

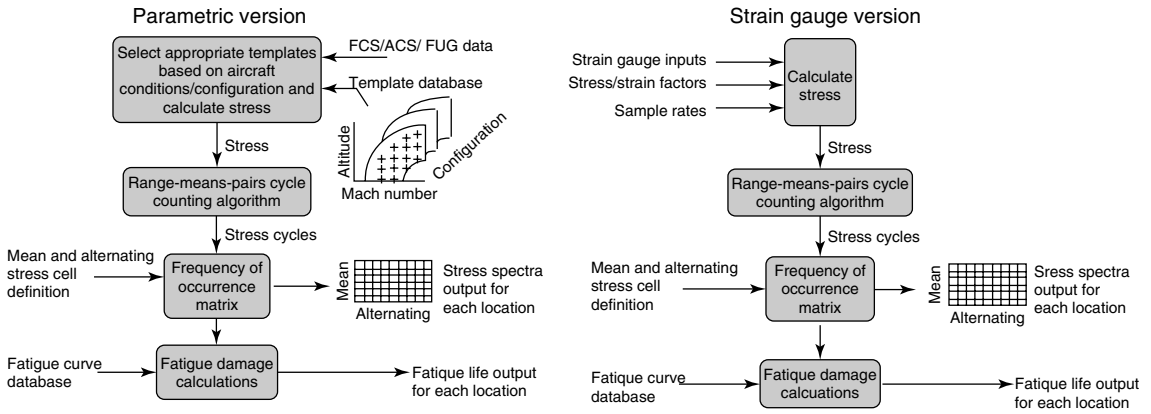


Figure 11. Baseline- and strain gauge–based fatigue calculation.

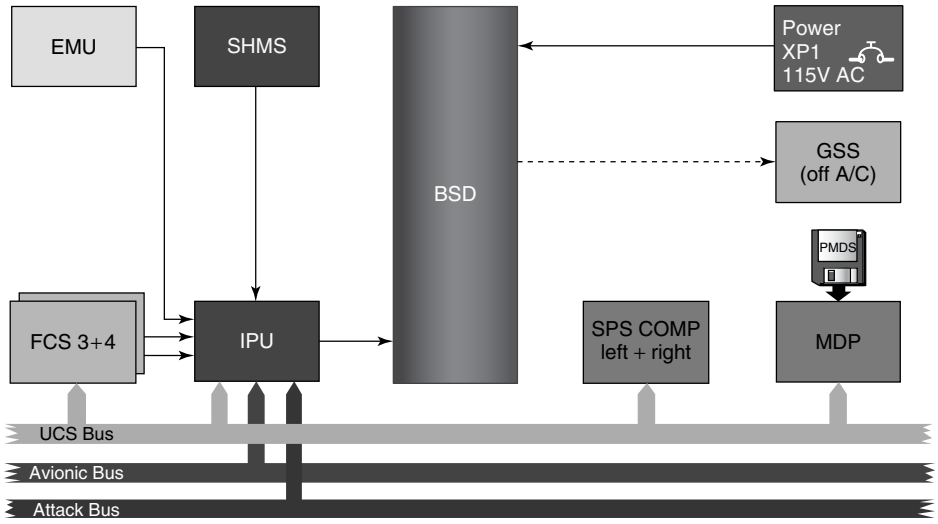


Figure 12. SHM and integral part of the integrated monitoring and recording system Eurofighter. EMU, engine monitoring unit; IPU, integrated processing unit; BSD, bulk storage device.

least 2048 samples per second and a maximum rate of 256 samples per second for any one channel. The sampling rate for each of the 16 monitored channels is set by means of the application software. The digitized data from the 20 channels are placed into memory prior to onward processing. The National Fit system is delivered with 16 sensor channel transducers installed for onboard fatigue damage calculations. The sensor sampling rates and strain-to-stress conversion factors per monitored location are all user-definable via the portable maintenance data store (PMDS).

5.2.3 Integrated Monitoring and Recording System

The SHM algorithms form part of the integrated monitoring and recording system (IMRS) software fitted to each Eurofighter (see Figure 12). The IMRS forms an integral part of the avionics suite on Eurofighter. Its main features are facilities related to the following:

- structural health monitoring
- mission data loading

- video voice recording
- mission data recording
- crash recording
- maintenance data loading
- limited configuration checking
- maintenance data recording
- special study recording
- warning handling
- build in test handling
- recording of consumables information and
- erasure of secure data.

5.2.4 Maintenance Data Panel

The maintenance data panel (MDP) is a fixed on-aircraft piece of equipment that displays information to the support personnel, allowing them to query on-aircraft systems data. SHM details available on the MDP show the total life consumed by each SHM-monitored location and information on SHM event messages that may have occurred on the previous sortie.

5.2.5 Portable Maintenance Data Store

The PMDS is a solid-state memory device approximately the same size as a cigarette packet ($\sim 100 \times 60 \times 25$ mm). The PMDS is used to transfer SHM engine and maintenance data to and from the aircraft as described for various cases in [16–19].

5.3 A400M aircraft

Monitoring systems onboard the aircraft may supplement or even replace periodic inspections that are part of the scheduled maintenance program. Automatic monitoring of the aircraft allows the replacement of certain tasks at predetermined intervals by on-condition maintenance, thereby extending the service life of some items by avoiding premature replacement. An assessment of aircraft status and the need for maintenance action in such cases will be determined by monitoring dedicated parameters to identify the need for maintenance action before an anticipated failure occurs, and monitoring performance or systems configuration degradation to enable maintenance to be undertaken before a critical loss of function occurs.

Degradation of a mission-critical system, otherwise imperceptible to the crew, is detected by the monitoring system, which predicts a schedule interval during which a maintenance action will be required. For the duration of this interval, maintenance action may be deferred to the most convenient time, and aircraft operation may continue with a high degree of confidence that a mission loss or further degradation will be avoided. Despite the fact that this might lead to the removal of a component that still has some potential, it allows the corrective maintenance activity to be performed when the aircraft is on standby or preferably during a scheduled inspection, thereby avoiding mission loss or removal of the aircraft from the flight line. The scheduled interval can be defined based on parameter or configuration requirements, as shown in the example of Figure 13, which should be self-explanatory.

On-condition maintenance, used for a long time on Airbus aircraft, is being systematically applied to all the systems and the structure of the A400M when flight safety allows it. For that purpose, the

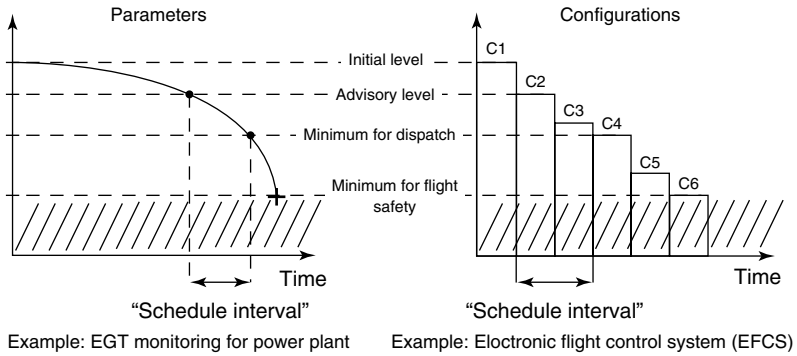


Figure 13. On-condition maintenance—examples of schedule interval definition.

integrated avionics of the A400M is equipped with an aircraft integrated monitoring and diagnostic system (AIMDS) that centralizes the control of the built-in test equipment (BITE) of all systems on each aircraft, detects system faults and provides failure messages in plain English, and collects and records engine, Auxiliary Power Unit (APU) and critical systems data. This data can then be analyzed using the maintenance data system (MDS), which is explained in more detail below, to enable prognostics, trend analysis, maintenance planning, and health and usage monitoring.

In addition, the extensive use of integrated modular avionics (IMA) and redundant architecture makes the continuation of operations with degraded configurations easier. The dramatic reduction of hard-time inspections and overhauls brought about by on-condition maintenance on modern aircraft and on the A400M is one key to reducing the operation and support cost of the A400M aircraft and to ensuring a high operational availability rate.

5.3.1 A400M Aircraft Integrated Monitoring and Diagnostic System (AIMDS)

In addition to an efficient system installation, which reduces the duration of components' replacement, maintenance time can be reduced by improving the diagnostic capabilities of the aircraft. As mentioned before, the AIMDS of the A400M is used to centralize the control of BITE, detect system faults, and provide failure messages in plain English, and collect and record engine, APU and critical systems data, thereby enabling trend analysis and maintenance planning.

This data can be analyzed either onboard the aircraft or more deeply on the flight line by using a portable multipurpose access terminal (PMAT) that allows ground crews to interrogate the AIMDS via plug-in points inside and outside the aircraft to facilitate a more in-depth analysis of failures. In addition, the PMAT will provide access to the Interactive Electronic Technical Publication (IETP) of the aircraft for troubleshooting and corrective actions.

In addition, the aircraft is equipped with a lifetime monitoring system (LTMS), to perform direct measurements via strain gauges to detail information of the usage and fatigue life consumption of the monitored components. This system will record aircraft structural loads, including overloads, hard landings, and total cycles accumulated, enabling operators to track aircraft utilization and associated fatigue. This will allow the operators not only to plan their on-condition maintenance more efficiently but

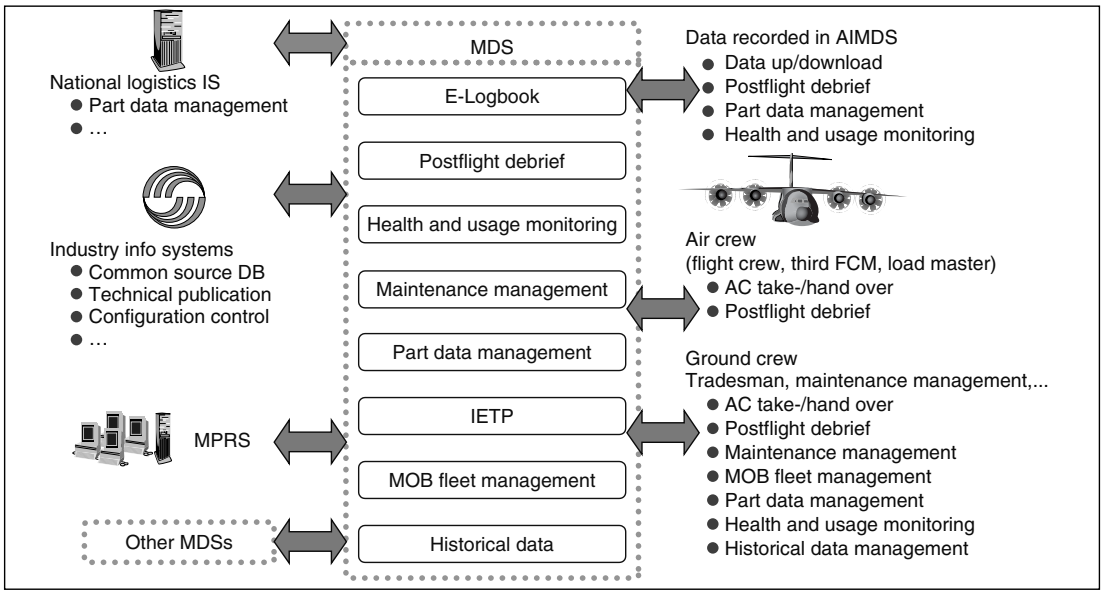


Figure 14. A400M MDS external interfaces and main functions.

also to review the scheduled maintenance program if required. The LTMS will work on the same principles as described in this article.

A schematic representation of the functionalities of the A400M MDS, showing its external interfaces and main functions, can be found in Figure 14.

The PMAT can download the AIMDS data and upload it into a ground support system (GSS) that includes both a mission planning and restitution system (MPRS) and a MDS. These systems allow the storage of the data in a database, manage the performance, fatigue and corrective actions on the aircraft, and schedule the preventive maintenance tasks. It presents the aircraft status in real time and performs other functions such as SHM. In addition, studies are being conducted to allow the transmission of maintenance data from the aircraft in flight to the ground via a data link, which would allow more efficient management of maintenance activities on the flight line and increase operational availability. Such methods are already in use by some civil airlines and military transporters [20].

6 CONCLUSIONS

The article shows the principles currently applied in the development and implementation of structural health and usage monitoring systems for military aircraft. The article also links the fatigue monitoring principles to a real application, e.g., Tornado aircraft fatigue monitoring and management.

The innovations in that field are exemplary, illustrated by the Eurofighter and A400M programs. The principles of fatigue and usage monitoring are more or less the same but many improvements in the information fusion, integration, processing, and information management technologies can be seen. These are considered as key enablers to manage structural integrity, aircraft/fleet availability, and the reduction of operation and support costs. In particular, operational availability cannot be dissociated from the related costs going along with corrective and scheduled maintenance actions. Significant operation and support reductions can only be achieved at aircraft level by having a comprehensive view about the aircraft system and structural performance. The more the information can be integrated into a maintenance information system, the more operational and cost benefits become obvious. This is a common requirement seen in the military air forces.

The next generation of military aircraft will hopefully have a usage and damage monitoring system implemented.

RELATED ARTICLES

Principles of Structural Degradation Monitoring

Loads Monitoring in Aerospace Structures

Risk Monitoring of Aircraft Fatigue Damage Evolution at Critical Locations

Military Aircraft

Usage Management of Military Aircraft Structures

Commercial Fixed-wing Aircraft

Agile Military Aircraft

Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft

Validation of SHM Sensors in Airbus A380 Full-scale Fatigue Test

REFERENCES

[1] Molent L, Aktepe B. Review of Fatigue Monitoring of Agile Military Aircraft. *Fatigue & Fracture of Engineering Materials & Structure* 2000 **23**(9):767–785.

[2] Schütz W. Fatigue Life Prediction for Aircraft Structures and Materials, *AGARD-LS-62*. Advisory Group for Aerospace Research and Development: Bordeaux, 1973, p. 10-1–10-32.

[3] Buderath M. Concepts for Life Extension Programmes of Ageing Aircraft. DLR Conference, Munich, 2001.

[4] Ward AP. The Development of Fatigue Management Requirements and Techniques. In *Proceedings of the 72nd Meeting of the AGARD Structures and Materials Panel*, Bath, April 1991, *AGARD-CP-506*. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1991.

[5] Sampath SG. Aging Combat Aircraft Fleets—Long Term Applications. *AGARD-LS-206*. Advisory Group for Aerospace Research and Development: 1996.

[6] Schütz R, Neunaber R. Operational Loads Data Evaluation for Individual Aircraft Fatigue Monitoring. *Proceedings of the 58th Meeting of the AGARD Structures and Materials Panel*, Siena, April 1984, *AGARD-CP-375*. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1984.

[7] Boller C, Buderath M. The Impact of Monitoring on Extending Aircraft Operational Life, CEAS Forum: Life Extension—Aerospace Technology Opportunities, Cambridge, UK March 23-24, 1999. Royal Aeronautical Society.

[8] O'Hara J. The Evolution of the BAe Hawk and its Structural Clearance. *Proceedings of the 17th Symposium of the International Committee on Aeronautical Fatigue: Durability and Structural Reliability of Airframes*, Stockholm, 1993.

[9] Ward AP. Tornado—Structural Usage Monitoring System (SUMS). *Proceedings of the 58th of the AGARD Structures and Materials Panel*, Sienna, April 1984, *AGARD-CP-375*. Advisory Group for Aerospace Research and Development: Neuilly sur Seine.

[10] Buderath M. Through Life Cycle Support—supported by Health Monitoring, NATO Symposium Brussels, 2005.

[11] Bochmann R. *Fatigue and Loads Monitoring Parametric versus Strain Gauge-based*. EADS Technical Report, 1995.

[12] Henkel C. *The Airframe Fatigue Monitoring Concept for the WS Tornado in the German Air Force*. EADS Technical Report, 2001.

[13] Buderath M. *A Feasibility Study to Manage Components of the Tornado Aircraft According to Damage Tolerance Principles*. EADS Technical Report, 2002.

[14] Molent L. A Unified Approach to Fatigue Usage Monitoring of Fighter Aircraft Based on an F/A-18 Experience. *Proceedings of the 21st Congress of the International Council of Aeronautical Sciences*, Melbourne, 1998.

[15] Hickethier H, Neumair M, Baudisch B and Ranft R. *Life Time Monitoring Concept for Transport Aircraft*. EADS-MAS Technical Report, 2004.

[16] Hunt SR, Hebden IG. Eurofighter 2000: An Integrated Approach to Structural Health and Usage Monitoring. In: International Committee on Aeronautical Fatigue: *Proceedings of the ICAF 97: Fatigue in New and Ageing Aircraft*, p. 481–498, 1997.

[17] Sanchez J, Franz M and Manco E. *Specification of Eurofighter Ground Support Station*. EADS-MAS Technical Report, 1999.

[18] Proceedings of the RTO-AVT Meeting on Enhanced Platform Availability through Advanced Maintenance Concepts and Technologies, Latvia, 2006, NATO RTO Report AVT-144 (RWS).

[19] Buderath M. *Combat Aircraft SHM Functional Desciption*. EADS-MAS Technical Report, 2000.

[20] Heuninckx B. MCIPS: Availability Improvements in New Transport Aircraft—The Case of A400 M. Enhanced Platform Availability through Advanced Maintenance Concepts and Technologies, Latvia, 2006, NATO RTO Report AVT-144 (RWS).

Chapter 102

Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance

Kenneth Pipe
Humaware, Petersfield, UK

1 INTRODUCTION

The first work attempt to use condition monitoring techniques for helicopters was a research program undertaken by the UK ministry of defense (MoD) in the late 1970s. The technical breakthroughs in the technology were made by a UK R&D company—Stewart Hughes Ltd—by pioneering the use of pattern matching and model reference techniques. This technology became the foundation of what is now known as a HUM system (*health and usage monitoring system*), which is described in this article. The acronym HUM was coined by Westland Helicopters to describe a range of technologies, which, when taken together became an integrated system that could be fitted to a helicopter to provide a defect forecasting capability. The first implementation was on the Westland WG 30 helicopter. Other manufacturers—Sikorsky with their mechanical components diagnostics system (MCDS) and MDDS programs, Eurocopter with the Euro-ARMS program, and Bell Helicopter with their BUCS program—also developed a capability during the 1980s.

The first HUM system was certified in November 1991 for operation in the North Sea. The system was one of two competing designs developed for meeting the oil companies' requirements for a HUM system following the HARP (helicopter airworthiness review panel) review of helicopter airworthiness. Both systems were designed to meet the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

same requirements, but each operated with significantly different functionality and operator interfaces. In the 10 years that followed, HUM systems for civil helicopters have appeared from Eurocopter and latterly from Bell, again ostensibly designed to meet the same North Sea requirements, but they are also distinctly different systems. Lately, in the military arena Intellegent Automation Inc (IAC), General Electric (GE) Aerospace, and Goodrich have produced HUM systems, which represent yet further variations in design.

There is guidance on what constitutes a HUM system published by the UK CAA (Civil Aviation Authority) Helicopter Health Monitoring Advisory Group [1], but this guidance has not led to the emergence of a consistent systems' design philosophy nor has it established criteria for the measurement of a HUM system's performance.

This article aims to explore how the performance of these systems of differing designs can be established and provides a comparison of their relative attributes.

The purpose of a HUM system is to provide a timely indication of the deterioration of the continued airworthiness of a component in order that maintenance can intervene and rectify the defect. It is a supplementary method to the various prescribed inspections and preventative maintenance actions contained in the helicopter's maintenance handbook to ensure the helicopter's continued airworthiness and thus enhance the margin of safety of helicopter operations. To achieve this improvement, HUM systems monitor (i) the usage of the lifed components, (ii) any exceedances from the operational envelope, and (iii) the health of the power train components. The health monitoring of components provides a goal keeper function, guarding against any failure of the maintenance procedures to preserve the airworthiness of the helicopter. The most important health monitoring feature is the vibration monitoring function, which utilizes techniques of varying degrees of sophistication to identify defects developing in the power train components.

There are secondary functions in HUM systems, which relate to obtaining maintenance credits (*see* **Experience with Health and Usage Monitoring Systems in Helicopters**). Principal amongst these is the balancing of shafts and rotors without employing specialized maintenance test equipment and in the case of the main rotor, eliminating the need for maintenance test flights following the rotor adjustments.

It is, though, the primary requirement of a HUM system to enhance airworthiness, which requires that the contribution of a HUM system to a helicopter's airworthiness can be measured.

2 MEASURING AIRWORTHINESS

Airworthiness can be defined as the risk to an aircraft of a catastrophic incident posed by its design or operation. The design criterion set in the Federal Aviation Authority's (FAA) JAR 29.547(b) and 917(b) is that for any failure mode in the dynamic train, the risk of it causing a catastrophic failure of a helicopter must be less than 1 catastrophic failure in 10^6 flight hours, which is a probabilistic measurement of a very rare event. The only practical estimates of this probability are from large fleets of aircraft operated over many years. Relating a helicopter's overall risk of failure to the performance of a particular HUM system function is not possible unless statistics from many billions of flight hours more than the fleet will fly can be gathered.

The Regulatory Authorities and the accident investigation bodies do look at the contribution, or otherwise, of a HUM system in the case of a catastrophic failure. Design assessments for certification of new helicopter design can assess the contribution of a HUM system to overall aircraft airworthiness. Neither of the above processes has produced criteria around which HUM system performance standards can be determined.

The Heinrich Principle states that accidents have incidents as their precursors. Analysis of the causes of the much more frequent occurrence of incidents can be used to provide data that would allow measures of risk to be made. This allows performance measures to be established that will by inference ensure the much less frequent accident rate is impacted in the desired direction.

The Regulatory Authorities maintain records of mandatory occurrence reports and the CAA, which is the only Regulatory Authority to mandate a HUM system [2], does record its contribution to reported events. Therefore, there is an experience base of HUM system operation that can provide some actuarial evidence of performance. The CAA has

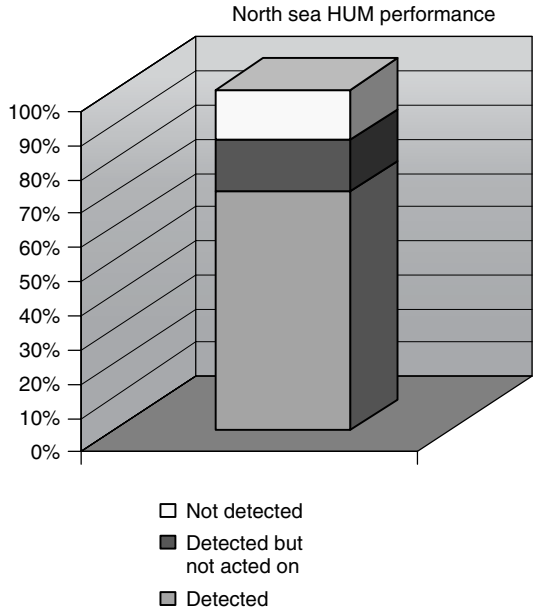


Figure 1. Success Rate of HUMS in North Sea.

published data on HUM system performance in the North Sea [3] as summarized in Figure 1. This data was informally updated in 2002 and the success rate was basically the same [4]. This North Sea experience documented by the CAA shows a 70% success rate in detecting defects. This can be used as a performance measure for the contribution of HUM systems to airworthiness.

The distribution of faults detected by North Sea HUM system is shown in Figure 2. This diagram shows that, with the exception of the planetary gear system, there is no justification for applying different performance standards for individual components of the dynamic train. Clearly, components where the technology has not been shown to be effective should be excluded from the performance measure, or a measure more applicable than the North Sea standard should be employed.

This actuarial measurement of performance is only viable for large fleets generating a large number of hours and where there is independent verification of the statistics. This is possible in aviation because of the monitoring of operations imposed by the Regulatory Authorities. It is unlikely that any vendor would accept such a performance criterion unless he is offering a mature system with a track record of success. This overall performance criterion does represent a differentiator for separating the wheat from the chaff in the market place.

The corollary to the success rate is the false alarm rate. Specifying a demanding target for the success rate could always be met in principle, but with a high cost in nugatory maintenance activity. Fear of this potential problem with HUM systems has been a major impediment to helicopter manufacturers

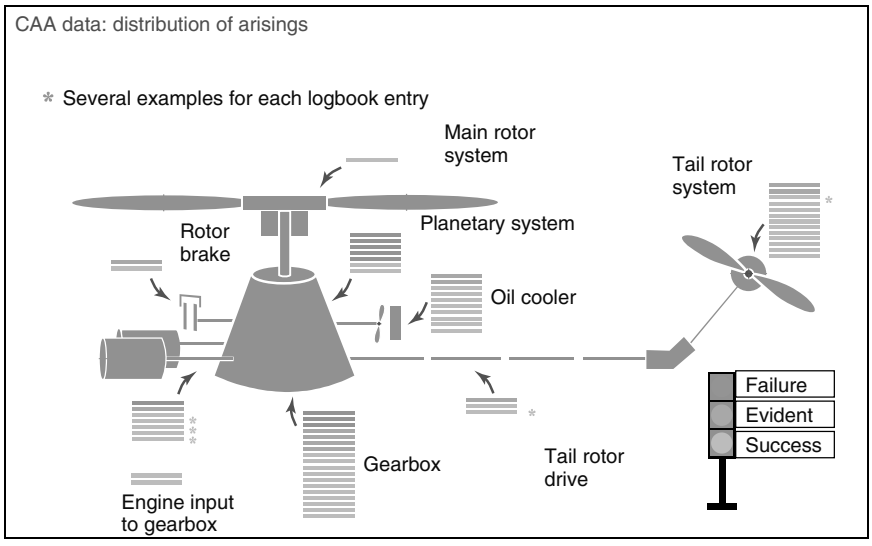


Figure 2. Drive Train Defect Coverage. [Reproduced with the permission of the UK Civil Aviation Authority].

incorporating HUM technology into their maintenance practices.

An acceptable definition of false alarm in the HUM system context is contentious. For much of the industry, any single alarm generated by a HUM system not shown to result in a defect being found is judged to be a false alarm, even though the alarm did not lead to any nugatory maintenance. This is a very demanding definition, particularly for alarms being set on random data such as those generated by vibration. It is usually found that setting of alarm levels with sufficient sensitivity to fault conditions leads to the generation of spurious alarms. If these are shown to be spurious and dealt with without incurring unnecessary maintenance actions, such as inspections, then this should be treated as part of the alarm processing, albeit processing by the engineer rather than the system. A better definition of a false alarm is an alarm that has caused nugatory maintenance action. Dealing with spurious alarms in a HUM system environment can be regarded as an operating overhead. This overhead cost should be measured and the system should have features that ensure that it is kept to a minimum. The time spent at the ground station and skill levels for alarm processing needs to be specified by the vendor.

The inverse of the successes are the missed faults, which is also an important measure for controlling the performance of the system. Determining the reason for a HUM system failing to detect a fault is necessary for quality control of the HUM system and for ensuring that any gaps in performance that are within the scope of the technology are closed.

These actuarial measures of HUM system performance can be monitored by any operator operating under the CAA's CAP 693 guidelines [5].

Any actuarial measurement of performance relies on ample evidence being available to generate the statistical measures at an acceptable confidence level. For a new HUM technology or a new helicopter design, these data are not going to be available and other approaches have to be taken.

3 BUILDING BLOCKS OF A HUM SYSTEM

The approach to a new system has to be to identify the components of a HUM system and establish the performance criteria for these components such that an estimate of the expected overall performance can be made. Figure 3 shows the processes that have to be performed for an alarm to be produced by a HUM system.

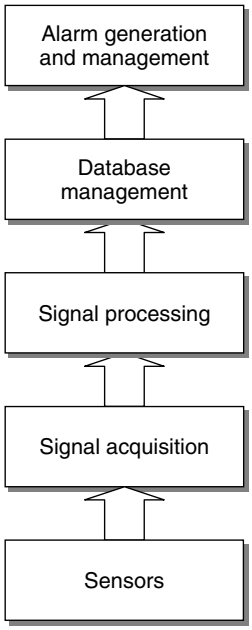


Figure 3. HUMS Architecture.

Sensors with the correct characteristics have to be fitted; the signals are then acquired and processed to produce variables that are stored in a database. There may be more than one database that has to store the variables without loss; and these variables are then processed to produce alarms, which have to be managed until either a defect or a false alarm is determined.

For the monitoring of data such as airspeed, the HUM system has to emulate the processing used for acquiring and displaying the data in the cockpit, with application of the same performance measurement criteria. There can be difficulties here in that usually the HUM system's signal acquisition and processing is not identical to the cockpit instrumentation and can produce results of differing accuracy and resolution. Establishing what the tolerable difference between the HUM system's data and the cockpit data, which is usually less accurate and of a lower resolution, can be a more strenuous exercise than first envisaged.

This problem can be further compounded when data is used in calculations such as those used to compute power assurance. These differences are often small but can cost the credibility of the system unless they are understood and accepted by pilots and engineers.

There are sensors and processing used in HUM systems for different purposes than those for which they are used in the aircraft. Tachometers are typical of these. These Tachometers are usually pulse devises, integrated to produce shaft speed, for HUM system, and they are also often used as a phase reference for vibration processing, so that the performance of the sensor in terms of producing a regular and well-defined pulse shape is more precisely specified. In most HUM systems, the processing of vibration data is highly dependent on reliable tachometer data, requiring that the performance and integrity of this system component is very thoroughly engineered.

Then there are the specialized sensors and processing that are peculiar to HUM system; these are principally the vibration processing for health monitoring of the shafts, gears, and bearings, and also the blade tracking systems combined with the vibration processing used for rotor track and balance. Vibration processing adds a great deal of cost and complexity to a HUM system and needs to perform well to justify the burden it imposes.

4 VIBRATION PROCESSING IN HUM

Vibration has been shown to be a very powerful tool in faultfinding in the helicopter's dynamic train [6]. Any deformation of a rotor or shaft produces an increase in its vibration. Faults in gearing actions, cracks in splines, and spalls in bearings all produce discernable changes in the vibration characteristics of the helicopter.

Shaft-induced vibration has been shown to be relatively straightforward to deal with reliably by HUM system and it produces a high probability of success at an acceptable false alarm rate. This is not so for the more complex analysis techniques where a more variable performance has been experienced in the field.

The difficulty with vibration monitoring is that one sensor integrates a large number of vibration signals from the components of the machinery it is monitoring. A typical vibration signature for a component on a helicopter transmission is shown in Figure 4.

The vibration spectrum for a helicopter can be divided into six regions. The lowest frequencies are quite discrete and are primarily due to the helicopter's rotors and their harmonics. As the frequency increases, the transmission shafts, engine shafts, and meshing frequencies come into play. All of these machinery-generated vibrations cause the structure to

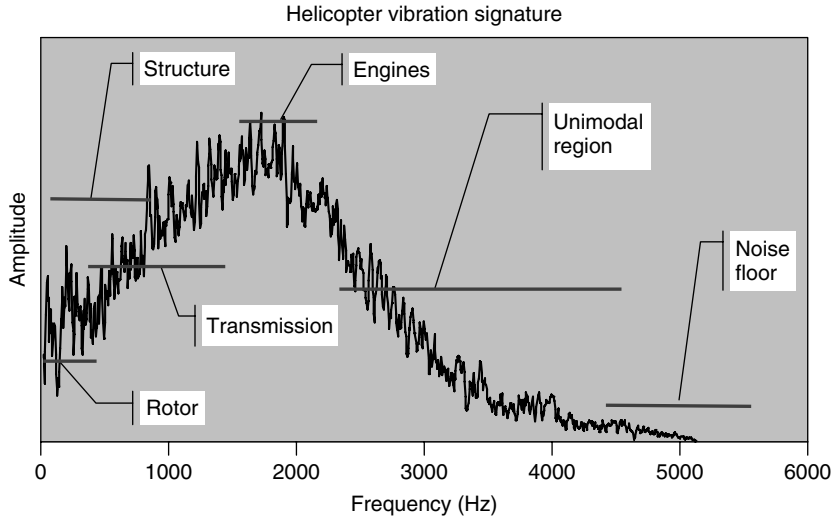


Figure 4. Helicopter Vibration Signature.

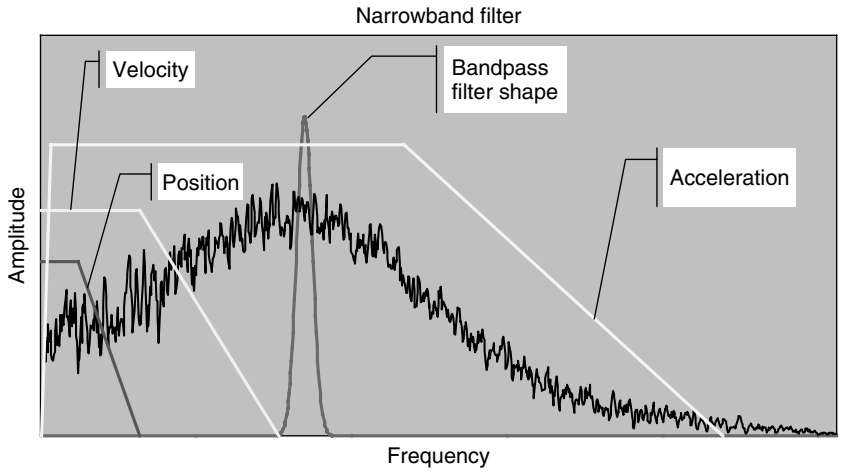


Figure 5. Narrow Band Filter.

vibrate, and as the sensors are mounted on the structure, the structural resonances or modes generate a range of tones throughout the vibration signature. All of these vibrations produce harmonics and the energy in the increasingly complex signature builds up into the characteristic bell shape as shown in the diagram, with the frequency discrimination increasingly becoming less distinct. After the primary shaft orders and major structural resonances frequencies have passed, the spectrum is then made up of only the higher harmonics, the energy in the spectrum then falls off, and any frequency discrimination disappears altogether. The lack of frequency discrimination leads to the term *unimodal* to describe this region. Finally, no spectral energy is discernable and the energy levels off, this is the noise floor of the sensor and processing.

Each of these regions makes different demands on the vibration processing and different technologies have been produced to provide useful analysis of the data. This has impacts on the sensors as well as on the processing capability of HUM system. These will now be considered in turn.

5 SENSORS

Performance of the sensors is clearly critical to the performance of the vibration processing in a HUM system. Vibration sensors originally measured position and were very restrictive in frequency range or bandwidth. The technology then progressed to velocity sensors, which offered higher bandwidths and tracking filters, as shown in Figure 5; these could be used to analyze higher frequencies such as engine spool vibration. It was the development of the accelerometer that brought sufficient bandwidth to cope with the whole of the useful spectrum.

Accelerometers do have a problem of roll-off at dc as well as at high frequencies. Specifying an accelerometer that can measure the low-frequency rotor orders as well as the higher frequency harmonics of the mesh tones can prove to be a demanding requirement. Often accelerometers with different sensitivities and bandwidths have to be used on the helicopter: one for structural vibration monitoring giving the required performance at lower rotor and structural resonant frequencies, and another type for the transmission and engine frequencies. Environmental conditions also play an important role in accelerometer specification.

At the start of the HUM system development, accelerometers were largely laboratory devices but great strides have been made in their design and constriction to make them suitable for use in the aviation environment.

The positioning and installation of an accelerometer is as critical as its bandwidth, range, and sensitivity for determining its performance. If the sensor is acquiring the lower frequency rotor and structural vibration, then the sensor positioning is less critical

as these vibrations can be sensed almost anywhere in the helicopter. For the higher frequencies, or for when the shaft frequency is the same, sensor positioning is more critical to produce the necessary discrimination between components. The sensor has to be installed as close as possible to the component so that its vibration dominates the spectrum.

Great care has to be taken over the installation of sensors to ensure that the sensor orientation is appropriate for the vibration that is being measured and that the installation bracket does not introduce structural responses that either swamp the signal or interfere with it.

Adequacy of the sensor positioning can only be demonstrated by the production of spectra that show that the characteristics of the monitored component's vibration signal is clearly discriminated in its normal operation such as the once per rev vibration and harmonics from the shaft and the mesh tones in the gearing.

The problem comes in determining whether components that do not produce a characteristic vibration in normal operation can be monitored. Bearings and splines fall into the class of components that only produce vibration when there is a fault present. The only practical method to determine if the component can be discriminated is by trials using seeded "marker" faults in the component. Only the helicopter manufacturer has the resources to perform this exercise, but it is expensive and is probably only practical to carry out during the development program's rig tests of components.

In the absence of this evidence, what can be done to establish performance? For external bearings, such as tail rotor drive bearings, placing the sensor on the bearing housing and checking the signature for the lack of other vibrations apart from the supporting shaft is a reasonable approach. It is important that where evidence of the component's characteristic vibration being discerned cannot be produced, then alternate techniques should be available in the HUM system to monitor the component.

The best evidence that the sensors have been selected, positioned, and installed correctly can be produced by performing trials to establish that the vibration signals can unambiguously discriminate the features required to substantiate the claims made for the defect monitoring. One sensor type is unlikely to meet all monitoring requirements. Specifying the characteristics of the sensor independent of its installation is not sufficient. Positioning and installation of the sensor are much more critical to a HUM system's overall performance and this aspect of the sensor's performance cannot be determined from the drawings.

6 SIGNAL ACQUISITION

Unlike process variables such as pressure and temperature, vibration does not have a single mode of operation to model to develop its performance criteria. Signal acquisition techniques for vibration again follow the development of the technology from RMS measurements of the overall signal energy, through tracking filters and finally the modern digital signal-processing techniques. In terms of technology, the digital acquisition of signals represents a complete break from the past, and using digital techniques to emulate the older RMS and tracking filter techniques is not generally possible.

Current processor technology and very low cost memory means that signal processing is now entirely software driven and signal acquisition is a matter of anti-aliasing filtering, analog to digital (A/D) conversion at a high frequency to a resolution of 16 bits or higher and then piping the data into memory (*see* **Data Preprocessing for Damage Detection**). This approach can have performance consequences downstream. For example, if a spectral resolution of 2 Hz is required to discriminate the rotor frequencies and a maximum frequency of 50 KHz is required to acquire the higher frequencies and if there are 16 sensors acquiring data once every 20 min, then the amount of data accumulated per flight hour is in excess of 150 MB. It is now not a problem to store or process this data, but it could take some time to download it from the aircraft.

If HUM system data is going to be useful for identifying meaningful trends, then most data needs to be gathered in steady-state conditions or in specific regimes. For a reasonable acquisition rate to be achieved, then specifying the regime recognition parameters that match the regimes the helicopter actually flies is important. This is a more difficult issue than it appears as there is always a conflict between reducing the range of the regime to minimize the variation in the data and increasing the range so that

it does not restrict the time for data acquisition and hence the number of acquisitions.

The number of acquisitions that are required per flight hour is the key parameter to specify for data acquisition, and not the number of channels. The acquisition rate needs to be sufficient to allow for the time the helicopter spends in maneuver. A single channel of acquisition can mean that for the number of sensors fitted in HUM system applications, the number of acquisitions per flight hour may not be sufficient to generate meaningful trends.

Quality and quantity of data are the features that drive the data-acquisition requirements in HUM system, and these are the items that need to be specified with clear performance criteria that relate to trending and other uses of the data.

7 SIGNAL PROCESSING

There is no single signal processing technology that is effective for all types of vibration analysis. The fast Fourier transform (FFT)-based spectrum analysis techniques that are effective at low frequencies where resonances are pronounced are ineffective in the midfrequencies where the spectral energy is dense and complex in structure and they are totally ineffective in the unimodal region where there is no tonal energy at all. In these cases, more sophisticated types of processing have to be employed. Key to the performance of a HUM system is that appropriate vibration analysis techniques are employed for the range of fault detection capabilities claimed.

All processing is now digital using the FFT in varying degrees of sophistication on the data piped to memory from the acquisition process. There is a problem in FFT processing of matching frequency resolution with frequency range as shown in Figure 6. It shows the number of spectral lines in each decade of frequency range for an analysis with a minimum frequency of 5 Hz and a maximum of 10 KHz, plotted on a logarithmic scale. In the lower frequencies, which contain most of the spectral data, there are the least number of lines, 2 in the 1–10 Hz range and 18 in the 10–100 Hz range. In the unimodal region where there is no spectral information, there are 1800 lines. FFT processing covering a wide frequency range will not produce the frequency resolution where it is required. There are "zoom" FFT processors that can locally increase resolution in the frequency band and this processing is essential if spectrum analysis is going to be used over a wide frequency range.

The most powerful technique developed for frequency analysis in the complex environment of the helicopter is synchronous signal, or time domain averaging (*see* **Statistical Time Series Methods for SHM**). This produces spectra where the resolution is a multiple of the period of one cycle of revolution of the shaft, and the frequencies are a function of the shaft speed. This separates the contributions of each shaft to the overall signal and automatically controls the FFT processing's resolution and bandwidth conflict. HUM system vibration processing

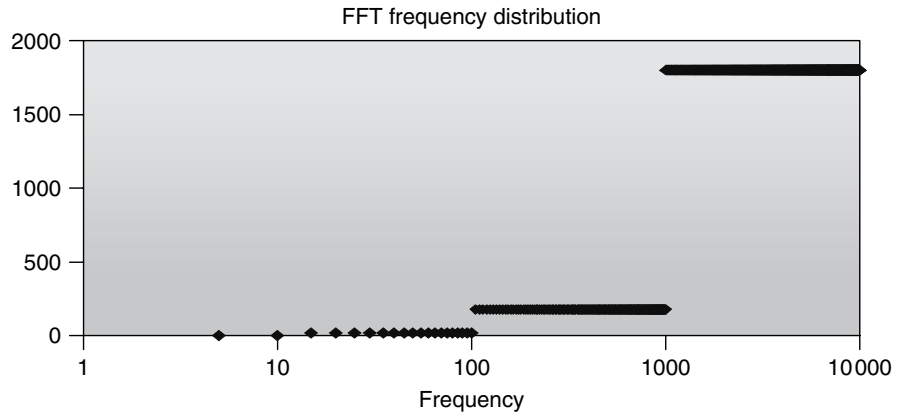


Figure 6. FFT Frequency Distribution.

without signal averaging is unlikely to be effective for the complex signals generated by the main transmission. This processing is more complex than asynchronous spectrum analysis and a system has to be specifically designed to accommodate it. It cannot be implemented postacquisition without utilizing very sophisticated signal processing techniques. Signal averaging represents a cost driver in HUM system design, but it is an essential feature of the system if gear box analysis is to be attempted.

The highest frequencies of the spectrum contain the vibration that relates to fatigue cracking and to pitting or seizing of bearings. To analyze this component of the signal, the lower frequency information has to be removed. This is normally achieved by band-pass filtering as shown in Figure 7. The filter removes both the lower frequency vibration and the noise floor. The energy of the resulting signal is measured or it is subjected to statistical techniques such as Kutorisis estimation to identify the impulsive characteristics of the signal that are related to faults. These techniques are sufficient where a single component's vibration dominates the signal. There is a number of proprietary techniques [6] utilizing pattern recognition (*see* **Statistical Pattern Recognition**) that are necessary in the more complex areas of the transmission.

It is essential to clearly establish the functionality of the vibration processing in order to establish its performance, and too many vendors try to hide behind inadequate descriptors such as "advanced" or "state of the art". The standard performance criteria for vibration processing of bandwidth, frequency, and amplitude resolution are meaningful for the basic spectral measurements of low-frequency features, but are not meaningful for the more complex analysis techniques frequently used in HUM system. The criteria for acceptable performance of the vibration processing are that the resulting spectra and other indicators clearly demonstrate that each of the shafts can be discerned from the background signal. For gearboxes, where signal averaging is implemented, the shaft frequency and mesh tone (dual mesh tone for epicyclics) should be clearly discerned. For systems claiming bearing fatigue and similar defects, the signal in the unimodal range should be free of tones and well above the noise floor. It is also important to demonstrate insensitivity of the data to the normal small variations in shaft speed, weight, and center of gravity that can be experienced during and between flights.

It should then be possible for any HUM system to establish a matrix of which components are monitored and for what type of fault, clearly delineating the signal processing techniques employed to accomplish the task as a basis for establishing performance criteria for the processing.

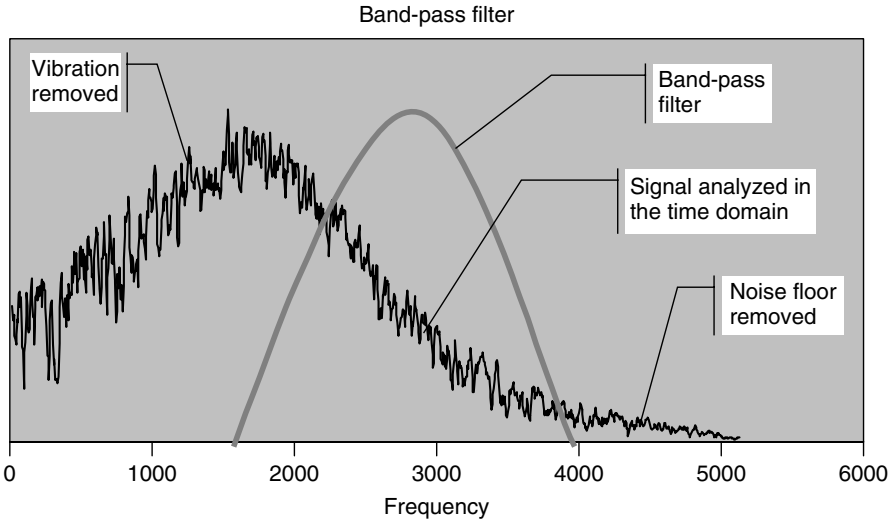


Figure 7. Band Pass Filter.

8 DATABASE MANAGEMENT

HUM systems have to store large amounts of data per flight, which over time build up into huge amounts of data. Very little of the data is ever reviewed. Avoiding data corruption and maintaining reasonable access times is not a straightforward design issue. The problems are compounded if the data is to be accessed by a number of users over a network.

The first process in managing the data is the postflight downloading of the data from the aircraft and in some cases preflight uploading of the configuration data. Clearly, performing these tasks should not interfere with the normal operation of the helicopter. Specifying the routine download times and pilot/engineer interaction with the system is critical. Nothing is more fatal to the acceptance of HUM system in the field than the system being perceived as obstructing the normal efficiency of flight operations.

At the ground station, there is a similar issue of the acceptable latency in transferring the data from the download device and the subsequent updating of the database. If care is not taken in the design, then after a few hundred flight hours of operation these latencies can become excessive. It is essential that the download times and latency in updating the database have limits specified. Data from a number of aircraft being downloaded simultaneously and nonsequentially can cause chaos. Difficulty in extracting data from aircraft is a real problem in HUM system ground station operation. Flexibility of download modes is as important as latency if the system is to gain acceptance at the flight line. Latency in retrieving data, in particular data for postflight pilot review, also needs to be specified. It is important that these latencies are specified for mature databases containing hundreds, not tens, of flights for a representative number of aircraft being supported by the ground station.

Failure to put practical hard limits on these issues has resulted in HUM system being designated as unusable. Software engineers tend to concentrate on functionality and make light of these performance issues. They must have performance standards to achieve if setting to work and operating a HUM system is to be a tolerable experience for the operator.

Database integrity is a difficult feature to specify, but if the data is to be trusted then integrity is essential. Error detection and correction features need to be included in the specifications, particularly for databases that are moved with the aircraft. Data loss needs to be very rare if confidence is to be achieved in the data and if its eventual application is to provide usage or maintenance credits (*see* **Experience with Health and Usage Monitoring Systems in Helicopters**).

9 ALARM GENERATION AND MANAGEMENT

Most HUM systems use a threshold method to generate alarms. In setting a threshold level, there is always the pressure to set the thresholds conservatively to reduce false alarms resulting in the threshold being set too high to be sensitive to a developing fault as shown in Figure 8. As the occurrence of faults is very rare, there is a lack of a counterpressure to balance the pressure from engineers to minimize the false alarm rate.

To guard against this, a target acceptable false alarm rate could be set across the fleet. If the observed rate is less than this target, then the thresholds should be reviewed to ensure that fault sensitivity is not being sacrificed. If it is more than the target, then the suitability as a HUM system indicator of the parameter needs to be reviewed and, if possible, an alternate more stable parameter used.

Vibration signals and noise on other signal types always produce “wild” points in the data. No system can maintain sensible threshold levels without an appropriate form of wild-point rejection in the processing. All of the wild points in Figure 8 exceeding the lower more sensitive threshold level would be removed by this processing and detection of the trend would remain unaffected.

The use of fleet statistics to set alarm levels can also have an impact on performance. In general, the higher the frequency of analysis is, the less likely that a group threshold is going to be valid. In the group of helicopters if the differences in the statistical means and variances of the data monitored are small, then the group threshold level is likely to perform well for all. If the means are well separated or one data trend has a dominant variance as shown in Figure 9, then a threshold set by the group statistics is only going to be valid for either the data with the highest mean or largest variance.

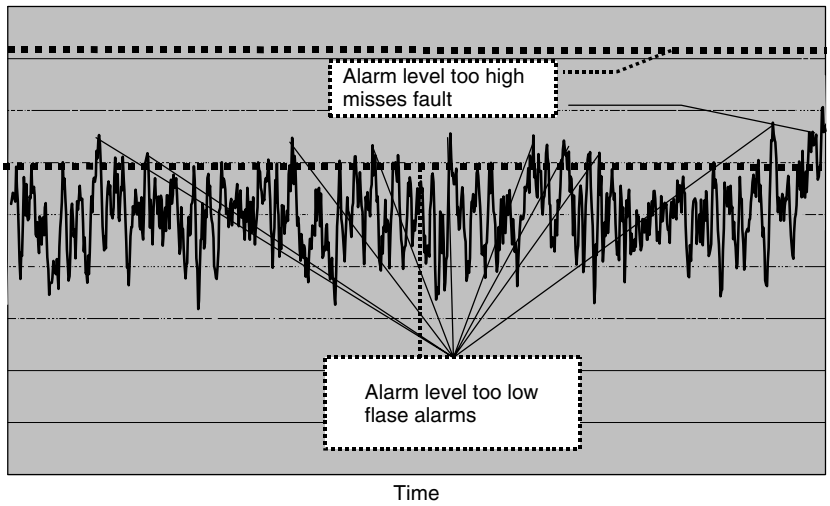


Figure 8. Classical Threshold Setting.

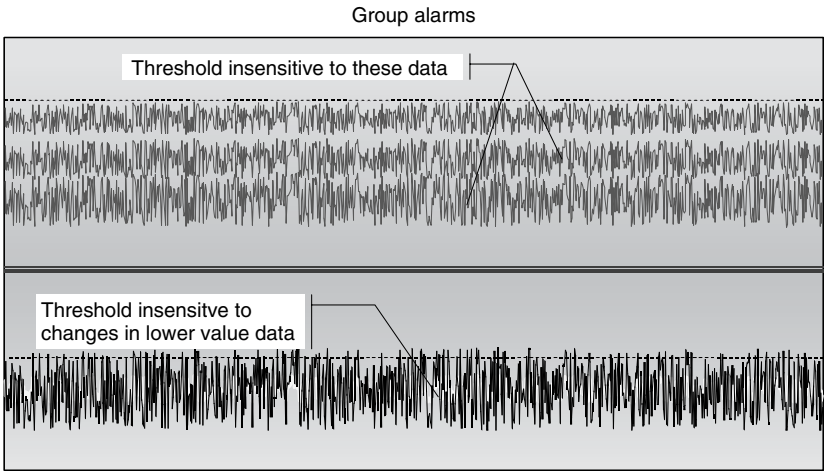


Figure 9. Group Threshold Setting.

Facilities to corroborate data are also essential to assist in determining whether an alarm is a result of a faulty sensor or just noise. It is rare in vibration that a shaft vibrating abnormally will be discerned on just one sensor.

Sensitivity to faults can only be determined if the HUM system has actually identified any. Establishing the range that is known to be sensitive to detecting a fault is important for assurance that the threshold is set appropriately. These data are only likely to be available for mature systems with sufficient flight hours behind them to have developed a track record of success.

Alarm management in HUM system as a proactive task goes beyond the process of checking and following through on any actions caused by an alarm being transgressed. To keep the system adequately sensitive, the alarm levels need to be subject to management scrutiny. Changes in weather, pilot, weight, and center of gravity of the helicopter all

change the characteristics of the HUM system data and can compromise some of the thresholds levels. Monitoring the performance of the alarms to determine the validity of all of the thresholds being applied is critical to maximizing the overall performance of the system.

The key performance measure for managing the threshold setting is the false alarm rate. In case of too few alarms, there is a problem with fault detection sensitivity, and in case of too many alarms, there is a problem with suitability of the data. Also, the probability of success is only maintained if the arisings that are identified by other means are investigated to determine if the fault was present in the HUM system data and why the alarm threshold was not triggered.

10 MAINTENANCE MONITORING

Vibration levels on shafts and performance monitoring of the engines are common maintenance functions where a HUM system replaces the existing maintenance test equipment. It needs to be ascertained that the HUM system really does replace the test equipment. For vibration, some test equipment utilizes older analog vibration measurements such as RMS or swept band filtering. It is not possible, without a great deal of effort, to reproduce these data with the digital signal processing techniques found in HUM system. Establishing the correspondence between handbook vibration limits and HUM system data is a key performance attribute for a system and this correspondence can usually only be established by the design authority.

The ability of HUM system to routinely monitor the rotor and produce recommendations for adjustment that keep the rotor in smoothest possible condition is an important function. Rotor track and balance systems usually contain proprietary software for determining the rotor adjustments. Specifying performance is not possible without the disclosure of the processing. Accuracy of sensors is not a guide to the overall performance of a rotor smoothing system. For rotor smoothing, the only effective check is to maladjust a smooth rotor well within its safety limits, and to check that the track and balance system predicts the correct adjustment to make the rotor return to the smooth condition. A correctly performing system should not confuse a weight adjustment with a track rod or a tab adjustment.

The key performance features here are (i) the average vibration level achieved and (ii) the reduction of effort required to tune the rotor. It is not unreasonable to require that the rotor vibration in all axes and flight conditions is maintained to be less than a specified percentage of the handbook limits for 95% of the time. This has to be on the condition that the adjustment recommendations produced by the HUM system are implemented. Also, this measure should exclude the adjustments following blade removal where a specification should be to bring the system to below the target vibration level in less than two test flights.

11 EQUIPMENT STANDARDS

HUM systems are peculiar in three respects when compared to other avionics: (i) the system's complexity, (ii) the split of functionality between airborne equipment and ground equipment, and (iii) the interaction of the system with the helicopter's design and operation.

If the equipment is specified as part of the helicopter's minimum equipment list (MEL), then many, but not all, of the large number of sensors that make up a HUM system greatly reduces its overall reliability. A long mean time between failure (MTBF) is essential for individual equipments in order that the overall system MTBF is reasonable, or the MTTRs are very short and adequate spares are provisioned so that the MEL requirements can be met.

The split of monitoring functions between the airborne and ground components of the system can cause difficulty in certifying the system even if the certification standard is no hazard/no benefit. The FAA's HUM system advisory circular (AC) [7] requires different software standards to be met for the ground station than those for the airborne because of the use of commercial off the shelf (COTS) software in most ground station applications. For a HUM system, other than no hazard/no benefit, the ground station software has to be of a higher software standard than that of the airborne component, or independence of the HUM system processing to the COTS software has to be demonstrated. Some further hurdles have to be cleared for credits to be possible,

such as the independent verification of data and the mitigation of the credit to ensure that airworthiness is enhanced. This level of functionality is beyond the scope of current HUM system design.

Where maintenance credits are to be sought, it usually requires that the design authority approves the changes to procedures. Only the design authority has the necessary knowledge to determine if an anomaly detected by a HUM system is damaging to the extent of risking the airworthiness of the helicopter. This requires that the helicopter manufacturer acknowledges and supports the data analysis produced by HUM system. Design authority approval, or at the very least the HUM system having no objection status, is an essential attribute of any HUM system.

12 CONCLUSIONS

Divining performance measures that relate directly to the purpose of HUM system requires the assessment of the contribution of the technology to helicopter airworthiness. The nearest approximation to this is achieved by building up a statistical knowledge of more frequent events such as maintenance arisings and other events reported to the Regulatory Authorities. The monitoring performed by the CAA shows that, generally, a 70% success rate of defect detection is achievable, and would provide an overall performance standard for HUM system.

Measuring the performance of the HUM system in service is most easily achieved by monitoring the false alarm rate. This should be randomly distributed amongst the components of the dynamic train and within a range that is acceptable to the operator. It is not the case for HUM system that no news is good news.

For a new HUM system type or a system that is to be fitted to a new aircraft design, the actuarial data for determining the performance of a HUM system does not exist. The performance of the individual building blocks of the system has to be specified such that confidence in the overall system performance will be adequate. For flight data, then, the normal system engineering practices of modeling the process so that the required performance of each component of the system can be determined, will suffice. But the major component of a HUM system is the vibration processing and it cannot be dealt with in such a straightforward manner.

Measuring the performance of a modern vibration analysis system based on digital signal processing is essentially a process of the vendor demonstrating that the features used for monitoring purposes can be clearly discerned in the measured spectra of the normal operation of the helicopter, or at the very least the data contains no phenomena that prevent the monitoring function from being effective. The features that count, if the performance of the system is to be unequivocally determined, are (i) a clear statement of the components that are to be monitored from the sensors, (ii) the techniques to be used for the monitoring, (iii) the acquisition rate, and (iv) the positioning of the sensors (viz., accelerometers).

The suitability of sensors and processing in terms of resolution, range, and bandwidth needs to match the monitoring techniques used. A clear statement of these attributes and their specific purpose in terms of the fault modes monitored in the helicopter's dynamic train components should be produced with evidence that the processing can discern the relevant attributes. Without this, any attempt to substantiate the efficacy of the processing cannot be made in any quantitative way.

Measurement and processing are not sufficient in themselves to determine the likely performance of a HUM system. The alarm processing and management functions of the system need to be flexible enough that the thresholds can be tuned to achieve the desired false alarm rate and maximize the system's sensitivity to faults. The features that count in the alarm system are the availability of management functions for setting and assessing the threshold levels so that the sensitivity to faults is optimized with respect to the acceptable false alarm rate and the cost of managing the spurious alarms by the engineer. A system for removing singleton alarms is essential if viable threshold levels are going to be set.

A HUM system is not a "fit and forget" system; management of the alarm threshold system is essential if the expected improvement in the safety margin is to be achieved. For convenience, it is all too easy for alarms levels to be too conservatively set, making the system insensitive to faults. The other performance check to determine that the thresholds are sensitive to faults is to investigate arisings that are not indicated by the HUM system and to identify whether the threshold is set at a suitable level or on the most appropriate data.

The system's operation has to be acceptable to both pilots and line engineers, and their management. Latency and integrity of the operation of ground station and its databases are the key features here. These need to be unambiguously specified and performance should be measured on mature databases.

Finally, care needs to be taken that the system can support the MEL requirements and that the software and HUM system AC requirements are met if maintenance credits are to be sought.

In short, do not take anything for granted, and ensure that the system is adequately and clearly specified, that acceptance testing includes performance measures that relate directly to the purpose of the HUM system, and that the operation of HUM system does not interfere with operational efficiency. Besides, for a specification that stipulates meaningful performance measures, the contract has to be sufficiently robust to ensure that these specifications are met and maintained.

ACKNOWLEDGMENTS

Figure 2 is reproduced with the permission of the UK Civil Aviation Authority.

REFERENCES

[1] United Kingdom Helicopter Health Monitoring Advisory Group, *A Guide to Health Monitoring in Helicopters*, UK CAA, 1990.

[2] CAA, *AAD 001-05-99*, London, 1999.

[3] Chapman DJ. The way forward on helicopter safety. *Presentation to the European Helicopter Operators Committee*, Estoril, May, 1996.

[4] Evens A. *Report Minuted in 35th HHMAG*, Gatwick, June, 2002.

[5] CAA, *CAP 693—Acceptable Means of Compliance Helicopter Health Monitoring CAA AAD 001-05-99*. London, 1 May, 1999.

[6] Larder B. Helicopter HUM/FDR: Benefits and Developments. *American Helicopter Society 55th Annual Convention*, Quebec, May, 1999.

[7] FAA, *Airworthiness Approval of Rotorcraft Health Usage Monitoring Systems (HUM system). FAA Advisory Circular Material to be incorporated in AC-27-1 and AC-29-2*, July 15, 1999.

Chapter 101
Unmanned Aerial Vehicles

Matthias Buderath
Product Support, EADS Military Air Systems, Ottobrunn, Germany

1 INTRODUCTION

With the newly emerging interest in unmanned aerial vehicles (UAVs), such as those used for surveillance missions, like Global Hawk or Predator, the new trend in military aeronautics is to clarify whether the UAVs' technology can be used not only for surveillance but also for combat missions.

The main difference between a cruise missile and an UAV is that the UAV will be recovered after its mission, while the cruise missile will not. This distinction is clearly made in the definition of UAVs in the Joint Publication l-02 of the Department of Defence (DoD) dictionary [1]:

> *A powered aerial vehicle that does not carry a human operator uses aerodynamic forces to provide vehicle lift, can fly autonomously or be piloted remotely, can be expendable or recoverable and can carry a lethal or nonlethal payload. Ballistic or semiballistic vehicles, cruise missiles and artillery projectiles are not considered to be UAVs.*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

UCAVs (unmanned combat air vehicles) are supposed to be the next generation of military aircraft. The DoD of the United States and many other countries claim that the combat aircraft of the future will be unmanned and that the last manned combat aircraft will be the Joint Strike Fighter. The main characteristics of a UCAV are as follows:

- It is capable of autonomous flight but can be remotely piloted, if necessary.
- It can be recovered after completion of mission or can even be expended with.
- Its operation is not dependent on or limited to human restrictions such as G force, mission and relaxation time, and combat readiness.
- It is predestined for hazardous missions and enemy region penetration.

The main characteristic of this weapon system is that the pilot is located outside the air vehicle, probably in a ground control station (GCS).

The idea is that the UCAV can fly autonomously but with a man in the loop, who can take over at any time. It is important that the communications cannot be jammed, so a high encryption will be needed to

ensure that the weapon system cannot be controlled by any unauthorized person.

UAVs will fully utilize the emerging information revolution. They are expected to take advantage of multiple, real-time data sources and secure communication networks to plan for, and respond to, the dynamically changing battlefield.

But why this interest in UAVs? UAVs are supposed to fulfill the most hazardous missions, where manned aircraft pilots have to risk their lives to accomplish their mission. One of these hazardous missions could be the suppression of enemy air defenses (SEAD).

So, not having a pilot onboard would permit the air vehicle to take on greater risks than a manned aircraft would.

A further motivating factor is that UAVs are supposed to be less expensive in their life-cycle cost than the average manned aircraft. UAVs are considered as items that can be stored and only brought out for service during times of conflict or that can be operated for 24 h and more. A limited number of UAVs are operated because of expected high mission capability, no need for training in peacetime and new maintenance concepts, e.g., maintenance-free operation periods. UAVs are expected to achieve savings in operation and support costs of 30% compared to traditionally 60% of the entire life-cycle cost.

2 OPERATIONAL REQUIREMENTS

Three key elements have been identified, which justify the development, integration, and use of a health management system for UAVs; these are

- improved decision support for autonomous mission execution;
- reduction of mission, operation, and support cost; and
- contribution to the certification process of the UAVs.

As already mentioned, to derive the functional and design requirements of a health management system that includes structural health management (SHM), it is essential to understand the mission and operational support requirements. Therefore, this article discusses, to some extent, the operational requirements that can serve as a guideline for the potential SHM technology provider.

Figure 1 describes the dependencies and trade-offs between the concepts of operation and operational support. The better one optimizes and consolidates the needs and performance of both concepts, the more easily one can achieve the following goals:

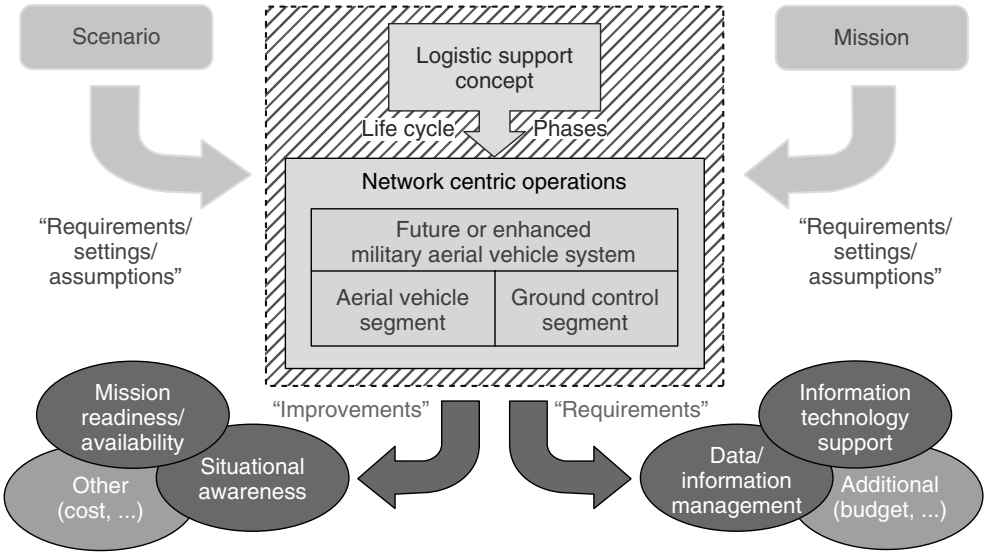


Figure 1. Operational support scenario.

- reliable autonomous mission execution;
- improved aircraft operability, which also means a control of mission and operational support cost;
- mission effectiveness/capability (the hours a fleet is available to perform missions versus the hours a fleet is possessed by a flying unit); and
- compliance with certification requirements.

An overview of the mission and thus operational effectiveness is provided in Figure 2.

2.1 Autonomous mission execution

There are different operational and thus autonomous mission approaches, but one common challenge to be met is the proper understanding of operational and operation support requirements of UAVs to ensure safe and reliable autonomous mission executions. Besides the top level requirement (TLR) of a high probability of detection through inspection, the integration of an SHM system into the avionics suite is mandatory. Integration into the avionics suite makes sense, since most of the SHM system is anyhow electronics-based and avionics *per se* is already well associated with built-in test equipment (BITE). Furthermore, health management for an UAV is a crucial requirement, and, as such, SHM will thus become a part of it. The expression of "health management" implicitly includes SHM throughout this article.

It is recognized that a vital goal for autonomous missions is to enable intelligent decision support. An operational risk assessment (ORA) in close relation to the mission plan has to be part of the decision support algorithm. A very important aspect of OR is determined by the health status or anticipated health status of the aircraft.

Two possible levels of integrating SHM with mission management are considered. The levels of integration are determined by the levels of autonomy that are intended to be achieved by the UAVs.

In the technical literature, six levels of autonomy are defined, ranging from level 0 (remotely piloting the vehicle) to level 5 (aircraft flies by itself; an operator may, however, interrupt what it is doing). In this article, levels 3 and 4 are considered. Level 3 is when the UAVs indicates its status and an operator decides on the basis of this information to command the vehicle to take a certain action. Level 4 is when the vehicle itself decides to take action based on its status. The operator may, however, command it to do otherwise.

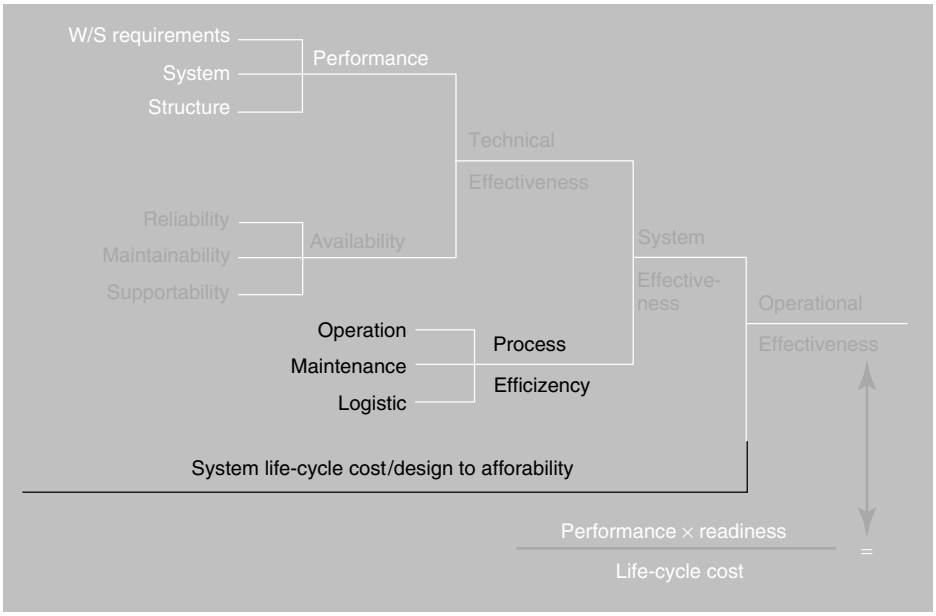


Figure 2. Definition of operational effectiveness.

2.2 UAVs with autonomy level 3: accept advice authorize action

In this case, the health management system provides the status/anticipated status to the operator and hence this must be available at the GCS. The operator then commands the aircraft to take a particular action. In keeping with goals, the information is provided as an ORA. The emphasis is on providing the prognosis of health, with risk level and consequences as a basis for the operator's decision. This drives the architecture as follows: no interaction between the health management system and the onboard mission management system (MMS) is necessary. The health management could have minimal functionality onboard and the prognostic functions could reside on the GCS. A possible approach could be to develop and integrate the functionality on the ground and then later migrate the functionality to onboard the aircraft in anticipation of achieving the next level of autonomy, described below.

2.3 UAVs with autonomy level 4: action unless revoked

In this case, the health management system provides the status/anticipated status to the onboard MMS. The MMS makes the decision of what action has to be taken, but it must first provide information to the operator, in order to give him/her a chance to revoke the action. The aircraft initiates its own action based on preloaded rules or decision algorithms. This drives the architecture as follows: a substantial interface is required between the health management system and the MMS: functions need to reside onboard. If anything, the interface to the GCS is minimal, as the operator is monitoring and only intervenes when it is felt necessary.

It is assumed that software will need to be developed, but that it can be integrated into the existing/newly provided hardware.

2.4 Aircraft operability, including control of mission and operational support costs

Since UAVs are operated in a network of systems in which mission effectiveness is dependent on mission capability of all platforms involved, the term *availability* or *operability* adds another dimension to mission cost.

Some governments, like that of the United States, have introduced in their acquisition regulations, directives such as the DOD 5000.2-R, "Mandatory Procedures for Major Defense Acquisition" [2], in which one of the major themes addressed is that "The acquisition process must consider both performance requirements and fiscal constraints. Accordingly, cost must also be an independent variable in programmatic decisions, with responsible cost objectives set for each Program Phase".

In this context, the translation of the top-level mission operational support requirements means we have to provide excellent health monitoring and management performance without increasing acquisition and support costs. In order to meet these, performance requirements, a health monitoring system should provide the capabilities and features listed below:

- a high probability of damage detection while determining the condition of the monitored component;
- integration of sensors into the structure that will provide more efficient information than is available today;
- consideration of low cost but reliable sensors that can be integrated into the structures and be more than sufficient;
- provision of power supply to the acquisition system without compromising weight and structural-system performance;
- processing of the high quantity of information generated at component/system level;
- health management within the health management architecture hosted in the onboard and on-ground data management platform;
- prognosis (link usage/load monitoring to damage monitoring) within the health management architecture hosted in the onboard and on-ground data management platform; and
- decision support (including replacement of the pilot detection and assessment capability in case of specific in-flight events) within the health management architecture hosted in the onboard and on-ground data management platform.

2.5 Certification

So far, SHM systems have been considered for monitoring the theoretical fatigue life consumption as a basis for fleet management and for scheduling of inspections and modifications. Owing to the used safety factors and scatter factors, the implication for the day-to-day airworthiness was a secondary factor.

With the development of high-performance aircraft, the issue of airworthiness has become more significant in postflight assessment with regard to an intact structural strength capability.

Nevertheless, the identification, assessment, and decision making in case of an in-flight event like a bird strike, lightning strike, battle damage, or damage during taxiing or landing caused by foreign object damage (FOD) always rests with the pilot. To allow for cases in which the pilot is not in a position to make a decision, additional advanced in-flight capabilities for data acquisition, state detection, health assessment, and decision support need to be provided.

Future requirements therefore include a reduction of UAV mishap rates and an integration of UAVs into national/international aerospace outside restricted areas. This will also raise issues of establishing rules and guidelines for UAV certification. These two items emphasize the necessity for an onboard real-time loads monitoring and damage detection system with a subsequent diagnosis and health assessment functionality to check the structural strength capability and structural performance after an in-flight event.

This will lead to increased demands for the use and integration of health management systems for which sufficient evidence and qualification must be provided to demonstrate their maturity level and the readiness for operation.

Figure 3 shows the top-level process and related requirements on an in-flight SHM event-monitoring function.

3 CONCEPTUAL APPROACH

Two different types of monitoring systems are currently under evaluation with conventional aircraft [3, 4] (*see* **Principles of Structural Degradation Monitoring**; **Usage Management of Military Aircraft Structures**; **Development of an Active Smart Patch for Aircraft Repair**.):

1. load/usage monitoring and
2. damage monitoring.

The aim of the load/usage monitoring system is the measurement of the loads (deflections, temperature, strains, and subsequent loads) during operation (*see* **Loads Monitoring in Aerospace Structures**; **History of SHM for Commercial Transport Aircraft**; **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**). Such monitoring systems serve as a database for real loads acting on the component to be monitored during operation and reduce the uncertainties of the operation scenario models. Still, accurate models for the prediction of the onset

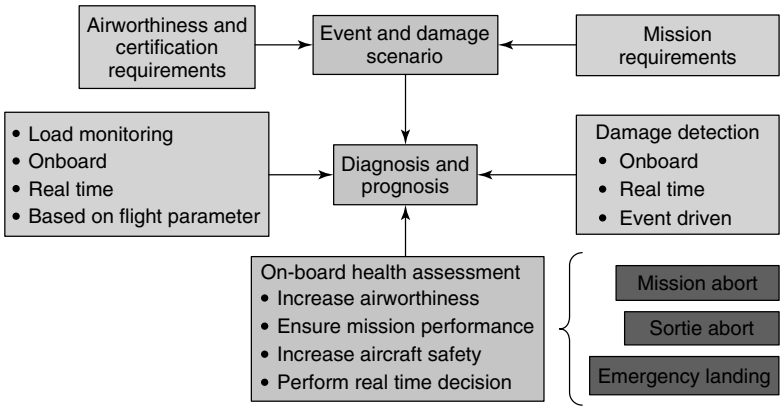


Figure 3. Design requirements and process.

and growth of damage and the remaining life of components are required (*see* **Modeling for Detection of Degraded Zones in Metallic and Composite Structures**). Especially in cases of composite structures, these models have a lot of uncertainties due to a large number of influencing factors, like production and handling procedures, calling for a number of safety measures. The aim of the damage-monitoring system is to determine the type of damage (e.g., fatigue crack growth in metals or impact delaminations in composites, location, and size) during operation (*see* **Military Aircraft**; **Use of Leave-in-place Sensors and SHM Methods to Improve Assessments of Aging Structures**; **History of SHM for Commercial Transport Aircraft**; **Flight Demonstration of a SHM System on a USAF Fighter Airplane**; **Comparative Vacuum Monitoring (CVM™)**; **Development of an Active Smart Patch for Aircraft Repair**; **Aerospace Applications of SMART Layer Technology**).

By combining both systems, together with a reliable materials model, the real load history and the actual damage distribution as input data, should, subsequent to the optimization of the decision support algorithm, result in more accurate predictions of the remaining life of the component. Therefore, the technology of SHM systems forms the basis for the realization of autonomous missions and contributes to implementing new maintenance strategies for UAVs, e.g., in the change from "time-based" to "condition-based" maintenance. Also, SHM systems for sensing and characterization of the structural condition of specific components have to meet the certification requirements (aircraft accident prevention). For example, NASA performed a study in 2002 [5] in which the requirements for such SHM systems, the characterization of a prototype structural sensor system, the development of sensor interpretation algorithms, and the demonstration of the sensor systems on operationally realistic test articles were addressed, in all of which the sensor data served as inputs to ARINC's Aircraft Condition Analysis and Management System (ACAMS) [6].

Usage monitoring systems serve as a database for real loads acting on the component to be monitored during operation and reduce the uncertainties of the operation scenario models. To guarantee availability and operability, the operator needs condition awareness to make decisions throughout the whole life-cycle process. Related to the monitoring of the structural strength capability of the aircraft, the load monitoring will provide the relevant details, for example, by comparison of the expected load against the measured loads, as a basis for the real-time assessment of the aircraft condition. In addition to the fatigue life monitoring, accurate models for the prediction of the onset and growth of damage and the remaining life of the components are required. Especially in cases of composite structures, these models have a lot of uncertainties due to the large number of influencing factors, such as those resulting from production and handling procedures, which require that high safety measures be imposed. An SHM system for UAVs should therefore cover a process, which is shown in Figure 4.

3.1 Operational requirements/operational parameters

The basic process to fix a problem is to step from an unknown condition of a component to a known condition, which may then require a dedicated maintenance action. To do this, operational parameters need to be evaluated as part of the condition state assessment and the health assessment. Today, the most common operational parameters are time, i.e., the number of flights, and results from visual inspections, such as cracks, corrosion, wear, and dust. To improve the diagnostic, prognostic, and decision support capabilities, more information is likely needed on what is the load impact that occurs and when (i.e., the real load sequence), the damage characteristics and location, etc. This requires a careful selection of sensing technologies that support the above-listed features (*see* **Principles of Structural Degradation Monitoring**; **Commercial Fixed-wing Aircraft**). Figure 5 gives an overview of operational parameters that are taken into account for monitoring followed by a subsequent assessment.

Since design principles in engineering are well established, and monitoring is just a consequence, the central question with regard to monitoring is as follows: what are the parameters that are lacking, which we need to assume, in regard to a cost-effective design with improved operability?

All structural design is based on loads (static, as well as cyclic) that we have to assume prior to configuring the structure. These loads do not

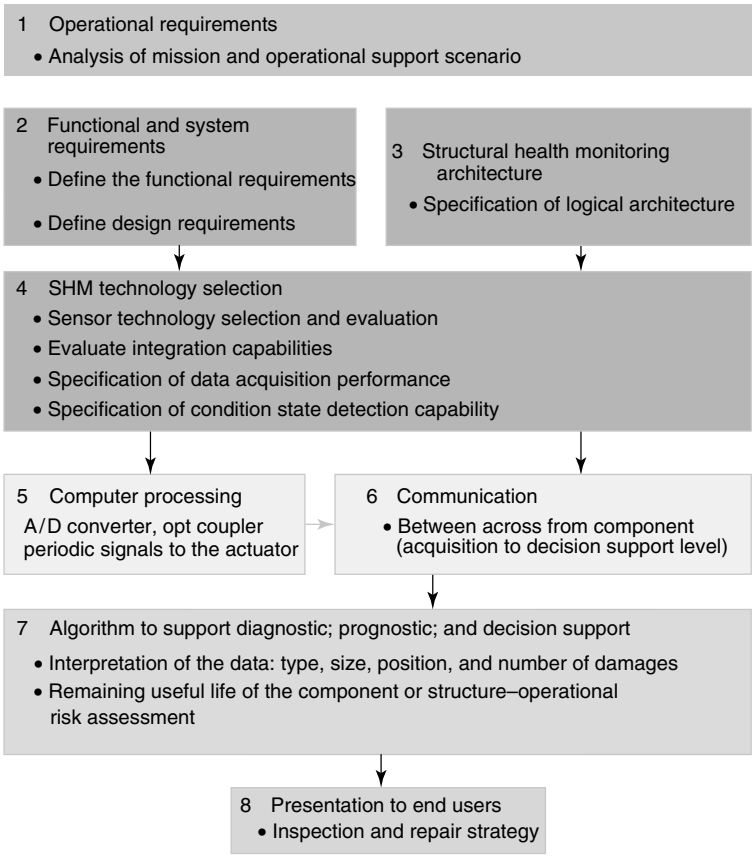


Figure 4. Elements of an SHM development and integration process.

have to be limited to mechanical loads only. They can also include other environmental loads, such as temperature, humidity, chemical corrosives, etc (*see also* **Environmental Monitoring of Aircraft** and **Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control**). To improve the diagnostic, prognostic, and decision support capabilities, we require more information on when each load occurs (i.e., the real load sequence), the damage characteristics, location, etc. This would allow us to manage operational risk and availability of resources more efficiently, without compromising safety.

The other factor that needs to be monitored and which is a consequence of aircraft design and usage is damage. The assessment of whether a structural condition is acceptable or not for aircraft operation is largely covered by design limits. Design limitations are different in nature, especially for different types of materials and structural concepts.

Whatever these limitations are and no matter how they are characterized, it is basically assumed that, for safe aircraft operation, the condition is allowed to change within certain allowable bandwidths, given by the design.

If the actual structural condition remains within the given design bandwidth, it can be assumed safe: "assumed actual condition safe". If the actual structural condition exceeds the given design bandwidth, it has to be assumed unsafe: "assumed actual condition unsafe".

The definition of "safe" is fundamental to the design process. Typically, it follows that the probability of extreme failure (and consequential loss of lives) must be very small. The overall aircraft system (including operational processes) must be designed

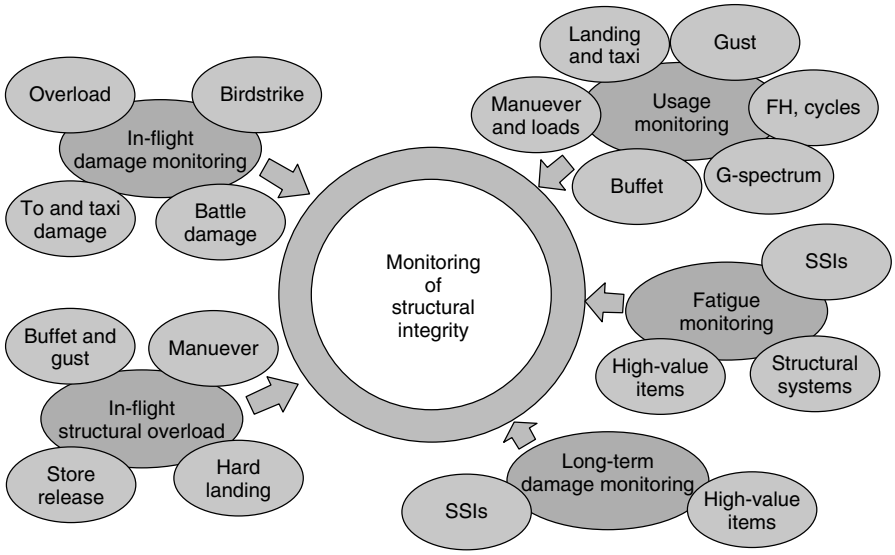


Figure 5. Operational parameters.

so as to ensure safety at all times. Whatever the operational optimization goal of the operator is and whatever infrastructure the operator has available in terms of maintenance alternatives, design limits have to be obeyed at all times.

Figure 6 schematically shows the complete bandwidth of degrading structural condition along which safety is not compromised. The left part includes the region where the actual condition cannot be measured due to limitations in knowledge and monitoring techniques while the right hand part includes the region where the condition is measurable and tolerable from the point of view of structural integrity and which is worth exploiting.

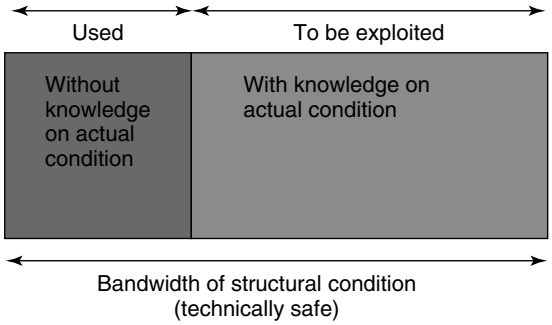


Figure 6. Illustration of achievable bandwidth of structural condition without compromising safety.

To exploit the available bandwidth, it is necessary to know with sufficient accuracy that the structural condition remains within the allowed bandwidth under operation at all times. This is the reason why structural condition monitoring and thus SHM for modern aircraft is of interest.

Structural condition monitoring is the continuous process of determining the condition of a structure. The primary objective of structural condition monitoring is to ascertain that the structural condition remains within the allowed bandwidth under operation. Of course, this requires greater accuracy the closer that it approaches the design limits.

A potential improvement on condition monitoring is the process of "condition-based planning" (see Figure 7). Therefore, the benefits of condition monitoring can be summarized as follows:

- **Full health exploitation**

Statistically, the individual damage-free period of a part is significantly (related to a "scatter factor") higher than is assumed during design and subsequently cleared for operation. Condition monitoring in combination with a detectable damage being below the critical damage in damage tolerance terms would be sufficient to exploit the available structural health without compromising safety or the need for unacceptable short-scheduled maintenance.

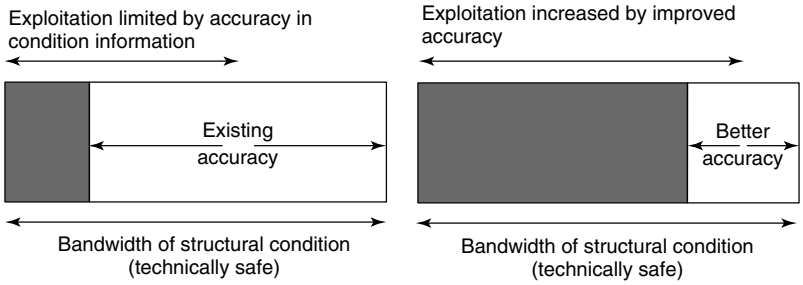


Figure 7. Illustration of limitation of exploitable bandwidth of structural condition due to available accuracy.

- **Improved risk reduction**

Improved condition monitoring continuously gives information about the actual condition of a structural item, because the frequency of monitoring can be significantly increased without increasing the inspection effort.

- **Condition-based planning**

The more accurate the structural condition monitoring is, the more reliably the operator can plan the necessary maintenance actions to avoid unscheduled actions or unnecessary originally scheduled actions.

- **Improved operational margin**

Improved condition monitoring increases the available time for planning maintenance alternatives and selecting the optimum. It helps to reduce unscheduled maintenance.

3.2 Functional and system requirements

The functional requirements needs to be derived from the functional view of the required health monitoring and management capabilities. Here, the diagnostic, prognostic, and decision support capabilities should be defined for individual component and aircraft levels. Figure 8 shows a typical overall functional view in which each function is specified in detail, e.g., to derive the input and output data for the models and the algorithm to determine the CPU requirements, etc. The logic provided here is strongly based on the Open Systems Architecture for Condition-Based Maintenance (OSA-CBM), which is explained in more detail in **Open Systems Architecture for Condition-based Maintenance**. The approach shown in Figure 8 is a structured monitoring data management approach that consists of the following seven layers:

- presentation (Pres)
- display selection (DS)
- prognostics (Prog)
- health assessment (HA)
- state detection (SD)
- data manipulation (DM) and
- data acquisition (DA).

In the functional view, the end-user requirements, like types of damage, critical damage size, and level of importance of the structural components to be monitored, include the layout, interaction, and redundancy depending on the level of reliability of the components of the SHM system. Depending on the application, different types of methods for optimized sensor placement, like "improved genetic algorithms", finite element model (FEM) methods or spectral finite element method have been used (*see also* **Optimization Techniques for Damage Detection**; **Modeling for Detection of Degraded Zones in Metallic and Composite Structures**; **Sensor Placement Optimization**). This includes the decision between on-line (real-time) or off-line (discontinuous) operation and the density of sensors versus damage resolution.

3.3 SHM architecture

The modern avionics system combined with remote interface unit (RIU) technologies have tackled the major challenges like

- **Weight**

The RIU, Aircraft Mission Computer (AMC), and bus architecture have eliminated the excessive wiring that would otherwise be required with multiple individual systems. This allows the SHM software and

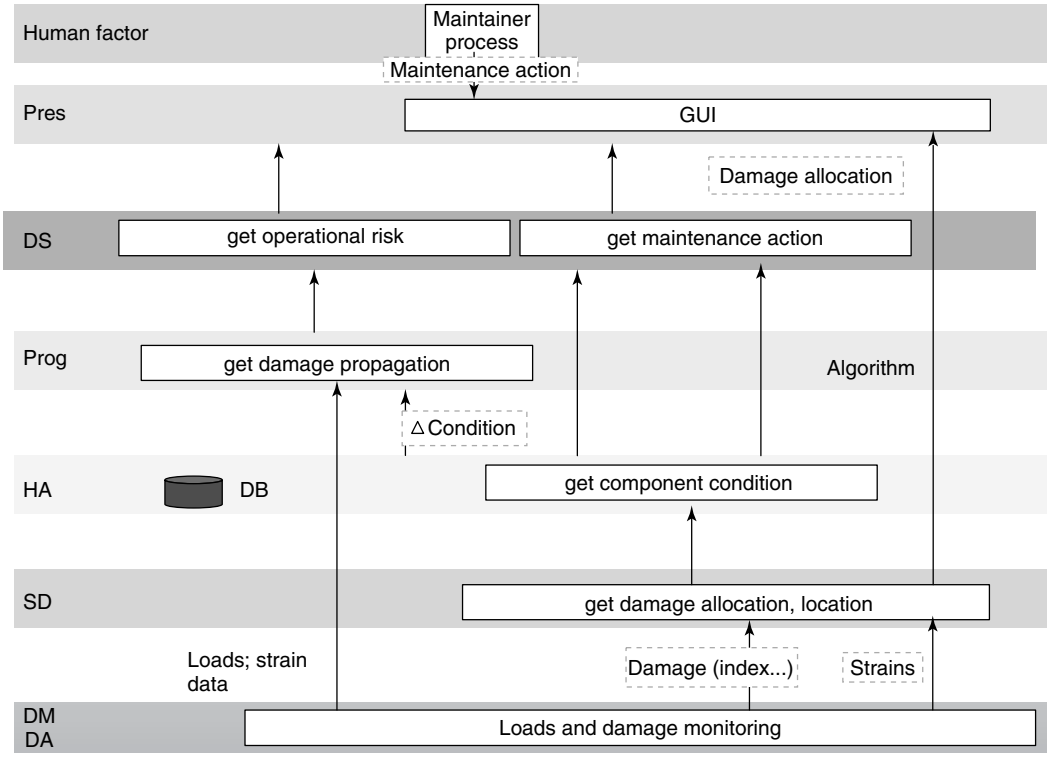


Figure 8. Functional view.

algorithms to be run alongside the other airborne software, avoiding the need for extended SHM hardware, even extending as far as utilizing the aircraft's normal communication methods.

- **Cost**

By utilizing the existing aircraft hardware in terms of both processing and network capability, the cost of installation now becomes one of the sensors, something that modern "virtual sensing" techniques are tackling, and software that can be installed without significant aircraft downtime.

- **Complexity**

By transmitting all the data over common buses and to a common core, the complexity is greatly reduced and this has been tackled even further by developments such as the OSA-CBM initiative (*see* **Open Systems Architecture for Condition-based Maintenance**). The maintenance technician can now expect one coherent interface to the entire maintenance system. Current research is looking at taking this even further to provide a "process-oriented" structure to the data displayed at the point of need.

In addition to this, it has also placed all of the data in one place, which has opened up the possibility to see the bigger picture of how the aircraft is performing; this truly is a case of "the whole is greater than the sum of the parts".

However, to realize a workable system still requires an open software architecture in which airframers, equipment suppliers, and specialist prognostic health management (PHM) companies can provide the tools to turn these data into information, knowledge and, finally, actions. This is where developments such as OSA-CBM are allowing the further realization of these systems.

The goal of OSA-CBM as shown in Figure 9 is "to facilitate the integration of PHM components from a variety of sources. OSA-CBM is striving to build a *de facto* standard that encompasses the entire range of functions from data collection through the recommendation of specific maintenance actions".

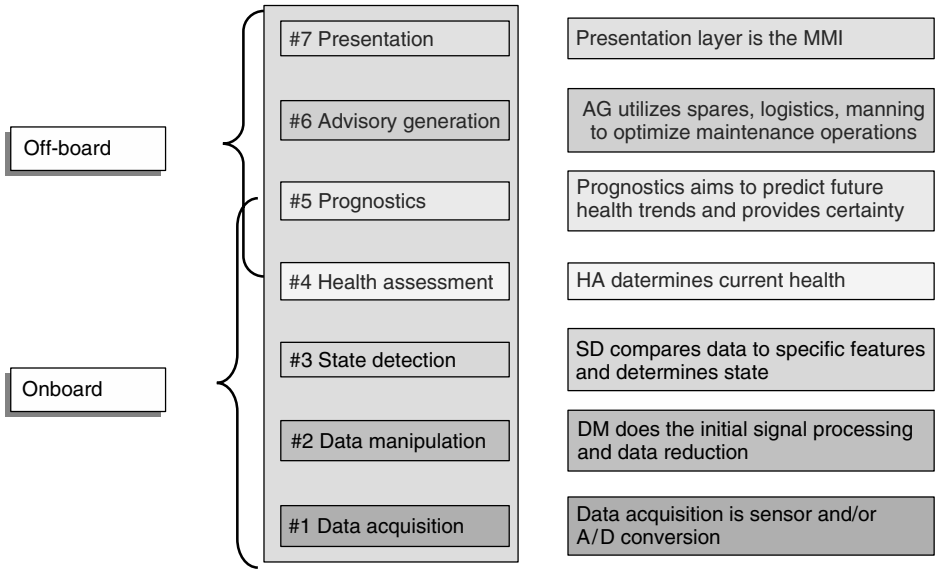


Figure 9. OSA-CBM.

3.4 SHM technology selection

Sensors meeting respective needs will allow monitoring of operational loads at various locations on the aircraft in much more detail, which will further allow calculation of consumed operational life more according to the real usage. This will be supported by loads determined at a higher degree of detailing and finite element models, which allow a more accurate calculation of loads being based on standard flight parameters. Further to this, there are now more and more sensors emerging that allow the monitoring of damage to structures *in situ*. This offers potential to help to meet the operational requirements.

The most important component within the SHM process is the identification of the relevant kind of damage depending on the material and load and its impact on the performance of the component. With metallic structures, designers and operators are mostly concerned with fatigue cracks and corrosion, while for composite materials, delamination and impact damage are more of a concern. In that regard, appropriate sensors have to be selected, for which a variety are described in Part 5 of this encyclopedia.

Selecting the appropriate SHM technology for a specific application can be a rather tedious process. Under the Techniques and Technologies for nEw Maintenance concepts (TATEM) project, the technology selection matrix shown in Figure 10 has been developed and used for the technology selection process under real conditions. This assessment of SHM technology is based on a cascading set of requirements. Starting from simulating future maintenance scenarios, TLRs have been defined, further followed by specific SHM system requirements. All of these requirements have then been summarized. It is the purpose of the procedure to work out and check a framework that allows prioritization and selection of new sensor systems and SHM technologies for air vehicles such as an UAVs, i.e., to determine those systems that allow the information recorded and processed to step from an unknown to a known condition of the item to be monitored. Figure 10 gives an overview on how to assess an SHM technology. The procedure starts by setting up the list of requirements, which can either be the TLRs *per se*, SHM system requirements or a combination of both. These requirements are then correlated with technology criteria that are described in further detail below. Finally these technology criteria are then used for correlation with the respective sensing (SHM) technology considered, resulting in a technology assessment matrix that allows different technologies to be prioritized on a numerical basis.

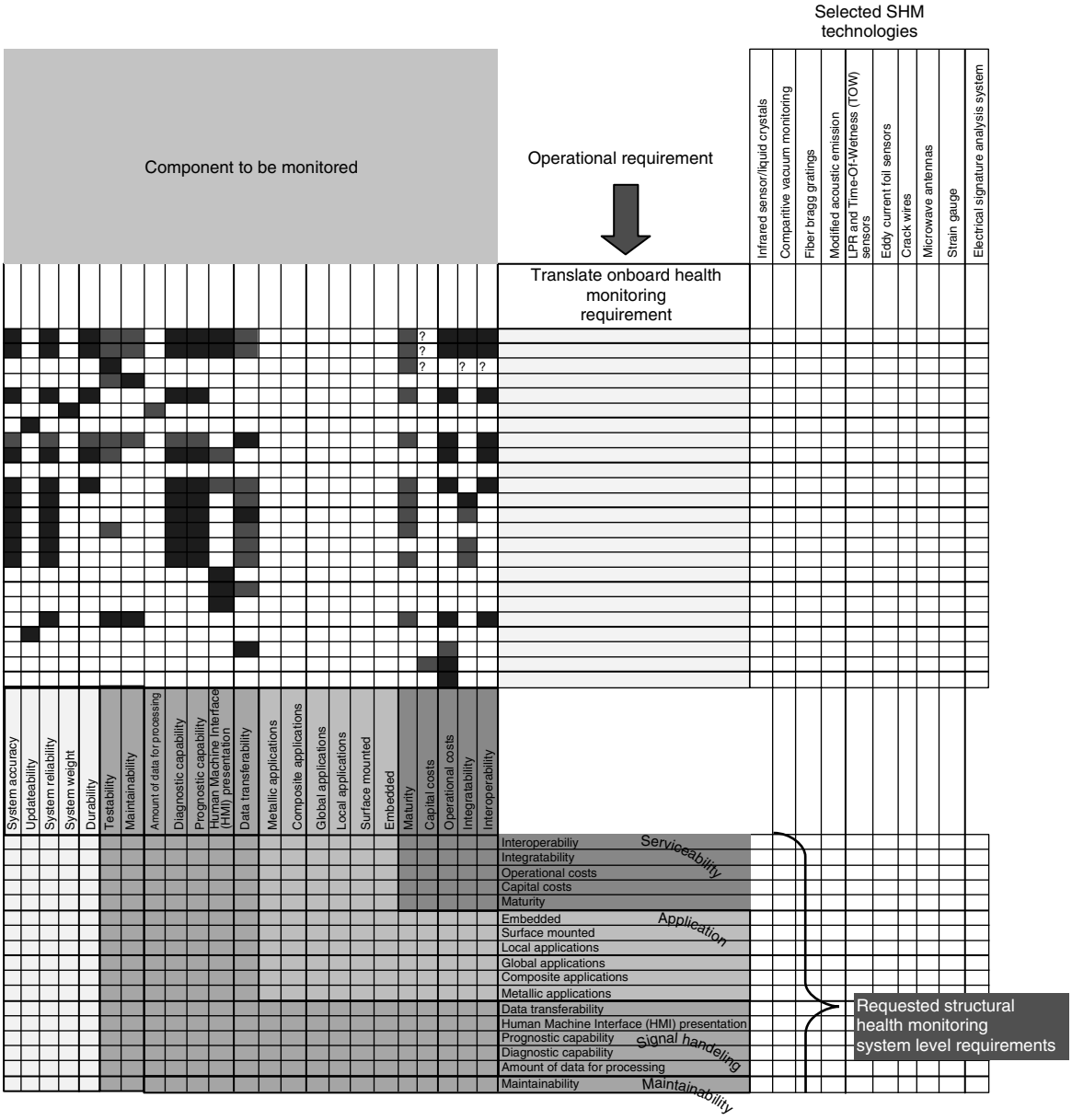


Figure 10. Technology selection matrix.

Technology criteria (also considered as operational handling requirements) in TATEM were derived from a broad variety of operational parameters that mainly looked at metals and composites, the two main classes of structural materials used in aerospace, and at the damage resulting on them. Technology criteria were split into five areas that include

- performance
- maintainability

- signal handling
- maintainability and
- performance.

The procedure was applied to identify promising SHM technologies to be applied for monitoring horizontal tailplane (stabilator) of an F/A 18 fighter airplane, composed of carbon/epoxy skins on an aluminum honeycomb core with an internal metal structure at the stabilator root and a longitudinal riveted aluminum joint of an Airbus A340 front fuselage. Similar applications could therefore be sought for UAVs.

3.5 Sensor signal processing

Several processing units are necessary to operate an SHM system. On the local level, a processor must interface with the sensors to acquire the data and convert the raw analog signals to digital ones. If it is an active system, such as with Lamb-wave methods, the processor must send instructions or waveforms to the actuator periodically. High data rates would be necessary for a set of either Lamb-wave or acoustic-emission sensors collecting data. It can easily be seen that data manipulation and data fusion in the sense OSA-CBM must be provided by the SHM system at the component level. As a result, a damage index or equivalent data is sent via the data bus to a host such as the AMC, which hosts the health assessment applications. Part 3 of this Encyclopedia describes several signal processing techniques. An overview of the health management architecture is provided in Figure 11.

3.6 Communication

A variety of communication models exist for communication between a network of components, which include: multicast, broadcast, and client–server. In the multicast model, the information supplier publishes his information to the network, addressed to a known list of recipients; this is considered to be an asynchronous approach. In the broadcast model, the information supplier publishes his information to all network listeners and the listeners must decide if they are interested in the content of the message. Finally, the client–server model pairs a client (who initiates communication) with a server (who is designed to respond to certain requests). The server implements interfaces that may be used by a client to request a service. A client can only request services available at the server's public interfaces. Data passing may be implemented by means of a single synchronous message, or through a pair of asynchronous messages.

Most of the components mentioned above require power to function. Piezoactuators, for example, operating at 15 kHz with 5-V peak to peak would draw

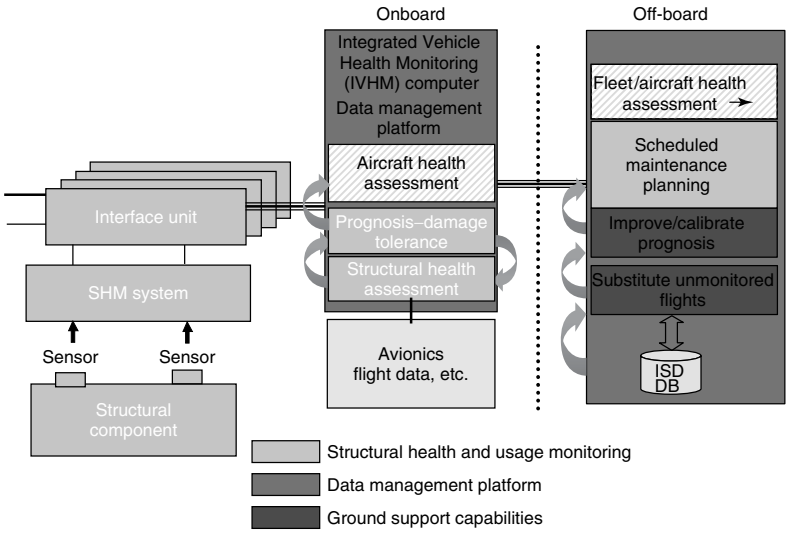


Figure 11. Proposed health management architecture.

24 mW each. A low power microcomputer to process the data would likely draw about 10 mW, and a short-range wireless device would require about 5 mW to function. Although the individual component power demands are low, this becomes challenging when there are many components distributed throughout the surface of the structure, some of which can even be embedded within the skin. The power supply should be provided via the vehicle power management. This does not answer the question of how the connection to the electrical data bus can be realized. There is a wide range of discussions from structure-integrated electronics to poor surface applications. Independent of what one can finally achieve, the power supply is the key enabler and requires efficient integration.

Comparison between the detectable damage size of one single sensor on a 1 m × 1 m composite panel as a function of sensor size, coverage area, and the required power for different SHM technologies has been provided by Kessler in [7].

3.7 Provision of an algorithm to support diagnostic, prognostic, and decision support

3.7.1 Algorithms

Algorithms are probably the most essential elements to an SHM system. They are necessary to decipher and interpret the collected data, and require an understanding of the operational environments and material thresholds. Examples of algorithms that have been used in this research include codes that perform modal analysis and wavelet decomposition. Other algorithms that could be embedded into an SHM system include codes that interpret the sensor data to specify the damage size and location, codes that calculate the residual strength or stiffness of the structure, or codes that predict failure based upon the measured damage.

Health monitoring and management encompasses the set of activities that are performed in order to identify, mitigate, and resolve faults and damages on the aircraft [8]. The activities can be grouped into the four phases, illustrated in Figure 12 and described below.

3.7.2 Condition state detection

There is a need to monitor the systems, equipment, and components that have the greatest impact on mission effectiveness, maintenance, and operation cost.

3.7.3 Health assessment and prognosis

The detection of failures or damage and isolation of the corresponding causes ("what is wrong with the component/structure") and the assessment of the actual aircraft status given any detected failures, damage, and actual degradation ("the overall fitness—i.e., nothing may be broken but the condition is still not 100%") and the prediction of the future

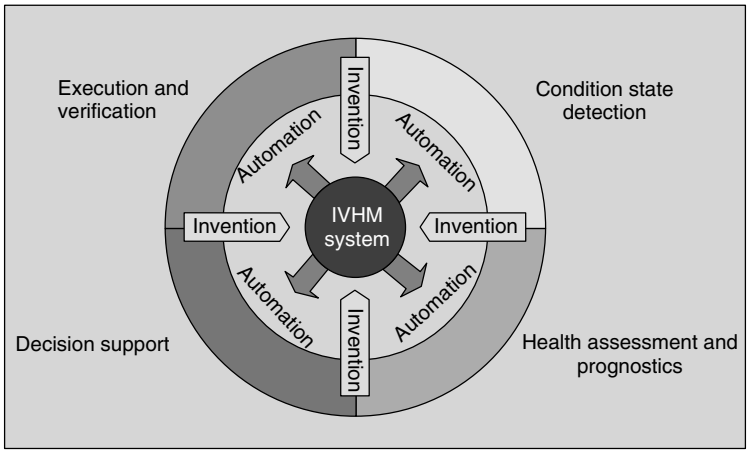


Figure 12. Modified health management activity model (proposed by Aaseng 2001).

condition of the monitored system subsystem or component to provide a remaining usable life (RUL) ("how much longer will the structure/component last").

3.7.4 Decision support

This provides advice for the maintenance engineer and planners to perform their jobs better in face of time pressure, complex scenarios, etc. ("What is the best course of action—monitor, repair, replace—and then how to go about it?").

3.7.5 Execution

This involves performing activities to replace or repair the failed or damaged components, to ensure that the repairs and inspections were performed correctly, and that the system has been returned to full operational status.

The development of the diagnostic, prognostic, and decision support capabilities as part of the aircraft health monitoring consists of the following steps:

- definition phase
- design phase
- development and qualification phase and
- usage phase.

The main parameters that have to be taken into account for diagnosis are as follows:

- Aircraft type
- Modeling of operational scenario
- Mission variability and area
- Usage parameters (environmental parameters)
  - event monitoring
  - damage monitoring
  - load monitoring (including dynamic loads)
  - usage monitoring
  - fatigue life monitoring and
- Critical areas
- Damage index or condition state
  - failure model
  - material properties
  - threshold parameter
  - aircraft and component qualification
  - etc.

A snapshot of the operational scenario modeling is provided in Figure 13. The scenario is divided into operational scenarios such as intelligence, surveillance, and reconnaissance (ISR), strike and suppression of enemy and electronic attack and combat, shown as the horizontal lines. Each of these scenarios is the result of different loading and damaging conditions, which need to be covered by the SHM system.

	In-flight damage monitoring and assessment	In-flight structural overload monitoring	Usage monitoring (UM) and fatigue life monitoring (FLM)	Long-term damage monitoring
ISR Long distance, high altitude	Birdstrike	G-monitoring Hard landing monitoring	Usage monitoring Gust, FH, cycles	Single cases
ISR Short distance, medium to low altitude	Birdstrike Battle damage	G-monitoring Hard landing	Usage monitoring G-spectrum, Gust, FH, cycles	Single cases
Strike/ suppression of enemy defense	Birdstrike Battle damage	G-monitoring Hard landing Store release	Extended UM Dedicated FLM	Dedicated areas Manuever loads Low-cycle fatigue
Electronic attack and combat	Birdstrike Battle damage Structural overload damage	Manuever Hard landing Store release Buffet	Extended UM Dedicated FLM	Dedicated areas Manuever loads Low-cycle fatigue

Figure 13. Snapshot of operational scenario modeling.

4 HEALTH ASSESSMENT

This new decision support process to be covered within the "operational support" will add a proactive function to the decision support algorithm, where a GO or NO–GO decision will be assisted by the health assessment function of the integrated vehicle health management of an aircraft. The "operational risk assessment" concept appears here—an extended function of the operational support that will be supported on the structural health information to develop predictions of the future maintenance-relevant events (e.g., component degradation–driven repair or replacement events) and their impact on the operational planning of the aircraft/fleet. On the basis of the ORA, short-term scheduled maintenance activities should be proactively defined and the long-term scheduled maintenance planning should be adapted. It is recommended that the ORA be carried out according to OSA-CBM architecture in the onboard data management platform.

As part of the health assessment, it is essential to generate a conditional view function that is responsible for providing the RUL prediction with associated confidence level at real operation with respect to the expected usage of the aircraft. This conditional view will provide a basis for operational risk estimation, together with other sources of information, such as operational constraints, economic/safety information, etc.

4.1 Prognosis

There are basically three types of information that may be the basis of the RUL prediction in prognostic approaches. At one end are the models based on statistics (reliability or failure data). Here, knowledge is based just on failure probabilities that can also be coupled with expert judgments. The "confidence" that may be associated with the estimation provided in this way is the lowest, although the applicability of this method within the aircraft is widest. At the higher end of reliability should be located those estimation approaches that are built on top of physical or mathematical models, usually validated physically at component or major airframe fatigue test level. Here, once the main input parameters are known, it is possible to estimate the system condition with great accuracy. Information for the prediction may be based on condition of usage monitoring that allows incomplete models of the degradation of monitored components to be derived, normally on the basis of identification of partial information within the model (trends, limits). In this case of model-based information, the RUL output can usually be interpreted as degradation information, whereas when only statistical or reliability information is available, the RUL estimation referred to is a perceived probability of failure (with no relation of the internal degradation of the piece).

In Figure 14, a possible approach to performing reliable diagnostic measurements and prognostics is shown, which has been derived from a concept initially proposed by Boeing [9] (*see also* **Commercial Fixed-wing Aircraft**) and adapted to the needs of UAVs. The basic idea is to link usage, damage, and damage monitoring of monitored components. This approach requires the development of a database including a wide set of data, e.g., damage index or condition state of all flights, failure model, material properties, threshold parameters, aircraft and component qualification, configuration, etc. A further prerequisite is synchronization, e.g., regarding time of usage and damage monitoring. This is not considered as critical in the case where damage monitoring is linked to the data bus. By means of the operational parameters and the accurate condition state determination, reliable diagnosis and prognosis appears to be feasible.

4.2 Decision support

The ORA in the scope of aircraft operational support calculates a set of discrete probabilities of operational interruptions over an assigned set of missions. The calculation is made under the terms of improved operational effectiveness optimization. The ORA in this context describes the calculation of probabilities of operational interruptions caused by unscheduled maintenance events.

The processing of ORA starts with the estimation of the survival rate for each monitored system or component, as well as for the complete aircraft for an assigned mission.

The estimation of managed components and aircraft-level survival rate uses an implementation

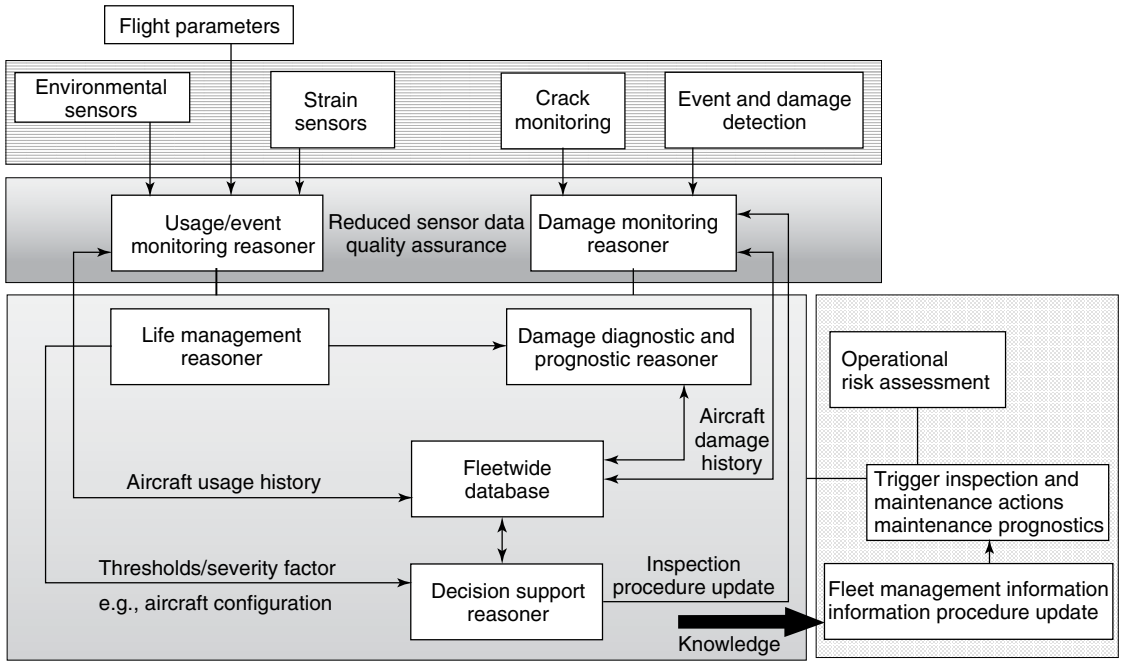


Figure 14. Conceptual approach to performing diagnosis and prognosis at aircraft level as part of an embedded health management system.

of the maintenance-free operating period (MFOP) approach. It allows the reliability assessment of complex technical systems. The primary goal is the calculation of the survival probability for a discrete time period to a given level of risk for availability of the technical system.

The parameter to be calculated in this assessment is the maintenance-free operating period survivability (MFOPS), the probability of survival for a lapse of time t_{mf} time units. t_{mf} is, in this context, the concrete MFOP cycle and defines the lapse of time in which the system will be able to perform its function accordingly.

5 CONCLUSIONS

The variety of activities mentioned in this article with regard to designing, monitoring, and integrating SHM systems into UAVs to support new maintenance strategies and support autonomous missions form a challenge. Significant activities have to be performed until SHM systems become part of the future design process, mainly due to the fact that the interface between fatigue monitoring systems, the subsequent management of aircraft life limitations and repair limits is still missing. Broader thinking in terms of life-cycle cost has become highly important.

Advanced sensing, materials, and data-processing technology are possibly the biggest challenges in revolutionizing the inspection process and in automating the procedure leading to the most common information of "no failure found". In case this automation can be achieved, then further, and possibly a much larger potential can be seen in the integration into the decision support algorithms for autonomous missions. Further pursuing this method will allow new maintenance concepts and strategies for UAVs and repair techniques that can be more easily certified and more efficiently applied. There is, therefore, much potential still waiting to be explored.

REFERENCES

[1] Fry S.A (ed). *Joint Publication 1–02: Department of Defense Dictionary of Military and Associated Terms*,

amended version August 31, 2005, US MoD, see also www.dtic.mil/doctrine/jel/doddict, 2005.

[2] Office of the Under Secretary of Defense for Acquisition Technology and Logistics Washington, DC. *Mandatory Procedures for Major Defense Acquisition (MDAPS) and Major Automated Information System (MAIS) Acquisition Programs*. US Ministry of Defence, Publ. DoD 5000.2-R, April 5, 2002.

[3] Boller C. Ways and options for aircraft structural health management. *Smart Materials and Structures* 2001 **10**:432–440.

[4] Trego A, Akdeniz A, Haugse E. Structural health management technology on commercial airplanes. *Proceedings of 2nd European Workshop on Structural Health Monitoring*. Munich, Germany, 2004.

[5] Abbott D, *et al. Development and Evaluation of Sensor Concepts for Ageless Aerospace Vehicles*, NASA/CR-2002-211773. 2002.

[6] Munns T, Palmer M. Health Management Technologies and Experiments for Transport Aircraft Landing Gear. *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Paper AIAA-2005-6355, San Francisco, CA, 15–18 August 2005.

[7] Kessler S.S. *Piezoelectric-Based In-situ Damage Detection of Composite Materials for Structural Health Monitoring Systems*. MIT, Department of Aeronautics and Astronautics: Cambridge, MA, 2002.

[8] Aaseng G, Gordon B. Blueprint for an Integrated Vehicle Management System. *IEEE 20th Digital Avionics System Conference*, Daytona Beach, FL, October 2001.

[9] Akdeniz A. *Structural Health Management Technology Implementation on Commercial Airplanes*, Boeing White Paper. 2005.

Chapter 116
Monitoring of Solid Rocket Motors

Gregory A. Ruderman
Air Force Research Laboratory, AFRL/RZSB, Edwards AFB, CA, USA

1 INTRODUCTION

Missiles are typically considered to be "wooden rounds", in that they are deployed in the field as all-up, ready to go devices, which require no maintenance and will function correctly when required, even if that time is many years in the future. During this time, the missile may be subjected to a wide range of inputs. Depending on the environment, these can include mechanical and vibrational loads from transportation and operational use as well as a broad range of thermal environments, which may occur in rapid cycles due to deployment on aircraft. In addition, motors are also subject to chemical attack, which may occur entirely internally to the motor, or could involve the external environment, often from moist or saline environments. The structure of the rocket motor itself contains a number of bondlines, which are often the location of structural flaws, and these may affect the integrity of the system. The energetic materials and binders in different regions may attack other materials, changing the local properties. And the entire system is enclosed in a pressure vessel, which must resist high thermal and structural loads from both internal and external environments during operational use.

Despite these challenges, solid rocket motors (SRMs) typically not only have extremely high reliabilities and mission success rates, which is a tribute to the design process and the design engineers, but they also often incorporate large factors of safety to ensure reliability. As budgets decrease, however, more is often requested from systems, whether it is improved performance or extended service life, which eats into that margin and requires both development of improved modeling techniques and sensor technologies, which will be able to accurately assess the condition of missiles.

2 FUNDAMENTALS OF SOLID ROCKET MOTORS

SRMs are generally simple devices, particularly from a structural point of view [1]. The propellant grain, a composite of energetic materials and a polymeric

 ISBN: 978-0-470-05822-0.

binder, is enclosed by a case, which serves as a pressure vessel. As the propellant burns, hot gases are generated and the thermal and pressure energy is converted into kinetic energy by a nozzle. Each of these components has their own structural issues, individually and with respect to the entire missile system that are discussed below. Figure 1 presents an idealized SRM for reference. This motor was developed by ATK Launch Systems, Brigham City, Utah, under the Air Force Research Laboratory's "Critical Defect Assessment" program [2]. While it does not correspond to any operational system, it possesses many of the aspects, which create difficulty for analysts and motor designers, including a segmented grain and a complex geometry composed of a combination of the central cylindrical bore and six radial fins, making it an ideal validation case for sensors and computational models.

SRMs come in a variety of shapes, sizes, and thrust levels, depending on the mission requirements. For the purpose of this article, they can broadly be divided into three classes: tactical, strategic, and boost/space launch, each with unique missions and issues relating to structural requirements and monitoring.

Tactical missiles are often small, ranging from 70 mm (2.75 in.) diameter, typically deployed on missile pods of helicopters, to 127 mm (5 in.) Sidewinder, 178 mm (7 in.) AMRAAM, and 254 mm (10 in.) Patriot missiles. These missiles are used all over the world and are, as such, exposed to a variety of environments from Arctic to desert to oceanic, and in the particular case of air-launched missiles, the missiles are often rapidly cycled between these environments, as an aircraft takes off, climbs to the cold upper atmosphere, and then is subjected to high heating at supersonic speeds.

Strategic missiles, broadly, are significantly larger (e.g., 211 cm (83 in.) for the Minuteman III and 234 cm (92 in.) for the Trident D-5) and are used to deliver long-range payloads such as nuclear devices. Unlike many tactical missiles, strategic missiles are deployed to a single location such as a silo or submarine and left in place, generally reducing the variability and severity of the environments they are likely to face.

Space launch motors can be considered as a separate category generally in the way they are treated. While many are of similar size to strategic motors, they are designed and manufactured specifically for launching payloads to space, and are generally treated very carefully and not allowed to age significantly before use. Common solid motors for this application are the Space Shuttle Reusable Solid Rocket Motor (RSRM) and the Atlas V Solid Rocket Booster. Upper stage motors are often also solids, but are significantly smaller and come in many shapes and sizes (e.g., the

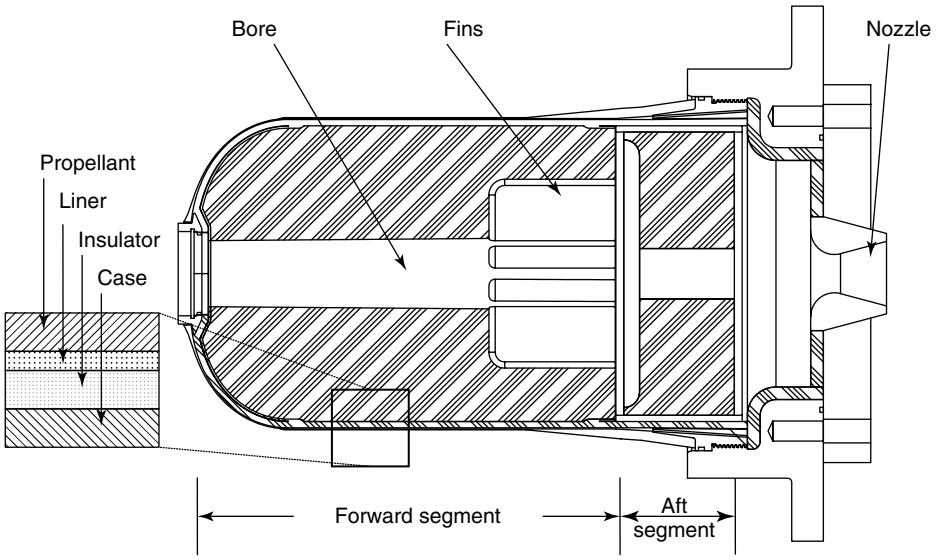


Figure 1. Schematic of an idealized solid rocket motor.

Orbus or Star motor series). A notable exception to the idea that these motors are of recent pedigree is the Orbital Sciences Corporation Taurus, Minotaur I, and proposed Minotaur IV, which will use decommissioned strategic missile stages.

3 MOTOR COMPONENTS

In the next section, individual components of the SRM are examined in greater detail, emphasizing the challenges and opportunities for structural health monitoring for each.

3.1 Cases

The case of an SRM is a pressure vessel, maintaining the integrity of the motor at high pressures during operation. Cases are typically made from materials such as D6AC steel, 6Al-4V titanium, or various fiber composites such as Kevlar, glass, or carbon fiber. Generally, the cases for larger motors have a greater potential to be composite, although this is driven by mission requirements rather than the size itself. For example, tactical missiles tend to have metal cases because of the need for longitudinal stiffness during maneuvering, while the space shuttle RSRM boosters are steel simply so they can be reused on future missions.

The use of composite materials, as in many other applications, is strongly driven by the high specific strength of these materials. Reduction in the inert weight of a system directly enables improved capability for the missile, usually in terms of increased payload, but the use of these materials also entails new risks. Unlike metal cases, composites can be fairly easily damaged by impact, and can be damaged severely that a catastrophic motor failure can result while still not being detectable to the naked eye. Composites, at least in comparison to the metal cases they replace, can allow diffusion of water into the motor, particularly if the matrix has been damaged by handling. In addition, due to the natural anisotropy of the wound bottles and complexity of design, fidelity analysis of composite pressure vessels is not as mature as that for metals. Composite cases are regularly overdesigned in the dome region to ensure failure in the cylinder, where predictive models are more accurate.

3.2 Propellant–liner–insulator (PLI) system

Working inward from the case, the first material encountered is the insulation, typically a rubber material such as ethylene propylene diene monomer (EPDM), which protects the case from the extreme thermal environment in the motor and maintains the integrity of the case as the burning surface of the propellant approaches the case wall. Often, the insulation is reinforced with Kevlar or other fibers to protect the case wall from mechanical erosion caused by the burning fuel particles. The insulator is adhesively bound to the case wall, but often has sections in the head or aft end (or both) where it is not bound. These "flaps" reduce the stress and strain on the propellant during motor operation by freeing the propellant from stretching as much as the case does under pressure. The next layer is the liner, a polymeric material, the main purpose of which is to bond the propellant to the insulator, although in many systems, this includes a "barrier coat" designed to prevent the diffusion of chemical components from the insulator into the propellant (or vice versa). The importance of these types of chemical processes to the structural aspects of SRMs is discussed in greater detail.

The next layer is then the propellant itself. The most common type of propellant, particularly for use in Air Force ballistic missile and space launch applications, is composite propellant. Typical composite propellants are composed of a crystalline oxidizer, often ammonium perchlorate (AP), ground to a nominal particle size of typically 200 μm, although many propellants have bi- and trimodal distributions of sizes to improve mixing properties and tailor burn rates. Aluminum powder is added to the mix as a fuel and both are held together by a polymeric binder. Commonly used binders include hydroxyl-terminated polybutadiene (HTPB) or polybutadiene acrilonitrile (PBAN). These materials are mixed, along with appropriate curatives and other chemicals such as burn-rate modifiers, cast into the motor and cured. Some tactical motors reduce or omit the aluminum fuel, as it generates a highly visible white smoke trail, which can easily be used to track back to the source of the missile. Many Army tactical and Navy strategic motors use a propellant type known as *double base*, composed of nitrocellulose and nitroglycerine, sometimes with aluminum and/or

nitramines (HMX or RDX added). Regardless of the type, rocket propellants are generally chemically and mechanically complex, nonlinear, viscoelastic, prone to damage, debonding of particles from the binder, and often nonuniform in properties due to material flow and particle segregation during the manufacturing process.

There are multiple different effects that make structural analysis and health monitoring challenging for SRMs. These include material property variation (spatial nonuniformity), defects (voids and inclusions), and material aging (chemical–mechanical link). Each of these is discussed in the following sections.

3.2.1 Material property variation

To understand the variability inherent in the mechanical properties of an SRM, it is necessary to examine the process by which the propellant is made and the motor cast. The propellant is manufactured in a large mixing bowl and then cast directly into the case. For motors with a bore or fins, a casting tool of the appropriate configuration is placed inside the case and the propellant flows around it. The uncured propellant has a similar consistency to thick cake batter and tends to flow nonuniformly as it fills the case. Larger motors can require multiple separate mixes of propellant (the largest mixer in common use in the rocket industry is 6800 liters), which may cause knit lines to form between the different portions of the cast. During the casting process, the particles in the propellant often segregate by size, with smaller particles remaining closer to the bore and larger particles comparatively richer in the bulk. This is a known effect that is most often seen in unexpected changes in the burning rate of the propellant often called the *burn anomaly rate factor* (BARF), but can also be seen in the resulting differences in propellant properties in those regions as well. The fully loaded motor is then placed into an oven, where the propellant is cured at elevated temperatures, for multiple days in the case of large motors. While care is taken, nonuniform oven temperatures can also occur, which may affect the cure process in different locations within the motor.

3.2.2 Material aging

Aging of motor materials is one of the greatest concerns for the long-term viability of a motor. Aging generally falls into three categories, which are generically, and somewhat inaccurately, called *aging*. In the first instance, the materials themselves may simply change chemically (and therefore mechanically) over time. For example, the binder material is generally lightly cross-linked and, as a result, is prone to oxidative cross-linking at exposed surface (e.g., at the bore), which generally results in propellant hardening. A second type of aging involves the influence of adjacent materials on each other, typically involving chemical diffusion at the various bondlines between the propellant, liner, insulator, and case. These chemical constituents can migrate between materials, driven by chemical gradients and accelerated by temperature. In some instances, the curative of one material has unexpectedly turned out to be antagonistic to the bonds in another material, eventually resulting in a bondline of zero strength and a motor exploding soon after ignition as the entire grain attempted to be extruded through the nozzle. The third type of aging is typically called *mechanical aging* or *mechanical damage* in the SRM community. This involves direct mechanical changes in the properties of the materials, typically as a result of thermal cycling, vibrational loads, or physical insult, which can result in the weakening of bonds, either between the particle and binder, or within the binder itself, which can lead to weak spots, fractures, or increased porosity.

3.2.3 Defects

Despite the best efforts of manufacturers, rocket motors often have unexpected features. Voids in propellant often occur as a result of insufficient settling of the propellant during the casting process. Trapped air bubbles are not fully eliminated and a small area is formed, which contains no propellant. If small enough, these are not typically of great concern, although if they are proximate to an interface or other high stress or strain region can contribute to the formation of cracks. Inclusions are objects that end up in the propellant, which should not be present. These can be large pieces of propellant ingredients or other

motor materials, but also include anomalous objects. Notable inclusions that have historically occurred in motors include lead shot and a crumpled paper cup. Regardless of the source, these objects are often poorly bonded to the propellant and cause perturbations to the stress/strain field of the motor in a similar fashion as voids. If the item is large enough or not fully consumed, then damage may result on the inside of the case and to the nozzle causing a major concern. Depending on the material in the inclusion, the combustion process in the region can be significantly changed. In some cases, this has been intentionally used to advantage, such as placing fine metal wires in the propellant to increase the burning rate by increasing thermal conduction and providing a path along which the flame travels more easily.

Cracks can occur throughout the motor, although they are often seen in the bore, particularly in motors that have undergone thermal cycling. When a crack occurs, there are two scenarios. In the first case, when the combustion surface reaches the crack, the flame speed exceeds the crack propagation velocity. In this situation, the crack tip is blunted by the burning and does not propagate, so the concern is simply the increase in pressure of the motor due to additional burning surface area. If the crack area is small compared to the surface area of the motor, the pressure will not be significantly increased and this will not be a major issue. In the case that the crack propagation speed is greater than that of the flame, the crack will propagate. In this situation, the burning surface is exposed deeper in the motor before it is expected. Since the insulation thickness is determined during the design process by the calculated time of exposure to the hot gases (with an appropriate factor of safety), early exposure can overwhelm the insulation, heating the case and creating an opportunity for failure. Cracks also occur in the propellant near the propellant–liner interface. This compounds the problem, as there is hot gas near the wall, and if the crack propagates, the motor grain is detached from the bonding surface.

Debonds are similar to the cracks described earlier, but result from insufficient or incomplete bonding between two of the propellant–liner–insulator materials. As with cracks, the concerns are augmented burning near the case wall and the structural impact of the decreased bonding.

4 CURRENT STATE OF THE ART FOR SOLID ROCKET HEALTH MONITORING

The air force is interested in developing health management technology for SRMs to reduce system life-cycle costs while improving safety and reliability. The ultimate goal is to be able to tell which assets have aged out or otherwise need to be replaced and which assets are still viable. Currently, a typical missile aging program takes a small number of motors, fires some, and dissects others for verification. This is performed periodically, and as long as the verification firings and dissection data are nominal, the entire lot is considered viable. Otherwise, following further investigation, the entire lot of missiles may be condemned and destroyed. While attempts are made to choose motors that are expected to be "bad", i.e., which have seen more time in service or excessive thermal cycling, often the precise history of any given motor is not well known, making such attempts problematic. Since a statistically meaningful quantity of verification motors are not used, there is a strong probability that viable motors will be destroyed and have to be replaced at a significant cost to the government, or that despite "successful" verification of the system, failures will occur, potentially causing mission failure, destruction of government property, or loss of life.

Changing to a condition-based paradigm has potentially enormous payoffs. Consider the following notional example (numbers are used for illustrative purposes and do not relate to any current policy or system): Assume a missile fleet of some size X. The fleet is considered nonviable at 17 years, at which time 2.5% (0.025X) will not successfully complete the mission due to age-out of a structural material. Currently, the entire fleet would be retired and replacements procured, rather than culling and replacing the bad missiles and leaving the rest. As those original missiles age, more will need to be retired more rapidly, but never as many as replacing an entire fleet in a short period to maintain readiness. Simple analysis assuming this 17-year service life and motors aging out in a Gaussian distribution means we have built approximately 2X missiles at the end of year 51. However, following the current paradigm, we will have built the original fleet and replaced it entirely three times (total of 4X missiles).

While this is a very simple example, it reveals some of the enormous benefits that could be realized.

4.1 Current use of IVHM for strategic systems

The major health monitoring activity for current strategic motors is the automated nondestructive evaluation system (ANDES 2) at Hill Air Force Base, Utah [3, 4]. ANDES 2 is the second generation of an non-destructive evaluation (NDE) data analysis system currently examining computed tomography (CT) data and capable of inspecting all three stages of a strategic system, detecting voids, inclusions, debonds, or other flaws as small as 0.25 mm (10 mils). One of the benefits of this second generation of ANDES is the capability to be "trained" on any type of NDE data, be it digitized film X ray, CT, or ultrasound. Any identified flaws are reported to the user and a recommendation is provided as to whether these meet the motor specification. These data are maintained by Hill air force base (AFB) and provide a zero-time assessment of motor structural state. The ANDES system has historically been used for inspection of the current Minuteman III fleet. Unfortunately, due to cost and logistical difficulties, these inspections are not performed regularly, but only when the system needs to be brought back to the depot for other maintenance.

Marks measured and evaluated by ANDES are converted into faceted surfaces, which can be transferred directly to the Air Force's Structural/Ballistic Analysis System (SBAS II). SBAS is an analysis code that solves coupled fluid-structural-thermal-ballistic problems [5]. Of particular interest to structural analysis, SBAS reads the marks detected by ANDES and can take a baseline motor mesh and automatically integrate the flaws, remeshing the model as necessary without user intervention. As the analysis proceeds, if integrated continuum failure or fracture propagation models determine that a crack will form or propagate, this too is performed automatically, significantly reducing the time required for an analysis.

Other nondestructive techniques are rarely used on deployed systems. Eddy current or ultrasound are sometimes used for quality control by the motor manufacturers during the manufacturing process, but are not typically used once a motor has been fielded. Embedded sensors have been frequently used in demonstrations on subscale articles, but are not used on deployed systems. Chemical sensors are currently being developed, but remain at a low technology readiness level (TRL). All chemical data used in aging models is acquired in a destructive fashion, usually by dissection of the motor, after which the desired properties (e.g., cross-link density, sol–gel, and chemical concentrations as a function of position) can be determined in laboratory experiments. Chemical aging models have been significantly improved in the last decade, allowing not only a high reliability prediction of the current chemical state of the propellant–liner–insulator (PLI) system but also a prediction of the mechanical state of the motor as a function of the chemical state. This type of modeling is critical to the prediction of motor service life, but is currently limited in functionality as the motor of interest is always destroyed in the process of acquiring the necessary data [6, 7].

4.2 Current state of the art in laboratory demonstrations

In the past decade, the various Department of Defense laboratories have demonstrated new sensors and data-acquisition systems, usually in conjunction with rocket motor manufacturers such as Aerojet and ATK Launch Systems in an effort to raise the TRL to the point where the technology can be transitioned to a program office for use on a new system.

So far, the embedded (bondline) sensor, which has shown the greatest promise, is the Micron Instruments dual bond stress temperature (DBST) sensor, a small steel stress sensor 7.6 mm in diameter and 2.0-mm thick. DBSTs have been used in laboratory demonstration assets such as the motor shown in the first section. They have been flown in tactical configurations and have been used in motors, which have been fired on test stands. Currently, the Air Force Research Laboratory possesses two motors with embedded sensors, including DBSTs, which have been set aside for long-term storage and examination to ensure reliability and stability [7]. The other area that has shown some promise in the laboratory environment is sensor networks to detect and quantify damage to

composite cases. These have taken various forms, including a network of embedded multiaxial fiber-optic Bragg gratings (*see* **Fiber Bragg Grating Sensors**) wound into the composite case performed by Blue Road Research [8], and a network of piezoelectric sensor/transmitters (*see* **Piezoelectric Wafer Active Sensors**; **Hybrid PZT/FBG Sensor System**) attached to the surface of the case, which has been demonstrated by Acellent Technologies [9]. In both cases, the idea is to provide a system which, when interrogated, can detect the invisible or barely visible impact damage, which could result in the complete loss of structural integrity of the composite case without needing the data and logistics overhead required by an accelerometer that is always on.

5 CURRENT CHALLENGES FACING SRM IVHM

Despite the research, which has been performed by Department of Defense (DoD) laboratories and many contractor organizations, no current SRM system in operation possesses an onboard health management system. In part, this is due to practical reasons. New systems are not developed often, and the best time to insert health monitoring is during the design process. Once a system has been designed, even if modifications are required (for example, due to material obsolescence issues), those changes are kept to the absolute minimum to reduce requalification costs and any potential issues that might crop up due to design changes. In essence, while the philosophy has generally been that any kind of health monitoring needs to "buy its way onboard", i.e., have a quantifiable payoff that (far) exceeds the challenges of implementation, the buy-in has the potential to be even higher for the existing systems.

Of course, this argument assumes that the TRL of the system and associated technologies is sufficiently high. Unfortunately, while this is the case for some individual sensors or systems, overall, we are not quite there. Following are some of the outstanding issues that need to be further addressed before use of this technology can be seriously considered.

5.1 Data acquisition, storage, and analysis

One of the greatest challenges in the implementation of a practical integrated vehicle health monitoring (IVHM) system is the sheer quantity of data that can be collected. This is made even more difficult if one considers the example of a tactical rocket motor. As previously described, these items may be shipped to various locations across the globe and experience extremes of thermal, chemical, shock, and vibration loading. Depending on the rate of data acquisition, even for just a single asset, this could very easily overwhelm any amount of storage available. Clearly, despite the relatively low (and rapidly shrinking) cost of data storage, simply acquiring raw data at a high rate would not be practical, except in rare situations.

The obvious solution to this issue lies in significant improvement in constitutive modeling for any life-limiting component of the rocket motor. An accurate model of the behavior of the propellant–liner–insulator system would enable both detailed simulation of the rocket under various loading conditions and a fundamental understanding of the limits of loads, which could be applied without causing permanent damage and reducing functionality. Models of this type have been pursued for many years with varying degrees of success. Phenomenological models have historically been used to describe material behavior, but due to the complexity of the materials have been of little predictive utility, ultimately resulting in the large factors of safety and margins mentioned in Section 1. For example, propellants are typically considered to be nonlinear viscoelastic materials with an associated damage function in which the damage may or may not be reversible to some degree. Integration of the aging process into these models has been problematic as well. More recently, mechanistic models, which describe the material behavior in terms of the interaction of the constituents, have been developed and are coming into use [2, 5]. However, these models are extremely computationally intensive and significantly require more expertise to develop and use correctly.

The current approach is twofold. The phenomenological models are continuing to be developed with more advanced capabilities, including chemical aging models, multiaxiality, and new failure theories. Likewise, the mechanistic models will also continue

being developed, taking advantage of increasing processing power, computational efficiency measures, and the like. Ultimately, some combination of the two approaches will likely be employed. A fast model embedded in the asset could make an initial, rough assessment, and if it calls into question the asset's reliability or capability, a much more detailed analysis could be performed at the depot or other location when time permits.

5.2 System longevity and reliability

Another challenge to implementation has to do with the long-term stability of the sensor system itself, and the effect it might have on the rocket motor. In general, the sensors that have been demonstrated on Air Force Research Laboratory programs have been made of inert materials (glass for fiber-optic sensors, ceramics for piezoelectrics, and steel or titanium for stress sensors), which have shown no chemical effect on the motor in accelerated aging tests. In addition, because of concerns that the placement of the sensors within bondlines could actually generate the flaws that we are attempting to avoid or detect, a large number of bondline test articles with different sensors have been tested. In every case, the articles failed away from the sensor, suggesting that the effect of the sensors is smaller than that of the natural nonuniformities in the material. Of course, this assumes that the placement of the sensors is performed correctly by not placing them in a region that would be very sensitive to small flaws or variations.

Long-term reliability and stability of the sensors themselves then becomes the problem. Assuming that the sensors survive the manufacturing process, the immediate concern is whether the sensors are still correctly calibrated, and if so, for how long will they remain so. So far, this remains an open question. While some embedded bondline sensors have shown excellent long-term stability, maintaining accurate readings in a laboratory environment for many months, others have shown a tendency to drift from their calibration. Without a reliable method for *in situ* recalibration and validation of the sensor data, the data will always have some level of question surrounding it.

In the very long term, for example, over an expected 35-year service life of a new strategic missile, we have no data on how any sensor system will survive. It is possible to make an excellent argument, however, that as long as the sensor fails gracefully, it does not really matter how long the sensor lasts. Even though motors continue to evolve over their entire lives, most of the aging changes (as opposed to the changes driven by external boundary conditions) occur in the first few months. *In situ* data in this time period would significantly improve any models and allow the operators to understand the pedigree of individual assets, which, in turn, will improve predictions of future behavior.

5.3 Lack of service life sensors

One of the more interesting issues in using health monitoring to determine service life of SRMs is that the sensors currently available, and the majority of those in the current development, cannot directly measure the phenomena of interest that would correlate directly to service life-limiting phenomena. For example, sensors that have been embedded in the propellant–liner–insulator interfaces measure stress or strain, but ultimately what is desired is a way of measuring the material moduli, as those are the intrinsic properties that change with time and environment. Complex sensors that can locally determine the modulus have been proposed, but are generally in very early stages of development.

With the development of mechanistic material models, sensors that can measure primitive variables in those models are greatly desired. For example, any method that could directly determine the cross-link density of the polymer (which is directly related to the modulus) would be of great value. Similarly, the idea of chemical sensors that can directly measure either the chemicals that diffuse from different regions of the motor or that are by-products of chemical aging processes has been frequently discussed, but has not been greatly successful. Fiber-optic Raman spectroscopy probes have been explored, but have a tendency to heat the propellant locally, influencing the process (and in the case of one experiment, nearly igniting the propellant) [6]. Other chemical sensors such as fiber optics with coatings that respond to the presence of certain chemicals have also been discussed, but concerns about whether the sensors themselves would act as chemical sinks and influence

the local behavior, or would become saturated and rapidly stop working has limited development.

5.4 Business case

While not a technical issue, the successful development of a legitimate business argument for the implementation of IVHM systems is probably the largest hurdle to overcome [10]. Such an argument must address the concerns of all the stakeholders, from the policy makers, to the program offices, to the end users, and explain why taking a relatively simple and generally reliable system can be improved by adding complexity.

The simplistic example previously provided begins the discussion, but does not address the details necessary to allay concerns. The addition of IVHM will increase the complexity of the manufacturing process, add to the logistics tail of the system by requiring personnel to acquire, store, and analyze the data, add to inert weight, and cost more. This must be balanced against the potential payoffs in terms of reliability, safety, and availability. The argument will have to be made on a system-by-system basis and involve all the stakeholders from the beginning of the process. Otherwise, the IVHM system, as is often the case, may be the first thing to go when a program has cost overruns or exceeds the planned weight.

6 CONCLUSION: A WAY FORWARD

Despite all the challenges, the expected payoffs of an IVHM system are great and will continue to be pursued. In the short term, a minimally invasive approach will likely be the most successful in being adopted by the user community. In tactical motors, this most likely means a compact environmental (temperature, humidity, and vibration/acceleration) datalogger, which can be stored in the missile's transport case. Analysts will have access to high fidelity information on the motor environment for the first time and will be able to use that data to both make predictions and improve the predictive models with real-world experience.

In larger motors, a similar concept with sensors integrated into the weather seal may be possible. Integrating the environmental monitoring system, potentially with gaseous chemical sensors in the bore, should be a necessary first step in any new system, and could even be easily integrated into existing systems without costly design changes.

In the longer term, embedded sensor systems need to be considered as part of the design process, along with external monitoring such as the case of damage-detection systems or future generations of the ANDES system. Continued development of physically based models is critical to the process, as are the advances expected due to the advancements in computational and storage technology. The long-term ideal would be a fully instrumented asset that can assess its own status and, if necessary, make adjustments or inform the users that it needs to be replaced or taken back to the depot for a more detailed inspection. While we are currently far from that kind of system now, the demonstrated success of such systems on commercial automobiles and other complex vehicles suggests that we are not as far off as we might think.

REFERENCES

[1] Sutton GP, Biblarz O. *Rocket Propulsion Elements*, 7th Edition, John Wiley & Sons, 2001.

[2] ATK Launch Systems for Air Force Research Laboratory, Propulsion Directorate, *Critical Defect Assessment Program*, Final Report, AFRL-PR-ED-TR-2007-0038, Edwards AFB, 2007.

[3] Hildreth JH. *Application of Automated NDE Data Evaluation to Missile and Aircraft Systems*, PL-TP-97-3002, December 1996.

[4] ATK Launch Systems for Air Force Research Laboratory, Propulsion Directorate, *Non-destructive Evaluation Data Processing Program*, Final Report, AFRL-PR-ED-TR-2008-0009, Edwards AFB, 2008.

[5] ATK Launch Systems for Air Force Research Laboratory, Propulsion Directorate, *Service Life Prediction Technology Program*, Final Report, AFRL-PR-ED-TR-2003-0050, Edwards AFB, 2003.

[6] ATK Launch Systems for Air Force Research Laboratory, Propulsion Directorate, *Sensor Application and Modeling Program*, Final Report, AFRL-PR-ED-TR-2007-0027, Edwards AFB, 2007.

[7] ATK Launch Systems for Air Force Research Laboratory, Propulsion Directorate, *IHPRPT Phase 3 Sensor Application and Modeling Program*, Final Report, AFRL-PR-ED-TR-2007-0056, Edwards AFB, 2007.

[8] Blue Road Research for Air Force Research Laboratory, Propulsion Directorate, *Fiber Grating Sensor System to Determine Motor Case Damage*, Final Report, AFRL-PR-ED-TR-2004-0090, Edwards AFB, 2004.

[9] Qing XP Beard SJ, Kumar A, Ooi TK, Chang F-K. Built-in sensor network for structural health monitoring of composite structure. *Journal of Intelligent Material Systems and Structures* 2007 **18**(1):39–49.

[10] Aerojet for Air Force Research Laboratory, Propulsion Directorate, *AMPT Task Order 1: IVHM*, Final Report, AFRL-PR-ED-TR-2007-0055, Edwards AFB, 2007.

Chapter 106

Comparative Vacuum Monitoring (CVM™)

Duncan P. Barton

Structural Monitoring Systems Ltd., Perth, WA, Australia

1 CVM™—A SHORT HISTORY

Comparative Vacuum Monitoring or CVM™, was invented by Ken Davey in the early 1990s. Ken had been a commercial pilot during the 1960s and had been on a Vickers Viscount the flight before its wing fell off on descent into Port Hedland New Year's Eve 1968, killing all onboard. The main wing spar had failed owing to a fatigue crack initiating from an incorrectly installed rivet. Ken knew the crew of the ill-fated aircraft and vowed to find a better way to monitor the structure of an aircraft.

Several years later, Ken eventually stopped flying due to health reasons and after trying a few different jobs, he eventually founded a company for repairing the picture tubes in televisions. The final step in the repair process requires the glass tube to be placed under a vacuum. Ken observed that if there were any small cracks in the glass the vacuum was quickly lost. Ken combined the earlier tragedy with this observation and in 1994 patented "the Davey System", the basis of what is now known as *Comparative Vacuum Monitoring* or *CVM*™.

Ken successfully demonstrated a prototype of the system to the Australian Defence Science and Technology Organisation (DSTO) during a series of laboratory trials in the late 1990s [1]. On the basis of these promising early results a company, Structural Monitoring Systems Ltd., was formed to develop the system for use in the aerospace industry. As the technology has continued to be developed, it has been used in numerous programs including the certification of Glare® for the A380 with Airbus and flight trials with Airbus, Boeing, and several Air Forces. The technology is now also finding uses in other industries, including automotive testing and monitoring of infrastructure.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 THE CVM™ TECHNOLOGY EXPLAINED

2.1 The basic principle

The principle behind CVM™ is uniquely simple; a vacuum contained in a small volume is extremely sensitive to any leakages [2]. The CVM™ principle relies on placing a sensor onto the surface of a component where damage is expected to occur (see Figure 1). The sensor contains a manifold of fine channels that are open to the surface. Once the sensor has been installed on the surface, the channels form closed "galleries" to which a vacuum can be applied. It is important to note that the surface of the component forms part of the sensor system, with the crack itself providing the leakage path for air into the vacuum galleries.

The sensor is connected to a vacuum source through an accurate flowmeter. Figures 2 and 3 show a schematic of the equipment used in laboratory trials, which allows continuous monitoring of a sensor, or several sensors connected in series or parallel (not shown). If there is no damage on the component, then the vacuum in the sensor will be approximately the same as the vacuum source (Figure 2). If, however, a crack develops, a leakage path will exist and the vacuum level will be reduced in the sensor manifold (Figure 3).

2.2 The sensors

A variety of sensors have been manufactured from a number of materials, for a range of applications. The various sensor types are described below.

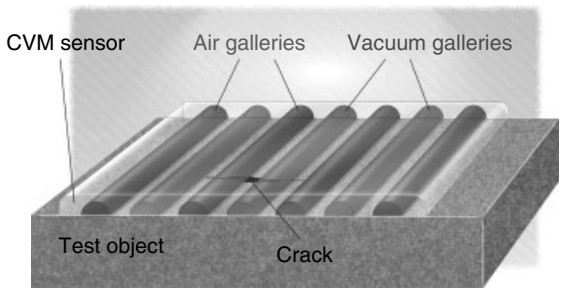


Figure 1. A schematic of a simple CVM™ sensor.

2.2.1 Sensors for metallic structures

The original sensors were manufactured using silicone with silicone or acrylic pressure-sensitive adhesive. The pressure-sensitive adhesive allows the sensor to be removed from the storage box and installed directly onto the surface of the component in much the same way as a postage stamp or tape. The silicone allows the sensors to conform to complex geometries with minimal internal residual stresses (Figure 4). Sensors are also manufactured from polymers such as polyimide and FEP for applications with different environmental and geometrical requirements. Polymer sensors can be manufactured to longer lengths, but are only able to conform to simple component geometries.

It is important to note that the sensors are the same whether used in a laboratory or an in-service component. This allows the response of the sensor configuration to be characterized in the laboratory for a specific component and then for these results to be transferred directly to an in-service application.

2.2.2 Sensors for composite structures

Sensors for composite structures are able to monitor barely visible impact damage (BVID) using standard sensors Figure 5a installed on the back face of an impacted structure and disbond using through-the-thickness or TTT sensors Figure 5b. By placing standard surface sensors, with detection galleries widely spaced on the back face of an impacted structure BVID can be reliably detected [3, 4]. The impact on the structure causes microcracks in the resin matrix, which can be detected using the CVM™ system.

The second form of composite sensor being developed is a TTT sensor. The sensor involves creating a small (diameter <1 mm) blind hole through the bonded structure and into the adhesive layer, which is then connected to a flow meter and vacuum source. If the adhesive layer fails, the damage forms a leakage path for air that is then detected by the flow meter. The size of the detectable damage is determined by the spacing of the holes in the structure. The sensors have been used to detect poor manufacturing processes and damage due to impact on the whole structure, with results confirmed using standard C-scans and thermography.

Figure 2. CVM™ system installed on an intact component.

2.3 The instrumentation

The CVM™ instrumentation consists of various derivations of vacuum source and a flow meter [5]. The laboratory variant allows users to gain experience with the new technology and understand its advantages and limitations. A periodic, off-board system allows the minimum level of instrumentation to be installed on an aircraft; essentially just the sensors, vacuum tubing, and a pneumatic connector. This system has enabled a history of robustness and durability data to be accumulated for the sensors. A cheap and simple CVM™ switch has been developed to meet the needs of the automotive testing industry and a fully integrated airborne monitoring system is currently in an advanced stage of design.

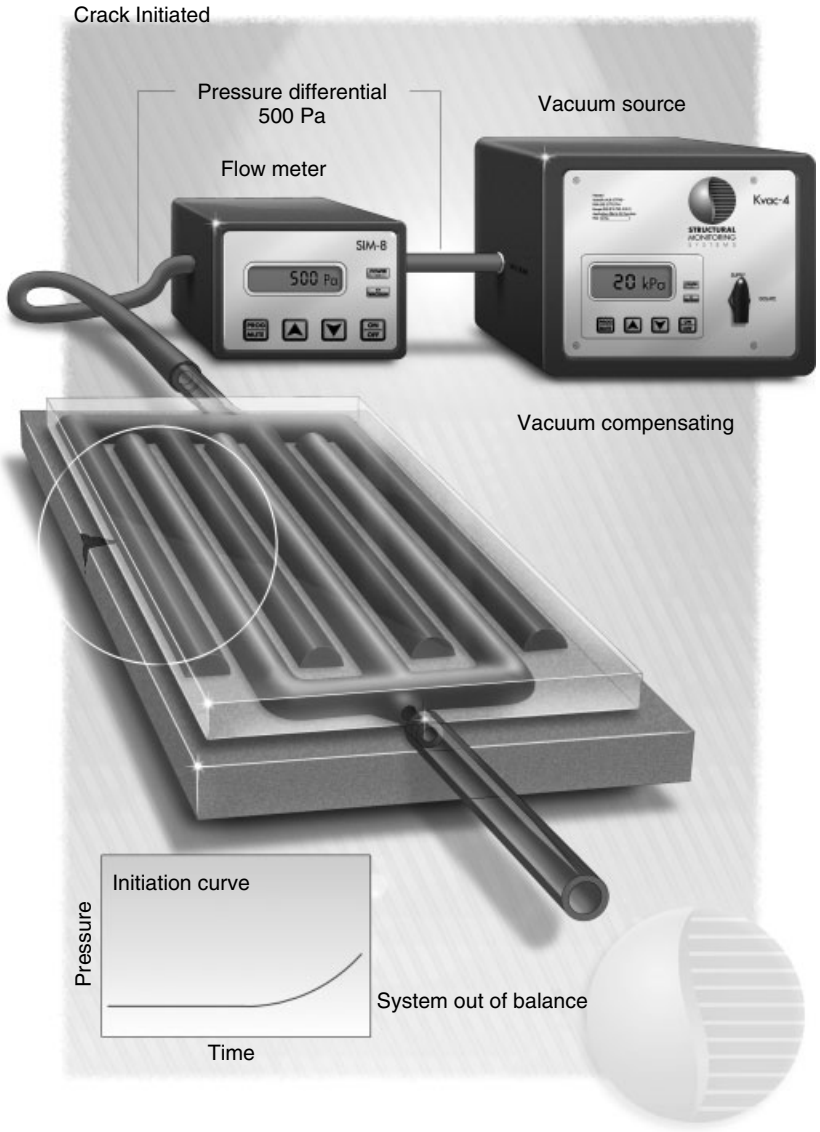


Figure 3. CVM™ system installed on a damaged component.

2.3.1 The laboratory kit

The CVM™ instrumentation was initially designed as laboratory equipment as shown schematically in Figures 2 and 3. The laboratory equipment continues to be used extensively in test programs from coupon level tests utilizing single sensors to full-scale fatigue test programs with several hundred sensors throughout the test structure and many tests in between.

The laboratory kit consists of a vacuum source, which can be reticulated to a number of flow meters, which can in turn be connected to a number of sensors. The whole system can be connected to a single computer to allow the data to be collected and stored in a standard database format. The system also includes standard analog and digital control output options to allow the instrumentation to turn off fatigue machines as soon as a crack has been detected. This

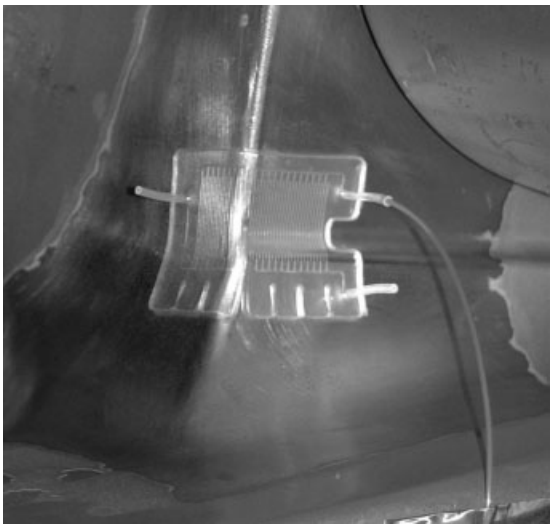


Figure 4. A silicone sensor used to detect crack initiation on a complex geometry.

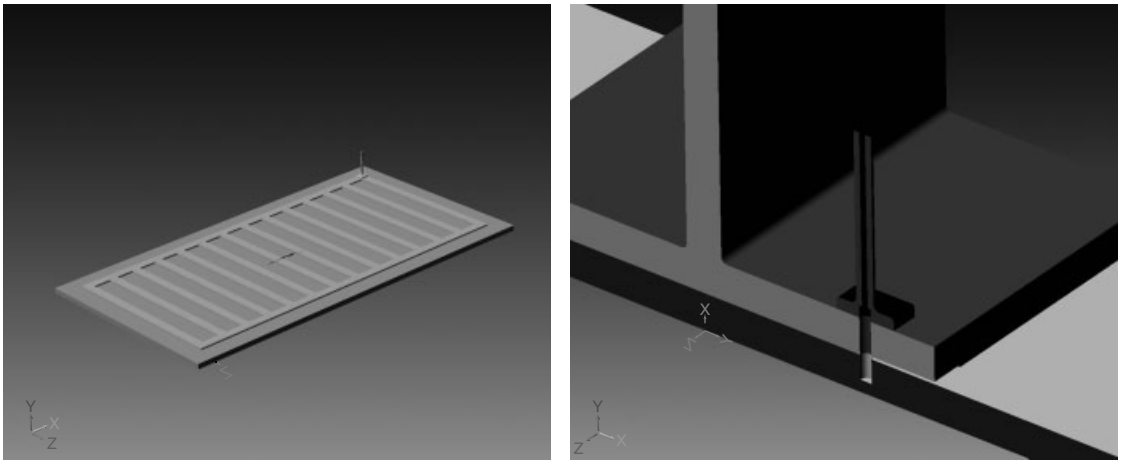
option, coupled with the fact that a sensor failure will also produce a signal to turn off the system (a large inflow of air on the rare occasion a sensor were to disbond), has allowed the system to be used to improve the productivity of a number of test laboratories that now rely on the CVM™ laboratory kit to turn off experiments left unattended during overnight testing.

2.3.2 The periodic monitor

In order to have sensors placed on aircraft quickly, an off-board periodic system of measurements was defined. Essentially, a very sensitive piece of ground equipment allows sensors permanently mounted on an aircraft to be monitored from a connector placed at a convenient inspection point. The system includes active noise dampening to remove many of the changes that can occur in a pneumatic system due to thermal and permeability changes of the air inside the tubes and of the tubes themselves and measures the conductivity (the inverse of impedance) of the sensor system in units of $m^3\,s^{-1}$.

The periodic system functionality is outlined in Figures 6–8. The sensor and connector are permanently installed on the aircraft, the sensor over a region where damage is likely to occur and the connector in a location with convenient access. The connector contains a passive microchip that is only activated when the instrumentation is connected to it. The microchip contains a serial number, and metadata including the aircraft (or other vehicle's) tail number, the sensor location, and a number of baselines required for the measurement.

When the instrumentation is plugged into the onboard connector, it automatically determines that it has been connected, downloads the data contained in the microchip, and begins the test procedure. In essence, the inspector simply needs to be able to turn



(a) (b)

Figure 5. Surface sensors for detecting BVID on composite structures (a) and a TTT sensor hole for detecting disbond in bonded composite structures (b).

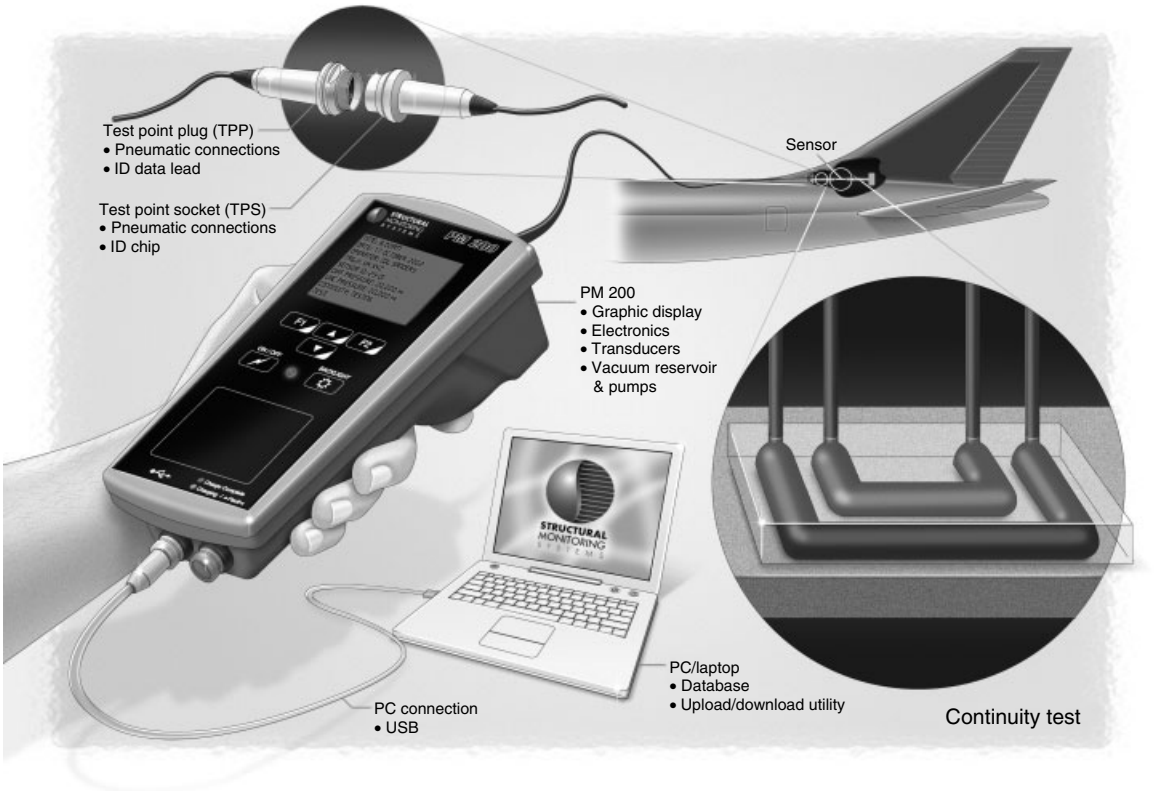


Figure 6. Continuity check of the periodic system—ensuring the sensors are not blocked.

on the instrument, select their name (pin protected), and be able to plug the system in; the instrument is otherwise fully automated.

The first test that the system performs is a "continuity" check; the vacuum is applied to one end of the sensor system and the other end is left open to atmosphere. If there is no blockage in the sensor system, air will flow through the system at a measurable rate. The actual rate is dependent on the impedance to flow of air by the sensor galleries and connecting tubing, the electrical equivalent being a number of resistors in series. The baseline value is determined at installation and a change in this value is indicative of potential system issue.

The second measurement closes the sensor system at the previously open atmosphere end, providing a sealed vacuum system (Figure 7). If the component is intact, the vacuum level will fall to the reference vacuum level obtained at installation, with a small flow of air through the system due to the permeability of the materials used.

The system is able to detect very small changes in the baseline conductivity value, which was determined at installation to allow very small cracks to be reliably detected (Figure 8). This has the required active noise dampening to remove changes in the conductivity measurement due to thermal noise (ideal gas law: $PV = nRT$ where P is the pressure (Pa), V is the volume in which the gas (air) is contained (m^3), n is the number of moles of the gas contained within the volume, R is the gas constant ($8.31441\,J\,mol^{-1}\,K^{-1}$), and T is the temperature (K)), and changes in permeability of the materials due to changes in environmental conditions. The instrumentation is able to discriminate real signals from the noise sources without operator involvement once the baselines have been determined as part of the installation process.

Figure 7. CVM™ check of the periodic system on an intact surface.

The requirement for such a high sensitivity is driven by the issue of crack closure in metallic structures. Small cracks can be closed to even the smallest flow of air due to residual stresses within the plastic zone and/or corrosion by-products if the crack is not being held open due to tensile loads in the structure. Crack closure delays the ability of the system to detect the crack, which has lead to the determination of $a_{90/95}$ probability of detection (POD) curves for the system for a number of metals (for example, see [6]). An $a_{90/95}$ POD curve is a measure of an inspection technique indicating the size of defect that will give at least 90% POD in 95 of 100 POD experiments under nominally identical conditions. The crack closure issue has not been observed in composite components; once the resin matrix has been damaged, the damage remains open with relatively large flow rates.

The instrumentation stores all the data from each measurement onboard. These data are able to be downloaded to a PC using an USB cable at the end of the shift into the system's data management software, or directly into an existing maintenance management database.

2.3.3 The CVM™ switch

The CVM™ Switch, shown in Figure 9, is the simplest of the instruments developed so far, and was originally designed to meet the requirements of the automotive test industry. The principle of the switch is identical to the rest of the CVM™ family of instruments, but the functionality has been kept to a bare minimum. The switch simply provides a factory preset alarm when a crack is detected. The alarm triggers a relay (which can operate in a latched or unlatched mode) to allow a fatigue machine to be turned off automatically and/or the time of detection to be recorded using an external data capture. The

Figure 8. CVM™ check of the periodic system on a cracked surface.

Figure 9. The CVM™ switch.

system has been designed to detect relatively large cracks (>5 mm).

A self-contained system with a CVM™ switch(es), small vacuum pump, general packet radio service (GPRS) transponder, and power supply has been adapted for monitoring the health of bridges and other major infrastructure.

2.3.4 Fully integrated CVM™ airborne monitoring system

SMS is currently working with a number of aircraft original equipment manufacturers (OEMs) to design a CVM™ system that can be fully integrated into their aircraft. The system is essentially a miniaturized laboratory system, with a regulated vacuum provided to measurement modules located close to the sensors. The closer the measurement modules can be to the sensor, the greater the amount of miniaturization

that can occur as the total volume to be monitored decreases proportionally.

The advantage of monitoring the airframe in real time, or at predetermined times during flight, is that the sensors can be monitored under ideal loading conditions, reducing or removing the effects of crack closure on metallic structures. The development program for the CVM™ airborne monitoring system will include the certification requirements for an aircraft system.

3 LABORATORY-BASED PROGRAMS

The CVM™ system has been proven in laboratory programs dating to the late 1990s [1]. The laboratory equipment has been designed for ease of use, to allow CVM™ sensors to be monitored in real time from basic coupon level tests to complex full-scale fatigue programs with Airbus and Embraer. The laboratory programs fall into three basic categories:

1. programs utilizing CVM™ to find damage in specific coupons or components;
2. programs to determine the functionality of the CVM™ system; and
3. programs to determine the robustness and durability of the CVM™ system.

3.1 Programs utilizing CVM™ systems

As CVM™ has gained acceptance as a technology, it has been utilized to detect cracks, delamination, disbond, or changes in base material properties in coupons and components in a variety of tests. Initially, many of these tests were also to determine the functionality of the system and its ability to withstand severe test conditions such as high vibration loadings.

The CVM™ sensors have been used to detect crack initiation in aircraft turbine blades [7], to detect crack initiation in Glare® lap joints [8] (Figure 10), and to monitor locations of interest in full-scale fatigue tests for both Airbus and Embraer, to name a few. The system has been used in laboratories in Australia, Europe, and United States and has allowed increases in productivity as the system can be safely allowed to run overnight; a sensor failure will switch the fatigue test off, as a disbond will cause a large amount of air to enter the system, triggering the alarm.

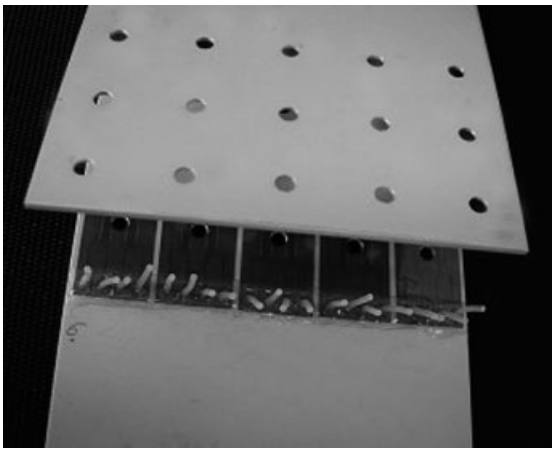


Figure 10. CVM™ integral sensors for detecting crack initiation within a Glare lap joint. [Reproduced from Ref. 8. © Airbus SAS, 2004.]

Sensors capable of being installed within lap joints to detect crack initiation from the bolt or rivet hole have also been developed [8]. The sensors were designed to detect sub-2-mm cracks required for the certification of the Glare® glass-fiber/aluminum laminate material used in the upper fuselage of the A380 aircraft (Figure 11). Standard inspection methods such as eddy-current and ultrasonic methods are not suited to such materials due to the number of reflection boundaries caused by the multiple aluminum layers. After a significant research and development program, sensors able to withstand the crushing forces below the titanium rivets used in the lap joints were developed and are now in regular use at Airbus in their test laboratories. Comprehensive testing by Airbus has shown that the sensors do not affect the shear strength, other mechanical properties or fatigue properties of the structures under test.

The CVM™ system has also been installed in programs to monitor full-scale fatigue tests (Figure 12), with large numbers of individual sensors (>100) being monitored in real time by a network of flow meters connected to a single computer.

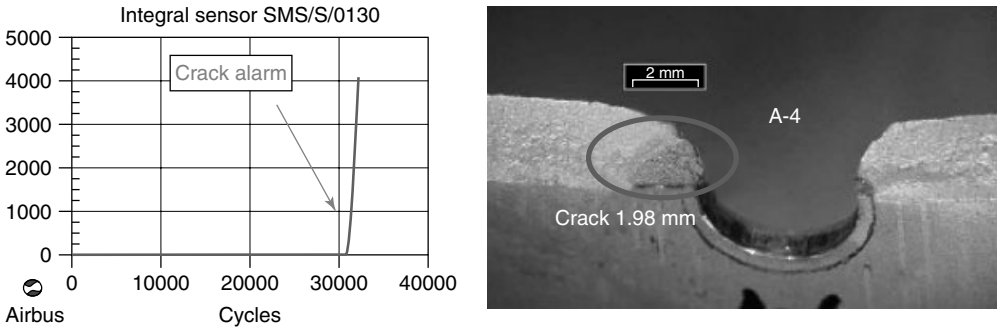


Figure 11. The output of the CVM™ system, and a 1.98-mm crack detected using the CVM™ integral sensors. [Reproduced from Ref. 8. © Airbus SAS, 2004.]

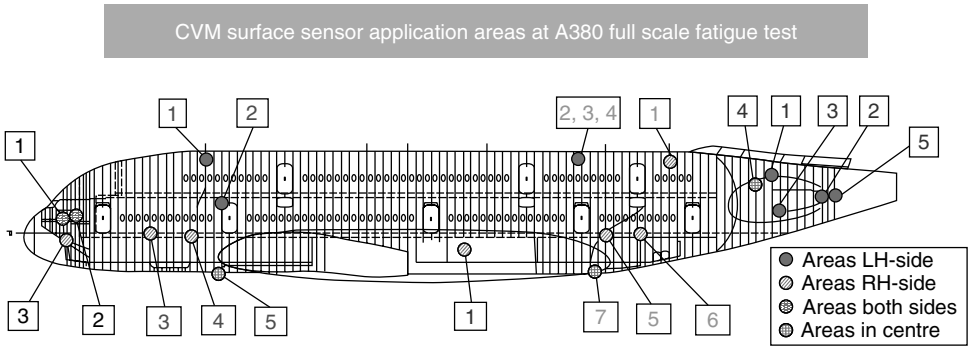


Figure 12. CVM™ sensor locations on the A380 full-scale fatigue test in Dresden, Germany. [Reproduced with permission from Ref. 8. *copyright* Airbus SAS, 2004.]

3.2 Programs to determine the functionality of the CVM™ system

A number of programs have been initiated to quantify the functionality of the CVM™ system. The real-time systems such as the laboratory equipment, the CVM™ switch and the airborne monitoring system do not suffer from the effects of crack closure in metallic structures. Many of the trials using these instruments have utilized the continuous monitoring characteristics to detect cracks in components under load and have continually demonstrated the repeatability of the system. Most of the programs remain proprietary; however [7–9] provide some useful examples.

The majority of recent functionality programs have focused on the development of a standard $a_{90/95}$ confidence curve for the CVM™ system when it is used in a periodic configuration. Crack closure will delay the detection of cracks under conditions of zero load or a compressive load. The effects are variable depending on the alloy used and the thickness of the component, but will result in the crack tip passing a small distance past the detection gallery before sufficient air can be drawn through the crack to measure the airflow with the instrumentation. Several studies have been undertaken at DSTO in Australia for the Australian Defence Force, Sandia National Laboratories for the Federal Aviation Administration (FAA) [6] and by Airbus to characterize the phenomena in a way that is easily recognizable within the aviation community.

Several test programs have also confirmed the ability of the CVM™ sensors to detect cracks in metals beneath a layer of paint. These programs have concluded that the CVM™ results are not affected if they are installed on bare metal, standard primer, and single layers of top coat. Hood and Nguyen [10] showed that thicker layers of top coat ($>60\,\mu$m) will cause a delay in the detection of the crack tip, but the measured value was less than 1.0 mm for a sensor installed on a 110-μm-thick paint system.

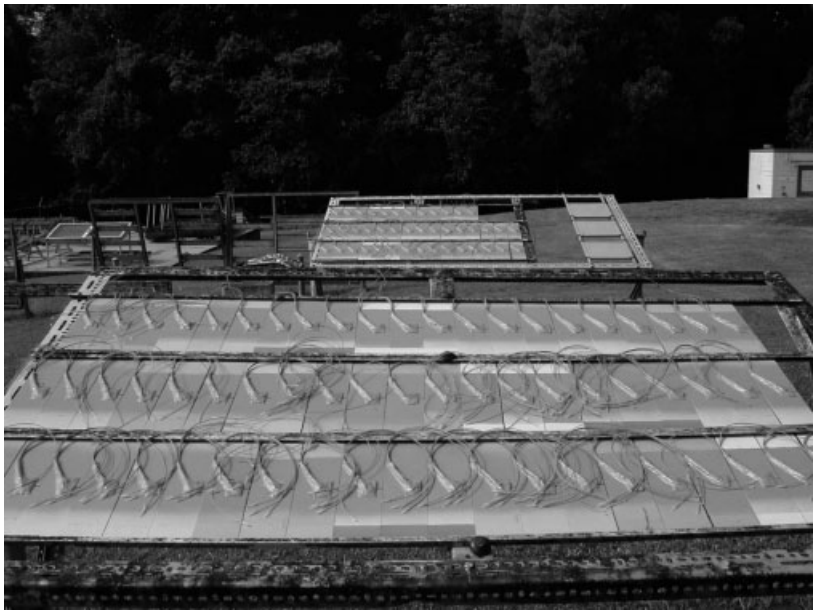


Figure 13. Test coupons experiencing hot/wet tropical conditions. [Courtesy of DSTO].

The measured delay was shorter if the paint system had aged.

The ability of CVM™ sensors to detect BVID, disbond, and delamination in composite structures has recently been explored through the SMIST and TATEM European Union Framework FP7 programs, and in projects in cooperation with the Cooperative Research Center for Advanced Composite Research (CRC-ACS) based in Melbourne, Australia [3, 11].

3.3 Programs to determine the robustness and durability of the CVM™ system

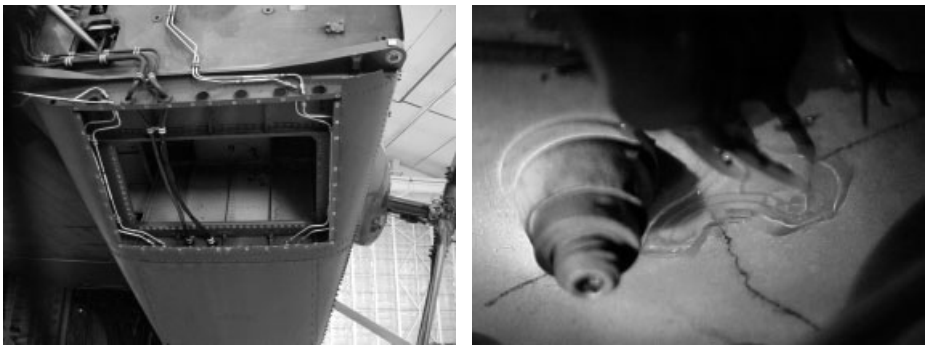
In order to ensure that CVM™ sensors were safe to be installed on aircraft and to meet future certification requirements, a number of laboratory programs have been undertaken to confirm that the CVM™ sensors meet the stringent requirements set by the aerospace community. These have usually included an in-service trial as well, which is discussed in the following section.

The first such trial program to determine if the sensors could remain functioning after exposure to aerospace conditions was conducted by the Australian DSTO [12]. Almost 250 sensors were put through a three-year program including cycling between -55 and $+100\,^{\circ}$C, immersion in hydraulic fluid and salt water, and exposure to high UV and humidity (tropical hot/wet) at a coastal environment (See Figure 13). All the coupons have been subjected to tensile loading to develop cracks detected by the sensors, to ensure full functionality of the CVM™ system at the conclusion of the exposure period.

Certification requirements have been developed through consultation with OEMs and regulatory authorities (military and civilian); the requirements for structural health monitoring (SHM) technologies to meet the requirements of the DO-160E [13] standard for environmental durability and robustness have become more evident. The various components of the CVM™ system have either undergone, or are undergoing, testing to show that they meet the various facets of the robustness and durability requirements as set out in the DO-160E standard.

4 AIRCRAFT PROGRAMS

CVM™ sensors were first installed on a US Navy H-53 helicopter in February 2002 as part of a trial program (Figure 14). The sensors were installed



(a) (b)

Figure 14. Crack location in the tail boom of the US Navy H53, with sensor installed.

in front of an existing crack in a location that required approximately 4 h to disassemble, inspect, and reassemble. This inspection was required every 25 flight hours on a large and heavily used fleet, but was reduced to approximately 5 min with the CVM™ system without any requirement to disassemble the aircraft. The system successfully detected crack growth on three separate occasions.

To date, only the periodic system has been installed on flying aircraft. The certification requirements for placing instrumentation related to the CVM™ system onto a flying aircraft are extensive, whereas only placing sensors and a small connector onto the aircraft and monitoring the system with ground equipment significantly reduces the requirements. This approach has allowed a large body of knowledge with regard to the sensor system to be developed through a number of military and civilian trial programs.

The CVM™ sensors have been installed and monitored on a number of military aircraft as indicated in Table 1.

Table 1. CVM™ military in-service trials

Country	Operator	Aircraft	Install date	Notes
US	Navy	H-53 Helicopter	February 2002	Successfully detected crack growth. One aircraft
Australia	RAAF	P3 Orion	September 2003	No cracks detected (or present) Functioned correctly for entire trial Three aircraft
Singapore	RSAF	A4	2004	One sensor installed in wheel well Functioned correctly for entire trial One aircraft
Singapore	RSAF	S-211	2004	Forty-two sensors installed throughout the aircraft Continues to function correctly One aircraft
Australia	Army	Blackhawk	April 2005	Two sensors installed per aircraft One-year trial completed successfully Three aircraft
UK	Navy	Sea King	June 2005	Eight sensors installed per aircraft Two-year trial completed successfully Three aircraft
UK	Air force	Nimrod	August 2005	Nine sensors installed on one aircraft Stress corrosion cracking detected if the fuselage is pressurized to 21 kPa (3 psi).

Table 2. CVM™ civilian in-service trials

Country	Operator	Aircraft	Install date	Notes
US	NWA	DC-9	2003	Six sensors installed in the fuel tank, and unpressurized empennage of the aircraft Trial ongoing
US	NWA	A320	2003	One aircraft
Germany	Airbus	A320	December 2006	Three sensors installed in wheel well One aircraft

As the civil aircraft OEMs have increased their interest in using the CVM™ system as a maintenance tool and, in the longer term, to potentially reduce the weight of their aircraft [14], their programs started to include installations on selected aircraft to gain experience on what is required to install the CVM™ sensors. Even on a trial basis, the paperwork to install the CVM™ sensor system can take several months and requires many of the tests outlined in DO-160E to be completed and documented. Table 2 shows a list of civilian in-service CVM™ sensor trials.

5 CERTIFICATION ISSUES

The case for SHM technologies in the aviation field has been well documented over the last decade or more and there now exist biannual SHM conferences in the United States, Europe, and Australasia. Airbus [14], Boeing [15], EADS [16], and the US Air Force [17] have all recently documented their broad requirements for SHM systems and the perceived advantages for maintenance of new and existing aircraft (*see* **Military Aircraft**; **Use of Leave-in-place Sensors and SHM Methods to Improve Assessments of Aging Structures**), as well as the possibility of introducing significant weight savings if SHM technologies can be implemented to improve structural efficiency (*see* **Design Principles for Aerospace Structures**; **Design Benefits in Aeronautics Resulting from SHM**; [14]). The primary driver in the short term appears to have improved maintenance regimes for both retrofit and new aircraft, with the goal of weight saving a medium to long-term benefit. An overview of the benefits for the use of SHM technologies in aerospace composite structures was well summarized by Scott *et al.* [18].

Although the need for SHM technologies has been demonstrated, the specific regulatory requirements for an SHM system remain vague and poorly defined. There does not currently exist a definitive list of requirements for certificating an SHM system or the organization that manufactures, installs, or supports it. The SHM system directly relates to the safety of the aircraft and as a result will need to comply with the most stringent regulatory requirements. An aerospace industry steering group (AISG-SHM) has been recently formed to address these needs and to define the requirements for SHM systems.

Prior to commencing any certification program, the airworthiness regulator must be closely involved. SHM technology maturity has, to date, been developmental; hence, regulatory involvement has been mostly by observation and professional interest. SMS is now moving from the developmental stage and has directly engaged with the Civil Aviation Safety Authority (CASA) to formally define the exact certification requirements for CVM™. There is a degree of interaction required on both the SHM-OEMs and the regulator, to ensure that both sides understand each other's requirements and a consistent certification plan is developed—one that supports the twin objectives of a business proposition on the part of the SHM-OEMs and safety on the part of the regulator. The earlier this relationship between the SHM-OEMs and the regulator is commenced, the more focused and hence more successful will be the development and execution of SHM technology.

5.1 Certification requirements

The fundamental principle of aviation certification is ensuring that aircraft and aircraft systems have

a uniformly acceptable level of safety. For civilian aircraft this level of safety is legislatively mandated around the world against a uniform worldwide standard. Military safety requirements are often either sourced directly, or closely aligned, with civilian practice. Safety, or inversely risk, is built around a total system philosophy not only controlling design requirements for aircraft and its components but also the human elements such as design, manufacture, operation, and maintenance. Consequently, to enter into the aviation industry, a new entrant must show the local aviation regulator that

1. the equipment being used is certified as meeting the appropriate airworthiness standards and
2. the entrant is certified as meeting the appropriate airworthiness standards for the types of operations being conducted.

It is only recently that papers have appeared considering the regulatory requirements for installing SHM systems on civilian and military aircraft. The papers have generally only considered the robustness and durability requirements of the system [19] using existing standards such as DO-160E. However, the certification of an SHM system requires the technology and its manufacturer to meet regulatory requirements in three distinct areas:

1. Functionality of the SHM technology
This is where it is shown that the intended purpose of the system does not adversely impact the inherent risk in the host aircraft's design over the period of installation (e.g., the SHM's crack-detection capability is acceptable to the aircraft's damage-tolerant design philosophy).

2. Fitment of the SHM technology to the aircraft structure
This is where it is shown that the presence of the system does not adversely impact the inherent level of risk in the host aircraft's design over the period of installation (e.g., the fitment of the system does not adversely affect the aircraft's weight and balance).

3. Certification of the organization
This is where it is shown that the SHM system designers, manufacturers, users, and maintainers can and do produce and maintain a product to an assured level of quality commensurate with the intended use over the period of installation. In the aviation arena, this is generally the ability of the SHM organization to design, manufacture, and support the SHM system to the level required by the airworthiness regulations.

Evidence would suggest that most SHM technology manufacturers have, to date, focused on part of the functionality of their systems without considering the equally important and necessary requirements outlined in points 2 and 3 above.

5.2 Certification of the equipment

5.2.1 Functionality of the SHM system—general requirements

In general, the certification path for any SHM system is built from a staged program consisting of three distinct steps. These steps are as follows:

1. Proving that the system works in detecting the intended target problems. This usually involves such activities as laboratory "POD" testing on simple test pieces.
2. Proving that the system works in detecting the intended target problems in the intended operating environment. This usually involves testing against environmental standards such as DO-160E [13] or MIL-STD-810F [20], "on wing" trials, and the like.
3. Proving that the system works in detecting the intended target problems in the intended operating environment for the life of installation. This involves reliability testing, system safety studies, and other similar activities.

It is proving the functionality of SHM that the understanding of the SHM concept of operations comes to the fore. An SHM system intended for use on a section of aircraft primary structure as a direct replacement for a mandated conventional nondestructive testing (NDT) technique, such as high-frequency eddy-current (HFEC) inspection, will have a much different functionality certification requirement set to an SHM system monitoring secondary structure or providing long-term research information on structural deterioration.

The relationship of SHM to standard "damage tolerance" or "safe-life" philosophies needs to be well

understood to ensure that the capability satisfies all the requirements of the different design philosophies. An interesting article by Easton (FAA) and Swift (CASA) [21], provides the most recent discussion by two regulators on how SHM technologies may fit into design and maintenance philosophies from a regulator's perspective. SHM systems are invariably compared to standard NDT techniques. This leads to a need to define their ability to detect cracks in terms of an $a_{90/95}$ POD. This may extend to requiring trials on varying thicknesses and geometries of these materials. Roach and Rackow [6] provide a summary of such a POD program evaluating three SHM techniques on a specific test coupon by Sandia Laboratory.

Combined with the (usual) SHM requirement for long term, no maintenance installation of SHM components, the random nature of system challenge (i.e., when a defect actually manifests relative to when inspections actually start), and SHM being used to support structural reliability requirements like FAR 25.1309 [22]; reliability determination is probably one of the most difficult certification challenges faced by any SHM system.

SMS is working toward certification of the CVM™ technology and the organization with the guidance of CASA. This decision was taken due to proximity of the Australian regulators to SMS' headquarters, CASA's familiarity with CVM™, and the existence of bilateral agreements between CASA, the FAA, European Aviation Safety Agency (EASA), and other regulatory agencies worldwide.

SMS has found the process of understanding and scoping the effort required to receive such organizational certifications especially enlightening. SMS has specifically employed engineers experienced in the aviation industry, who are used to operating within the aerospace organization and with relevant processes, because to outsiders its requirements can seem daunting, overly bureaucratic and complex.

6 OTHER (NONAEROSPACE) APPLICATIONS OF CVM™

CVM™ is not limited to the aerospace industry. Its sensors can be easily adapted to other industries where detection of surface-breaking flaws and damages is required. The CVM™ switch was specifically designed to meet the need of automotive testing and was launched in conjunction with a manufacturer of brake calipers to provide a significant improvement of efficiency in their quality assurance/quality control (QA/QC) test procedures. Figure 15 shows a fully instrumented caliper that is subjected to a battery of tests.

The CVM™ system is being used by a major automotive firm for track testing of their cars, providing real-time monitoring of crack initiation and growth

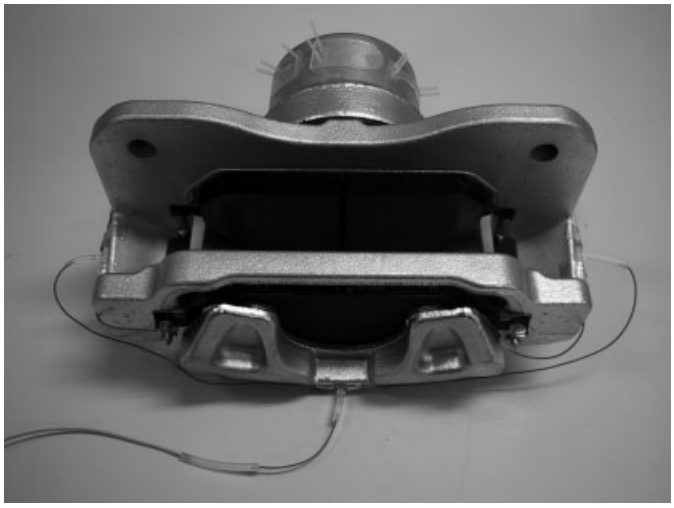


Figure 15. An instrumented brake caliper for improved QA/QC testing using CVM™.

previously unavailable to its test department. The system has been used to monitor failures of punched rivets in rail cars and could feasibly be adapted to remotely monitor structures such as bridges and ships once a business case has been identified.

7 CONCLUSION

This article provides an overview of CVM™, the vacuum-based monitoring technology originally invented by Ken Davey in 1994. The technology has been developed to monitor metallic and composite structures in the aviation industry. It has been used extensively in laboratory and full-scale fatigue programs by both military operators and civilian OEMs and has been flown in a number of in-service trial programs since 2002. The technology is close to meeting the certification requirements to allow it to be used in civil aircraft after an extensive test program including robustness and durability, functionality, and audit of the manufacturing process and facilities.

REFERENCES

[1] Sharp PK, Clark G. Evaluation of a novel NDE for surface monitoring using laboratory fatigue specimens. *Proceedings of ICAF*, Toulouse, 7–9 June 2001.

[2] Wishaw M, Barton DP. Comparative vacuum monitoring: a new method of in-situ, real-time crack detection and monitoring. *Proceedings of the AINDT Conference*, Brisbane, September 2001.

[3] Walker L. Real time structural health monitoring—is it really this simple? *Proceedings of the Sampe Conference*. Long Beach, CA, May 2004.

[4] White C, Orifici A, Bannister M. *Preliminary Assessment of Production Methods for Composite Compatible CVM Galleries*, CRC-ACS TM 08007, Cooperative Research Centre for Advanced Composite Structures, Melbourne, March 2008.

[5] http://www.smsystems.com.au/content/products/to_monitoring.asp, 2008.

[6] Roach D, Rackow K. Health monitoring of aircraft structures using distributed sensor systems. *Proceedings of the Aging Aircraft Conference*, Aging Aircraft 2006-Atlanta, Atlanta, 11–16 March 2006.

[7] Lacivita KJ. *Informal Evaluation of Vacuum Based Crack Detection Sensor*, Report No. AFRL/MLS 01–076. Air Force Research Laboratories, 28 September 2001.

[8] Stehmeier H, Speckmann H. Comparative vacuum monitoring (CVM™) of fatigue cracking in aircraft. *Proceedings of the 2nd European Workshop on Structural Health Monitoring*. Munich, 7–9 July 2004.

[9] Petitjean B, Simonet D, Choffy J-P, Barut S. SHM technology benchmark for damage detection. *Proceedings of the 2nd European Workshop on Structural Health Monitoring*, Munich, 7–9 July 2004.

[10] Hood R, Nguyen M. *CVM™ – Demonstration over Aircraft Finish Systems*, Report No. CR ADCO 2001-020-01 Rev 3, Royal Melbourne Institute of Technology, May 2002.

[11] Kousourakis A, Mouritz AP, Bannister MK. Compressive properties of polymer laminates containing internal sensor cavities. *Proceedings of the 3rd European Workshop on Structural Health Monitoring*. Granada, 5–7 July 2006.

[12] Loader C. *Durability of Comparative Vacuum Monitoring Sensor for ADF Applications—Commissioning Report*, Report No. DSTO-CC-Q1694/01. Defence Science and Technology Organisation, Melbourne, 2004.

[13] *RTCA/DO160E Environmental Conditions and Test Procedures for Airborne Equipment*. RTCA Paper No. 111-04/SC135-645, Washington, DC, 2005.

[14] Schmidt H-J, Telgkamp J, Schmidt-Brandecker B. Application of structural health monitoring to improve efficiency of aircraft structural structure. *Proceedings of the 2nd European Workshop on Structural Health Monitoring*, Munich, 7–9 July 2004.

[15] Trego A, Clark GJ. Structural health monitoring system: from collection to analysis. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford, Paulo Alto, CA, 7–9 July 2005.

[16] Buderath M. Review the process of integrating SHM systems into condition monitoring based maintenance as part of the structural integrity programme. *Proceedings of the 2nd European Workshop on Structural Health Monitoring*. Munich, 7–9 July 2004.

[17] Derriso MM, Olson SE. The future role of structural health monitoring for air vehicle applications. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford, Paulo Alto, CA, 7–9 July 2005.

[18] Scott M, Bannister M, Herszberg I, Li H, Thomson R. Structural health monitoring—the future of advanced composite structures. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford, Paulo Alto, CA, 7–9 July 2005.

[19] Kessler SS, Amaratunga K, Wardle BL. An assessment of durability requirements for aircraft structural health monitoring sensors. *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford, Paulo Alto, CA, 7–9 July 2005.

[20] MIL-STD-810F Department of Defense Standard for Environmental Engineering Considerations and Laboratory Tests, *United States Department of Defense*, January 2000 (original), November 2000, August 2002 and May 2003 (change Notices 1–3).

[21] Eastin R, Swift S. Rough diamond: two regulators review damage tolerance. *Proceedings of ICAF*. Hamburg, 7–9 June 2005.

[22] *FAR 25.1309 Equipment Systems and Installations, Federal Aviation Regulations, Part 25 Airworthiness Standards: Transport Category Airplanes, Subpart F: Equipment Amendment*. 25–41, Effective 9 January 1977.

Chapter 107

Development of an Active Smart Patch for Aircraft Repair

Nik Rajic

Defence Science and Technology Organisation (DSTO), Fisherman's Bend, VIC, Australia

1 INTRODUCTION

The strong imperative for weight reduction in aircraft design typically leads to highly optimized structures that often have an increased vulnerability to fatigue failure. The risk of in-flight failure, while real, is kept acceptably low through a structural management philosophy that is prescriptive, rigorous, and expensive. For aging aircraft especially, these costs are substantial and factor significantly in the decision to retire an aircraft. The ongoing costs associated with nondestructive inspection and preventative maintenance, in particular, have fueled strong interest recently in *in situ* structural health monitoring (SHM) technology and its potential to foster a more cost-efficient, condition-based approach to aircraft structural management.

Of the many sensor technologies that have emerged with potential for application in structural health monitoring, ultrasonic approaches (*see* **Ultrasonic Methods**) using piezoelectric materials (*see* **Piezoelectricity Principles and Materials** and/or **Integrated Sensor Durability and Reliability**) have arguably received the most widespread interest. Most of these are adaptations of standard inspection strategies used routinely in medical and structural ultrasound, and differ primarily in the use of ultrasonic source and receiver elements that are permanently attached to the structure. With appropriate powering and communication functionality, a network of such elements offers a potentially powerful and inexpensive basis for continuous autonomous monitoring of structural condition.

Compared to conventional ultrasonic inspection practice strategies employing a known, stable, and fixed source of ultrasound offer important diagnostic advantages. One such strategy is acousto-ultrasonics

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 Commonwealth of Australia. ISBN: 978-0-470-05822-0.

(AU), which is described in [1] as an inspection technique where ultrasound produced by a fixed source reminiscent of a spontaneous structurally sourced acoustic emission is sensed at a separate receiving location and used to identify material anomalies within the wave path. The term has since broadened from this early definition to encompass almost any technique involving the use of elastic waves in the ultrasonic regime transduced by elements fixed to or embedded in a host. The advantage of a stable ultrasonic source is twofold. It eliminates probe-coupling variations that can produce spurious response signals. But more fundamentally, it allows for the assessment of the acoustic response on a relative rather than an absolute basis. That is, a permanently attached source/sensor pair with stable transduction characteristics allows for the response to be compared to a baseline measurement corresponding to a structurally sound state (*see* **Signal Processing for Damage Detection**). The noise mitigation advantages of a differential measurement are well known.

Interest in this form of structural health monitoring strategy has grown rapidly in recent years, both within the scientific community, and amongst fleet operators. While key underpinning technologies are yet to reach commercial readiness, there is increasing confidence that this will occur. Indeed, major aircraft manufacturers like Boeing and Airbus have expressed the intention of incorporating SHM systems in next-generation commercial aircraft. Importantly, work has recently begun on the in-flight evaluation of prototype acoustic-based systems with technology demonstrations underway on both civilian [2] and military [3] platforms.

This article describes the application of *in situ* AU technology to an aircraft repair. Repairs are an important class of problem for SHM technology for two key reasons: (i) the technology provides a basis for assessing the performance of the repair system and in the case of bonded repairs can potentially address long-standing structural certification issues; and (ii) repairs are an ideal application to foster the transition of SHM technology into the field. The latter point is worth expanding. Aircraft repairs typically present a well-defined structural problem where the parameters required for the design and development of an SHM system are known. These include the failure mechanisms to be assessed by the diagnostic regime, the mechanical loading and environmental conditions experienced in service, and the target zones of diagnostic interest, which are often localized. Information of this type is critical to a systematic development and to the demonstration of a compelling business case for the technology.

The application considered here involves a composite bonded repair developed for a critical wing structure in the F-111C aircraft [4]. The development of an SHM *in situ* structural health monitoring capability for this repair is yet to be completed, but is sufficiently advanced to underscore both the inherent promise of the technology and the broader engineering and scientific challenges that still need to be resolved.

2 ACTIVE SMART PATCH REPAIR FOR THE F-111 LOWER WING SKIN

The F-111C is a high-performance strike reconnaissance aircraft that has been flown by the Royal Australian Air Force (RAAF) since 1973. With the help of the Australian Defence Science and Technology Organisation (DSTO), the RAAF has always maintained a high degree of self-reliance in airframe support, which has led to the development of several novel and cost-effective mitigation strategies for structural problems in the aircraft. One of the more serious cases involves fatigue cracking in a critical region of the lower wing skin (LWS). The cracking initiates at a stiffener depression (Figure 1),

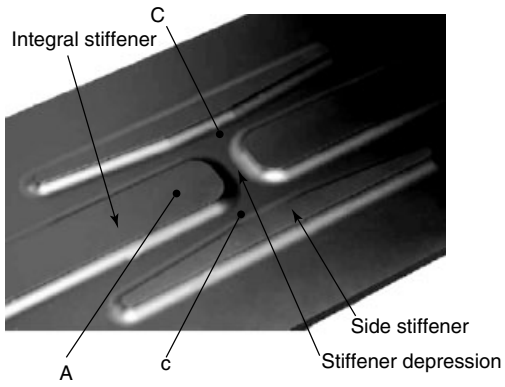


Figure 1. Structural detail in the FAS281.28 region of the F-111C lower wing skin (interior). Cracking initiates in the stiffener depression. Locations of interest labeled for reference.

conceived at the design stage to facilitate fuel flow and drainage between adjacent bays of the wing-box fuel tank. High local tensile-stresses were found to cause premature crack initiation and accelerated crack growth. If a repair could not be made, the only alternative would be to replace the wing—a very costly step.

A strong economic case made the development of a composite bonded repair for the damaged wing attractive. It was, however, an ambitious application since it involved a critical flaw in a primary or flight critical structure. Structural certification of the repair was consequently a key issue. Eventually, its compliance to strict airworthiness requirements was demonstrated through a stringent substantiation testing program [5]. However, the process proved costly and time-consuming, emphasizing that certification requirements would, in general, pose a serious impediment to the broader application of composite bonded repair (CBR) technology. This concern remains despite the long record of success of CBR technology in the field.

The certification issue led eventually to the concept of a "smart" composite repair. This envisioned a composite patch with an integrated network of diagnostic elements capable of furnishing a useful measure of the structural health and performance of the patch. Its development took several years, and, in 2006, culminated in a pioneering flight trial [6] on an aileron hinge in a RAAF F/A-18 aircraft. Structural health was inferred from strain relaxation measured in the patch using integrated strain sensors. Although effective in this application, where the patch was relatively small, the technology was understood to be generally useful only in circumstances where the location of deterioration was *a priori* known, since the strain sensors could then be placed accurately to ensure adequate coverage of the critical areas. This limitation led to the investigation of a more advanced form of smart repair called the *active smart patch* (ASP) which achieves broad-field coverage through the use of elastic waves produced by integrated piezoelectric elements.

Two diagnostic roles were envisioned for the piezoelectric network in the ASP: (i) to furnish information on the structural integrity of the bondline—the key issue for certification of composite bonded repairs, and (ii) to monitor crack growth in the parent structure. Although the first aim targets the certification issue directly, initial development of the ASP focused only on the cracking problem for two reasons. Firstly, it was thought to be a technically simpler problem partly because of the smaller area to be inspected. Secondly, despite its effect in profoundly reducing rates of crack growth, the patch does not alleviate the requirement for nondestructive inspection (NDI) of the damaged region, and indeed makes the inspection problem more difficult. Removing this requirement would both eradicate a large ongoing support cost and potentially improve the probability of crack detection by allowing a comparative basis for and removing human factors in the inspection process, important objectives in their own right. Diagnostic coverage for bondline degradation would be added as the next step in the development of the ASP.

This staged approach to the development of diagnostic functionality in the patch was deemed essential, given the limited resources available, as effort could then be made on concurrently addressing key engineering issues vital to a useful diagnostic capability, such as

- the mechanical and environmental durability of transducers;
- the structural impact of sensor embedment on the host;
- network powering and communication; and
- hardware reliability and robustness.

The ensuing discussion addresses these and other factors in relation to the development of the ASP for the F-111 LWS application.

2.1 Transducer development

One of the critical requirements of an ultrasonic transducer is that it should exert sufficient actuation authority to generate a measurable wavefield in the host. Piezoceramic materials are attractive in this respect, as is a relatively stiff mechanical coupling between the transducer and structure. This raises at least two structural implications. One is the impact of a stiff inclusion on the structural integrity of the host. Studies on laminates (see [7, 8]) have found that embedded sensors are largely benign to the host. Composite bonded repairs, however, are a separate class of problem as the elements may be embedded within the adhesive bondline, as in the particular example considered here. In general, it is

thought that the structural impact of an inclusion could be managed by ensuring that elements are placed within zones of damage tolerance and that the element dimensions are kept below a threshold size. The second point relates to the mechanical durability of a piezoceramic transducer under typical service loading. Since little, if any, scope will exist for replacing failed transducers embedded within a structure, transducer performance will, in general, need to be assured for the life of the platform. This is likely to mandate an extremely strict certification basis for the system.

Tensile strains at candidate transducer locations in the repair zone of the wing skin approach 2200 $\mu\varepsilon$ at the design limit load (DLL) [9]. At strain levels of this order, the fatigue resistance of piezoceramic materials is open to question. For example, Ref. 10 reports a deterioration in piezoelectric performance for elements embedded in a carbon–epoxy laminate loaded cyclically at a strain amplitude of 2000 $\mu\varepsilon$. Part of the deterioration in that study was attributed to failure of the adhesive bond between the piezoceramic element and the electrical interconnect, which is arguably the most vulnerable part of a wired piezoceramic transducer.

Notwithstanding concerns about the fatigue resistance of piezoceramic materials at high strain levels, which is a matter for further investigation, effort in developing a piezoceramic transducer for the ASP application focused chiefly on the creation of a robust electrical interconnect. A discourse on the development of the transducer is beyond the scope of this article. It suffices here to summarize the basic construction, which consists of a hard piezoceramic (Pz27) disc attached to an electrical interconnect layer comprising a polyimide film with conductive tracks.

Mechanical durability tests completed thus far have shown encouraging results. These tests involved bonding elements to a metal coupon that was loaded cyclically at constant strain amplitude. Electrical impedance spectra were measured prior to commencing the test to establish a baseline for the transducers, and then at periodic intervals during the test. Figure 2 shows the evolution of the impedance spectrum for a single typical element. It reveals largely stable behavior across the 25 million loading cycles despite strain amplitudes reaching 3000 $\mu\varepsilon$.

Closer inspection of the figure reveals some perturbation in the spectrum at around 20 million cycles. The precise nature of this change is clarified in Figure 3 which compares impedance spectra measured after 14 and 25 million load cycles to the baseline data. The main difference is a slight variation around the fundamental lateral and longitudinal resonances, at approximately 500 kHz and 3.5 MHz, respectively. However, the chief electromechanical indicators of failure in a piezoceramic material are a change in capacitive reactance and/or a frequency shift in the resonances, which are not represented in the data. The observed increase in strength of the lateral resonance suggests, instead, a possible

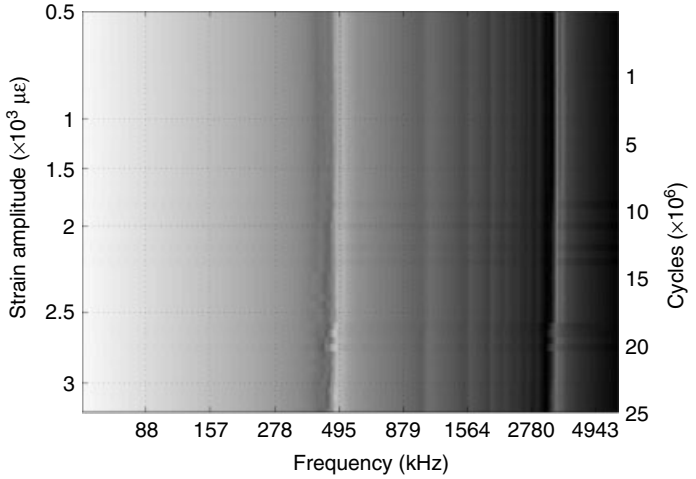


Figure 2. Evolution of the electrical impedance magnitude spectrum as a function of load exposure.

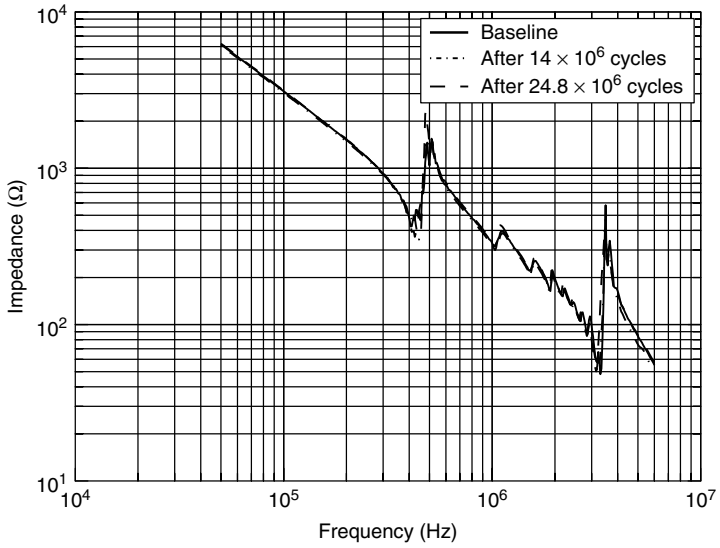


Figure 3. Electrical impedance magnitude spectrum.

deterioration of the bond between the element and the host; allowing the element to resonate more freely. Although a stable spectrum is a good result, it is important to stress that the electrical impedance spectrum provides only an indirect measure of the ultrasonic performance of a piezoceramic element. An in-depth study of fatigue resistance is currently underway that aims to measure directly the ultrasonic transduction efficiency of the elements as a function of exposure to cyclic loading.

Previous work [11, 12] has shown that factors like the transducer geometry and the properties of the adhesive bond can have an important bearing on the performance of structurally integrated piezoelectric elements. It was considered instructive therefore to examine the influence on transduction efficiency of the compliant electrical interconnect layer used in the transducer construction. To do this, an experiment was conducted, in which the strength of the elastic wavefield produced by the packaged transducer was compared to that produced by the bare constituent piezoceramic element. The elements were disc shaped, 10 mm in diameter, and were adhesively bonded to a 1-mm-thick sheet of aluminum alloy. The transfer efficiency was evaluated from the ratio of the amplitudes of the drive voltage and the out-of-plane plate velocity, measured 150 mm from the source using a laser vibrometer. The ratio is shown in Figure 4 as a function of drive frequency. With one exception, the transfer efficiency is lower for the layered element. Although undesirable, the decline is considered an acceptable trade off for the engineering benefit of a compliant and durable electrical interconnect.

2.2 Hardware

From a purely functional viewpoint, conventional laboratory instrumentation provides an adequate hardware platform for most AU inspection tasks and would do so for the current application as well. However, such arrangements are only partially effective in demonstrating the true practical potential of SHM technology. A set of ambitious but realistic design specifications were laid out for an ideal dedicated AU hardware platform, including performance, cost, and size objectives. A survey revealed that no off-the-shelf technology was available that could meet the specification, necessitating the manufacture of custom hardware. Local industry was engaged in a DSTO-led effort to develop a suitable system. The outcome was a novel device called the *AUSAM* (acousto-ultrasonic structural health monitoring array module), shown in Figure 5. The module contains two output and four input channels but is designed to operate synchronously with other units and so can accommodate larger transducer networks.

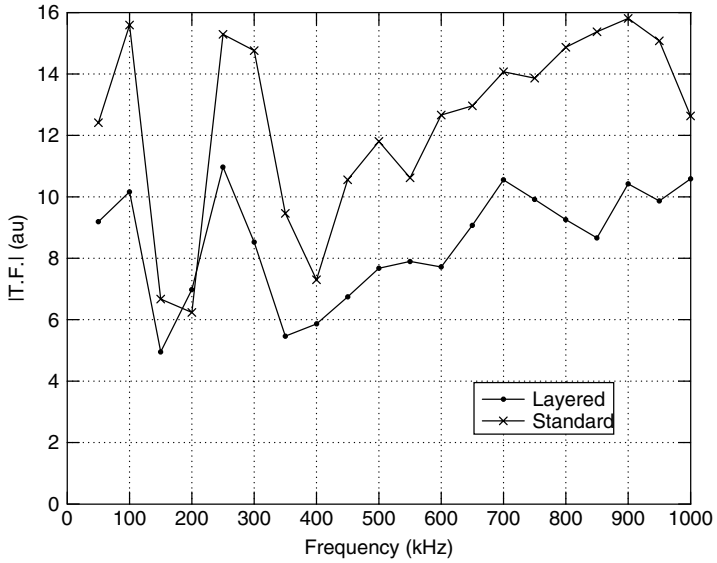


Figure 4. Measured electromechanical transfer function for a layered and bare piezoceramic disc of 10 mm diameter.

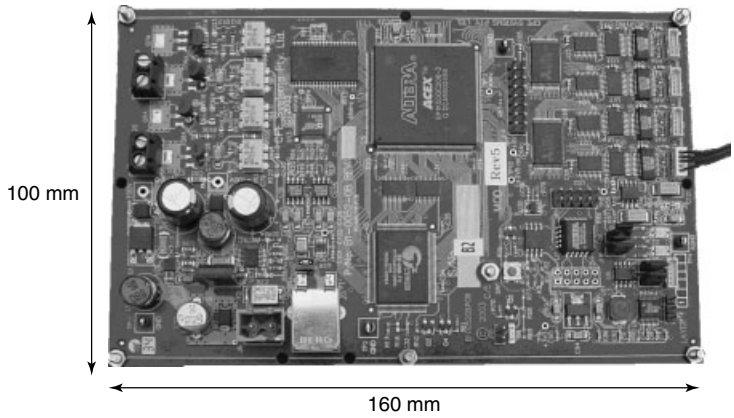


Figure 5. AUSAM device.

It links to a host (notebook or PC) through a USB connection, which also supplies all electrical power required by the device. The output channels produce a narrowband drive signal at frequencies up to 2 MHz. The drive amplitude varies with load capacitance, but is typically of the order of 100 V for a disc element 20 mm in diameter and 0.5 mm thick.

Proprietary software written for the AUSAM device allows for the autonomous control of drive and sense functions, with acoustic response data supplied periodically to the host through the USB. The transferred data is processed on line with real-time display of information.

3 FEASIBILITY TESTING OF A PROTOTYPE ASP

To address the initial diagnostic objective laid out for the ASP concept, a prototype patch was developed to investigate the feasibility of detecting simulated crack growth in a reinforced metallic coupon. The coupon comprised a 400 mm length of 6060-T5 aluminum

alloy bar stock 32 mm wide and 6.2 mm thick. The patch was produced from carbon-fiber prepreg in the layup $[+45, 0, -45, 90]_{2s}$, yielding a quasi-isotropic laminate approximately 2-mm thick. Transducers were bonded to the precured patch in depressions formed by dummy elements installed during the cure process. These were located symmetrically about the midpoint of the specimen, 170 mm apart.

The primary impetus for the development of wireless technology for sensors [13] comes from the practical challenges posed by wired installations. Problems include a susceptibility to electromagnetic interference (EMI), vulnerability to environmental and mechanical degradation, and the addition of structural mass (*see* **Energy Harvesting and Wireless Energy Transmission for SHM Sensor Nodes**). Notwithstanding the obvious benefits of a wireless approach, funding and time constraints dictated the use of conventional copper wiring in the current version of the ASP.

A key step in the design of a wired system is to identify an optimal path for the wiring through the patch. Two options were examined in this work: (i) along the bondline with an egress at the patch edge and (ii) through the patch with an egress from the top surface, avoiding the bondline entirely. While the through-ply access required for the latter option could potentially provide an ingress point for moisture, it was considered the better alternative as it would allow the entire transducer system to be contained within the damage-tolerant zone of the patch. Also, the shorter conductor length was considered beneficial in reducing vulnerability to EMI. Small holes were made in the precured patch to accommodate egress of the lead wires. Figure 6 shows an example of this wiring arrangement for an ASP installed on an aluminum coupon. In this case, the lead wires are fixed to a solder tab to allow for easy connection to instrumentation. Obviously, a less conspicuous and more robust termination would be required for in-flight applications.

The growth of a crack in the coupon was approximated by a notch, formed using an abrasive disc cutter 1 mm wide and 20 mm in diameter. Figure 7 compares, for a 600 kHz tone-burst excitation, the baseline receiver response to the response measured for a 2-mm notch depth. The pulse centered at $t = 0$ is EMI from the drive signal and provides a reference marker for the elastic waves, the first of which

Figure 6. Photograph showing egress of lead wires from a piezoelectric transducer embedded in the ASP. Wires are shown connected to a solder tab.

appears after a delay of about 35 μs, indicating a velocity of just under 5000 m s^{-1}. Reflections from the ends of the coupon are expected to arrive at the receiver after 80 μs, and were excluded from analysis by windowing the signal. The time period 60–80 μs was particularly instructive on the influence of the notch. The peak signal power deduced from a wavelet transform showed an increase of approximately 60% after the introduction of the notch. For comparison, the system noise floor accounts for a variation of about 2%. Figure 8 traces the evolution of signal power across the full range of notch depths. The growth in signal at short notch depths and the overall nonmonotonic trend are nonintuitive features that underscore the complex nature of wave–defect interaction. In light of this behavior, the development of a robust quantitative diagnostic assessment capability will require either a thorough understanding of the relevant wave dynamics or a careful calibration of the system. Both approaches involve some challenging problems. The case also highlights the value of a continuous monitoring regime in catering for system complexity. For instance, the behavior shown in Figure 8 would pose a difficult uniqueness problem under a conventional noncontinuous inspection strategy.

4 STRUCTURALLY DETAILED SPECIMEN

Compared to the idealized coupon in the previous example, the FASS281 structure constitutes a far

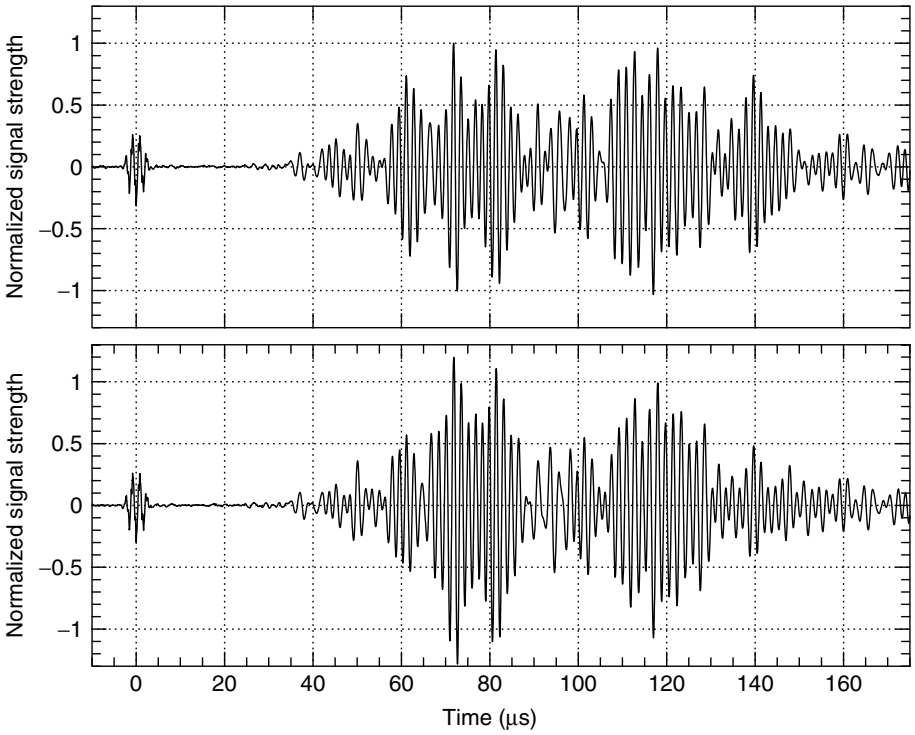


Figure 7. Receiver signal for a 600-kHz excitation. Upper trace shows the baseline response, while below is the response to a notch 2 mm deep.

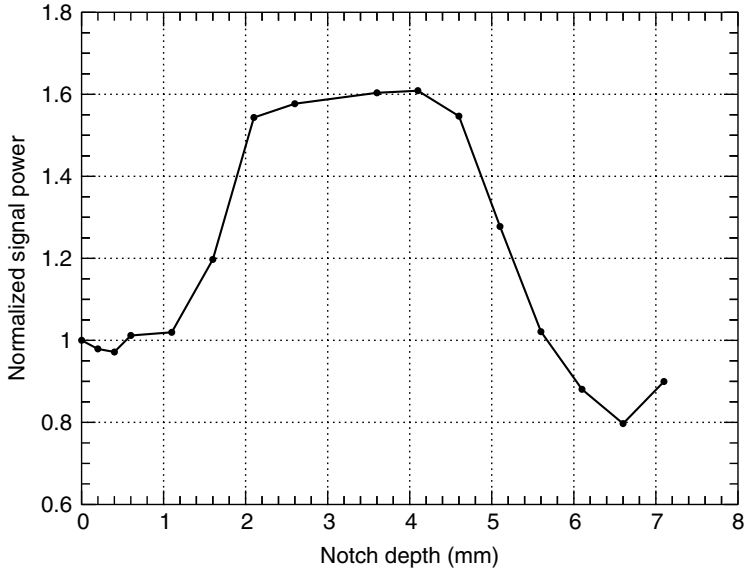


Figure 8. Variation in signal power as a function of notch depth for a 600-kHz drive frequency.

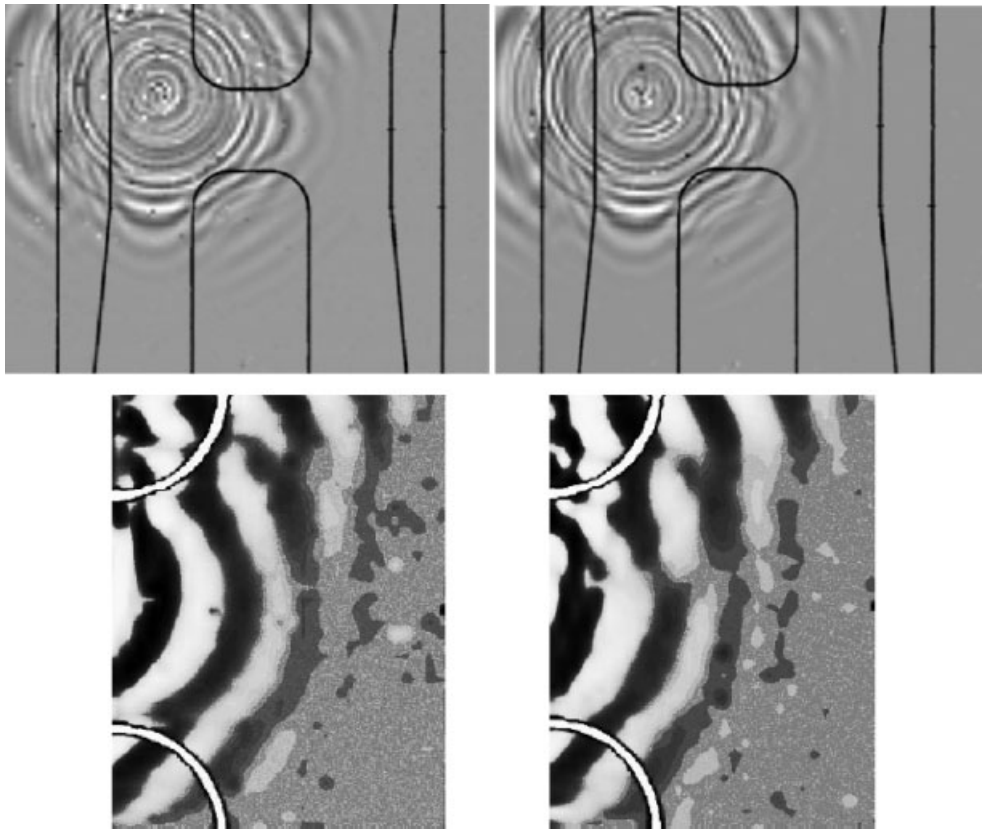


Figure 9. Velocity wavefield (a) in the absence of a notch and (b) with a notch. The bottom frames show zoomed and contrast-enhanced views of the region adjacent to the notch. The ultrasonic source frequency is 1.1 MHz.

more challenging problem, involving not only a far more complex geometry but other factors like crack closure and fuel ingress.

To help develop an understanding of the elastic wave dynamics in the LWS structure, laser- scanning vibrometry was applied to the outer (flat) surface of a structurally detailed sample for two candidate ultrasonic source locations: (i) on the main stiffener and (ii) adjacent to the depression where the panel thickness is approximately half that at the stiffener. These are respectively marked **A** and **C** in Figure 1. The panel was scanned twice, once in a baseline undamaged state and again after cutting a small notch in the depression to simulate the presence of a fatigue crack. The notch was semielliptical in profile, 2 mm deep, 1 mm wide, and with a length of 20 mm.

Of the two source locations, position **C** produced the stronger wavefield in the stiffener depression. The scan shown in Figure 9 corresponds to this source location where the excitation comprised a five-cycle narrowband tone-burst signal at a frequency of 1.1 MHz. The stiffeners produce strong scattering of the wavefield, which is fostered in part by the relatively high excitation frequency. The impact of the notch is not pronounced, but some scattering is evident in the enlarged and contrast-enhanced view of the region adjacent to the notch (Figure 9). An example of the perturbation in velocity response at an arbitrary point within this zone of scattering is given in Figure 10. The reduction in signal strength caused by the notch is more than 40%. Interestingly, the same analysis applied for the case of a 750-kHz drive frequency yields an increase in signal of about 30%. This marked difference in behavior reinforces the remarks made earlier about the complexity of the interaction of elastic waves with real structural

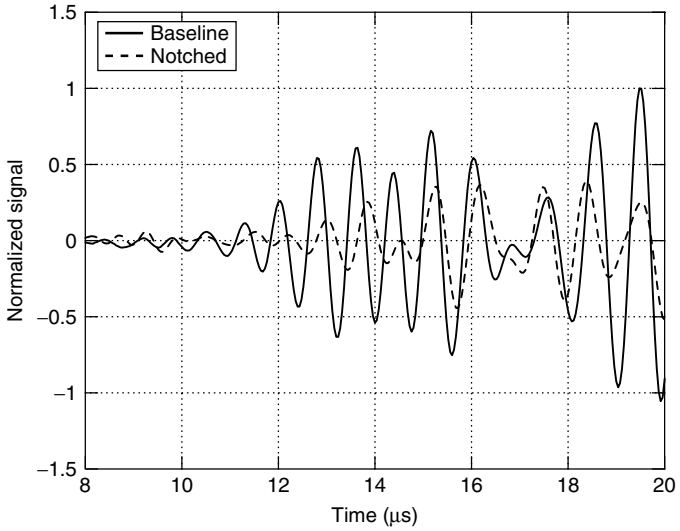


Figure 10. Perturbation in velocity response caused by the notch in the region of scattering shown in Figure 9.

features, and highlights the importance of sensor location for optimal sensitivity to crack growth.

Vibrometry scans are a useful resource for investigating candidate locations for sensor placement. In general, an optimal location will correspond to where the scattering of waves by a defect is at a maximum. The scattered field is obtained by subtracting from the measured field a baseline corresponding to the reference structural condition, which can be expressed as

$$\delta_p(i, j) = \frac{1}{\Delta} \sum_{k=k_a}^{k_b} (v^n_{i,j,k} - v^u_{i,j,k})^2 \qquad (1)$$

where v^n and v^u are the measured plate velocities for the notched and baseline conditions respectively, k is a time index, and the subscripts a and b relate to the time gate which has a width of $\Delta = (k_b - k_a)\,\mathrm{d}t$, where $\mathrm{d}t$ is the temporal sampling interval. An example of a scattered field calculated from measurements taken for a 1.1-MHz source frequency is shown in Figure 11. Three regions of apparent strong scattering are revealed; however, the two located in the left half of the image are artefacts caused by poor surface reflectivity, for example, at the coordinates (74 mm, 70 mm), where the piezoelectric source is attached. Genuine scattering is observed in the shadow zone of the notch. The elongated pattern is broadly intuitive, but is relatively narrow. Optimal placement of a sensor, in this case, would prove difficult without the guidance of a scattering map such as this. The distribution and strength of the scattered field varies significantly with the source frequency as this influences both the modal composition of and the dominant wavelengths in the elastic wavefield. Consequently, an optimal sensor location for one drive frequency is unlikely to be optimal for another. The scattering pattern will also vary with flaw geometry, further qualifying the optimality of any one sensor location.

By ignoring the presence of a repair patch on the studied panel, the analysis is only approximate. Since the wavefield in the repaired structure is expected to exhibit strong coupling between the adherends, the scattering in an unrepaired damaged panel is likely to be different from that in an equivalent patched panel. Consequently, measurement of the wavefield in the patch may not furnish a useful description of scattering from a flaw in the parent structure.

This technical limitation, as well as the laborious nature of the empirical approach provide a strong impetus for the development of a predictive elastic wave modeling facility that could provide guidance on optimal transducer placement. If modeling can account accurately for wave scattering caused by representative defects and other structural detail like thickness variations, adhesive layers, and composite reinforcements, the design process for integrated AU

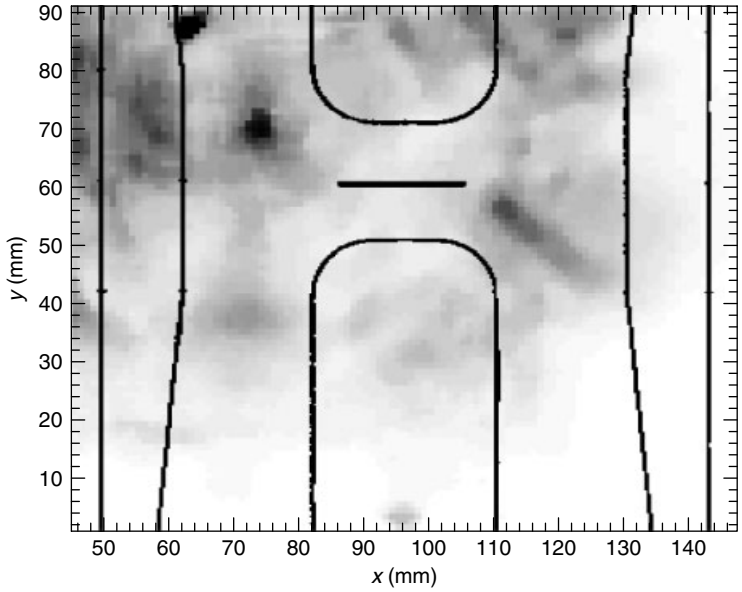


Figure 11. Scattered wavefield in LWS structure for a 1.1-MHz drive frequency.

inspection systems could be made far more efficient and would result in improved performance. With this aim in mind, finite element models of the LWS structure have been developed [14] and tested against experimental observations. The findings indicate that reconciling wavefield predictions with experimental data is difficult, and that more work on validation is needed. An example of an elastic wavefield prediction obtained from a finite element model of the LWS structure is shown in Figure 12.

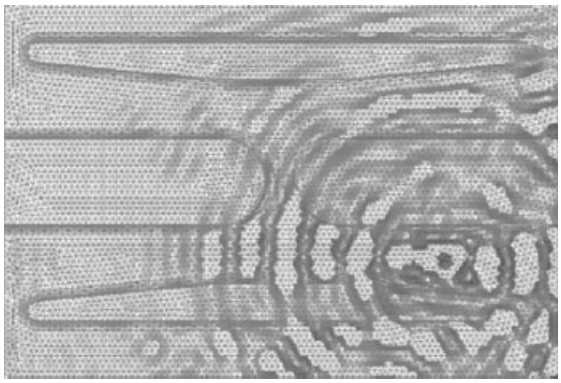


Figure 12. Finite element simulation of the displacement wavefield in the LWS structure for a circular source of ultrasound (200 kHz narrowband pulse).

5 SUMMARY AND FURTHER WORK

The development of the ASP for the F-111 wing skin application has brought into focus a range of challenging technical problems. One of these, and, arguably, the key impediment to the commercial acceptance of systems like the ASP, is transducer durability. The F-111 application is an ideal test case in this respect since it involves a realistic loading spectrum. A program is underway at the DSTO to systematically evaluate the performance of transducers developed for the ASP application under F-111 spectrum loading. Part of this effort involves an evaluation of load-mitigation strategies like the introduction of residual compressive stresses in the transducer.

Another important consideration is the potential impact of the embedded elements on the structural performance of the patch itself. While it was remarked earlier that restricting embedment to the damage-tolerant zones of the patch would likely minimize any deleterious influence, experimental work is needed to properly assess the impact of these inclusions.

One of the advantages of transducer embedment in the context of an in-flight application is the

protection it affords from the harsh environmental conditions encountered during flight, which include exposure to temperature extremes, moisture, and erosion. However, a system that interfaces with external instrumentation through a physical connection still has a point of vulnerability. Efforts are now underway to replace the wired interface of the ASP with a capacitively coupled configuration that eliminates the need for a connector. The approach is facilitated by the use of boron, an electrically nonconducting material, in the manufacture of the F-111 patch however, the approach is also adaptable to conductive materials like carbon–epoxy. Another obvious benefit of the approach is that it leaves the surface of the patch aerodynamically clean.

6 CONCLUSION

This article has given a short account of the application of AU to the development of an "ASP" repair for the F-111C, a high-performance strike reconnaissance aircraft flown by the RAAF. Laboratory testing has established that the approach is likely to yield sufficient diagnostic capability to resolve damage in the wing skin, and, in this sense, the program has produced a good outcome. However, a successful flight demonstration will require operation under both a relatively severe loading spectrum and harsh environmental conditions, issues that apply to almost all aircraft SHM systems. If the ASP can be shown to perform satisfactorily under these conditions, the scope of potential application for the technology is broad.

ACKNOWLEDGMENTS

The author acknowledges Mr Ian Powlesland for his instrumental role in creating the AUSAM device, Dr Sami Weinberg for developing the software interface, Dr Cedric Rosalie and Dr Kelly Tsoi for contributions to experimental work and data analysis, and Dr W. K. Chiu for his input to the numerical modeling.

REFERENCES

[1] Vary A. *The Acousto-Ultrasonic Approach*, Technical Memorandum 89843. NASA, 1987.

[2] Kearns J, Peña-Macias J, Criado-Abad A, Southward T, Evans D, Malkin M. Development and Flight Demonstration of a Piezoelectric Phased Array Damage Detection System. *Proceedings of the 6th International Workshop on Structural Health Monitoring*. Stanford University, Stanford, CA, 2005.

[3] Butkus LM, Guijt CB, Mazza JJ, Stargel DS. Selected U.S. Air Force Efforts in Bonded Repair of Aircraft Structures. *Proceedings of the 2006 SAMPE Symposium and Exhibition*. Society for the Advancement of Material and Process Engineering, Covina, CA, 2006.

[4] Baker AA. Bonded composite repair of metallic aircraft components: overview of Australian activities. *Proceedings of the 79th meeting of the AGARD Structures and Materials Panel*. Seville, October 1994, AGARD-CP-550; p. 1.1–1.14.

[5] Boykett R, Walker KF. *F-111C Lower Wing Skin Bonded Composite Repair Substantiation Testing*. Defence Science and Technology Organisation: Australia, Melbourne, DSTO-TR-0480, 1996.

[6] Galea SC, van der Velden S, Powlesland IG, Nguyen Q, Ferrarotto P, Konak M. Flight Demonstrator of a Self-Powered SHM System on a Composite Bonded Patch attached to an F/A-18 Aileron Hinge. *Proceedings of the First Asia-Pacific Workshop on Structural Health Monitoring (APWSHM 2006)*. Yokohama, 2006.

[7] Mall S. Integrity of graphite/epoxy laminate embedded with piezoelectric sensor/actuator under monotonic and fatigue loads. *Smart Materials and Structures* 2002 **11**:527–533.

[8] Paget CA, Levin K. Structural Integrity of Composite with Embedded Piezoelectric Ceramic Transducer. *Proceedings of SPIE, Smart Structures and Integrated Systems*. Newport Beach, CA, 1999.

[9] Callinan RJ, Sanderson S, Keeley D. *Finite Element Analysis of an F-111 Lower Wing Skin Fatigue Crack Problem*, Technical Note 0067. Defence Science and Technology Organisation: Australia, Melbourne, 1997.

[10] Paget CA, Levin K, Delebarre C. Actuation performance of embedded piezoceramic transducer in mechanically loaded composites. *Smart Materials and Structures* 2002 **11**:886–891.

[11] Liu T, Veidt M, Kitipornchai S. Modelling the input-output behaviour of piezoelectric structural health monitoring systems for composite plates. *Smart Materials and Structures* 2003 **12**:836–844.

[12] Rajic N. A numerical model for the piezoelectric transduction of stress waves. *Smart Materials and Structures* 2006 **15**:1151–1164.

[13] Galea SC, Powlesland IG, Moss SD, Konak M, van der Velden S, Stade B, Baker AA. Development of Structural Health Monitoring Systems for Composite Bonded Repairs on Aircraft Structures. *Proceedings of SPIE, Smart Structures and Integrated Systems*. Newport Beach, Vol. 4327, 2001.

[14] Wong CK, Chiu WK, Rajic N, Galea SC. Can stress waves be used for monitoring sub-surface defects in repaired structures. *Smart Materials and Structures* 2006 **76**:199–208.

Chapter 109
Fiber-optic Sensors

Peter Foote
BAE Systems, Advanced Technology Centre, Bristol, UK

1 INTRODUCTION

The rationale for structural health monitoring (SHM) of civil and military aircraft structures is being closely examined by aircraft manufacturers and those operators or owners who shoulder the life-cycle liabilities of these vehicles. Consensus is emerging that SHM offers potential advantages [1–3], but the choices and trade-offs involved in realizing those advantages are a subject of hot debate. Clear understanding of the "value proposition" of SHM must be achieved for individual manufacturers and operators before wide-scale adoption of the technology happens.

The trades inherent in this process range from high-level commercial and operational considerations down to choices in individual technologies and equipment. At the front end of SHM technology is the sensor or sensing principle that allows structures to be monitored in the manner required. The big ideas in SHM are that the structure can "self-inspect" and also record how it has been used and what environment it has been exposed to. The challenges for the sensor specialists, therefore, is to come up with devices that can detect accidental damage as it occurs and track its progress; detect environmental damage such as corrosion and ingress of moisture; monitor the stresses and strains during use as well as record the exposure to potentially harmful (e.g., corrosive) environments.

In this article, the role of optical fiber sensors for SHM is described. An overview of the main principles involved is provided followed by a comparison with other, nonoptical techniques highlighting respective advantages and disadvantages. Some examples of recent trials of fiber-optic sensing for SHM in aerospace are presented with an emphasis on actual aircraft installations. Finally, an opinion regarding future trends and needs for fiber-optic sensors is offered.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 WHAT ARE FIBER-OPTIC SENSORS?

Optical fibers were first developed in the 1970s, especially by Corning Inc., as a means of transmitting data over long distances without the need for electrical wires. Most readers will be familiar with fiber-optic cable installations in offices or as part of the spread of cable networks to homes for entertainment, Internet broadband, and phone links. However, almost as soon as optical fibers arrived, their exploitation as sensors had begun [4]. These sensors work by making use of the influence of light on the fiber by the surrounding conditions. The passage of light along the fiber can be affected by many factors such as temperature, strain, vibration, and chemical influences. The "pain" of the telecommunications technologist who would strive to avoid these is the "gain" of the sensors practitioner.

From a physicist's point of view, there are only really two properties in the fiber that can be changed by external influences, viz., optical path length and optical absorption. Nevertheless, a multitude of sensor techniques, which are relevant to most engineering disciplines and even medical applications, have emerged in the last three decades. Detailed descriptions of the operations and qualities of many of these techniques can be found elsewhere in this encyclopedia (*see* **Fiber-optic Sensor Principles**; **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**; **Fiber Bragg Grating Sensors**; **Novel Fiber-optic Sensors**).

3 AEROSPACE STRUCTURAL MONITORING

For aerospace structural monitoring applications, the requirements for sensors are driven by the same factors, irrespective of the sensors technology (i.e., electrical or optical); however, the optical approach offers some unique advantages as described in the next section.

For structural monitoring, there are three main categories of threat to structures that the sensor systems must address: fatigue, accidental damage, and environmental damage.

3.1 Fatigue

The measurement of strain at key points in an airframe allows the usage (or abuse) of the aircraft to be monitored [5] (*see also* **Loads Monitoring in Aerospace Structures**; **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**). Stress levels can be calculated from strain and on this basis, it can be determined how much "life" remains in the structure before it is in danger of developing fatigue damage (cracks). Aircraft in military service have used this approach for some years, using electrical strain gauges (*see* [6], **Agile Military Aircraft**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**). Fiber-optic sensors could also perform the task.

Unexpected changes in strain across a structure can also give warning that damage has taken place. For example, if a strut or stiffener has broken away, the load will be taken up by other parts of the structure leading to a change in the distribution of strain.

3.2 Accidental damage

Aircraft can be damaged through operational misuse (overload) but more commonly by accidental collision with objects such as birds and hail when airborne, and vehicles (such as baggage trucks) when on the ground. *Hangar rash*, a term used to describe damage while the aircraft is being moved, stored, or maintained on the ground, is also common. A major thrust in SHM is the development of damage-detection methods that perform the equivalent function of detailed inspection currently undertaken manually.

Sensing methods based on the interaction of sound with structural defects such as cracks in metals and delaminations in composite materials have been intensively studied (*see* **Ultrasonic Methods**; **Guided-wave Array Methods** for more details). Fiber-optic sensors can be made sensitive to ultrasound by acting as tiny microphones. Their susceptibility to sound is simply an extension of their susceptibility to pressure and strain. Acting in combination with a source of ultrasound [7], the optical fiber can pick up sound waves traveling in the structure. If

a crack occurs, the sound waves are scattered and this change can be detected by the sensors (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**; **Hybrid PZT/FBG Sensor System**; **Development of an Active Smart Patch for Aircraft Repair**).

Since the fibers can act as microphones, they are also able to "listen" to the structure. When metal or composite material cracks, it sends out pulses of sound called *acoustic emissions*. These may also be detected by the sensors [8] (*see also* **Acoustic Emission**; **Applications of Acoustic Emission for SHM: A Review**). Besides cracks, impacts caused by birds striking wing leading edges or engine cowlings, or baggage trucks colliding with cargo door surrounds could also be picked up [9].

3.3 Environmental damage

Corrosion is probably the biggest problem with aging aircraft structures. In the US Air Force alone, corrosion has been estimated to cost nearly $1.5 billion annually [10]. Although the aluminum alloy from which most airframes are constructed is carefully treated during manufacture, any corrosion prevention eventually breaks down. Sensors can be used to sniff the environment to judge how much exposure to corrosive conditions a structural item has seen (*see* **Environmental Monitoring of Aircraft**; **Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control**). Then, by using an understanding of the chemical and the physical processes of corrosion, an assessment can be made of the likely level of damage in the monitored area. Alternatively, the sensors themselves can be constructed to experience the same degree of corrosion as the surrounding structure by making them from similar materials [11]. As these sacrificial sensors rot away, their characteristics change in a precalibrated way, hence indicating the likely condition of the surrounding structure (*see* **Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control**).

Optical sensors can be designed both as environmental monitors [12] detecting parameters such as moisture and acidity/alkalinity (pH), and as sacrificial devices [13]. The sensor proposed in [13] works by encasing the sensor in a prestrained alloy package. As the package disintegrates through corrosion, the strain on the optical fiber is relieved, hence changing its light-transmission properties.

Composite materials using carbon or glass reinforcing fibers can also be affected by moisture, which can work its way into the material's layers through surface damage or manufacturing defects. As with environment sensors, optical fiber can be designed to sense moisture by using special coatings that react to moisture, in turn, affecting the light transmission of the fiber.

4 WHY USE OPTICAL FIBERS INSTEAD OF ELECTRICAL SENSORS?

Electrical sensors have been providing measurements for engineers since the nineteenth century. A wealth of experience and a bewildering array of devices and methods have built up during the 100 plus years since the birth of electrical engineering. Why is there a need to change to optics?

In many cases, no change is needed. Electronic sensors are constantly evolving (*see* **Nanoengineering of Sensory Materials**; **Miniaturized Sensors Employing Micro- and Nanotechnologies**) and many SHM methods push these sensors to the limits (e.g., acoustic emission detection sensors, *see* **Acoustic Emission**). However, the special features of fiber optics may prove irresistible in cases where they can clearly offer advantages of cost, weight, and measurement ability. In the highly safety conscious world of aerospace engineering, change is a costly and time-consuming business. So one can be sure that fiber-optic sensors will never be chosen in preference to tried and trusted electrical sensors unless they can offer at least one (or more likely all) of these advantages.

Here are some of the features that make fiber-optic sensors special (*see also* **Fiber-optic Sensor Principles**):

- **Immunity to electromagnetic interference**

Optical fibers are made from glass, a nonconductive dielectric material. It conducts no electrical current and is unaffected by radio or microwaves from nature (e.g., electrical storms and solar activity)

or from manmade sources. Light can travel along the fiber in electromagnetically "dirty" environments and be completely unaffected. Many aerospace environments, especially in the military, are saturated with radar and lower frequency electromagnetic interference. Electrical systems have to be carefully designed and shielded before they can function. This usually means added weight, cost, and complexity. Some environments are so bad that no electronic sensors can be used (e.g., power transformers and high-voltage switchgear systems). Fiber-optic sensors can clearly offer an advantage over electrical sensors in these conditions.

- **Low ignition hazard**

Having no electrical current flow, fiber-optic sensors present little or no hazard in combustible or highly inflammable zones such as fuel tanks and associated systems. Stringent standards for intrinsic safety must be met by electrical devices operating in such environments. Again, this adds to weight, cost, and complexity, which could be avoided if fiber-optic sensors are chosen for the job.

- **Light weight**

A typical optical fiber weighs about $30\,\mathrm{mg\,m^{-1}}$. A 1-mm-diameter copper conductor weighs nearly 80 times as much. Even when packaged, similar differences in weight between fiber optics and electrical cables are prevalent. Electrical cable often needs shielding from electromagnetic interference, which adds still further to weight. Fiber-optic cables and sensors need no such shielding. The advantages, especially for highly weight sensitive applications, such as aerospace, are obvious.

- **Small size**

The typical optical fiber is $125\,\mu\mathrm{m}$ in diameter, similar to a human hair. Many electrical sensors used for structural monitoring are bulkier than this. For example, the humble strain gauges used in current load monitoring systems aboard military jets, although based on thin film tens of microns thick, have surface areas of a few square millimeters. The vanishingly small size of optical fibers has led to their incorporation within the body of composite materials such as glass- and carbon-reinforced plastics [14]. In these cases, the optical sensor becomes part of the structure itself, realizing a "smart structure".

- **Multiple sensor capability**

This is sometimes known as *multiplexing capability*. This means that many sensors can be arranged along just a single strand of optical fiber. The most extreme examples of this are the so-called distributed sensors [15], where the entire length of a fiber-optic cable acts as an ensemble of thousands of sensor points. Temperature and strain can be measured at any point along a fiber that may stretch for thousands of meters around a structure. The measurement location along the fiber can be judged by accurate time measurement of the reflections from short light pulses as they travel along the fiber.

- **Reliability**

Optical fiber sensors, especially Bragg gratings are extremely resilient to fatigue. The silica from which they are made is a supercooled liquid and, provided it is adequately protected, does not fatigue readily. Electrical sensors are metallic and prone to fatigue just like the metal structures they are used to monitor. Sensors such as Bragg gratings also encode strain information as wavelength changes. This means that if the fiber is damaged or a connector degrades, provided there is still some light transmitted, a measurement can be made. This robust format is akin to digital signals in electronic systems. Electrical sensors such as strain gauges can give false readings if circuit voltages vary for any reason other than strain change.

Figure 1 illustrates the multisensor capability of fiber-optic Bragg grating strain sensors contrasted with electrical strain gauges.

If fiber-optic sensors are so good, why are they not being used more widely for SHM? Inevitably, they have their "downside". In addition to the natural conservatism found in safety-critical industries like aerospace, the optical sensor systems are often much more expensive than their electrical counterparts. The balance in cost per measurement point only tips in favor of fiber optics when large numbers of sensors are needed per system.

Connecting fiber optics together in dirty and dusty conditions is also a challenge. Alignment between fiber cores to a precision of less than $1\,\mu\mathrm{m}$ is called

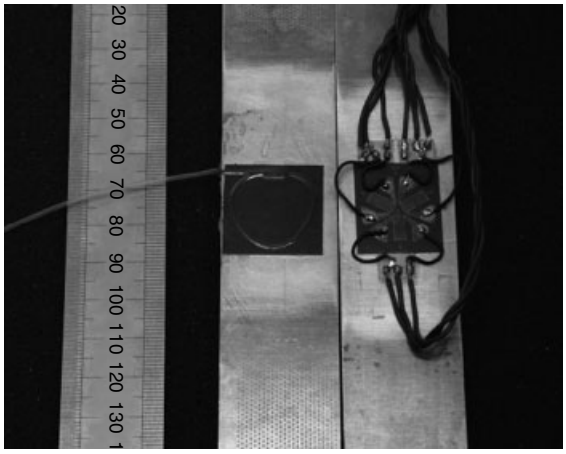


Figure 1. Comparison of a 3-axis, optical fiber strain gauge rosette (left) with an equivalent electrical strain gauge (right). Of note, the optical fiber device makes three independent strain measurements plus a temperature measurement using a single fiber connection. The electrical device requires nine connections and has no temperature measurement.

for and joining and repairing optical fibers needs specialist skills.

Much of the supporting hardware for optical sensor systems is in the form of specialized kits comprising semiconductor, laser light sources, and relatively delicate optoelectronic components. As yet, there are no industry standards or norms by which different suppliers' solutions can be compared. A discussion of specification standards and testing and application guidelines for optical fiber sensor systems is given in **Reliable Use of Fiber-optic Sensors**. The technology is still emerging, so the supplier base is patchy and typically consists of small start-up companies, often spun out of universities. This does not encourage long-term users/supplier partnerships.

However, the prospects for availability and cost of equipment are good, since most fiber-optic sensor technology draws heavily from the telecommunications business. The insatiable demand for bandwidth is encouraging the capacity of fiber networks to expand at a rate faster than Moore's law. The cost and availability of fiber-optic components has reduced by orders of magnitude over the last decade and the trend shows no sign of slowing. The commercial forces at work in the telecommunications industry will inevitably benefit the fiber-optic sensors industry.

5 SOME EXAMPLES OF FIBER-OPTIC SENSORS FOR STRUCTURAL HEALTH MONITORING

5.1 Strain sensing for operational loads and fatigue monitoring

Optical fiber strain sensors have been at the focus of many demonstrations and proposals for SHM. The hairlike nature of fibers has been exploited in trials of "smart structures" in which sensors are embedded into composite materials [14]. Behind this lies the desire to monitor loads in aerospace structures during operation to enable predictions of aircraft fatigue-life consumption.

Aircraft programs, especially in the military, are increasingly dependent on some form of operational loads monitoring to enable fleets to be managed in the most efficient way [5] (*see also* **Loads Monitoring in Aerospace Structures**; **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**). The usage of individual aircraft can be monitored allowing them to be deployed around fleets so as to even out airframe fatigue-life consumption and thereby maximize the usable life of the aircraft.

Two types of fiber-optic sensors have emerged as having the most potential for this use. They are extrinsic Fabry–Perot interferometer (EFPI) and fiber Bragg grating (FBG) sensors. The EFPI sensor is essentially an optical fiber that has been cut. The cut ends are then rejoined within a sleeve, leaving a small gap between them. The gap acts as an optical interferometer. A full description of these devices is given in **Intensity-, Interferometric-, and Scattering-based Optical-fiber Sensors**. FBGs are modified regions [16] (typically a few millimeters long) in an otherwise standard optical fiber. The regions are exposed to ultraviolet radiation patterns to form the FBG diffraction gratings. These reflect

light of different wavelengths depending on the state of strain or temperature in the fiber at their location. A full account of these is given in **Fiber Bragg Grating Sensors**.

5.2 Relative merits of fiber Bragg gratings (FBGs) and extrinsic Fabry—Perot interferometric (EFPI) sensors

FBG sensors are an intrinsic part of the optical fiber. They require no other components to form the sensor. Methods for producing them have advanced to the point that they can be created cheaply using premade fiber without disrupting the fiber in any way [17]. This means that the fiber retains all its intrinsic strength. The EFPI devices, on the other hand, need either some form of additional components such as a housing to contain the gap between fiber faces or they require the fiber to be cut and rejoined. These processes reduce their overall strength when compared with FBGs.

Many FBGs can be multiplexed in the same fiber [18]. The only limitation in this respect is the available spectral bandwidth of the light sources used to interrogate them and the wavelength range of the fiber. EFPI sensors are more difficult to multiplex since they are not inherently wavelength selective, at least in the manner that allows them to be easily distinguished from one another. Multiplexing can be achieved but with more complex methods [19].

Both types of sensors are susceptible to temperature changes as well as strain changes. Each can be used as temperature sensors in their own right, adding to the range of structural monitoring applications. More often though, temperature susceptibility is a problem [20]. Methods can be used to overcome this in strain measurement systems by mechanically isolating sensors within an installation, so that they experience temperature change but not strain. When combined with readings from nearby sensors subject to both temperature and strain, the temperature effects can be calibrated out. EFPI sensors that have evacuated or air-filled gaps are less sensitive to temperature than FBGs.

FBGs emerge as winners in comparison with most of the other interferometer style sensors. For this reason, there is a burgeoning market for sensor systems that use them and a growing number of applications of these in engineering.

5.3 Damage detection sensors

One early use of fiber optics for damage detection in composite materials relied on the fracture of the fiber when the structure was subjected to impact. The (visible) light in the fiber would bleed from the damaged fiber and could be seen at the point of impact through the translucent glass/resin of the structure [21, 22].

While limited in practical applications, these early demonstrations fired the interest in fiber sensors for SHM for aerospace applications.

Two major categories of fiber sensors for damage detection are those based on strain changes caused by damage and those used as acoustic sensors.

5.4 Strain-based damage detection

These methods rely on the redistribution of loads in a structure that occur when it is damaged. “Far-field” strain in a structure is usually only changed by large-scale damage. Damage detection requires measurement very near the area of damage. For this reason, methods using fiber-optic strain sensors are best suited to structural “hotspots” where high stresses are expected. Bonded or fastened joints are such regions where there are high local concentrations of stress. The linear nature of these structural features also suits the long-gauge measurement features of optical fibers very well.

In one such use of Bragg gratings, a fiber containing a series of gratings has been attached to a structure along a bonded stiffener made of carbon fiber composite material [23]. As the bond between the stiffener and skin breaks, the local stresses around the gratings distort. In this case, as the break reaches each grating, its reflection wavelength is blurred or distorted in a measurable way. This indicates that the crack face has reached that particular sensor and hence the onset and progress of damage can be tracked.

A similar technique uses the effect of transverse pressure on an optical fiber in which the polarization of light in the fiber is changed by that pressure [24].

As with the Bragg grating example given above, the sensor fiber is laid into a composite structure. When permanent deformation is induced in the structure by impact damage, the forces acting on the fiber change and locally affect the polarization of light in the fiber. Using a type of interferometer to analyze the light transmitted by the fiber, the location of this disruption along the length of the fiber and hence the position of the damage can be determined.

5.5 Acoustic sensors for damage detection

As mentioned earlier, damage in structures can be detected by using acoustic sensors either to "listen" for cracks or as receivers of ultrasound in conjunction with a suitable sound source. Optical fibers react to sound pressure fields in much the same way as they react to strain; only the effects are much smaller. Acoustic fields induce much smaller and higher frequency changes in the phase or intensity of light passing along the fibers. Sound pressure levels associated with damage-detection methods such as ultrasound and acoustic emission result in equivalent strain amplitudes in optical fiber of the order of picostrain (millionths of a microstrain). The techniques for strain measurement already described are not generally appropriate for such small effects, nevertheless, a range of optical methods have been developed for acoustic detection.

The changes in optical transmission in optical fibers caused by phase changes can be enhanced simply by increasing the length of the fiber in the sensor. By coiling optical fiber around spools, many meters of fiber can be contained in a small package. During the 1980s, there was much development of hydrophones using optical fibers designed to detect extremely faint, ocean-borne sounds. These devices were exceedingly sensitive and could detect sounds at great distances [25]. Hydrophones are designed for sound detection in the low-kilohertz frequencies. SHM sensors based on acoustic methods often operate at ultrasound frequencies (100 kHz to a few megahertz), nevertheless, they can use principles similar to those used in the hydrophone devices.

Ultrasound systems using sound generators are usually based on piezoelectric transducers and fiber-optic "microphones". For these systems, the fiber-optic sensors are often simply lengths of plain fiber, which form one arm of an interferometer [26]. Since the ultrasound signals are well-defined tone bursts of fixed frequency, the signals from the interferometer are easy to distinguish from background noise.

Acoustic emission requires greater sensitivity since the emissions from, for example, cracks in metals are generally broadband and of lower energy than the signals used in ultrasound systems, however, optical sensors that are capable of detecting these signals are under development [8]. Although more difficult to detect, acoustic emissions require no accompanying artificial source of ultrasound, which generally requires electrical power to function. Damage-detection methods using just acoustic emission can be performed using optical sensors alone.

6 TRIALS OF FIBER-OPTIC SENSORS FOR SHM IN AEROSPACE APPLICATIONS

6.1 Beginnings

As yet (at least to this author's knowledge), there are no SHM systems in service that use fiber-optic sensors either in civil or military aircraft. To date, the trials that have taken place have been experimental systems to evaluate performance in ground-based tests or in the flying environment. However, the abundance of research literature, including that of leading aerospace manufacturers, portrays a consistent and growing interest in the topic. Ironically, as the interest intensifies, less may be heard in public as the work becomes more proprietary and commercially sensitive.

The ease with which fiber-optic sensors can be embedded in composite materials and attached to structures captured the imagination of sensor researchers and aerospace engineers alike in the late 1980s. Large arrays of sensors using fibers weighing just a few milligrams can be achieved. This led to the notion of "smart structures" where sensors are embedded in structural materials such as glass and carbon-reinforced plastic at the point of manufacture and act as artificial nervous systems (see Figure 2 as an illustration of the smart structure concept). As long ago as 1989, Measures [22] provided early demonstration of how simple optical fibers could be

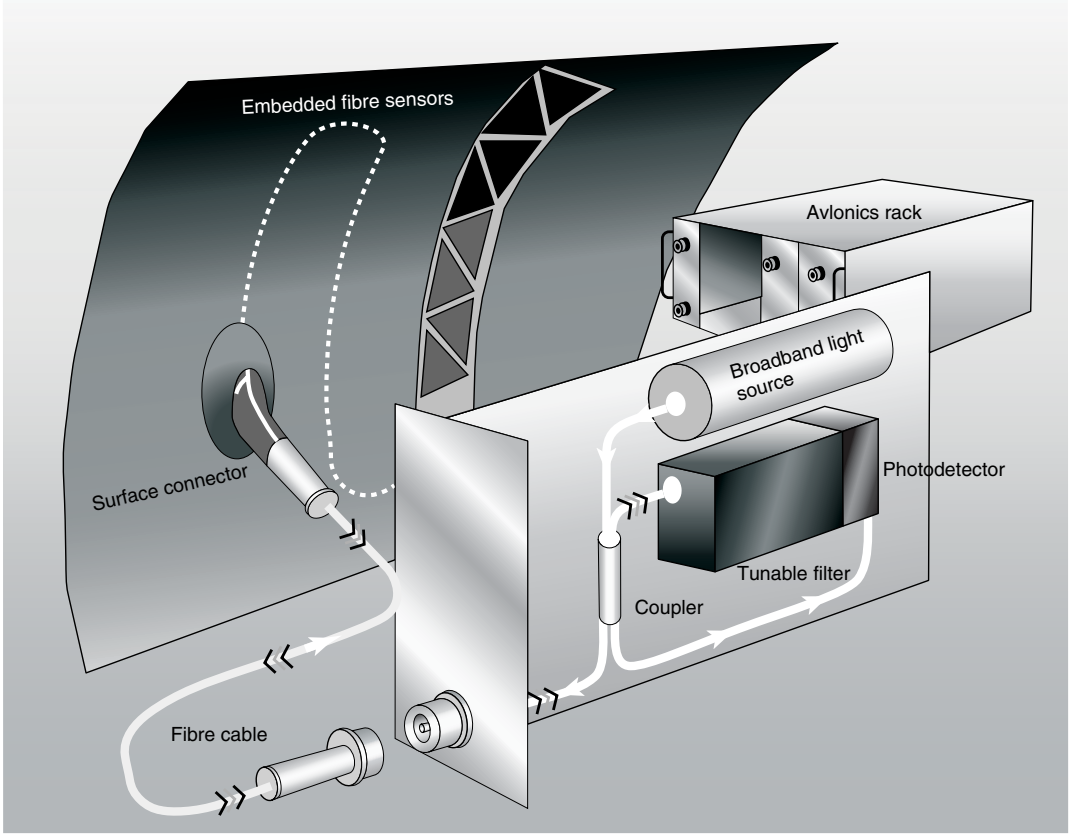


Figure 2. Smart structure concept for aerospace using structurally integrated, embedded optical fiber sensors.

used to detect the occurrence of damaging impacts in laminated glass fiber material of the sort used in aircraft construction. The sensors were modified optical fibers, buried within the material. When there were impacts of sufficient strength to damage the structure, the fiber fractured, spilling red laser light that could be seen through the translucent material in the region of damage. This was a very graphic illustration of structural damage detection using optical sensors and did much to spur on further development.

Udd [27] was also an early exponent of the smart-structure concept who, among others, proposed extending the structural monitoring principle to process control during the manufacturing stage. After the fiber was embedded, the optical sensors could detect the temperature, strain, and pressure conditions within composite materials during their manufacture and cure. This would invest such structures with "cradle-to-grave" self-monitoring capability.

Fürstenau *et al.* [28] claimed the first flight trials of fiber-optic strain gauges using a light aircraft as a flying test bed. A carbon fiber composite plate, on which optical fiber interferometric sensors were bonded, was attached to a Cessna C207A main wing spar. The strains in the plate were monitored during a series of test flights entailing varying flight maneuvers. The optical sensor performance was compared with electrical strain gauge data, thus demonstrating the ability of such sensors to operate in flight.

Slightly earlier, Murphy *et al.* [29] reported the installation and operation of a pair of EFPI optical strain gauges on a F-15, full airframe fatigue test. In this ground-based test, the sensors were bonded to the underside of the aircraft's wing, approximately 1 m from the root. The F-15's wings were deflected under static load to simulate up to 7g-loading during maneuver. Again, the optical sensor performance was

compared with conventional electrical strain gauges demonstrating the viability of the methods.

Demonstration and validation of new technology in the aerospace industry is costly, since even simple flight test equipment must still meet minimum certification and qualification requirements. The majority of aerospace trials with fiber-optic sensors are therefore ground based, but may still involve actual aircraft structures under varying conditions of flight simulation.

6.2 Developments from the 1990s to the present

Research and development activity on sensors for aerospace has slowly but consistently increased during the last two decades with active groups across the United States, Europe, Japan, and Australia. Many experiments report combined efforts from industrial and collaborative partners, which aids dissemination of results and also makes costs more manageable.

Significant development and demonstration of fiber-optic structural monitoring occurred during the NASA space plane studies for X-33 and X-38. Actual flight tests of fiber-optic strain sensing systems comprising both sensors and instrument units have been reported during the last decade. In one example [30], a combination of SHM technologies were flown aboard the NASA Systems Research F-18 aircraft. A number of Bragg grating sensor arrays were attached to this aircraft; these were positioned along the wing, including leading edge locations. Sensors were also placed in a purpose-built flight test structure attached to the underside of the aircraft. The system underwent a series of flight trials, which were part of the technology development for the X-33 reusable space plane project.

Ecke *et al.* [31] reported extensive development and testing of rugged FBG sensor installations and instrumentation for the NASA X-38 space plane program. The X-38 was NASA's prototype emergency crew return vehicle for the International Space Station. The vehicle was intended as a life-boat enabling crew to return to earth in the event of an emergency. The sensors would be required to withstand the harsh environment of space and the high loads, vibrations, and temperatures during reentry. A rugged instrument unit based on wavelength multiplexed FBG sensors was developed and partially qualified. A technique of packaging the sensors in a specially adapted organoceramic coating suitable for exposure in space environments was also developed.

Froggat *et al.* [18] demonstrated aerospace structural monitoring using FBGs in extraordinarily large numbers. Four optical fibers, each containing 500 FBG sensors were deployed on a large composite wing test. The researchers used FBG sensors with very low reflectivities and hence were able to multiplex them in large numbers along a single fiber. Each sensor could be interrogated by use of an optical frequency-domain method similar in principle to an electrical signal technique used to locate faults in long cables. Sensors of 5-mm length placed every 10 mm along the fiber were created and deployed on the test structure by bonding along structural features in the composite wing such as cutouts. A detailed map of strain distributions was achieved while the structure underwent static loading.

In Japan, extensive research into the use of optical fiber sensors in the aerospace industry is underway. Takeda [9] reports efforts in Japanese industry to advance composite structural damage detection and strain monitoring using FBG technology with small diameter optical fibers. The smaller, 40-μm-diameter fibers are more suited to embedment in composites, since they constitute a smaller inclusion than the standard 125-μm fibers and result in no loss of material strength due to their presence. Dense meshes of sensors have been implanted in composite specimens forming grids for damage detection and location. The close proximity of sensors in this arrangement allows the detection of damage by measurement of localized strain deformation. The same University of Tokyo team have also participated in flight tests of a reusable, vertical takeoff, and landing rocket vehicle. FBGs were attached to the liquid hydrogen tank and were monitored during ground test firings and also in flight. The flight test also included wireless sensor data telemetry.

Besides detection of damage and fatigue monitoring in aero structures, optical sensors can also play a role in structural repair monitoring. Metal-skinned aircraft are routinely repaired by blending out corrosion or drilling crack arrestor holes, for example. More complex repairs can involve building and fitting patches that are secured by fasteners. An alternative is to repair the damage by using

bonded patches. Qualification of this type of remedy presents problems because bonds may deteriorate, hence weakening the repair. This often restricts their use to secondary structure, hence limiting their usefulness. Repeated inspection is one way to overcome this difficulty, but this is labor and time intensive, especially if the repair is not easily accessible. Optical sensors can be used to monitor strain and hence load paths either within the body of a bonded repair patch or within the bond layer. In either case, if the bond between repair and structure fails, the load carried through the patch and its bond line will change and should be detectable using strain sensors. Davies [32] reports advances in low-cost monitoring techniques for such patches using FBG sensors. The technique uses a reference sensor bonded close to the patch but on the structure itself (Figure 3). The monitoring sensor is carefully matched (in terms of reflection wavelength) and bonded to the patch. When the repaired structure is put under load, a simple intensity-based optical system is able to detect relative changes in the strains between reference and sensor FBGs, hence indicating any deterioration in bond quality.

A challenging obstacle to the realization of fiber-optic smart structures is any standardized design and engineering guidelines for the fiber embedment processes. One aspect of sensor embedment frequently overlooked is the provision of appropriate connector technology. The point of egress, where the buried fiber emerges to the surface or edge of a composite panel leaves the fiber extremely vulnerable during the material fabrication process and during subsequent handling. Commercial fiber connectors, although very well developed for the telecommunications industry are not suited to the high temperatures, pressures, and contaminants encountered during composite manufacture. Specific component development is needed to meet these demands. Read [33] reported the development and testing of single mode, fiber-optic connectors designed to be compatible with the manufacture of aerospace grade, autoclave cured, and carbon fiber composite material. The connectors were preassembled with attached fibers and laid into the composite layers during manufacture. The components and tooling could be used by nonspecialist fabricators. Two styles of connector were trialled: a surface-emitting connector that allows fiber to be connected anywhere on a surface and side-emitting connectors that are accessed from panel edges (Figure 4).

The major commercial aircraft manufacturers, Boeing and Airbus, periodically report activities involving the use of fiber-optic sensors for SHM.

Airbus [34] have reported installation of FBG on aircraft wing and fuselage structures. An A340/600 test aircraft has been fitted with a loop of sensors around the rear fuselage, which has undergone loading during ground tests. The authors report that eight strain sensors and six temperature sensors (strain-isolated FBGs) were used. The optical sensors

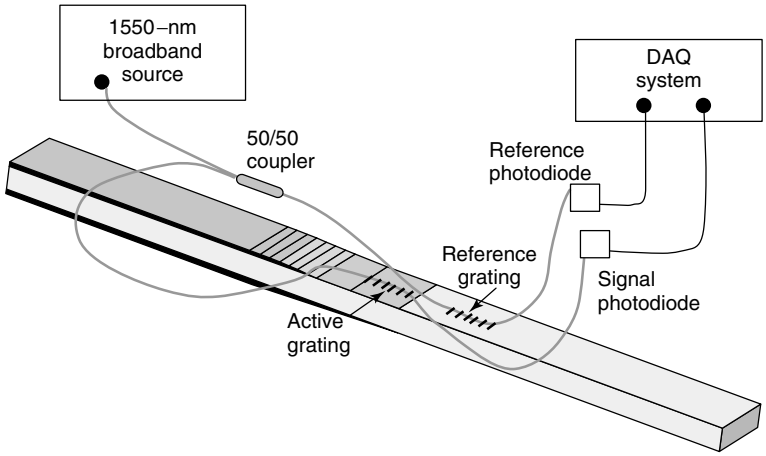


Figure 3. Composite-bonded repair patch concept using fiber Bragg grating sensors. [Reproduced from Ref. 33. © Stanford University, 2005.].

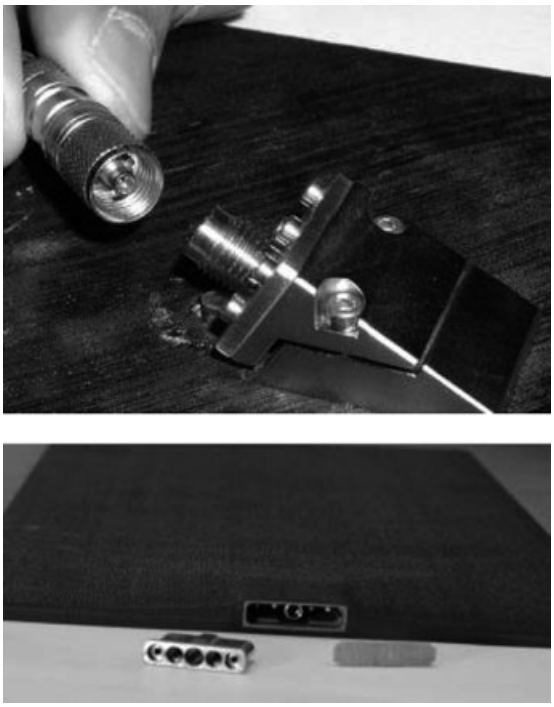


Figure 4. Optical connectors designed for fiber sensors embedded in high-performance composite material.

were bonded to the aircraft structure and the measurements produced compared with electrical strain gauge readings from comparable locations.

Boeing has reported aircraft installation of fiber-optic sensors for the detection of environmental damage in aging aircraft [12]. The fiber-optic sensors were designed to measure humidity, pressure, and temperature and were operated from a self-contained, battery-operated instrument unit. The installation was on board an aging Delta 767-300ER jet. The sensors were a mix of FBG-based devices and EFPI sensors, but the instrument unit was designed to operate with any sensor functioning on a spectral modulation principle (this includes most types of interferometric sensor as well as FBGs).

FBGs have also been evaluated in an A340/600 engine fan cowl test structure manufactured from carbon fiber reinforced plastic [23]. The sensors were embedded at the interface between the skin and a stiffener. When the structure was damaged, the gratings could be used to monitor the progression of disbond of a stringer from the main structure.

In similar trials, FBGs were attached to a large test structure comprising a fuselage barrel section based on the design of the Airbus A321 [35]. The section was made from carbon fiber composite material for the European Framework Project TANGO (Technology Application to the Near-term business Goals and Objectives of the aerospace industry). The test structure underwent static, fatigue, impact, and pressurization loading. The sensors were configured to detect impact on the structure and also to monitor any residual strain created by the impact event and thereby detect damage.

FBG strain sensors have been used to validate stress models on military aircraft airframe components within the Australian defense industry. An F/A-18 stabilator spindle (a component of one of the aircraft's control surfaces) was instrumented with both optical sensors and electrical stain gauges. The component was tested *in situ*, in ground-based airframe load tests [36].

Other flight trails were performed on a BAE Systems Jetstream 31 aircraft [37] in which patches containing FBG strain and temperature sensors were bonded to the aircraft's wing at a location on the lower leading edge. The sensors measured strain and temperature during a series of flight trials. More recently, a fiber-optic sensor system has also been flown in trials onboard a BAE Systems Hawk Jet in a flight test pod containing a range of SHM technologies (Figure 5).

A composite damage-detection system has been demonstrated for helicopter applications, this time using polarization effects in birefringent optical fiber [38]. This demonstration used polarization-maintaining fiber and a ground-based interferometer to check for permanent deformation in sandwich composite structure caused by impact damage. The optical fiber sensors were embedded in a rear deck cowling inspection door of the NH90 helicopter.

For damage detection using acoustic emission, highly sensitive transducers are needed (*see* **Acoustic Emission**). Currently, the best devices are piezoelectrical. If similar performance could be extracted from optical devices, then all of the advantages described earlier could apply to acoustic sensors as well. In an attempt to realize this performance, sensors have been built [39] and evaluated using an EFPI design. The sensor is probed

Figure 5. Recent flight tests of optical sensors aboard a Hawk military jet. A group of SHM systems, including fiber-optic strain gauges were contained in an underwing pod during flights.

by two lasers operating at differing wavelengths allowing the interferometric measurement to be decoupled from the effects of quasi-static strain. The technique is also immune to changes in temperature, since thermal effects occur on timescales much slower than the perturbation caused by ultrasound. The sensors were used to detect emissions from a barrel section structure made from aerospace carbon fiber composite after it had been damaged by an impact. After impact, when the structure was loaded in compression, the sensors detected acoustic emission from the damaged region. Similar optical sensors have also been reported [40], which use a specially made tapered and fused fiber coupler as the sensor element. Some development of techniques is still necessary to match the sensitivities of electrical sensors using these approaches.

Besides the NASA F/A-18 flight tests, a number of other programs aimed at future space vehicles have taken place. The European Space Agency Future Launcher Preparatory Program has evaluated the use of multifunctional FBG sensors to measure strain, temperature, and the presence of hydrogen [41].

EFPI sensors have also been evaluated in a program [42] targeting high-temperature measurement for reusable space-vehicle applications. The sensors were used to measure temperature and strain at up to 500 $^{\circ}$F (260 $^{\circ}$C) and were chosen because of their lower sensitivity to thermal effects and hence lower cross talk between temperature and strain.

7 CONCLUSIONS—FUTURE PROSPECTS FOR FIBER-OPTIC SENSORS FOR SHM IN AEROSPACE

The current picture of fiber-optic sensors for aerospace applications is one of unrealized potentials. Many of the unique attributes of optical fiber sensors (e.g., weight, size, and nonelectrical nature) described in this article are well suited to aerospace environments and yet they are still not in routine use. Part of the reason is the safety criticality and the ever-increasing cost sensitivity of the business. These factors often create hurdles to the introduction of new technologies such as SHM. Even though SHM has already been established in military applications, the more traditional electrical gauge devices are a first choice since they are the familiar technology.

For fiber-optic sensors to break through in this market, they must buy the manufacturer or operators a clear advantage in cost and performance. Inevitably, a "Catch-22" situation exists in which emerging business (typically small companies) will not invest in the expensive process of airworthiness certification

for their technologies until they can see a market large enough to provide a return on investment. Likewise, aircraft manufacturers will not select technology that has no proven track record or suitably qualified equipment from multiple sources. This implies a leap of faith or circumstances where the benefits of optical sensors are so compelling that manufacturers have no option but to commit.

The positive signs are that fiber-optic sensors companies are establishing markets in industries such as oil and gas, civil engineering, and wind power. This is leading to rapid maturation of the technology and consequent reduction in equipment costs, which will lower the hurdle to insertion in the aerospace business.

REFERENCES

[1] Boller C, Buderath M, Speckmann H. Measures for assessing structure-integrated damage monitoring systems in aircraft. In *Proceedings of the 1st European Workshop on Structural Health Monitoring*, Balageas D (ed). DEStech Publications: Chatillon, July 2002, pp. 853–860.

[2] Deririso M, Olson S, DeSimo M, Pratt D. Why are there so few fielded SHM systems in aerospace structures? In *Proceedings of the 6th International Workshop on Structural Health Monitoring*, Chang F-K (ed). Stanford University: Stanford, CA, 2007, pp. 44–55.

[3] Trego A, Akdeniz A, Haugse E. Structural health management technology on commercial airplanes. In *Proceedings of the 2nd European Workshop on Structural Health Monitoring*, Boller C, Staszewski W (eds). DEStech Publications: Munich, July 2004, pp. 317–323.

[4] Dakin J, Culshaw B (eds). *Optical Fibre Sensors I: Principles and Components*. Artech House, 1988.

[5] Hunt S, Hebden I. Validation of the Eurofighter Typhoon structural health and usage monitoring system. *Smart Materials and Structures* 2001 **10**:497–503.

[6] NATO Research and Technology Organization. Exploitation of structural loads/health data for reduced life cycle costs: NATO Research and Technology Organization document No: RTO-MP-7 AC/323(AVT)TP/4. *Papers Presented at the Specialists' Meeting of the RTO Applied Vehicle Technology Panel (AVT) Held in Brussels*, Belgium, May 1998.

[7] Betz D, Staszewski W, Thursby G, Culshaw B. Structural damage identification using multifunctional Bragg grating sensors: II. Damage detection results and analysis. *Smart Materials and Structures* 2006 **15**:1313–1322.

[8] Read I, Foote P, Murray S. Optical fibre acoustic emission sensor for damage detection in carbon fibre composite structures. *Measurement Science and Technology* 2002 **13**:N5–N9.

[9] Takeda N. Towards damage and structural health monitoring of aerospace composite structures using fibre optic sensors. In *Proceedings of the 3rd European Workshop on Structural Health Monitoring*, Guemes A (ed). DEStech publications: Granada, July 2006, pp. 34–45.

[10] Kinzie R. *2004 USAF Direct Costs of Corrosion*. Airforce Corrosion Prevention and Control Office, 2004.

[11] BAE Systems. *BAE Systems News Release: Advanced Sensing Equipment Offers Huge Savings to Military Budgets*, Farnborough, 18 June 2007; Ref. 184/2007.

[12] Elster J, Trego A, Catterall C, Averett J, Evans M, Jones M, Fielder B. Flight demonstration of fiber optic sensors, SPIE Smart structures and materials 2003. *Smart Sensor Technology and Measurement Systems* 2003 **5050**:34–42.

[13] Udd E, Haugse E, Trego A. *Fibre Optic Grating Corrosion and Chemical Sensor*. US Patent 6144026, 2000.

[14] Udd E (ed). Items. In *Fibre Optic Smart Structures*. John Wiley & Sons, 1995.

[15] Dakin J (ed). Distributed optical fibre sensor systems. In *Optical Fibre Sensors II: Systems and Applications*, Artech House, 1989, Vol. II, pp. 575–596.

[16] Bennion I, Williams J, Zhang L, Sugden K, Doran N. UV-written, in- fibre, Bragg gratings. *Optical and Quantum Electronics* 1996 **28**(2):93–135.

[17] Othonos A, Kalli K. *Fibre Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing*. Artech House, 1999.

[18] Froggatt M, Childers B, Moore J, Erdogan T. High density strain sensing using optical frequency domain reflectometry. In *14th International Conference on Optical Fibre Sensors*, Mignani A, Lefevre H (eds). CNR—Florence Research Area, Italy: Venice, October 2000, pp. 249–255.

[19] Kersey A. Multiplexing techniques for fibre optic sensors. In *Optical Fibre Sensors IV, Application,*

Analysis and Future Trends, Dakin J, Culshaw B (eds). Artech House, 1997, pp. 369–408.

[20] Vengsarkar AM, Michie WC, Jankovic L, Culshaw B, Claus RO. Fiber-optic dual-technique sensor for simultaneous measurement of strain and temperature. *Journal of Lightwave Technology* 1994 **12**:170–177.

[21] LeBlanc M, Measures R. Fibre optic damage assessment. In *Fibre optic smart structures*, Udd E (ed). John Wiley & Sons, 1995, pp. 581–613.

[22] Measures RM, Glossop NDW, Lymer J, LeBlanc M, West J, Dubois S, Tsaw W, Tennyson RC. Structurally integrated fiber optic damage assessment system for composite materials. *Applied Optics* 1989 **28**(13):2626–2633.

[23] Menendez J, Guemes A. SHM using fiber sensors in aerospace applications. *18th International Conference on Optical Fibre Sensors*, Cancun, MX, October 2006.

[24] Lloyd P. Structural health monitoring evaluation tests. In *Health Monitoring of Aerospace Structures*, Staszewski W, Boller C, Tomlinson G (eds). John Wiley & Sons, 2004, pp. 207–259.

[25] Dandridge A. Acoustic sensor development at NRL. *Proceedings of the Acoustic Society of America Annual Meeting*, Miami, FL, November 1987.

[26] Wade J, Zerwekh P, Claus R. Detection of acoustic emission in composites by optical fibre interferometry. *Proceedings of the IEEE Ultrasonics Symposium*, IEEE, 1981, pp. 849–853.

[27] Udd E. Fiber optic smart structure technology. In *Fiber Optic Smart Structures*, John Wiley & Sons, 1995.

[28] Fürstenau N, Janzen DD, Schmidt W. In flight strain measurement on structurally integrated composite plates using fibre optic interferometric strain gauges. *Smart Materials and Structures* 1995 **2**:147–156.

[29] Murphy KA, Gunther MF, Vengsarkar AM, Claus RO. Fabry-Perot fiber optic sensors in full scale fatigue testing on an F-15 aircraft. *Applied Optics* 1992 **31**(4):421–433.

[30] Schweikhard K, Richards W, Theisen J, Mouyos W, Garbos R. *Flight Demonstration of X-33 Vehicle Health Management System Components on the F/A-18 Systems Research Aircraft*, NASA Document NASA/TM-2001-209037, December 2001.

[31] Ecke W, Latka I, Willsch R, Reutlinger A, Graue R. Optical fibre grating strain network for X-38 spacecraft health monitoring. *Proceedings of the14th International Conference on Optical Fibre Sensors*. SPIE: Venice, October 2000, Vol. 4185.

[32] Davis C, Baker W, Moss S, Jones R, Galea S. In situ health monitoring of bonded composite repairs using a novel fibre Bragg grating sensing arrangement. *Proceedings of SPIE's International Symposium on Smart Materials, Nano-, and Micro-Smart Systems, Smart Materials II Conference*. SPIE: Melbourne, December 2002, Vol. 4934, pp. 140–149.

[33] Read I. Development and testing of connectors for optical fibres embedded into high strength composite materials. In *Proceedings of the 5th International Conference on Structural Health Monitoring*, Chang F-K (ed). Stanford University: Stanford, September 2005, pp. 1521–1529.

[34] Betz D, Trutzel M, Staudigel L, Schmuecker M, Huelsmann E, Czernay U, Muehlmann H, Muellet T. Fibre optic smart sensing of aviation structures. *Proceedings of the 3rd International Workshop on Structural health Monitoring*. Stanford, CA, 2001, pp. 306–315.

[35] Weis M, Hoflin J, Diemel P, Drechsler K. Evaluation of impact tests on the TANGO Barrel by means of fibre Bragg Grating Sensor measurements. In *Proceedings of the 3rd European Workshop on Structural Health Monitoring*, Guemes A (ed). DEStech publications: Granada, 2006, pp. 266–274.

[36] Davis C. *Strain Survey of an F/A-18 Stabilator Spindle Using High Density Bragg Grating Arrays*. Australian Government Department of Defence, Defence Science and Technology Organisation, Doc. Ref: DSTO-TN-0615, 2005.

[37] Read I, Foote P. Sea and flight trials of optical fibre Bragg grating strain sensing systems. *Smart Materials and Structures* 2001 **10**:1085–1094.

[38] Salomon J, Magnin P, Kauffmann C, Turpin M, Dumont B. Damage detection in sandwich structure by optical fibre sensors. In *1st European Workshop on Structural Health Monitoring*, Balageas D (ed). Paris, 2002, pp. 869–876.

[39] Read I, Foote P, Murray S. Optical fibre acoustic emission sensor for damage detection in carbon fibre composite structures. *Measurement Science and Technology* 2002 **13**(1):N5–N9.

[40] Doyle C, Porada S, Fernando G. Development of a new type of fiber optic sensor for the detection of acoustic emissions in composites. *Proceedings of the 1st European Workshop on Structural Health Monitoring*, Balageas D (ed). DEStech publications: Paris, 2002.

[41] Iannetti A, Ramusat G, Boggiato D, Francesconi D. An overview of the future launcher preparatory programme technology developments in structures health monitoring for the European next generation launcher. In *Proceedings of the 3rd European Workshop on Structural Health Monitoring*, Guemes A (ed). DEStech publications: Granada, July 2006, pp. 151–158.

[42] Richards L, Piazza A, Hudson L, Parker A, Carman G, Mitrovic M, Lee D, Steward A. Fibre optic sensor development for the SHM of reusable launch vehicles. In *Proceedings of the 3rd International Workshop on Structural Health Monitoring*, Chang F-K (ed). CRC Press: Stanford, CA, 2001, pp. 133–143.

Chapter 98
Agile Military Aircraft

Loris Molent[1] and Jason Agius[2]

[1] *Air Vehicles Division, Defence Science and Technology Organisation (DSTO), Melbourne, VIC, Australia*
[2] *Directorate General Technical Airworthiness, RAAF Williams, VIC, Australia*

1 INTRODUCTION

One application where the science of fatigue prediction reaches fruition is in the management of airframe structural fatigue. Fatigue management is now critical in aircraft operations, owing to the increased production costs of many newer models exerting pressure on operators to extract as much life out of their aircraft as possible. Furthermore, inspections, modifications, repair, and aircraft replacements are all expensive activities that are often a direct result of fatigue problems. Consequently, there is much incentive for operators to have efficient structural health monitoring (SHM) systems as part of structural integrity management programs in place. The primary component of the SHM system is the individual aircraft fatigue monitoring program (*see* **Fatigue Monitoring in Military Fixed-wing Aircraft**).

The fatigue management of an aircraft starts in the design process with the application of a design philosophy, stress spectra, material data, and a damage theory to estimate the fatigue life. This estimate is then certified through a structural fatigue test, following which (or sometimes before) the aircraft operator collects service loads data [1] and puts together a management policy [2]. The process of collecting service load data is termed *fatigue monitoring* and airworthiness regulations require all fighter-type aircraft to be fitted with an onboard usage monitoring or operational loads monitoring (OLM) system [3]. In addition to specific design requirements, the United States Air Force (USAF)

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 Commonwealth of Australia. ISBN: 978-0-470-05822-0.

institutionalized the requirements for aircraft structural integrity program (ASIP) management through the development and issue of MIL-STD-1530A in the mid 1970s.

Fatigue monitoring serves a number of purposes:

- Fulfill airworthiness requirements to ensure aircraft are not operated beyond an acceptable level of risk.
- Determine the fatigue life status of a fleet of aircraft throughout its life based on an operational spectrum.
- Determine the actual service load history (since many operators have found that operational usage of an aircraft is significantly more severe than the design spectrum) to ensure that aircraft are not operated beyond the fatigue damage accumulation threshold for various components as demonstrated through full-scale testing.
- Improve or optimize the structural integrity management of the fleet (when done in conjunction with a program based on tracking each aircraft in the fleet). The assertion here is that the utilization of each aircraft is different and that using an average value is inaccurate when monitoring the whole fleet.
- Detect occurrences of structural overloads in a timely fashion, thus enhancing fleet safety; and
- Assist in the definition of a flight load spectra for new aircraft of the same type.

This article presents a summary of Royal Australian Air Force (RAAF) fatigue monitoring philosophies, systems, fatigue models, and practices for agile fighter-type aircraft. Current processes are presented and comprehensively examined, and, where appropriate, the benefits and drawbacks of the respective methods are stated. The history of fatigue management is briefly presented as an introduction followed by an outline of usage programs currently used by operators. It examines the issues of strain gauge utilization and calibration, collection of flight parameter data, data integrity, data handling, comparisons with fatigue test results, and fatigue damage models. The article also includes a discussion on the problems that have arisen in the last decade due to high angle of attack capabilities and redundant structures of fighter aircraft (for more details, see [4–8]). Discussion in this article is delineated to fatigue monitoring of fixed wing fighter-type aircraft, where military transport type aircrafts are discussed in **Fatigue Monitoring in Military Fixed-wing Aircraft** and **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**.

2 HISTORY OF FATIGUE MANAGEMENT OF AGILE AIRCRAFT

Many Air Forces have experienced their share of fatigue problems. For example, in the RAAF, an MB326H suffered a wing fatigue failure that lead to the loss of that aircraft [9], while two Royal Air Force (RAF) Buccaneers [10] and an USAF F-111 experienced catastrophic fatigue failure [11]. The catastrophic wing failure of the F-111 occurred in 1969 after only 100 h of flying. More recently, wings on the RAF Hawks were replaced at about two-thirds of their design life [10].

The USAF experience in 1958 with B-47 fatigue failures [12] initiated the development of an ASIP [11], Air Force Regulation 80-13 in 1976 [13], incorporating damage tolerance requirements as per MIL-A-83444 [14]. The ASIP was intended to ensure that structural integrity is a consideration from design to throughout the service life of each new aircraft entering service with the USAF. This led to the mandatory utilization of usage monitoring systems.

In the early days of fatigue management of fighter aircraft, the only means of managing the fleet was through documenting the number of flight hours or landing cycles. When the aircraft reached a certified number of hours, they would be retired. Advances in the science of fatigue allowed the development of cycle counting methods [15] that related loads and stresses to fatigue damage. Subsequently, the peak-count method (of both maxima and minima) led to the concept of the fatigue meter. Fatigue meters (also called *g-meters*) compile a count of preset positive and negative g-levels exceedances during service. Consequently, low amplitude cycles that fall between two discrete levels are not counted. Fatigue meters (or counting accelerometers) and strain range counters were developed in 1952 [16–19] and received widespread use on UK military aircraft post-1954 [20, 21]. This method was extended to range-pair

or hysteresis loop counting, which considered both the amplitude and the mean of the load [22] and pairs turning points into cycles that relate to closed stress–strain hysteresis loops. The first range-pair counter was developed in Australia in the early 1970s [23–26].

Also velocity–normal acceleration (V–g) "slides" were used to generate gust statistics for aircraft fatigue design. In the early 1950s, velocity–normal acceleration and altitude (V–g–h) recorders came into use in the United States [27–29]. Later, swing-wing aircraft identified the need for more sophisticated recording systems than the V–g–h recorder [30].

"Scratch strain gauges" were developed in the early 1970s [31, 32] as self-contained mechanical extensometers capable of measuring and recording total deformation (and thus average strain) over the effective gauge length of the member to which it is attached.

Fatigue meters are still in widespread use with many aircraft types; however, they are being superseded by modern computers and recording systems. Direct derivation of stress using strain gauges or parametric equations or artificial neural networks [33, 34] has developed in more recent times. Recently, fiber-optic strain gauges have also been applied to fatigue monitoring [35], though, not at a military fighter aircraft level.

Current fatigue usage monitoring tools are summarized in Table 1, along with their advantages and disadvantages. Manufacturers continue to develop digital systems and sensors that record more flight parameters at higher frequencies than ever before and can be or are used in fatigue monitoring. The remainder of this article critically reviews these philosophies, tools, data processing procedures, damage models, and the interpretation of fatigue test results and their application to fleet management.

Military airworthiness authorities have long understood the potential detrimental impacts of fatigue to safety, availability, and cost of ownership and have hence taken steps through structural design standards to ensure that fatigue can be tracked and managed throughout the service life. The major military structural design standards are the UK MOD DEFSTAN 00–970 [3] and the US DOD JSSG-2006 [36]. These standards provide structural design and management requirements for SHM systems and are summarized as follows:

> **DEFSTAN 00–970**:
> 3.2.21: Every aircraft in the fleet shall be provided with instrumentation for the purpose of estimating the fatigue life consumption of fatigue critical structure and validating the assumptions made during substantiation. Provision shall be made for this instrumentation during production. For those components that are not individually monitored by an advanced direct strain measuring technique, a continuous or periodic Operational Loads Measurement (OLM) programme is a condition of compliance with these requirements and shall be agreed with the relevant Service Policy Authority.
> Fatigue Safe Life Substantiation—Leaflet 35: Section 6.1: It is emphasised that all aircraft must be fitted with basic instrumentation and that a continuous or periodic OLM program is a condition of compliance with the requirements of Clause 3.12. Section 6.2: According to whether most fatigue damage is accumulated towards the top or bottom of the S-N curve, a factor of 1.5 on life or 1.2 on stress must be applied to unmonitored structure.
>
> **JSSG-2006**:
> **A.3.15 Force management.** Force management will be applied to the airframe structure during operational use and maintenance of the air vehicle. A data acquisition system is required that collects, stores, and processes data which can be used to support the force management systems/program.
>
> **A.4.15 Force management.** Verification of 3.15 and subparagraphs shall be accomplished by analyses and tests to ascertain that all requirements are met.
> Analyses. Analyses which support the force management and maintenance concepts of the procuring activity are required to verify, for each fatigue critical location, that the individual aircraft tracking (IAT) methodology is updated and well correlated to full scale durability, damage tolerance, and flight load test results.
> Tests. Demonstration tests shall be performed to verify that the data acquisition system records and processes all required aircraft systems and flight parameters necessary for the IAT methodology.

These military design requirements protect the safety, availability, and cost of ownership of a platform. The discovery of unforeseen cracking through in-service failure or via inspection to the level where residual strength is compromised has a direct effect on the ability of an air vehicle to operate and hence

Table 1. Monitoring tools

Tools	Advantages	Disadvantages
Flight hour, flight/landing cycle counting	Minimum equipment needed Simple and cheap	Assumes each aircraft flies identical spectrum (no mission variability) Cycle counting is only applicable to landing and pressurized structure Additional conservatism required to compensate for lack of fidelity
Fatigue meter (N_z-based counting accelerometer—normally augmented by pilot recorded flight time, mission type and stores information. Weight is assumed to be constant for entire flight)	Simple and cheap Lightweight Robust Minimal postprocessing required	Relatively low accuracy Only components affected by N_z can be monitored N_z is the normal acceleration usually recorded at a fixed nominal center of gravity Difficult to validate data Difficult to account for missing data Asymmetric loads not considered Fixed N_z "trigger" levels Time history is lost, hence sequence effects cannot be accounted for Weight and point-in-the−sky must be assumed (conservative) Transfer function between N_z and stress at critical location required
Range-pair counters	Relatively cheap Some data processing conducted onboard	Time history lost PITS must be assumed Difficult to validate data Difficult to account for missing data Sensor calibration difficult due to data format
Multichannel recorders (parametric systems, neural networks)	Can monitor many flight parameters Time history retained Can be used for other investigations (incidents, overstressing) May record data from other sensors like strain gauges. Allows automation of health checks Can potentially be used to tailor flying operations to minimize damage	Large loads development program required (numerous flight conditions required and equation development is time intensive and intricate) Accuracy of loads estimated outside original data set is questionable Abrupt maneuvers, gust, and buffet loads may not be accurately accounted for Expensive and normally production interfaced with flight computer and related software Prone to data spikes and errors Software and postprocessing intensive Data validation needed
Strain gauges	Directly monitors principal load component (wing root bending moment) Responsive to abrupt maneuvers, gust, and buffet loads. Directly comparable to fatigue test Accounts for weight changes during flight Fiber Bragg grating sensors, *see* **Fiber Bragg Grating Sensors** [35]: Insensitive to electromagnetic interference Higher reliability than electrical resistance strain gauges High strain resolution Not prone to electronic drift	Difficult to determine gauge locations Gauge installation and maintenance is difficult Gauges require calibration Reliability of strain gauge and amplifiers can be poor Software and postprocessing intensive Electrical resistance strain gauges are sensitive to electromagnetic interference Fibre-optic gauges need further development

impacts capability. Appropriately designed and verified SHM systems are fundamental to the retention of capability, maximization of availability, and minimization of cost of ownership. As detailed in DEFSTAN Leaflet 35, to assure safety, the omission of an appropriate tracking system reduces the service life or inspection interval by a third. The availability and cost of ownership implications of not conducting fatigue tracking are generically presented as follows:

- **Availability**

For a transport fleet or fighter-bomber aircraft, the opportunity for structural inspections is only during deeper maintenance (DM) servicing. The time spent in DM can be estimated as approximately 20% of service life. Hence, for a fleet of 24 aircraft, 5 are in DM for structural inspections at any one time. If there is no fatigue tracking system for these aircraft, then the intervals are reduced by a factor of 1.5, increasing the time spent by each aircraft in DM to 30% of the service life meaning that seven aircraft are unavailable. From an aircraft availability perspective, the penalty for not tracking results is 10% of the fleet not being available for day-to-day operations.

- **Cost of ownership**

If the basic unit price for a fighter aircraft is assumed to be US$50 million and estimating that in-service costs are approximately twice that of the initial purchase cost, then, for a fleet of 100 aircraft operated over 30 years, the total cost of capability can be estimated as

$$3 \times \$50\,\text{million} \times 100 = \$15\,\text{billion}$$
$$\text{or } \$500\,\text{million per year} \qquad (1)$$

Now assume that this fleet of fighter aircraft is based on safe-life and no SHM system exists. Assuming no reduction in availability or capability then, either more aircraft would be needed or the fleet retired at an earlier date and another purchased. As such, the total cost of capability can be estimated as

$$3 \times \$50\,\text{million} \times 100 \times 1.5 = \$22.5\,\text{billion}$$
$$\text{or } \$750\,\text{million per year} \qquad (2)$$

On the basis of RAAF experience, the cost of tracking over 30 years can be estimated as follows:

$$\begin{aligned}&\text{Initial purchase} + \text{RAAF verification}\\ &\quad + \text{in-service cost} = \$1\,\text{million} \times 100\\ &\quad + \$0.5\,\text{million} \times 30 + \$5\,\text{million}\\ &= \$120\,\text{million} = \$4\,\text{million per year}\end{aligned} \qquad (3)$$

Hence, the cost of not tracking is approximately $250 million per year, compared to the cost of tracking of $4 million. While this example is simplistic, it indicates that the benefits of fatigue tracking far outweigh the cost by approximately one to two orders of magnitude.

3 FATIGUE MANAGEMENT PHILOSOPHIES

Operators do not follow one standard method of fatigue management as no detailed specifications exist. Design philosophies that feed into fatigue management programs are varied, fatigue test results are interpreted in different ways and different scatter factors are applied to the fatigue test spectra and fatigue test result. Operators continue to "experiment" with a number of fatigue monitoring tools as the technology changes rapidly. Some collect raw data while others process the data on board the aircraft. Others calibrate the data and the fatigue damage model to determine the crack lengths or fatigue indices and few operators use the same fatigue damage model [4].

An objective of fatigue or structural integrity management is to ensure that the life of the type of an aircraft at least meets the operator's planned withdrawal date [9], under normal operating loads and within approved flight limitations without collapse or unacceptable deformation. The philosophy to be followed to achieve this depends in part on a number of factors such as the ability to inspect and repair or replace the component and the result of complete failure of a component. The fatigue management process starts with a design philosophy that incorporates these factors.

As an example, RAAF aircraft structural integrity (ASI) management incorporates a combination of safe-life and damage tolerance philosophies for the

various aircraft as described in each aircraft's aircraft structural integrity management plan (ASIMP), e.g., [37]. In the case of the F/A-18, a safe-life philosophy is used, with a safety-by-inspection approach applied to extend the service life of inspectable structure in the aft fuselage structure. The F-111, which began service in Australia in 1973, was initially managed on a safe-life basis, but later, a safety-by-inspection approach was justified through analytical calculations, a durability and damage tolerance analysis, and proof load testing [9]. (The safety-by-inspection philosophy is equivalent to a damage tolerance philosophy.)

The aircraft design philosophy, however, is but one aspect of the overall fatigue management process. The fatigue management process should also

- consider that fleet aircraft cannot be operated beyond the factored equivalent damage accrual demonstrated in a fatigue test and any life extension must be substantiated by further fatigue tests to determine the next critical location (appropriate repairs followed by testing to failure is required);
- seek to manage fleet structural integrity based on fatigue test results;
- incorporate a load monitoring program on each aircraft to routinely measure load cycles in primary structure (as opposed to "hot spots" [8]);
- employ an economic and reliable fatigue monitoring system;
- ensure data integrity;
- include the calibration of operational data with fatigue test data;
- consider the method of processing fleet data (i.e., either raw data collection for ground-based processing or onboard processing);
- include a damage model that provides an accurate estimation of fatigue accrual on a scientifically robust basis; and
- provide the operator with regular feedback.

These elements are considered in the following sections.

4 INDIVIDUAL AIRCRAFT MONITORING PROGRAMS

Among other factors, the variation in the operational loading experienced by a fighter-type aircraft throughout its life and the need to identify operational overloads make IAT programs necessary. Furthermore, to assess the consumed fatigue life of an aircraft structure, knowledge of the actual load experienced by that structure is essential [38]. And even where a safe-life may be stipulated, some aircraft are retired at a different number of flight hours due to their calculated rate of fatigue damage accumulation being higher or lower than the target rate because of operational variations.

Prime factors driving IAT are the unique combination of loads experienced by different aircraft in the fleet and the availability of a good onboard monitoring computer. Traditionally, it was assumed that if the fleet N_z load factor exceedances matched or were within that of the design spectrum, the aircraft could safely be operated until the original equipment manufacturer (OEM) promulgated design life. Today, however, each operator of modern aircraft is likely to have a different usage spectrum to the design spectrum. The wing root bending moment (WRBM) is the primary factor to monitor instead of N_z (due to nonlinear aerodynamic and adaptive controls) and a fleet-wide average load spectrum is not viewed as being accurate enough for agile combat aircraft [39, 40], unless severely limiting unmonitored factors are applied.

While heavy military transport aircraft have very strict mission profiles, agile fighter, trainer, or attack type aircraft are well known to experience substantial variability in their missions (see [41] and the next section). Therefore, they cannot be tracked on the basis of mission hours alone and it is the authors' view that an IAT program is necessary for agile combat-type aircraft. For the RAAF F/A-18 fleet, IAT is conducted with every F/A-18 in the fleet instrumented with the same basic operational usage system, this being the maintenance signal data recording system (MSDRS) [40].

One of the greatest benefits of an IAT program is that loads monitoring can take place without a prior knowledge of the exact critical location. Ideally, provided that a sufficient number of primary load carrying structures are routinely monitored, stresses at all critical locations could be determined from strain surveys and/or finite element modeling, with a transfer function relating the monitored load to the critical location stresses. Therefore, a change in the critical location can be accommodated through

the development of a new transfer function to the new critical location.

Some of the benefits gained from the RAAF IAT program is that it enables

- a comparison between design and usage spectra for each aircraft;
- an estimation of the fatigue life or damage status of major components on each aircraft based on loads monitoring in the primary structure of that aircraft and related to fatigue test results;
- planning maintenance actions according to fatigue life estimates;
- modifying operations to stabilize the rate of fatigue life consumption;
- building an operational load database, in conjunction with flight trials, for application to a fatigue test and to compare with early fatigue test data, e.g., [42];
- identifying the variability in response between aircraft in the fleet under the same flight conditions through the assessment of mission severity, effects of stores, and point-in-the-sky (PITS) affects;
- gaining a better understanding of the loading environment (in conjunction with flight trials data); and
- obtaining a better understanding of issues introduced by buffet and structural redundancy at vertical tails [8].

Data obtained from IAT programs can also be used to

- better design future aircraft or be smart buyers in the acquisition of new aircraft for the same role; and
- define (in conjunction with flight trials data) which parameters might be measured on new aircraft or new systems for the same aircraft to allow more accurate calculation of the fatigue life of critical structural components.

4.1 Fleet usage variability

Once critical locations are identified in the design stage and in fatigue tests, IAT programs are used to accumulate and analyze load data from each aircraft in the fleet to predict the damage status at the critical locations. Hence, the fatigue life status of each aircraft throughout its life, based on its own operational load spectrum, is determined. From this information, the amount of fatigue life consumed and the remaining life for each aircraft in the fleet may be calculated independent of other aircraft in the fleet.

Calculating a life based on individual spectra, reveals a wide spread in the rate of fatigue usage as shown in Figure 1 for RAAF data collected over 135 000 operational hours on over 70 F/A-18 aircraft. The fatigue accumulation rate is the individual aircraft fatigue damage value, calculated using

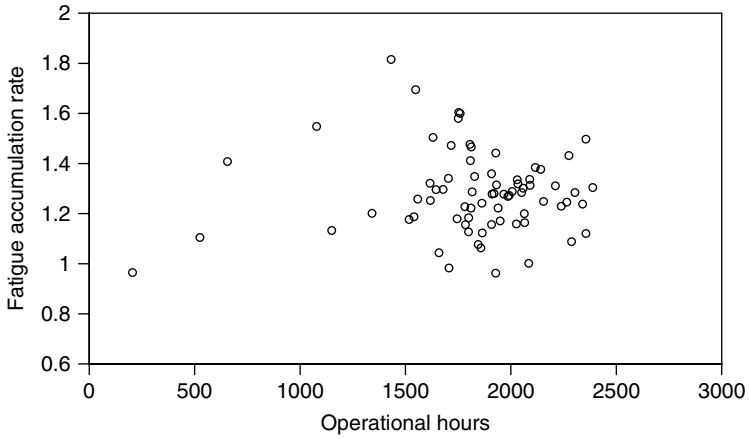


Figure 1. Rate of fatigue damage accumulation for a fleet of aircraft. (The fatigue accumulation rate is the individual aircraft fatigue damage value normalized by the aircraft's operational hours.)

the standard RAAF F/A-18 method, and then normalized by the aircraft's operational hours. It can be seen that using a fleet average would be unwise since some aircraft accrue fatigue damage at almost twice the rate of others. The figure also shows that left unchecked, these trends do not "average out" over the life of the fleet. Of significance, Figure 1 provides an opportunity to assess the 1.5 unmonitored factor required by DEFSTAN. The mean accumulation rate in Figure 1 is approximately 1.3, and applying a factor of 1.5 to this figure provides $1.3 \times 1.5 = 1.95$, which captures the outliers in this sample. As such, the scatter observed in Figure 1 supports the 1.5 factor provided by DEFSTAN.

4.2 Comparison between design and usage spectra

It has previously been stated that "if differences in mission mixture between aircraft remain systemic and significant, there is a case for individual airplane tracking" [41]. This systemic difference is now common and very significant in agile fighter aircraft. In fact, it is rare for two agile aircraft of the same type to experience identical loads for the same type of mission; hence, the need for IAT to examine usage spectra is justified.

New aircraft are serving multiple roles and expectations of enhanced performance are leading to higher operational demands being placed on them. Hence, the operational spectrum of a new aircraft type may be expected to be more severe than the same aircraft type just retired from the fleet. The experience of many operators is that the average usage spectrum is more severe than the design spectrum [41], as was the case initially experienced in the RAAF F/A-18 fleet, as indicated in Figure 2.

Operational loads spectra may be more severe than the assumed design spectra due to variations in the role, the way the aircraft is operated (pilot technique), weight growth, or more severe maneuvers being experienced for the same given mission. Their definition can be useful in identifying trends in aircraft usage, to determine whether the flying has become more benign or more severe, and to schedule operations accordingly.

4.3 Maintenance action

IAT programs can further be used to establish the inspection and modification requirements, and schedules for fleet management, to reduce the cost of unscheduled repairs or extend servicing intervals. In addition to a safety related structural failure, the last thing a military operator wants is an unanticipated detection of "damage". In RAAF experience approximately 10 years is required to plan and implement a fleet-wide structural modification program. Since IAT allows individual rates of fatigue usage or crack growth rates to be estimated, inspections, repairs, or any other maintenance action can be carried out based on accumulated fatigue values or crack lengths

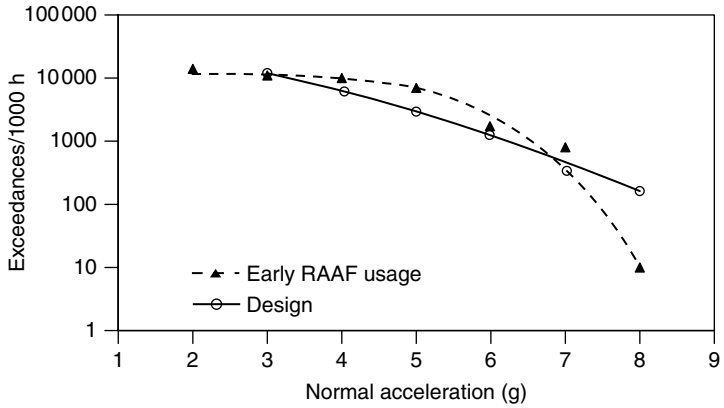


Figure 2. Comparison of design and operational usage for early flying on a RAAF combat aircraft. [Reproduced from Ref. 4. © Wiley-Blackwell Publishing Ltd., 2000.]

instead of flight hours or other simplified usage monitoring. If operational usage is found to be less severe than design estimates, the incorporation of structural modifications, repairs, or inspections based on design certification testing can be delayed. IAT programs can also highlight when operational limits are exceeded and identify the need for maintenance action.

4.4 Modify operations

IAT is particularly useful if large variability exists between squadron operations, between missions, and perhaps between pilots. With agile fighter aircraft, missions of the same type will lead to the accumulation of different amounts of damage [41]. For example, current RAAF F/A-18 operations are broken down into 44 different types of missions. A breakdown into mission type has revealed that the variation in fatigue damage accrual rate within a mission type is substantial (at least one order of magnitude). In fact, the variation seen within one mission type can be as large as that between missions [41]. Hence, it cannot be assumed that a change in mission from one type to another will necessarily result in less fatigue damage for fighter aircraft.

IAT can be used to determine how the structural life of an aircraft varies with aircraft operations. These can be customized (by varying the PITS flown) to meet operational and maintenance needs or to determine the cost of specific operations. Particularly damaging flight regimes may be identified and their occurrences may be reduced.

IAT programs also allow for identification of usage trends over time at fleet, squadron, mission, or pilot levels. The effect of changes in roles, mission types, and mission content on the fatigue life can all be examined and appropriate changes to aircraft operations can be made if warranted.

4.5 Operational loads measurement

While the IAT program means that all aircraft are fitted with the same standard equipment, it is also beneficial to have at least one aircraft in the fleet equipped to perform a loads development or strain survey program. In the RAAF, the F/A-18, Hawk 127, and F-111 fleets have at least one aircraft each fitted with the standard IAT equipment as well as additional strain gauges, accelerometers, and sophisticated data acquisition systems for loads development work. In the case of the F/A-18, one of the fatigue critical locations are the three center fuselage bulkheads, which absorb the wing bending loads, and it is worthwhile collecting flight data at these locations to verify the loads used on the fatigue test.

4.6 Aerodynamic buffet

A major performance improvement to fighter aircraft over the last two decades has been the increased angles of attack that they have been able to achieve. This improvement has given rise to unsteady aerodynamic buffet loads that excite the flexible modes of the wing and empennage. This has led to structural problems with the F-111 TACT [43], F/A-18 [44], F-15 [45], Jaguar [46], Hawk [47], T-45 [48], and numerous other aircraft. IAT and flight test programs can also be used to examine phenomena such as outer wing and empennage buffet and their effects on the fatigue life of critical structure. With sophisticated fatigue monitoring systems such as those on the F/A-18, an extensive database was developed to identify the conditions at which these phenomena occur and to further investigate the problem.

Fatigue monitoring of the vertical tail can be difficult due to complexities such as buffeting, a redundant structure and nonlinear relationships with the normal acceleration at the aircraft's center of gravity [49]. Strain gauges have been fitted to the F/A-18 empennage for the purposes of fatigue monitoring. Time spent in certain fatigue-damaging angle of attack and dynamic pressure regimes have also been examined [44] to quantify buffet affects.

5 FATIGUE MONITORING SYSTEMS

Historically, substantial effort has gone into system design, manufacture, and data collection in fatigue monitoring systems. However, the rapid improvement in the computing power of structural fatigue monitoring systems has also led to a sharp increase in the amount of data collected and thus in the costs involved with data processing, software development,

and data analysis. Importantly, the integration of the fatigue monitoring systems with aircraft software can be problematic, as even simple changes to the flight critical software are expensive and difficult to implement. Therefore, to minimize the effort required in modification and data processing, there is an incentive for the operator to choose the right monitoring system at the outset.

In fleet operations, the accuracy of the fatigue life or crack length prediction depends primarily on two factors, viz., the fatigue monitoring tools that are used in the IAT and the accuracy of the model being used for the prediction. Usage monitoring, based solely on recording "administrative" parameters such as flight hours, mission type, mission duration, pilot name, configuration, takeoff, and landing weight have been used in the past. However, the advent of sophisticated data acquisition systems has lead to more accurate methods being developed. Flight hour or N_z counting are poor options (Table 1) for modern Air Forces operating technologically advanced fighter aircraft. The individual and combinations of tools used vary greatly among aircraft and even among operators of the same aircraft [4]. The most popular combination consists of a strain gauge–based system supplemented by flight parameters as recommended in references [4, 8].

When considering the tools to be used for fatigue monitoring, aside from the cost, perhaps the most important considerations are the volume and accuracy of the data. Other factors such as maintenance of the system, data compression, data integrity, data retrieval, upgrade cost, size, and weight must all be considered. Modularity of the system, the number of channels, memory, programming, and data sampling frequency must also be given consideration. Many aircraft today are undergoing avionics upgrades, and fatigue monitoring systems are being reviewed with these upgrades. Sampling rates of the systems are increasing and "megasamples per second" are common. Parameters should be sampled at sufficiently high rates to account for dynamic loading. Fleet structural integrity managers must take into consideration possible upgrades in computer systems, and hence data handling compatibility between different systems is a significant concern. When data are not transferable between systems, it becomes difficult or impossible to accurately account for data from early periods of flying. This difficulty in filling in missing data and other problems associated with midlife upgrades highlight the importance of getting it right at the time the aircraft is introduced into service.

The direct method of loads monitoring using strain gauges is the method advocated by the authors. However, these should be complemented by the indirect or flight parameter–based method to "fill-in" missing or corrupt data and to validate and calibrate strain gauge data [8]. Other advantages of this combination include the ability to analyze flying on a PITS basis and the option of using a parameter-based secondary system to validate data from the primary system [40].

Commonality in the ground-based processing across all aircraft types, for each Air Force, is highly desirable albeit probably uneconomical and impractical. While it may not be necessary for all the systems to be identical, similarity in the systems can lead to cost savings through commonality in ground-based software.

5.1 Strain gauges

Historically, concerns with the inability to monitor stress activity near the wing root by a fatigue meter alone lead to the development of strain measuring devices capable of responding primarily to the WRBM. Strain gauges were installed near the wing root to monitor the effects of weight changes with fuel burn and weapons release during flight, etc.

Today, judicious placement of the strain gauges can account for these effects at various PITS constituting the flight envelope. The location of the strain gauge must be such that its response is predominantly influenced by the principal loading inducing the fatigue damage at the relevant critical locations. In particular, care must be taken to ensure that the location of the strain gauge is

- able to be calibrated to the damage inducing load;
- dominated by the principal load (e.g., WRBM) and insensitive to other loading actions;
- in an area of low stress gradient;
- able to be directly related to the stress at critical structural locations (preferably by a linear relationship for both positive and negative loads);

- not prone to gauge "drift" (varying response to a nominal load over time—the F/A-18 wing root lugs are an example of this);
- not subject to load redistribution due to redundant load paths or structural changes such as bushing migration;
- accessible for easy replacement;
- positioned as close as practicable to a backup strain gauge in the advent that the primary strain gauge fails or drifts;
- replicated at a "mirrored" location to estimate the asymmetrical component of the loading;
- replicated on the fatigue test article so that direct comparisons can be made (often overlooked in many IAT programs); and
- accurately positioned and protected from the environment and service wear.

Strain gauges (or equivalents) have the advantage of being sensitive to load, and thus aerodynamic phenomena, and provide an indication of the loads the structure experiences. The magnitude of the effects of phenomena such as buffet and gust loads can only be accurately measured by strain gauges or accelerometers [48, 49] and generally not by flight parameters or fatigue meters. The installation of a gauge must be done precisely with a template (location and orientation are critical) and the gauge must not be fragile or erratic. Procedures must be in place to frequently check the condition of the gauges and erroneous gauges must be found and replaced quickly. Ideally, both (left and right) sides of the attachment locations (especially the wing root) should be monitored [40]. Operational data have shown that the accumulation of fatigue damage on the two sides of the aircraft may not be even, as was demonstrated by left and right F/A-18 wing root strain being different depending on the maneuver [4].

The number of channels available on the data acquisition system may restrict the number of gauges that can be placed. Currently, about seven gauges appear to be standard, but this number may vary in future aircraft [4].

Critical point or "hot spot" strain measurement is still common practise, for example, see [4], but is not recommended for IAT [6, 7]. The major problem with hot spot gauges is that they are placed in regions of nonuniform strain that make calibration and replacement difficult. A good example of the former problem was with the F-16 mechanical strain recorders (MSR) where a variation in strain from 85 to 155% was observed over the length of the MSR [50] for a given load case. (The MSR is 203-mm-long with a gauge length of about 13 mm and is installed on the lower flange of a center fuselage wing carry-through bulkhead.) Furthermore, a high strain gradient and the relatively large gauge length implies that the maximum strain is not recorded as uniform strain through the strain gauge is not present. Finally, hot spots may not be known or anticipated until after the conduct of the full-scale fatigue test.

While the benefits and drawbacks of "hot spot" monitoring have been mentioned [8], the authors' views are that strain gauges used in IAT programs should be for structural load monitoring only. In that application, the loads measured by the strain gauges are related to stresses at a critical point via a transfer function, instead of being used directly for maximum stress measurement. Hence, the aim is not to place gauges to determine their lower or upper limits, but to measure loads in the main paths leading to the critical areas.

Gauges should be sampled at frequencies of about 10 times the natural frequency of the *highest most damaging resonant mode* of the structure for areas that are suspected to be dynamically affected. This will ensure that the maximum peak and valley of each cycle are captured.

5.1.1 Strain gauge calibration

Since the fatigue usage of a military aircraft is normally calibrated against the damage accumulated on a fatigue test article, calibration of strain gauges located in nominally identical locations to those on the fatigue test article is essential in order to obtain an accurate estimate of the fatigue life. That is, the gauges are calibrated such that loads derived from operational strain gauges are directly related to loads derived from the equivalent strain gauge on the fatigue test article. To verify the fatigue test loading, the test article gauges may also have been calibrated against the response of a loads development aircraft.

Furthermore, two gauges placed at nominally identical locations, but on different airframes, may not respond equally to a nominally equal global load due to slight differences in airframe build quality, strain gauge alignment, adhesive thickness, and gauge

factor or gauge/amplifier sensitivity. Multiple load paths in a redundant structure may also cause varying gauge response arising from differences that are "built-in" before delivery. In the extreme case, this variability has been observed to be as much as 50% in vertical tails of the RAAF F/A-18 fleet [51].

Calibration is also necessary to account for drift in the strain gauge reading. With the F/A-18, the wing root strain gauge is known to drift as a result of the wing pin attachment bushings causing a redistribution of stress near the strain gauge [40]. This strain gauge is calibrated by comparing operational data with that produced by a reference WRBM applied to the appropriate fatigue test article [8, 40].

Analytical predictions of the calibration factor should be adopted because it is very expensive to physically conduct a ground calibration of each aircraft. While the RAAF F/A-18 fleet of approximately 70 aircraft is relatively small, a major effort would be required to calibrate each aircraft (as was done in RAF Tornado [52]). Hence, analytical methods involving the identification of similar operational PITS and configurations, were developed and validated by ground calibration of 10 fleet aircraft from various squadrons [51–54]. The ground calibration involved application of a distributed or point load to the structure in question and the simultaneous recording of the strain experienced by the strain gauge. This procedure was used to identify the strain per root bending moment (from regression analysis) for the wings, vertical tails, and horizontal stabilators to validate the analytical methods. Alternatively, gauges may be calibrated in flight, under certain configurations and regimes that are flown often. For example, the 1-g trimmed condition under a common stores and weight configuration could be used. The major advantage of this method is that it can be automated and postprocessing efforts reduced.

5.2 Flight parameters

Many military aircraft today have a sophisticated computerized control system that relates flight parameters to control surface deflections. These control systems, together with fatigue monitoring systems, are sometimes integrated into the mission computer. With flight parameter–based systems, loads in the major load carrying members are calculated from flight parameters using regression techniques or neural network techniques [33, 53–56]. These loads, in turn, are related to stresses at critical locations via transfer functions. The load equations are often developed for a certain range of strain (i.e., separate equations for tensile and compressive loads) and for symmetrical or asymmetrical flight, supersonic, and subsonic conditions. Further studies have shown that separate equations are also required for different stores configurations [53–56]. Flight parameters should be integral to an IAT system and may be used to

- calibrate strain gauges;
- validate strains and estimate strains when data are corrupted;
- produce aircraft utilization statistics;
- determine significant loads; and
- provide an independent check of the damage calculated via the strain gauges, as recommended in [40].

In order for flight parameters to be used in the first two cases, sufficient synchronously monitored parameters are required to estimate the recorded strains to a desired level of accuracy. For example, it has been shown [53] that for empennage strain gauges, the following parameters (among others) are significant:

- angle of attack, α;
- stabilator deflection, δ_{elev};
- rudder deflection δ_{rud};
- trailing edge flap deflection, δ_{TEF};
- yaw rate, r;
- pitch rate, q; and
- aileron deflection, δ_{ail}.

6 DATA HANDLING AND PROCESSING

With the growing volume of data being captured by the monitoring systems, data handling procedures that are efficient, inexpensive, and simple must be in place. While much of the data handling procedures are being outsourced by operators, it is important for the operator to determine the level of involvement they have in the overall process. The level of involvement feeds back into the decision as to whether an

aircraft should have onboard data manipulation and analysis software to produce a final damage value for each flight or only capture data with all processing being executed on-ground by the operator or a contractor. It is imperative that the operator has access to the unprocessed data in order to independently validate the accuracy and reliability of the system.

6.1 Onboard versus ground-based processing

The amount of onboard processing may vary. As a minimum raw N_z, strains and flight parameter data may be recorded. A form of onboard data compression is the storage of only peaks and valleys of the signals (where low amplitude or low mean cycles are "discriminately" omitted). If only the peaks and valleys are stored, then it is highly recommended that each peak and valley trigger be "time-stamped" to enable data checking at a later date [57]. Typical onboard processing today includes data checking routines, a stress calculation for each location, cycle counting, damage calculation, and result storage [58]. An example of a fighter aircraft where onboard real-time fatigue calculations are conducted is the Eurofighter 2000 (which is known as *the SHM system, see* **Fatigue Monitoring in Military Fixed-wing Aircraft** and [59]).

At the other end of the spectrum, the F/A-18 is an aircraft where minimal processing is carried out onboard and extensive processing is done on-ground. Although onboard processing appears attractive, it has many significant pitfalls [4]. Data that are collected onboard but compressed cannot be easily verified, validated, or calibrated after the flight. On the basis of the experience of the authors, onboard damage calculation cannot be recommended if raw data is not stored with the final damage values. Onboard systems are good if they work, but present a dilemma if they provide clearly nonintuitive results and the maintainer has no ability to directly evaluate the raw data.

The frequency of data downloading and the time spent in downloading is a major maintenance consideration and downloads after every flight are not desirable as this consumes much time. A download frequency of about once every 50 h appears acceptable, depending on the reliability of the system. For unreliable systems, it is recommended that downloading and checking occur at reasonable intervals such that system malfunctions can be detected and rectified in a timely manner.

Fleet reprocessing may sometimes be required to account for errors or improvements in the software. In such cases, it may be necessary to identify the status of the fleet (from the date of acceptance) using the improved software. For example, RAAF experience has shown that reprocessing the raw data for a fleet of 70 aircraft can be completed in approximately a fortnight and, on average, will be required between 2 and 4 times over the service life.

6.2 Data integrity and fill-in methods

Recording systems are affected by external factors that lead to a loss of data or to the recording of spurious data. It is common for data losses to be between 10 and 20% [30]. About two decades ago, this figure was in the order of 50% [60]; hence it may be expected that the current figure will decline to half its current value in another decade. Data errors may have various sources:

- instrument malfunction, faulty sensors or unserviceability errors;
- recording system failure leading to no data being recorded for portions of or for complete flights;
- data download errors leading to loss of data;
- recording errors in the system that lead to data spikes;
- system input errors that lead to excessive data (e.g., too many turning points in a particular time being captured due to a discriminant being set too low); and
- other reasons that lead to corrupt data (where the data recorded is unrealistic, such as where data is duplicated across various portions of a flight).

Hence, for each parameter or combination of parameters, the following checks should be conducted:

- range operational envelope limit checks;
- maximum rate of change;
- excessive recording;
- data cutting out in the middle of a flight (continuity);
- spikes;
- data repetition;

- initialization; and
- synchronization between parameters (for time lags).

Spurious data are found on every system and lost or bad data from a fraction of a second or a whole flight must be accounted for. As an example, with the RAAF F/A-18, single bad points in the wing root strain gauge are accounted for (filled-in) using V–g–h parametric methods, while whole flights are filled-in using a method based on the typical damage accumulated by the type of flying conducted [40]. Owing to the variability in missions stated earlier, the fill-in method should be conservative in its estimate of the life (i.e., predict a shorter life value) to ensure safety of the aircraft. The conservatism can be introduced by factoring the mean accrual of that mission type.

7 DAMAGE MODELS AND FATIGUE TEST RESULTS

A purpose of any fatigue monitoring program is to determine the fatigue life status of a fleet of aircraft based on their operational spectrum. All fleet structural integrity programs are established on the results of analytical studies and full-scale fatigue tests. However, with a difference between operational and design spectra, interpretation of fatigue test data and application to the fleet can be difficult.

Full-scale fatigue tests seek to [61]

- identify the most critical parts of the overall structure, which are susceptible to fatigue damage;
- compare analytical design data with fatigue test data;
- substantiate a life extension program;
- determine the safe-life or damage tolerance limits;
- determine the onset of widespread fatigue damage; and
- determine crack growth characteristics and accordingly formulate inspection and maintenance schedules.

The results of the fatigue test are required in order to implement a fatigue monitoring system. It is then the fatigue behavior at each critical location that fatigue damage models seek to simulate. It should be noted, as highlighted in this article, that the "damage model" is only one component of the overall monitoring system. Each component contributes to the overall accuracy of the monitoring system. Regardless of the basis of the damage model, be it crack initiation, total life, or crack growth, the other components should be common.

These fatigue models should be calibrated using the full-scale fatigue test results complemented by material coupon test, component tests and/or from in-service defects. It must be shown that the damage model can scale between the fatigue test result and the extremes of fleet usage. Therefore, the spectrum applied to a fatigue test must be accurately interpretable using the fatigue damage model chosen for IAT purposes. In the case of the F/A-18, the damage model was validated against a coupon test program, which tested upward of five F/A-18 spectra, ranging from benign to severe, and at five stress levels per spectrum [62].

Many aims of an OLM or an IAT program can only be achieved through the conduct of a fatigue test. These aims include identification of fatigue critical locations, substantiation of analytical test lives, and the identification of potential services failures due to high loads. Hence, there is a strong relationship between the full-scale fatigue test result and the IAT program.

8 CONCLUSIONS

A review of the philosophies and requirements in fatigue monitoring of agile aircraft has been presented, examining systems and tools, fatigue models, and fatigue test interpretation. Experience with Australian fatigue monitoring programs has been drawn on to highlight deficiencies in certain practices and forecast future trends.

It has been shown that due consideration in the management of fighter aircraft fatigue must be given to the application of fatigue test results to fleet data, an IAT program, a reliable and economical fatigue monitoring system, validation of damage models and data calibration.

It has been shown that IAT has been beneficial in comparing operational and design usage, in the planning of maintenance action, in modifying operations, and in the understanding of structural problems. Importantly, structural design and

management standards mandate the requirement to conduct OLM or IAT, or incur penalties, which substantially increase cost and reduce availability over the service life. The various options for fatigue monitoring systems have been presented and a way forward using a combination of direct and indirect methods has been recommended.

In summary, the fatigue management program should not be an afterthought to the design. Careful consideration must be given to the design philosophy, the monitoring system, the fatigue test, and the application of its results to the fleet, early in the process.

REFERENCES

[1] Schütz W. *Fatigue Life Prediction for Aircraft Structures and Materials*, AGARD-LS-62, ICAF-Doc-693. Advisory Group for Aerospace Research and Development: Bordeaux, 1973.

[2] Jones DJ, Duffield MJ, Holford DM. *Future Fatigue Monitoring Systems for Fixed Wing Aircraft. Proposals for a New Policy and a Strategy for the Way Ahead*, DERA/AS/ASD/CR/97600/1.0. Defence Evaluation and Research Agency: Farnborough, 1998.

[3] Ministry of Defence, *Defence Standard 00–970 Issue 5 Part 1 Section 3 Structure*, London, 2006.

[4] Molent L, Aktepe B. Review of fatigue monitoring of Agile Military Aircraft. *Journal of Fatigue and Fracture of Engineering Materials and Structures* 2000 **23**:767–785.

[5] Molent L, Inan S. *Recommendations for an Individual Aircraft Tracking System for an Agile Fighter Type Aircraft*, DSTO-TR-1102, Melbourne, 2001.

[6] Aktepe B, Molent L. Management of airframe fatigue through individual aircraft loads monitoring programs. *Proceedings 8th International Aerospace Congress*. Adelaide, September 1999.

[7] Molent L. Proposed specifications for an unified strain and flight parameter based aircraft fatigue usage monitoring system. *Proceedings USAF ASIP Conference*. San Antonio, TX, December 1998.

[8] Molent L. A unified approach to fatigue usage monitoring of fighter aircraft based on F/A-18 experience. *Proceedings 21st Congress of the International Council of Aeronautical Sciences*, ICAS Paper 98-5.1.3. Melbourne, 1998.

[9] Wilson ES. Developments in RAAF aircraft structural integrity management. *Proceedings 18th Symposium of the International Committee on Aeronautical Fatigue (ICAF): Estimation, Enhancement and Control of Aircraft Fatigue Performance*. Melbourne, 1995.

[10] Render MEJ, Stevens JE. Aircraft Fatigue Management in the RAF. *Proceedings 72nd Meeting of the AGARD Structures and Materials Panel*, Bath, April-May 1991, AGARD-CP-506: Fatigue Management. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1991.

[11] Lincoln JW. Life management approach for USAF aircraft. *Proceedings 72nd Meeting of the AGARD Structures and Materials Panel*, Bath, April-May 1991, AGARD-CP-506: Fatigue Management. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1991.

[12] Negaard GR. *The History of the Aircraft Structural Integrity Program*, R-680.1B. Aerospace Structures Information and Analysis Center: Washington, DC, 1980.

[13] *Aircraft Structural Integrity Program: Research and Development*, AFR 80-13. Department of the Air Force: Washington, DC, 1976.

[14] *Military Specification: Airplane Damage Tolerance Requirement*, MIL-A-83444. Department of the Air Force: Washington, DC, 1974.

[15] Byron RAV. *Fatigue in Aircraft*, Report 1981/AM/1. University of New South Wales: Sydney, 1981.

[16] Taylor J. Accelerometers for determining Aircraft flight loads. *Engineering* 1952 **173**:473–507.

[17] Taylor J. *Design and use of counting accelerometers*, R & M No. 2812. Aeronautical Research Council: London, 1954.

[18] Scott CE, Rowland WD. *A Transducer for Measuring and Recording Several Levels of Strain*. Proceedings of Society of Experimental Stress Analysis: London, 1970.

[19] Sturgeon JR. *Increasing the Operational Effectiveness of Military Aircraft by Flight Data Acquisition*, RAE-TR-72169. Farnborough, 1972.

[20] Sturgeon JR. The use of accelerometers for operational loads measurements in aircraft. *Proceedings Conference of Stresses in Service*. London, 1966.

[21] Ward AP. The Development of Fatigue Management Requirements and Techniques. *Proceedings 72nd Meeting of the AGARD Structures and Materials Panel*, Bath, April 1991, AGARD-CP-506:

Fatigue Management. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1991.

[22] Matsuiski M, Endo T. Fatigue of metals subjected to varying stress. *Proceedings Kyushu Meeting of Japan Society of Mechanical Engineers*. Kyushu, 1958.

[23] Ford DG, Patterson AK. *A Range Pair Counter for Monitoring Fatigue*, ARL-Struct-Mat-TM-195. Aeronautical Research Laboratories: Melbourne, 1971.

[24] Fraser RC. *Fatigue Damage Estimation for the BAe Aircraft Fatigue Data Analysis System*, ARL-Struct-TM-297. Aeronautical Research Laboratories: Melbourne, 1979.

[25] Goodrige MJ, Woods LE. AFDAS—an aircraft fatigue data analysis system. *Proceedings of Conference of Institution of Engineers*. Adelaide, 1980.

[26] Finney JM, Denton AD. Cycle counting and reconstitution, with application to the Aircraft fatigue data analysis system. *Journal of the Institute of Mechanical Engineering* 1986 **1**:231–240.

[27] Taback I. *The NACA Oil-Damped V-G Recorder*, NACA-TN-2194. Washington, DC, 1950.

[28] Richardson NR. *NACA VGH Recorder*, NACA-TN-2265. Washington, DC, 1951.

[29] Sturgeon JR. *Flight Data Acquisition for Fatigue Load Monitoring and Conservation*, RAE-TM-799. 32nd AGARD Structures and Materials Panel, Advisory Group for Aerospace Research and Development: Neuilly sur Seine, France, 1971.

[30] Schütz R, Neunaber R. Operational Loads Data Evaluation for Individual Aircraft Fatigue Monitoring. *Proceedings 58th Meeting of the AGARD Structures and Materials Panel*, Sienna, April 1984, AGARD-CP-375: Operational Loads Data. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1984.

[31] Haglace TL. *Flight Test Evaluation of a Scratch Strain Gauge*, AFFDL-TR-69-116. Wright Patterson Air Force Base: Dayton, OH, 1970.

[32] Wood HA. In-flight strain measurements using the Prewitt Scratch strain gauge. *Proceedings ASTM Committee E-9*. Toronto, 1970.

[33] Reed SC. Development of a parametric-based indirect aircraft structural usage monitoring system using artificial neural networks. *Aeronautical Journal* 2007 **111**:209–230.

[34] Levinski O. *Prediction of Buffet Loads using Artificial Neural Networks*, DSTO-RR-0218. Defence Science and Technology Organisation: Melbourne, 2001.

[35] Blaha FA, Grenier L. A fiber-optic loads monitoring system for the CL-600 challenger aircraft. *Proceedings USAF ASIP Conference*, WL-TR-96-4030. Dayton, OH, 1996.

[36] *Joint Services Specification Guide (JSSG)*. Aircraft Structures, US Department of Defence: Washington, DC, 2006.

[37] Ward EJ. *Hornet Aircraft Structural Integrity Management Plan, Issue 1*. RAAF Logistics Systems Agency: Melbourne, 1995.

[38] de Jonge JB. The monitoring of fatigue loads. *Proceedings 7th Congress of the International Council of Aeronautical Sciences*, ICAS Paper 70-31. 1970.

[39] de Jonge JB. Load experience variability of fighter aircraft. *Proceedings of the 3rd Australian Aeronautical Conference*. Melbourne, 1989.

[40] Molent L. A review of a strain and flight parameter data based aircraft fatigue usage monitoring system. *Proceedings 1996 USAF ASIP Conference*. Dayton, OH, 1996.

[41] de Jonge JB. Assessment of service load experience. The 12th plantema memorial lecture. *Proceedings 15th Symposium of the International Committee on Aeronautical Fatigue: Aeronautical Fatigue in the Electronic Era*. Jerusalem, 1989.

[42] Graham AD, Symons D, Sherman D, Eames T. ARL F/A-18 IFOSTP full scale fatigue test. *Proceedings 5th Australian Aeronautical Conference*. Melbourne, September 1993.

[43] Coe CF, Cunningham Jr AM. *Predictions of F-111 TACT Aircraft Buffet Response and Correlations of Fluctuating Pressures Measured on Aluminium and Steel Models of the Aircraft*, NASA-CR-4069. COE Engineering: Washington, DC, 1987.

[44] Conser DP, Keys GL. *F/A-18 Production ASPJ Vertical Tail Dynamic Fatigue Test FT98 Test Spectra Development*, MDC 91B0424. McDonnell Douglas Corporation: St. Louis, MO, 1992.

[45] Johnston JT, Pinckert RE, Melliere RA. The F-15 Flight Loads Tracking Program. *Proceedings 58th Meeting of the AGARD Structures and Materials Panel*, Sienna, AGARD-CP-375: Operational Loads Data. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1984.

[46] Holford DM, Sturgeon JR. Operational loads measurement—a philosophy and its implementation. *Operational Loads Data*, RAE-TR-84031, AD-A-149-445, AGARD-CP-375. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1984.

[47] O'Hara J. The evolution of the BAe Hawk and its structural clearance. *Proceedings 17th Symposium of the International Committee on Aeronautical Fatigue: Durability and Structural Reliability of Airframes*. Stockholm, 1993.

[48] Burnham JK. Predicted dynamic buffet loads from limited response measurements: T-45A horizontal tail, AIAA Paper 95–1338. *Proceedings 36th AIAA/ASME/ASCE/AHS/ASC, Structures, Structural Dynamics and Materials Conference*. New Orleans, LA, April 1995.

[49] Aktepe B, Molent L, Graham AD, Conser D. Buffet loads and structural redundancy considerations in vertical tail fatigue monitoring programs. *Proceedings 8th International Aerospace Congress*. Adelaide, September 1999.

[50] Spiekhout DJ. Re-assessing the F-16 damage tolerance and durability life of the RNLAF F-16 aircraft. *Proceedings 15th Symposium of the International Committee on Aeronautical Fatigue Symposium: Aeronautical Fatigue in the Electronic Era*. Jerusalem, 1989.

[51] Aktepe B, Hewitt K, Ogden RW, Molent L. *Ground Calibration of RAAF F/A-18 Onboard Fatigue Strain Gauges*, DSTO-TR-641. Defence Science and Technology Organisation: Melbourne, 1999.

[52] Ward AP. Tornado—Structural Usage Monitoring System (SUMS). *Proceedings 58th Meeting of the AGARD Structures and Materials Panel*, Sienna, April 1984, AGARD-CP-375: Operational Loads Data. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1984.

[53] Molent L, Polanco F, Ogden R, Ooi YG. *Development of Parametric Strain Equations for Fatigue Sensors on the RAAF F/A-18*, DSTO-TR-140. Defence Science and Technology Organisation: Melbourne, 1995.

[54] Molent L, Ogden RW, Ooi YG. *Development of Analytical Techniques for Calibration of F/A-18 Horizontal Stabilator and Wing Fold Strain Gauges*, DSTO-TR-0205. Defence Science and Technology Organisation: Melbourne, 1995.

[55] Holford DM. *Generation of Loads Equations from Flight Parameters for Use in Fatigue Life Monitoring*, RAE Technical Report 81133, Farnborough. 1981.

[56] Lamarre F, Major S, Couture M. *IFOSTP Loads Model Documentation*, IFO-0044. Bombardier: Mirabel, 1994.

[57] Aktepe B, Ogden RW. *A Comparison of Strain Measurements from the AFDAS and MSDRS Fatigue Monitoring Systems Using RAAF F/A-18 Operational Flight Data*, DSTO-TR-640. Defence Science and Technology Organisation: Melbourne, 1998.

[58] Cazes RJ, Defosse P. Aircraft tracking optimization of parameters selection. *Proceedings 16th Symposium of the International Committee on Aeronautical Fatigue Symposium: Aeronautical Fatigue Key to Safety and Structural Integrity*, Tokyo, 1991.

[59] Hunt SR, Hebden IG. Eurofighter 2000: an integrated approach to structural health and usage monitoring. *Proceedings 19th Symposium of the International Committee on Aeronautical Fatigue: Fatigue in New and Aging Aircraft*. Edinburgh, 1997.

[60] Johnson AH, Dubberly MJ. Navy operational loads data sources and systems. In *Proceeding 58th Meeting of the AGARD Structures and Materials Panel*, Sienna, April 1984, AGARD-CP-375: Operational Loads Data. Advisory Group for Aerospace Research and Development: Neuilly sur Seine, 1984.

[61] Mann JY. *Fatigue Testing-Objectives, Philosophies and Procedures*, ARL-Struct-Mat-R-336. Aeronautical Research Laboratories: Melbourne, 1972.

[62] Dickinson T, Molent L. *Validation of Fatigue Damage Models used for F/A-18 Life Assessment using Fatigue Coupon Test Results*, DSTO-TR-0940. Defence Science and Technology Organisation: Melbourne, 2000.

Chapter 113
Video Landing Parameter Surveys

Thomas DeFiore[1] and Richard P. Micklos[2]

[1] *US Federal Aviation Administration, Atlantic City International Airport, NJ, USA*
[2] *PE, Warminster, PA, USA*

1 INTRODUCTION

One of the ongoing challenges faced by the aircraft loads community is to assess the validity of the assumptions used in the aircraft design process. Do these assumptions accurately reflect operating conditions? This article describes a program established to assess the validity of aircraft landing impact conditions.

Although commercial jet transport landing loads are evaluated as part of the test and certification program, no known program exists to analyze operational service loads imposed during commercial operations. Prior to this program, there was very little operational landing data available for commercial aircraft design, and none for wide-bodied jet transports. The implementation of survey procedures using new video technology permits the Federal Aviation Administration (FAA) to determine structural load information for commercial aircraft operations.

For both military and civil applications, visual survey methods have the following advantages over using onboard instrumentation:

- The analysis requires no installation of equipment on the observed aircraft.
- Records are obtained without interference with normal operations.
- Records yield a large number of aircraft approach and initial contact parameters.
- The system enables the study of a large number of landings at minimal cost.
- Permanent records are available for reference purposes.

Since the primary goal of all surveys is to collect statistical information on actual operations, the identity of individual operators, flight numbers, and dates are deliberately omitted from published reports. Landing performance is analyzed only on the basis of aircraft category, model type, weather, and other

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

runway conditions. This operational landing data collection program has proved to be a highly valuable resource for conducting fatigue and damage tolerance assessments of landing gear and its support structure, and in developing design and certification requirements for future jet aircraft.

2 BACKGROUND

The use of image data to evaluate the landing performance of aircraft has been used since jet aircraft were introduced into military service. The US Navy developed a system to characterize the typical carrier-landing environment and develop and implement procedures to make carrier-arrested landings safer [1]. The Navy developed a system to acquire aircraft landing and approach data from the tracking and analysis of recorded 16-mm film images of the arrestment. The basic concept was developed in 1947. The National Aeronautics and Space Administration (NASA), in 1954, developed a similar system using a 35-mm camera and conducted a number of surveys of commercial airplanes, the last one in 1959 [2–7]. Only one NASA survey contained actual jet operations. The significant difference between the two systems was that the Navy photographed from a head-on aspect along the runway apron or aircraft carrier catwalk, while NASA's camera was positioned perpendicular to the runway, approximately 274.32 m from the runway centerline.

In 1967, the Navy [8] enhanced its system by replacing the 16-mm cameras with 70-mm cameras. This provided considerably greater image resolution and, consequently, greater accuracy. Using these film systems, the Navy conducted over 30 landing parameter surveys and has an active carrier-landing survey program.

The Navy's survey techniques required access to the edge of the runway or landing deck to adjust and reload the film cameras. Federal Aviation Regulations require equipment positioned near the runway apron to be both frangible and less than 22 in. above the surface. Subject film system was neither.

Since the film system data reduction was very labor intensive, the Navy, in the late 1980s, replaced their film cameras with specialized high-resolution video cameras. This system permitted the collection of image data remotely and significantly reduced the need to access the edge of the runway. Documentation that the performance and accuracy of the video camera system is comparable to the Navy's 70-mm film system is provided in [9] and [10]. In addition, the system's automated image-tracking capability greatly expedited the data-reduction process. This system technology change enabled the use of the Navy's techniques at commercial airports. Subsequently, a joint FAA/US Navy research effort to collect landing impact parameters at commercial airports was established. Figure 1 shows a camera in operation on a commercial runway. The FAA funded the modifications of the Navy's equipment from a single video camera to a multiple-camera arrangement. This was necessary to increase the camera's coverage area to account for the anticipated wider scatter in touchdown distances from the runway

Figure 1. Video camera in operation during commercial landing parameter survey.

threshold for commercial operations as opposed to those of the military.

3 SYSTEM DESCRIPTION

The development of video technology permitted the Navy to transition its landing parameter data analysis system from one using photographic film to one recording video images. The Navy video system, used for both its own surveys and the FAA surveys is known as the *Naval Aircraft Approach and Landing Data-Acquisition System* (NAALDAS). The system consists of high-resolution frame grab video camera(s), a laser disk recorder, and a computer control unit. The data collection equipment is shown in Figure 2. The key to the NAALDAS system is a highly modified video camera. The camera's enhanced vertical resolution (double that of standard video formats) permits highly accurate measurement and tracking of aircraft position data. The camera is supported by an image analysis system using image processing technology. Particular image features (landing gear wheels, wing tips, flaps, or engine inlets) are tracked in consecutive images, and used to determine the relative motion of the aircraft. These image feature positions are used to derive position–time curves for each image feature. Either linear or second-order regression routines are applied to the data as appropriate. Velocity information is derived from these position–time curves, which are evaluated at the last image frame prior to touchdown. Derivatives of aircraft wheel height curves provide sink rates and derivatives of aircraft range–time curves provide closure speed to the camera. Figure 3 shows the landing height versus time plots for one of the more dramatic landings observed during the survey at New York's John F Kennedy Airport. The other parameters reported for these surveys are derived from these basic measurements. The combination of camera resolution and image processing technology permits the location of image features to be determined within 0.1 pixels. This technique is as accurate, but more efficient than the Navy's 70-mm film system.

The original NAALDAS design used a single camera, which covered the restricted touchdown area on an aircraft carrier. To support the commercial application, the FAA funded the design and development of a modified, multiple-camera configuration of NAALDAS using multiple video cameras situated along the apron of the runway and usually in line with the runway edge lights. The images from these cameras are recorded sequentially as the aircraft passes through their field of view. This modification expands the system coverage area to approximately 609.6 m along the anticipated touchdown

Figure 2. NAALDAS video system, data-acquisition hardware aboard aircraft carrier.

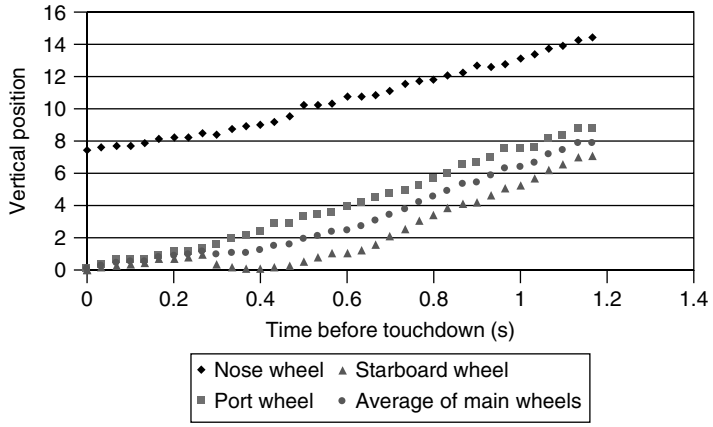


Figure 3. John F Kennedy Airport Survey, landing # 293, landing gear position versus time curve.

region of the runway. Some of the more recent surveys used as many as seven video cameras, thus increasing the camera coverage even further. Fiber-optic signal cables are used to eliminate interference and line losses between the cameras and the recording station.

NAALDAS video camera(s) are installed facing the approaching aircraft; the camera's aim is fixed and does not track the aircraft. The NAALDAS video cameras have a fixed field of view and each camera is aligned and calibrated against targets that are placed temporarily on the runway. These targets are placed in surveyed locations, and the target images are recorded as a calibration sequence (Figure 4). This sequence is processed to generate a transformation matrix, which relates image measurements to the airplane's position on the runway.

During Navy surveys, the NAALDAS data recording system is operated from a compartment near the aircraft carrier's port catwalk. The camera is located in the port catwalk facing the flight deck landing area. A typical camera installation on an aircraft carrier is shown in Figures 5 and 6.

For field surveys, both Navy and commercial, the NAALDAS is operated from a vehicle parked at a safe distance from the runway edge, yet in the vicinity of the touchdown region of the surveyed runway. Temporary cabling is run from the vehicle to the cameras and the vehicle remains in the chosen location during flight operations. The system is

Figure 4. NAALDAS video system calibration targets.

Figure 5. Video camera installed onboard aircraft carrier.

Figure 6. View of carrier flight deck as seen by video camera.

powered entirely with portable electrical generators. NAALDAS is limited to the coverage of one end of a runway and cannot be relocated to accommodate runway changes. This restriction exists since the cameras must be precisely aimed and recalibrated if they are relocated, which requires that the runway be closed. Most camera installation and calibration is conducted during overnight hours.

Video images of each airplane's touchdown are captured and stored on an optical laser disk recorder for subsequent analysis on the NAALDAS analysis system workstation. Approximately, 60 landings can be stored on one disk. An identity number is assigned to the disk, and event numbers are assigned to each video sequence. The use of video disks eliminates film processing cost and time.

Image enhancement and automatic data point tracking are performed using the analysis workstation. The Navy video system introduced the use of these technologies, replacing the manual data point tracking required using photographic film. The analysis procedures remained the same. Image features on the aircraft, usually landing wheel positions, are tracked and the airplane's position and range from the camera are determined for each film or video frame. Figure 7 shows the track windows used in tracking the landing gear wheels of a transport category aircraft. Knowledge of actual aircraft dimensions is necessary to determine the aircraft's range in each image frame, and geometric corrections account for apparent image foreshortening in the video image. These corrections are established during a system

Figure 7. Video image of transport aircraft showing tracking windows used to extract analysis data.

calibration and are dependent on the survey location and conditions. Details of the procedures used to determine these landing parameters are documented in [1, 8, 11].

The analysis station, shown in Figure 8, consists of a Sun computer workstation with an image processing board, laser disk player, computer monitor, high-resolution monitor, and associated power regulator and cables. The station operator automatically tracks the video image features during the landing sequence. By positioning windows over the desired image feature, the operator prepares the system to track that feature through the entire sequence. Multiple-image features can be tracked simultaneously using multiple windows. The operator has the capability to select image threshold levels, image enhancement formats, and algorithms. The operator can also select the type of tracking (edge or centroid) to be used. These selections allow the system to automatically track the image, thus eliminating the errors in data reduction, which were inherent in the manual tracking procedures used with the 16-mm or the 70-mm

Figure 8. NAALDAS video system analysis station hardware.

film systems. The centroid-tracking algorithm enables the system to locate image features with subpixel accuracy.

Once the image sequence is tracked, the pixel information is transformed, digitized, and entered into the landing parameter analysis software. This software takes image position information, determines the change in image feature position of successive frames at a rate of 30 frames/s, and generates position time curves for the feature.

In addition to the video images, from which the ground contact parameters are derived, other data describing each landing are collected during the video survey to determine which set of geometric data to use in the analysis. An anemometer, temporarily installed near the survey site, collects wind speed and direction for each landing. The operators provide an estimate of the touchdown landing weight. After all of the landings on a survey are processed, the landing parameter data is presented as statistical summary tables as well as listings of the parameters for the individual landings. Table 1 is a sample statistical data summary and Table 2 is a partial table of individual landings. Other forms of data presentation include scatter plots, histograms, and "box-and-whisker" plots. The results of all the FAA surveys published to date are available in [12] and [13].

4 OBSERVATIONS FROM MEASURED USAGE DATA

Figure 9 presents a plot of carrier cable engaging speed versus landing weight. As the landing weight increases, the typical engaging speed increases as well. Since carrier landings operations specify a fixed glide slope, the touchdown sink speed increases along with both landing weight and engaging speed. Figure 10 presents a frequency histogram of carrier-landing sink speeds. Sink speed as high as shown in this figure are well within those specified in the military specification.

Figure 11 presents a box-and-whisker diagram of the DeHavilland Dash-8 sink speeds at three commercial airports: London City (LCY), Philadelphia (PHL), and John F Kennedy (JFK). "Box-and-whisker" plots contain the minimum and maximum values of a frequency distribution along with the median, 25th percentile, and 75th percentile. These are used to present a visual comparison among multiple statistical distributions. Clearly, sink speeds

Table 1. Statistical summary of landing parameters results from video landing parameter survey

Philadelphia International Airport landing parameter survey				
Aircraft model: DeHavilland Canada DHC-8				
Parameter	Mean value	Standard deviation	Measurement units	Number of landings
Sink speed				
Port wheel	1.52	0.87	Feet per second	127
Starboard wheel	1.88	1.17	Feet per second	127
Average of main wheels	1.76	1.0	Feet per second	127
Closure speed (measured to camera)	92	7.3	Knots	127
Approach speed	96	7.1	Knots	127
Wind speed				
Head wind	3.8	3.67	Knots	127
Cross wind	1.4	4.89	Knots	127
Pitch angle at touchdown	4.8	1.19	Degrees	127
Roll angle at touchdown	0.8	1.47	Degrees	127
Yaw angle at touchdown	0.4	1.31	Degrees	127
Distance from touchdown to runway threshold	1171	273	Feet	127
Off-center distance at touchdown	0.9	2.96	Feet	127

Table 2. Individual landing parameters results from video landing parameter survey

Philadelphia International Airport landing parameter survey											
Aircraft model: DeHavilland Canada DHC-8											
Landing no.	Power approach airspeed (knots)	Closure speed	Sink speed at touchdown			Ramp to TD distance (ft)	Runway off center (ft)	Pitch angle TD°	Roll angle TD°	Wind parallel (knots)	Wind perpendicular (knots)
			Port (ft s^{-1})	Starboard (ft s^{-1})	Average (ft s^{-1})						
4	98.7	94.7	0.3	2.7	2.3	1339	0	6.6	2.2	4.0	−0.3
5	84.7	79.1	0.4	0.2	1.1	1192	−2	4.0	1.6	5.6	−2.1
8	102.8	95.3	1.0	1.8	1.4	1600	4	3.0	−1.1	7.5	−2.7
9	89.4	82.4	2.0	2.0	2.0	1047	1	3.3	1.8	7.0	0.0
10	98.8	96.5	2.0	1.8	1.9	1413	−4	4.8	−0.6	2.3	−1.9
11	86.8	83.9	0.7	2.3	1.5	972	−2	7.3	−1.8	2.9	−4.1
12	98.7	93.8	3.0	2.7	3.4	1362	−2	5.5	−0.4	4.9	−0.9
13	86.6	82.8	1.3	1.7	1.5	1700	0	6.9	0.5	3.8	−1.4
14	97.4	91.8	1.3	3.5	2.6	1652	5	5.0	3.0	5.6	−2.1
15	77.8	73.8	0.3	2.7	1.5	1223	1	8.4	1.0	4.0	0.3

TD-touchdown.

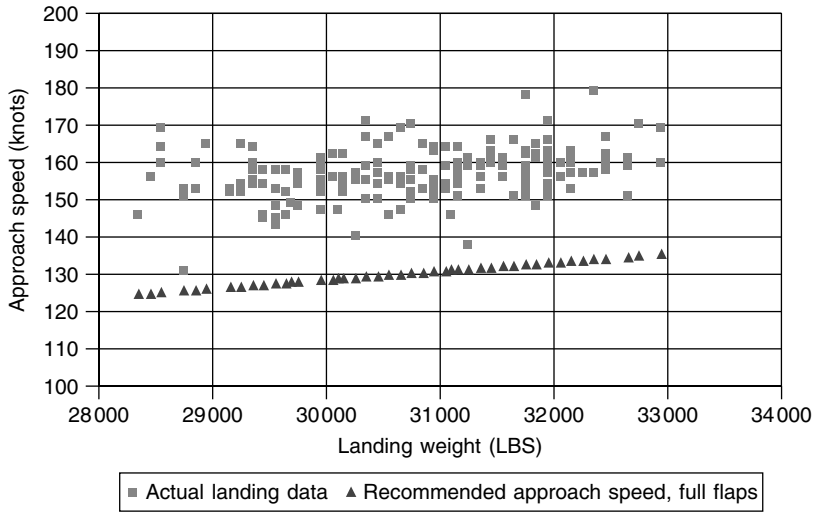


Figure 9. Scatter plot of typical approach speed versus landing weight curve, carrier landings.

at smaller airfields are typically higher than those at longer airfields. Sink speeds at airfields using a higher glide slope are higher than the lower glide slope results.

MIL-A-8863 contains a distribution of sink speeds, which are typically used for military transport aircraft. This probability distribution is compared with the probability distribution for KC-10 aircraft on Figure 12. MIL-A-8863 probabilities and those of the KC-10 are a close match, and thus MIL-A-8863 sink speeds are reflected in actual usage.

The same is not true for civil aircraft. Figure 13 presents a probability distribution of NASA usage spectrum for narrow body civil transport jets. Aircraft

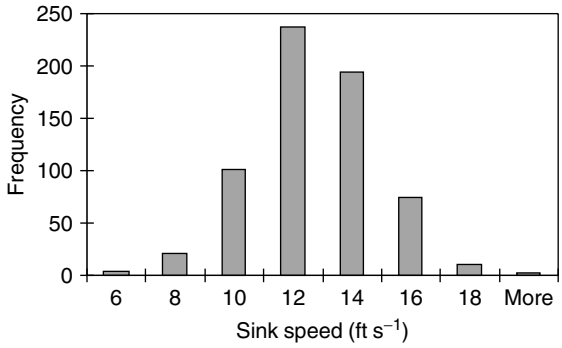


Figure 10. Typical sink-speed histogram, carrier landings.

manufactures have typically used sink-speed data from this NASA report [5] as their sink-speed fatigue spectrum for both testing and fatigue and damage tolerance analysis. If we compare the measured commercial wide-body aircraft usage with the NASA spectrum and the MIL-A-8863 for transports, one can easily observe that the MIL-A-8863 spectrum is more suited for design than the old NASA values. Even narrow body probability distribution levels of Figure 13 are higher than the NASA levels, especially those for narrow body aircraft operating into shorter runways.

5 CONCLUSIONS

The findings from both military and civil video landing parameter surveys are as follows:

- For military transports, the sink-speed fatigue distribution (mean sink-speed of 1.097 m s^{-1},

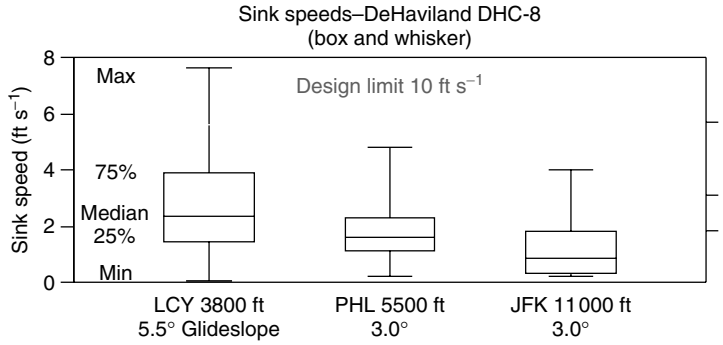


Figure 11. Typical box-and-whisker plot.

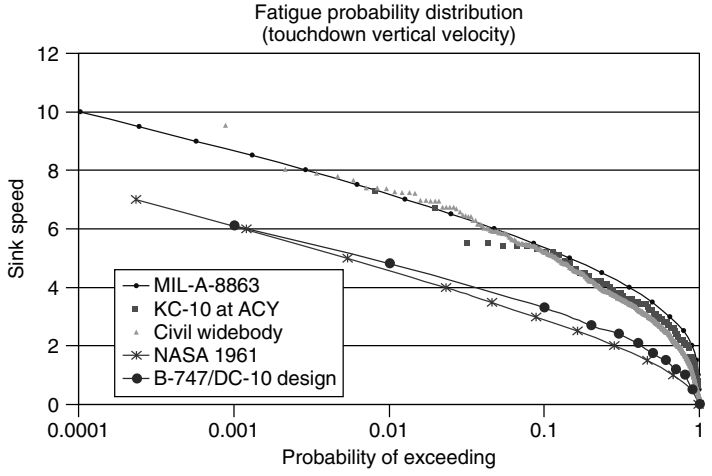


Figure 12. Sink-speed fatigue spectrum—widebodies.

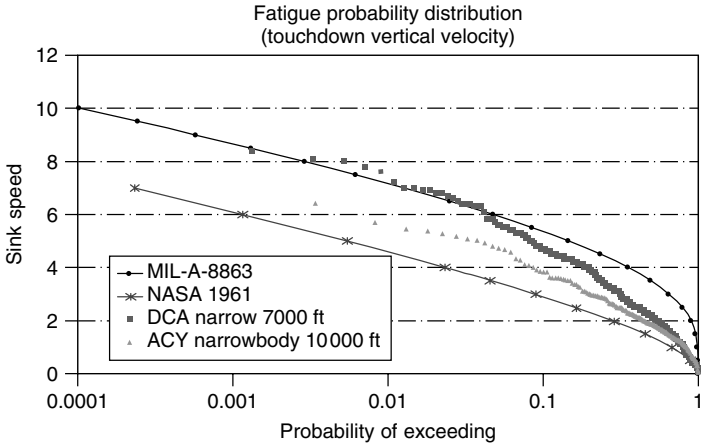


Figure 13. Sink-speed fatigue spectrum—narrowbodies.

standard deviation of $0.405\,m\,s^{-1}$ and skewness factor of 0.5 specified in MIL-A-8863) accurately characterizes typical operations of military transports.

- For civil transports, the limit load of $3.048\,m\,s^{-1}$ for 14 CFR Part 25.473 is accurately reflected in the operations for jets. For turboprops, the $3.048\,m\,s^{-1}$ would appear to be conservative.
- For civil transport jets, the sink-speed fatigue probability distribution shown in NASA TN D 4529 and typically used for design by airframe manufacturers is highly unconservative. Commercial aircraft survey results indicate that the transport aircraft fatigue probability distribution is similar to MIL-A-8863 and more accurately represents typical civil jet operations.
- For civil transports, landings occurring on shorter runways or runways with higher than normal glide slope produce higher sink-speed probability distributions.
- Military surveys allow the US Navy to evaluate the impact of procedural and equipment changes on the landing parameters of carrier aircraft. The validity of operational guidance and standard procedures can be verified.
- Results from military landing parameters surveys has been used to calibrate aircraft flight recorder acceleration data to establish "hard landing criteria" for military aircraft.
- Navy survey sink-speed data was used to document the difference in fatigue loads on aircraft operating from various airfields and aircraft carriers. This information was subsequently used to adjust airframe maintenance schedules.
- In unusually high wind conditions, a Navy landing parameter survey identified unexpectedly high aircraft closure speeds (ground speed) during numerous carrier landings. This data identified the cause of excessive loads on the ship's aircraft arresting gear [14].
- Military landing load survey results have been utilized in preparing detailed specifications for new or modified naval aircraft [14].

It is very important for aircraft manufacturers, operators, military, and regulatory authorities to have a good understanding of the usage of aircraft used in both civil and military service. Analytical models exist, which estimate stress from measured loads. The stress histories are subsequently analyzed to calculate remaining structural life. Without a robust structural health monitoring (SHM) program, operational stresses have to be estimated, consequently, the operator and manufacturer have no way of knowing if the in-service loads are consistent with original design. Continued operation of their airplanes without SHM will occur with higher levels of risk.

RELATED ARTICLES

Landing Gear

REFERENCES

[1] Naval Air Development Center. The Standard NAES Photographic Method for Determining Airplane Behavior and Piloting Technique During Landing. *Naval Air Development Center Technical Report, ASL NAM-DE-210.1*. 26 Sept. 1947, 07Jul. 2008 175152.

[2] NACA-TN-3050, *A Photographic Method for Determining Vertical Velocities of Aircraft Immediately Prior to Landing*, US National Aviation and Space Administration (NASA) formerly NACA, January 1954.

[3] NASA Rep. 1214, *Statistical Measurement of Contact Conditions of 478 Transport-Airplane Landings During Routine Daytime Operations*, US National Aviation and Space Administration (NASA) formerly NACA, 1955, 478 Landings (T-Prop).

[4] NASA, Jewel & Stickle, *Landing Contact Conditions for Turbine-Powered Aircraft*, 1958, 304 Landings, (T-Prop).

[5] NASA TN D-527, *An Investigation of Landing Contact Conditions for a Large Turbojet Transport During Routine Daylight Operations*, US National Aviation and Space Administration (NASA) formerly NACA, October 1960, 103 Landings (Jet).

[6] NASA TN-D-899, *An Investigation of Landing-Contact Conditions for Two Large Turbojet Transports and a Turboprop Transport During Routine Daylight Operations*, US National Aviation and Space Administration (NASA) formerly NACA, May 1961, 100 Landings (T-Prop).

[7] FAA Flight Standards Service, *Statistical Presentation of Operational Landing Parameters for Jet Transport Airplanes*, June 1962, 183 Landings (Jet).

[8] Naval Air Development Center, *The Standard ASD Photographic Method For Determining Airplane Behavior and Piloting Technique During Field or Carrier Landings*, Naval Air Development Center Technical Report, NADC-ST-6706, Jan. 27, 1968.

[9] Naval Air Warfare Center Aircraft Division, *Naval Aircraft Approach and Landing Data Acquisition System (NAALDAS) Video Landing System Shipboard Performance Evaluation*, Technical Report 941034-60. Warminster, PA, 4 Sept. 1994.

[10] Naval Air Warfare Center Aircraft Division, *Naval Aircraft Approach and Landing Data Acquisition System (NAALDAS) Video Landing System Land Based Evaluation*, Technical Report 93004-60. Warminster, PA, 15 April 1993.

[11] DOT/FAA/CT-93/7, *Methods for Experimentally Determining Commercial Jet Aircraft Landing Parameters from Video Image Data*, US Department of Transportation Federal Aviation Administration Technical, August 1993.

[12] DeFiore T, Micklos R. DOT/FAA/AR-96/125, *Video Landing Parameter Survey–John F. Kennedy International Airport*, US Department of Transportation Federal Aviation Administration Technical, July 1997.

[13] Barnes T, DeFiore T, Micklos R. DOT/FAA/AR-04/47, *Commercial Aircraft Video Landing Parameter Surveys, Summary Report–London City Airport, Philadelphia International Airport and Atlantic City International Airport*, US Department of Transportation Federal Aviation Administration Technical, December 2004.

[14] Micklos R. Naval Aircraft Approach and Landing Data Acquisition System, NAALDAS, WL-TR-93-4080. *Proceedings of the 1992 USAF Structural Integrity Program Conference*, September 1993.

FURTHER READING

Barnes T, DeFiore T, Micklos R. DOT/FAA/AR-97/106, *Video Landing Parameter Survey–Washington National Airport*, US Department of Transportation Federal Aviation Administration Technical, June 1999a.

Barnes T, DeFiore T, Micklos R. DOT/FAA/AR-00/72, *Video Landing Parameter Survey–Honolulu International Airport*, US Department of Transportation Federal Aviation Administration Technical, May 2001.

DeFiore T, Micklos R. DOT/FAA/AR-07/53, *Video Landing Parameter Survey: London Heathrow*, US Department of Transportation Federal Aviation Administration Technical, November 2007.

Chapter 115
Monitoring of Aircraft Engines

Visakan Kadirkamanathan and Peter Fleming

Rolls-Royce University Technology Centre, Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK

1 INTRODUCTION

Aircraft engines are complex systems composed of highly advanced mechanical engineering systems and tightly coupled electronic control systems that operate in harsh environments. Though the reliability of the engines is very high, engine performance degradations due to the wear and failure of components will be inevitable over the operational lifetime of the engines. The cost of in-flight engine shut downs and unscheduled engine removal and repair combined with increased safety have been strong drivers for aircraft engine health monitoring systems to be developed. Condition-based maintenance or predictive maintenance, where unnecessary additional downtime and maintenance costs are reduced, requires a functional engine health monitoring system to support it.

Engine condition or health monitoring systems have two main functionalities:

- *Diagnosis:* detect the presence of faults, identify location, and estimate the nature of the fault;
- *Prognosis:* predict the performance degradation, detect early symptoms of a fault, and estimate the timing of maintenance.

The methodologies employed in the construction of the monitoring system depend on the nature of the engine variables measured and the detectability of the fault from those measured signals. The difficulty in obtaining all the necessary measurement signals from aircraft engines and the poor repeatability and detectability of the faults of interest have seen a suite of different approaches being proposed for condition monitoring.

Civil aircraft engine health monitoring systems have historically suffered from high false-alarm rates. Stringent aviation authority certification issues, cost of implementing, installing onboard monitoring systems, and use of unreliable methodologies are also additional factors.

The challenges, therefore, are to derive and develop better fault diagnosis and prognosis methods that are

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

robust, reliable, and can detect the subtle signature patterns associated with weakly detectable faults.

The engine condition monitoring techniques employed can be categorized into model-based methods, model-free methods, and hybrid methods. The model-based methods are based on explicit or derived knowledge of the engine behavior whereas model-free methods are mostly data driven. A different categorization of the methodologies is to divide them into conventional statistical methods and intelligent systems approaches that involve the use of neural networks, fuzzy logic, and evolutionary computation. Often, intelligent systems are used in the hybrid methodologies, which combine model-based and model-free methods in providing information-fusion-based diagnosis and prognosis.

2 ENGINE OPERATION, MEASURED SIGNALS, AND FAULTS

Development of health monitoring approaches depend on how the system being monitored functions, what signals are available from which to infer the system state, and the nature of the system faults to be detected and predicted.

2.1 Engine components and operation

The gas turbine engine is the most economical jet engine used to power civil passenger aircraft. It is analogous to a four-stroke internal combustion piston engine. There are four phases in the engine process—induction, compression, combustion, and exhaust, which are continuous in gas turbine engines as seen in Figure 1. The engine can be divided into three main sections: compressor, combustion chamber, and turbines.

The basic operation of the gas turbine engine is as follows [1]: air enters the air inlet, passes to the compressor where the air-stream pressure is increased through compressor rotor blades and stators, discharges into the combustion chamber where it is mixed with fuel and combusted. The hot, high-pressure gas enters the turbine providing kinetic energy to rotate the turbines with the remaining air ejected via the exhaust. The aircraft motion results from the thrust generated in all three parts of the engine.

Figure 1. Aircraft gas turbine engine cross section.

The compressor, with its large number of rotating components, is a part that is associated with a number of potential engine failures [2]. The high temperatures in the combustion chamber can lead to serious failures in this section of the gas turbine engine. The turbine section consists of fast-rotating blades in very high temperature environment and are, therefore, the most vulnerable for failures. Each stage of the turbine has a stator with stationary vanes and a rotor with rotating blades (turbine wheel) that absorbs energy from the hot high-pressure gas and converts it to mechanical shaft power.

Engine performance parameters define the performance levels of an engine and are therefore a suitable set of parameters that define the health of an engine. Degradation of these parameters may indicate the presence of a fault or of an impending fault. Some of the overall engine parameters of interest are the efficiencies of the compressor, combustion chamber, and the turbine, air mass flow rate, specific fuel consumption, and total pressure loss.

2.2 Measured signals and monitoring

Monitoring from measured signals require a suite of sensors mounted on the aircraft engine that depend on the complexity of the engine, identification of the key variables, and their measurability. A number of key engine variables are measured for the purposes of electronic engine control and for cockpit monitoring. They are [3]

- *temperatures*—inlet temperature, ambient air temperature, exhaust gas temperature, compressor temperature, turbine temperature, bleed air temperature;
- *pressures*—inlet pressure, compressor pressure, discharge pressure, lubrication oil pressure and bleed air pressure;
- *speeds*—spool speeds from the engine stages;
- *vibration*—rotor vibration, shaft vibration, bearings vibration, and accessories vibration;
- *oil system*—oil quantity, oil consumption, oil debris, and oil contamination;
- *life usage*—operating hours, flight cycles;
- *others*—exhaust pressure ratio, fuel flow, throttle position.

An example of on-engine sensors and measurements is shown in Figure 2.

Unique identification of faults in a complex system such as aircraft engines is challenging because of the difficulty in measuring the aerothermal parameters directly. Given the lack of knowledge of many of the symptoms associated with failures, monitoring

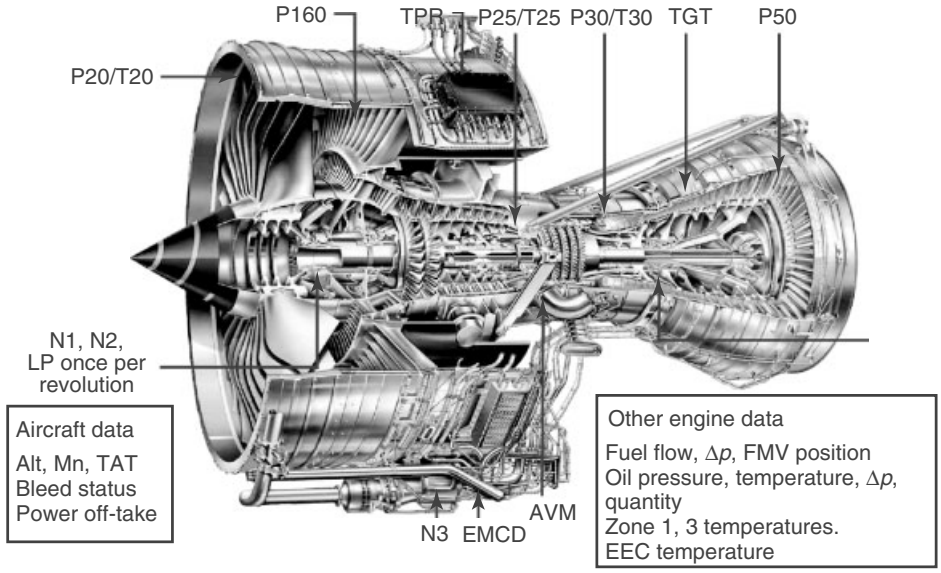


Figure 2. Sensors mounted on an aircraft gas turbine engine.

systems have focused on anomaly-detection methods that discriminate between the nominal operational characteristics and faulty characteristics, particularly when the faults are incipient and slowly developing.

The onboard monitoring system is typically contained within the full authority digital engine control (FADEC) where all measured signals are collected. The data are collected at high sampling rates during flight for onboard on-line monitoring and control, but are not stored for off-line ground-station analysis. The onboard system performs built-in tests (BITs) that traditionally have led to high false alarms resulting in the problem of no fault found in the subsequently removed line-replaceable units. If the onboard monitoring based on limit checking detects an event, then the performance data are recorded for on-ground analysis [4]. For civil aircraft, typically, only a snapshot of data at a small number of operational conditions such as during takeoff, cruise, and landing are transmitted to the ground station for condition monitoring.

2.3 Engine faults

Engine performance changes can be mapped to abrupt faults (rapid short-term deterioration) or incipient faults (gradual long-term deterioration). Analyses of abrupt faults are the functions of a diagnostic system, whereas the analyses of incipient faults are the functions of a prognostic system.

The traditional maintenance strategy involves periodic inspections after a specific number of operational hours and/or flight cycles, often determined from the life limits of critical parts such as compressors, turbine blades, and disks. Unscheduled maintenance can arise from the detection of specific events during engine operation. Some of these are [2]

- *Foreign-object damage* can lead to small scratches or to complete destruction of the engine. It is typically associated with vibration signature patterns and changes in nominal operating parameters. Damage to compressors or turbines often result in an increase in exhaust gas temperature, decrease in engine pressure ratio, and a change in the ratio of shaft speeds.
- *Over-limit operation for temperatures* is often caused by malfunction of the engine fuel control or a malfunction in the engine and results in high exhaust gas temperature at start-up. Operating the engine at excessive temperatures can lead to cracks, burning, metal distortion, and metal loss.
- *Over-limit operation for speeds* impacts on the rotating assemblies of the engine and can cause fan blade and vane damage and fan rotor damage.
- *Engine stall* is also a condition that can cause damage to the engine.

However, the ability of engine health monitoring to predict, rather than detect, these and other impending failures, where possible, is what is required of a condition monitoring system. The measured signals that are required to determine aerodynamic performance include engine pressure ratio, fuel flow, shaft speeds, and the exhaust gas temperature. The parameters needed to evaluate mechanical performance include vibration and oil consumption.

The engine failures that are amenable to detection and prediction via condition monitoring systems are [2]

- **Failures due to air leakage from compressor cage**

A number of compressor section failures may be due to the failure of bleed air duct external to the engine, a stuck overboard bleed valve, or failure of the engine casing. The drop in engine pressure ratio due to this failure is compensated by the increased fuel flow and hence the increase in rotor speeds and other parameters.

- **Compressor contamination**

Contamination due to operation near salt water, use of impure water, oil leak leading to dust accumulation on blades etc., can degrade engine performance. The reduced compressor efficiency again leads to increased fuel flow to maintain required engine pressure ratio and hence to increased shaft speeds and other parameters.

- **Mechanical failures**

These failures involve only a few blades or vanes, unlike compressor contamination, though it also affects the compressor efficiency, albeit by a small value. Severe failures, however, show up during high power operation during which exhaust gas temperatures are very high and/or compressor stalling occurs. Downstream foreign-object damage may also occur

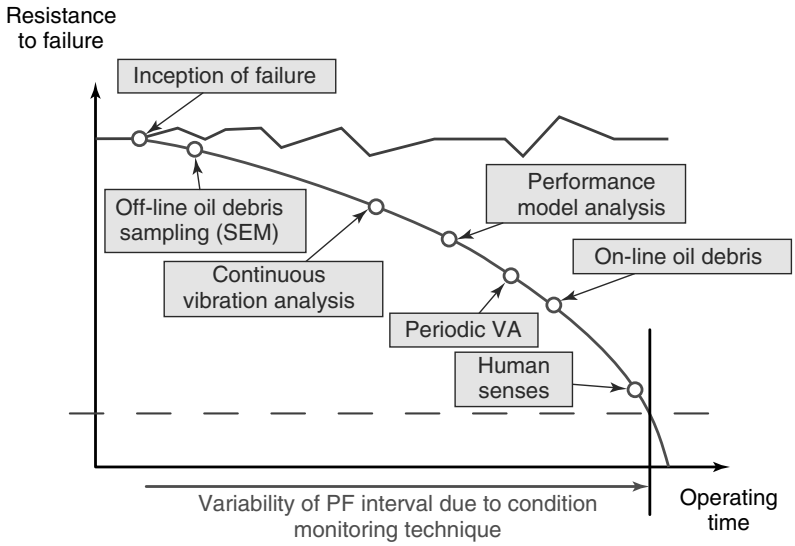


Figure 3. Aircraft engine condition monitoring methods and their failure detection times.

from flying parts. Bearing malfunctions and broken blades lead to signature vibration patterns.

- **Combustion section**

Failures are due to blocked fuel nozzles, fuel line leaks, and the failure of the burner. Since obtaining measurements from this section is difficult, detection of failures must be carried out from parameters from other sections. This, in turn, makes the detectability of these failures very weak.

- **Turbine failures**

Loss of turbine efficiency due to broken blade or seal erosion is the primary failure of interest in this stage. Loss of efficiency increases fuel flow to maintain required engine pressure ratio while the ratio between shaft speeds is likely to have changed.

3 AIRCRAFT ENGINE CONDITION MONITORING

Engine health monitoring systems, whether onboard or off-board, divide the analysis into four distinct categories requiring different approaches to be used:

- Mechanical parameters monitoring—vibration;
- Gas path analysis and performance parameters trending;
- Oil and debris monitoring;
- Remaining useful lifetime estimation.

The four different methods enable detection of faults at various times from the onset, as illustrated in Figure 3.

Advanced health monitoring systems usually perform information and decision fusion from the different category monitoring systems.

3.1 Vibration monitoring and analysis

The vibration signals measured are typically high bandwidth accelerometer signals attached to the casing of the engine. Principles of piezoelectric sensors for accelerometer signals are given in **Piezoelectricity Principles and Materials**. Alternative candidates such as eddy current (*see* **Eddy-current *in situ* Sensors for SHM**) and acoustic emission sensors are also available to monitor such mechanical parameters. With the aircraft engine containing rotating components, the vibration signals contain distinct frequency components. For monitoring and analysis, the time domain accelerometer signals y_t are therefore converted to the frequency domain through fast Fourier transform (FFT) or other spectral

estimation techniques:

$$y(f) = \mathcal{F}(y_t) \tag{1}$$

for $f = 1, \ldots, F$ frequency points. Spectral analysis, in the context of general structural health monitoring, can be found in **Statistical Time Series Methods for SHM**.

Since the engine rotational speeds can vary during the operation, the spectral patterns also vary. However, the frequency components in the vibration signals are primarily the rotational frequency and its higher and subharmonic components. The amplitudes of the signal at these frequencies of interest define the signature pattern for the vibration signal and are referred to as tracked orders [5]:

$$\mathbf{x} = \left[y(f_1), \ldots, y(f_n)\right] \tag{2}$$

These features are excellent candidates for detecting mechanical faults and an example of a spectral pattern is given in Figure 4. The tracked order signal shows interesting patterns associated with an anomalous event.

In addition to tracked orders, broadband energy amplitude, side bands energy, resonance and jumps could also be used as features for monitoring vibration signals. Health monitoring based on vibration signals analyze the tracked orders of the engine for changes from its signature. However, these signature tracked orders vary with the engine speed and, therefore, the engine signature is represented through a series of tracked orders at specific engine speeds. Other frequency-domain feature extraction methods such as wavelets can also be applied [6].

The extracted vibration signal features consist of information sufficient to characterize normal engine behavior across the fleet of engines. However, no explicit mathematical model exists that relates engine parameters to the specific vibration signatures. Hence, model-based approaches cannot be used for monitoring vibration signals. Vibration engineering expert knowledge and experience has led to the proposition of simple heuristic rules in the form of fault signature matrices relating specific changes in vibration

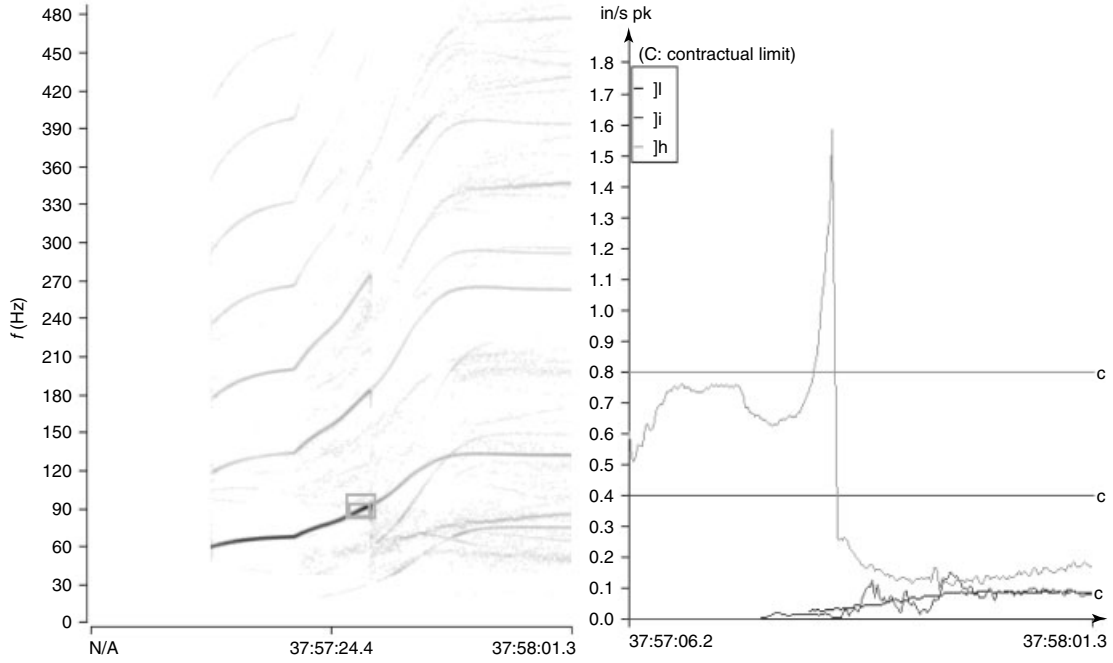


Figure 4. Spectral patterns showing high energy in resonant frequencies (left) and the tracked order signal amplitude (right).

signature pattern to specific faults, but these matrices are very sparse and imprecise. Vibration signal monitoring methods have therefore been based on data-based methods, mostly addressing the problem of anomaly detection [7]. Slowly developing engine faults such as blades and bearing failures can often show early symptoms in vibration patterns. It is therefore important to detect deviations in these vibration patterns through anomaly-detection approaches. Statistical, machine learning and dynamic system identification approaches have been used for novelty detection in vibration signals [5, 7]. Often, they are also combined with visualization methods based on dimensionality reduction for decision support to the engineers [5].

Even though these anomalies cannot be associated with any specific fault at the time of detection, subsequent engine overhaul and analysis can lead to specific faults being associated with the vibration patterns that are precursor symptoms to that fault. A library of such vibration signature patterns can be created and these can then be used to analyze historical engine data using data-mining approaches. A high-performance neural network time-series pattern matching of the tracked orders has been developed for monitoring specific vibration characteristics changes in a fleet of aircraft engines using such a library [8]. This pattern-matching system is integrated with a case-based reasoning intelligent system that learns the association between the signature patterns and the underlying faults.

3.2 Gas path analysis

Gas path analysis is also known as *module performance analysis*. The key idea is to estimate the level of degradation in the main stages of the engine in the gas path, based on the signals measured [9, 10]. It is based on the basic notion that engine component faults that are the changes in engine parameters will lead to changes in the measured signals. These measured signals are very noisy and some trending analysis is required to detect the incipient changes in the performance measures, an example of which is shown in Figure 5. Gas path analysis methods have proved effective in detecting faults in aircraft engines.

The first gas path analysis methods devised were model based, utilizing mathematical models derived from physical processes. The earliest gas path analysis is the parameter sensitivity analysis in which

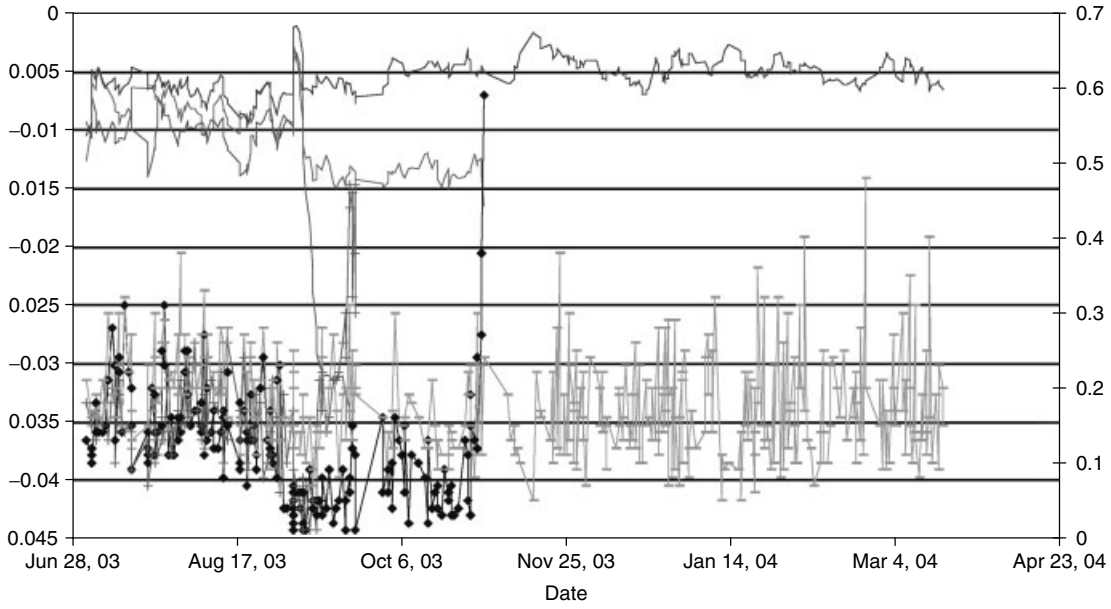


Figure 5. Compressor efficiency performance parameter of engines with and without faults.

linearized models around the operating conditions are used in determining the sensitivity matrix between the engine parameters and the measured signals [11]:

$$\Delta \mathbf{y} = \mathbf{H}\Delta \mathbf{x} \tag{3}$$

where $\Delta \mathbf{y}$ is the change in measured signals and $\Delta \mathbf{x}$ is the change in engine performance parameters. The sensitivity matrix $\mathbf{H}$ is known. Then, the problem of estimating the fault is determining the inverse function $G(\cdot)$ of the model:

$$\Delta \hat{\mathbf{x}} = G(\Delta \mathbf{y}) \tag{4}$$

While it may be possible to isolate single and multiple faults, it may not be possible to estimate the level of fault in all cases due to the nonuniqueness of the inverse model. However, additional constraints can be imposed to improve this accuracy, such as flow capacity, and efficiencies deteriorate over time. A particular advantage of the sensitivity approach, and other model-based approaches, is that the estimated parameters such as combustion efficiency and turbine efficiency are physically meaningful.

To overcome the limitations of the parameter sensitivity analysis, dynamic model-based methods were devised. The physically derived mathematical models describing the dynamics and the measurements are represented as state space models and the problem of estimating the faults are cast as state estimation problems. This allows advanced state estimation filtering-based methods such as Kalman filters for engine health monitoring [11]. Instead of Kalman filters, other state estimators such as observers and even a full complex simulation model [12] can be used.

3.3 Oil and debris monitoring

Surface failures lead to debris formation, which also leads to oil contamination. By measuring the size and quantity of particles in oil, unacceptable wear and fatigue failures of engine components can be detected. Off-line debris particles are analyzed with X-ray fluorescence instruments or energy dispersive scanning electron microscopes, which can identify the material type, thereby providing fault isolation [13]. On-line debris monitoring, sometimes included in gas path analysis, uses magnetic and electric chip detectors. Faults such as blade rubs, nozzle-vane erosion, and combustion-chamber burn can be detected by monitoring the electrostatic charge associated with the debris in the gas turbine engine exhaust gas. By monitoring the lost metal particles in the oil, degradation trends can be identified for early bearing fault detection. It is considered a promising technology for detecting precursors to failures. The Joint Strike Fighter aircraft engine has engine distress monitoring based on the debris monitoring principle [14].

3.4 Life usage monitoring

The life usage monitoring approach traditionally relied on the simple principle that critical engine components have an average lifetime based on the engine operation in-flight hours or in-flight cycles [15]. The average lifetimes are derived from historical reliability statistics and do not take into account a single engine condition. The key assumption here is that each engine is operated in a typical expected flight profile. A better estimation of the remaining useful time of an individual component can be made with improved damage accumulation modeling and use of information from measured signals. For example, component degradation may impact on the engine efficiency, which is reflected in the maximum engine speed and maximum turbine temperature. Additional measurements such as blade-tip deflection and torsional vibration can improve remaining life estimation. Life usage monitoring is particularly important for tracking damage due to fatigue, oxidation, and creep [14]. The algorithms are derived using damage models that are empirically adjusted from operational experience.

4 METHODS FOR ENGINE MONITORING

The standard method of monitoring parameters is for anomaly detection based on simple limit checking of the deviations from nominal operating values. In the absence of knowledge of faults and their symptoms, the limits are usually kept tight, leading to a high rate of false alarm. Performance parameter changes exceeding limits and vibrational amplitudes at the relevant frequencies exceeding average base line

signatures are events that point to developing faults. This functionality can be achieved via a multitude of approaches, some of which are given here. Dimensionality reduction is a powerful tool for diagnostic decision support systems that aids visualization for the engineer. Automatic diagnostic systems require methods that can build on the available knowledge and/or learn from data. Intelligent systems approaches and model-based approaches are two of the main class of methods for aircraft engine condition monitoring.

4.1 Dimensionality reduction

Visualization is a powerful approach to aid operators monitoring the condition of an aircraft engine. The multidimensional measurement signals, extracted features, and performance parameters must be projected down to a lower dimension to aid visualization. Dimensionality reduction can also be used for extracting features that contain most of the informative part of the measurements.

One of the most popular approaches to dimensionality reduction is principal component analysis (PCA), which is a linear projection method. It is a multivariate statistical method that identifies the orthogonal directions of largest variance in the data. Let the observed data be normalized (centered and scaled) to be zero mean and unit variance and be denoted by $\tilde{\mathbf{y}}_t$ for $t = 1, \ldots, T$. Then the PCA projects the data on to the lower dimension via,

$$\mathbf{x}_t = \mathbf{T}\tilde{\mathbf{y}}_t \tag{5}$$

where $\mathbf{T} = [\mathbf{u}_1 \cdots \mathbf{u}_n]^{\mathrm{T}}$ is the transformation matrix with $\mathbf{u}_i$ as the ith eigen vector (ordered by decreasing eigen values) of the data covariance matrix $\mathbf{P}$,

$$\mathbf{P} = \frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{y}}_t \tag{6}$$

By representing the normal variation in the monitored parameters, it is possible to visualize deviations that are anomalous and thereby identify faults in engines [16]. Additional anomaly detection procedures may be required to ensure that abnormal behavior is detected. Linear projection methods will not capture nonlinear correlations that are present in the multivariate data.

An example of nonlinear projection method is the Sammon's mapping, which attempts to preserve the distance in the high-dimensional multivariate space to be the same in the projected low-dimensional space. It seeks to identify the projection data by minimizing the sum of squared error,

$$E = \frac{1}{\sum_{i<j}\delta_{ij}}\sum_{i<j}\frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}} \tag{7}$$

where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ and $\delta_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|$ are the low- and high-dimensional space distances between the i and j data points. Neuroscale algorithm is a radial basis-function neural-network-based extension of the Sammon's mapping, which has been used for visualization of engine vibration parameters [5]. Further details of dimensionality reduction for structural health monitoring can be found in **Dimensionality Reduction Using Linear and Nonlinear Transformation**.

4.2 Learning with intelligent systems

4.2.1 Expert systems

Obtaining detailed and precise mathematical models for complex systems such as aircraft engines is difficult. Where expertise and experience of engine behavior, faults and their symptoms are available, rule-based expert systems can be used for engine diagnostics. The rules mimic the fault manual information and/or the rules applied by a troubleshooting expert. Rule-based systems may be used to verify whether the evidence for a specific fault is supported using the backward-chaining algorithm or to identify the rules that best match the observed evidence using forward-chaining algorithm. The advantage of the rule-based systems is the transparency offered in how the diagnostics was arrived at, like knowledge capture from experts [17]. However, knowledge transfer from expertise and experience to a set of rules for complex systems can be fraught with challenges including conflicting rules, tracking the internal state of the system, and completeness and correctness of the derived rule base. They are also difficult to adapt to changes over time. Use of expert systems for gas turbine engine prognostics and diagnostics can be found in [18].

4.2.2 Case-based reasoning

Case-based reasoning is a knowledge-based problem-solving paradigm that resolves new problems by adapting the solutions of previous similar problems [19]. It allows consolidation of the rules extracted from knowledge and uses a probabilistic matching-based reasoning. Case-based reasoning permits the development of a diagnosis system incrementally and allows multiple information sources to be integrated, making it a technology suitable for reasoning with complex systems. It is ideally suited to fault diagnosis when the problem being addressed is poorly understood and data are available to characterize a range of operating conditions and faults.

The case-based reasoning system consists of a case base of data from historical cases and a reasoning engine, which has four key steps:

- *Retrieval:* Given a new problem with observed attributes, retrieve the best past cases from the case base.
- *Reuse:* Adapt solutions from matched cases to form a new solution to the present problem.
- *Revise:* Evaluate the outcome of the solution and provide reasoning. Continue until an acceptable solution is found.
- *Retain:* Incorporate the new and successful solution into the case base.

The case base represents a knowledge repository for engine faults, symptoms, maintenance actions and outcomes. Best practice of maintenance experts and experienced engineers are also represented in this case base. Even though case-based reasoning systems can be applied in an autonomous fashion to diagnose faults and health conditions, it is also a powerful interactive data-mining tool for an engineer. For example, restricting searches to relevant attributes and time windows may lead to knowledge discovery about recurring faults and the identification of relevant symptoms. An example of engine cases associated with a specific problem within a case-based reasoning system is shown in Figure 6.

Case-based reasoning systems have been used in a portable PC-based flight-line maintenance advisor to correlate and integrate fault indicators from the engine monitoring systems, built in test equipment (BITE) reports, maintenance data, and dialog with maintenance personnel to allow troubleshooting of faults. It has also been applied to the Rolls-Royce and general electric (GE) gas turbine aircraft engines for engine maintenance decision support and in fault identification [20].

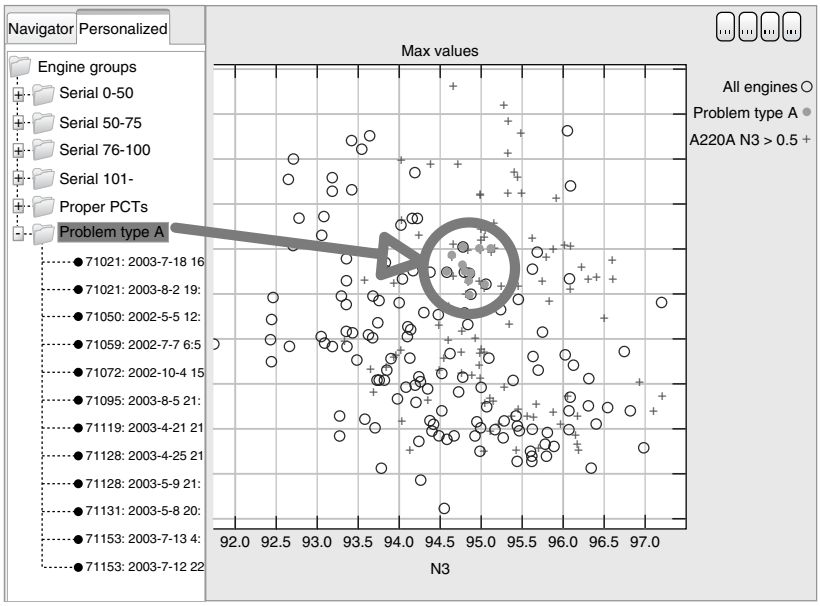


Figure 6. Condition monitoring with case-based reasoning.

4.2.3 Neural networks

Neural networks have been applied to engine health monitoring in two different ways: as function maps that find association between parameters and as classifiers [14]. Neural networks construct nonlinear functions of the form

$$h(\mathbf{x}) = \theta_0 + \sum_{k=1}^{K} \theta_k \phi_k(\mathbf{x}) \tag{8}$$

where $[\theta_0\ \theta_1\ \cdots\ \theta_K]$ are the set of weights and $\phi_k(\mathbf{x})$ are the basis functions, which depend on the form of the neural network and themselves are parametrized. The common forms of neural networks are the multilayer perceptrons, radial basis-function networks, and support vector machines [21, 22]. The universal approximation property of the neural network allows it to map arbitrary nonlinear functions to a specific level of accuracy. Neural networks are trained from a collection of input–output data with the training algorithm determining the weights and the basis-function parameters.

The key principle for the approach based on classification is that different faults lead to differences in the measured signals or extracted features. Faults can be mapped to partitions in this signal or feature pattern space. Neural networks are used as discriminant functions $f(\mathbf{x})$, which create nonlinear partition boundaries and improve classification performance over general linear classifiers. Neural networks are constructed or trained using example fault signature patterns from known faults. Their application to gas turbines have shown comparable performances to Kalman filter-based fault-isolation methods [11]. They can also be used for anomaly detection by identifying boundaries of normal operation. Conditions associated with anomaly lead to measurement signals or features falling outside the boundary [7, 23]. Novelty detection and neural networks for structural health monitoring can be found in **Artificial Neural Networks**; **Novelty Detection**.

4.2.4 Fuzzy systems

Fuzzy logic approaches have proved valuable in engine health monitoring because of the fact that the available knowledge and information about engine characteristics and fault symptoms are incomplete and/or imprecise. The fact that fuzzy systems are universal function approximators like neural networks, but with the advantage of being expressed in linguistic terms allowing easy interpretation has made this approach powerful and yet transparent. Accurate gas turbine engine fault isolation has been demonstrated [24]. The approach utilizes the linearized sensitivity model to generate the fuzzy rules with isolation being carried out by determining the fault with the highest degree of membership for the given set of measurements. The imprecision in the linearized model and the uncertainties associated with the random variations are accounted for by the fuzzy inference process, resulting in a robust fault-isolation scheme in relation to the traditional sensitivity analysis.

4.3 Model-based tracking, detection, and identification

Model-based methods rely on a descriptive dynamic model characterizing the behavior of the system and the measurement model for the engine performance parameters (*see* **Model-based Statistical Signal Processing for Change and Damage Detection**). Often, the model is linear and is described in state space with states $\mathbf{x}_t$, inputs $\mathbf{u}_t$, and measurements $\mathbf{y}_t$:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t \tag{9}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \tag{10}$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are the model parameters assumed known and $\mathbf{w}_t$, $\mathbf{v}_t$ are random disturbance and noise respectively. Other forms of models can be in the form of input–output models such as autoregressive (AR) models, transfer function models, or complex physical models for system simulation, shown in Figure 7.

The general dynamic model-based condition monitoring systems employ the approach of residual generation followed by residual analysis. The residuals are generated by differencing the actual measured signals from the model predicted output. If the system is faulty, then the residual signal pattern can be associated with specific faults. The accuracy of these approaches are critically dependent on the accuracy of the derived models and the detectability of the faults.

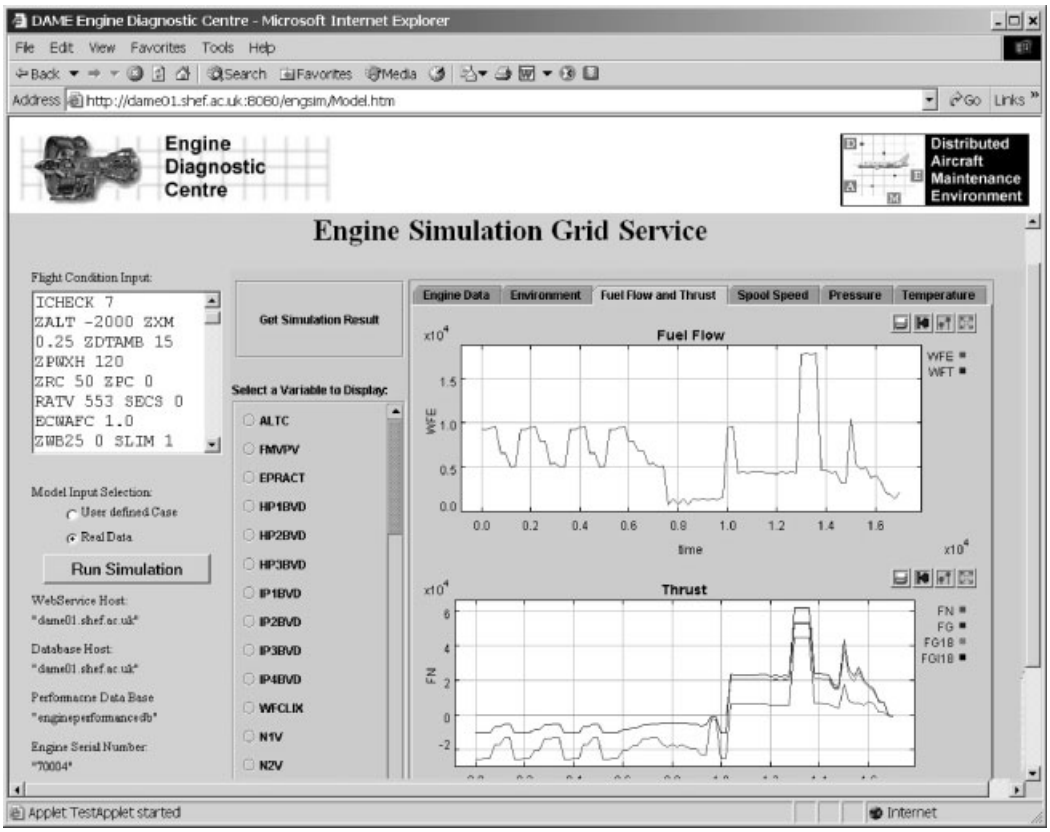


Figure 7. A simulation of key parameters from a detailed empirical model of an aircraft engine.

4.3.1 Filtering and residual generation

Model-based methods offer the greatest promise for prognosis in which engine performance parameter changes have to be tracked. This is a filtering and smoothing problem that is best approached from a model-based perspective. Kalman filters are the minimum mean square optimal linear filters and are popularly used in tracking time-varying parameters [25]. The states or parameters of interest are updated as

$$\hat{\mathbf{x}}_{t+1} = \hat{\mathbf{x}}_t + K_t e_t \tag{11}$$

where K_t is the Kalman gain, which multiplies the residual prediction error $\mathbf{e}_t$ (difference between the actual and model predicted measurements) to provide the update for the states,

$$e_t = \left(\mathbf{y}_t - \hat{\mathbf{y}}_t\right) \tag{12}$$

An example application of the Kalman filter for aircraft engine prognosis is found in [11].

Kalman filtering is also used for fault detection and diagnosis. When the aircraft engine is normally operating without any faults, the model prediction will closely match the actual engine measurements. Thus, the residual prediction error will typically be white noise. By monitoring the sum squared residual over a specified time window, by using generalized likelihood ratio test or other residual-based change-detection methods, emergence of an abrupt fault can be detected [26]. The use of a model in this context is known as *matched filtering approach* where the model or filter is matched to the system operating normally. Kalman filters can be replaced by others such as observers and parity space approach to residual generation and fault detection [27].

One approach to model-based fault diagnosis is based on the multiple-model approach [22]. First,

a number of models each matched to a specific fault in the system (including the normality conditions) are constructed and residuals generated from each. By comparing the residuals and using detectors based on, for example, the generalized likelihood ratio, fault isolation and estimation can be performed as in [26]. This approach with a bank of Kalman filters have been used for aircraft engine fault diagnosis [28].

4.3.2 System identification

One of the popular model-based techniques is the system identification-based condition monitoring. These methods abandon the physics-based models and represent the system dynamics in terms of parametrized black-box models such as the autoregressive models with exogenous inputs (ARX), state space models, or neural networks and estimate these model parameters from observed signals. They offer advantages over other model-based approaches, since the identified models are constructed from measured data. This removes the effect due to model imperfections that can reduce performance of other model-based methods derived under homogeneity and ideal conditions. Once the model is identified, fault detection is then carried out either by monitoring the changes in the model parameters or by monitoring the residuals formed from model predictions. An application of this approach to aircraft engines in which a state space model is identified for fault diagnosis can be found in [5]. Use of nonlinear models based on neural networks can also be identified [29].

5 CURRENT AND FUTURE DIRECTIONS

The challenges faced by the problem of aircraft engine health monitoring is in developing robust, reliable engine health monitoring systems with improved detection, diagnosis, and prognosis capabilities. The developments toward achieving better diagnosis and prognosis have seen increased instrumentation and data gathering from aircraft engines, test beds, and overhaul and repair facilities. The abundance and variety of data and engine fleetwide information have not only made the development of sensors and condition monitoring algorithms challenging but have also placed an emphasis on the system architecture and infrastructure.

Aircraft engine health monitoring requires new sensing mechanisms that permit the estimation of the engine performance parameters more accurately and which are sensitive to the engine faults of interest. A network of integrated sensors with embedded microprocessors offer a smart sensing technology that can aid fault detection isolation [30]. Development of novel fiber-optic sensors that can provide data on the distribution of variables, such as pressure along the engine, facilitates better diagnosis while creating a need for new analysis tools (*see also* **Novel Fiber-optic Sensors**). New developments in microelectromechanical systems (MEMS)-based wireless network of sensors permits direct health monitoring, for example, the temperature monitoring of bearings to predict failure rather than using indirect analysis based on oil analysis. On the algorithmic front, recent advancements in model-based diagnosis and prognosis can be applied to aircraft engine condition monitoring such as particle filters [26]. The physics-based models are also being improved in their sophistication such as finite element models of microstructures for monitoring cracks in structures [14]. The paradigm of semisupervised learning [31] also promises improvement in data-based approaches, which traditionally tend to suffer from the fact that faulty data are scarce, particularly for the highly reliable aircraft engines. Along with the improvements in methods for the different monitoring areas such as vibration and performance parameters, integrating and fusing the different decisions from condition monitoring is a challenge that is beginning to be addressed [32].

The Distributed Aircraft Maintenance Environment (DAME) project [8, 20], a UK e-Science pilot project involving Rolls-Royce, developed a grid computing architecture-based diagnostic workbench, which seamlessly integrates geographically disparate data and analysis tools. The distributed grid architecture creates a virtual organization environment for collaborative decision-making between multiple stake holders such as engine manufacturers, engine operators, and maintenance analysts and experts. This framework includes wireless mobile handheld devices providing access to the distributed system

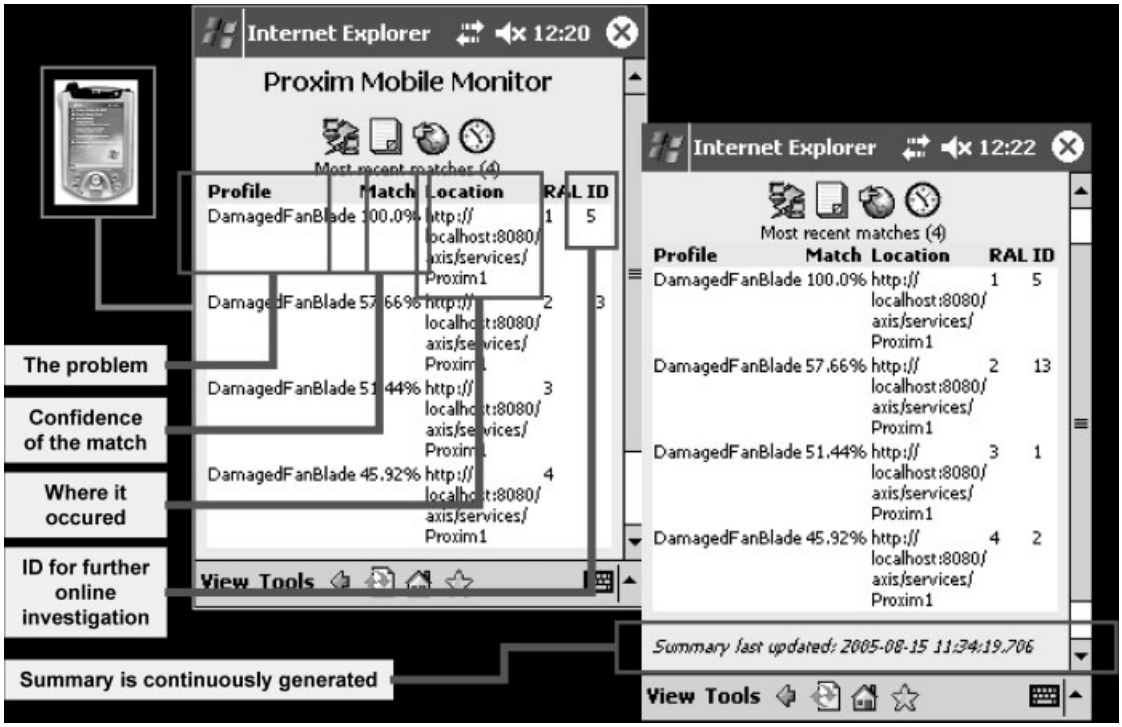


Figure 8. Mobile distributed aircraft engine condition monitoring.

enabling full diagnosis and prognosis capability while working in remote locations, an example shown in Figure 8. The DAME system consisted of diagnostic services for vibration and performance parameters feature extraction using novelty detection methods, data management, and data mining of time-series patterns using neural networks, physical model-based simulation, and case-based, reasoning-based decision support [8]. Such a distributed system with heterogeneous monitoring tools, user interaction, and automated analysis with collaborative decision-making is a powerful system that holds great promise for the future.

ACKNOWLEDGMENTS

The figures used in this document have been mainly sourced from the EPSRC DAME project, thus acknowledging the contributions of Dr Graham Hesketh and Dr Steve King (Rolls-Royce plc.), and the DAME project team including York, Oxford and Leeds Universities.

REFERENCES

[1] Huenecke K. *Jet Engines: Fundamentals of Theory, Design and Operation*. Airlife Publishing: Shrewsbury, 1997.

[2] Tumer IY, Bajwa A. Learning about how aircraft engines work and fail. *Proceedings of the 1999 AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, AIAA-99-2850. Los Angeles, CA, 1999.

[3] Tumer IY, Bajwa A. A survey of aircraft engine health monitoring systems. *Proceedings of the 1999 AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, AIAA-99-2528. Los Angeles, CA, 1999.

[4] Cue RW, Muir DE. Engine performance monitoring and troubleshooting techniques for the CF-18 aircraft. *ASME Journal of Engineering for Gas Turbines and Power* 1991 **113**:11–19.

[5] Hayton P, Utete S, King D, King S, Anuzis P, Tarassenko L. Static and dynamic novelty detection methods for jet engine health monitoring. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 2006 **365**:493–514.

[6] Turso JA, Lawrence C, Litt JS. Reduced order modelling and wavelet analysis of turbofan engine structural response due to foreign object damage (FOD) events. *ASME Journal of Engineering for Gas Turbines and Power* 2007 **129**(3):814–826.

[7] Hayton P, Scholkopf B, Anuzis P, Tarassenko L. Support vector novelty detection applied to jet engine vibration spectra. In *Advances in Neural Information Processing Systems 13*, Leen TK, Dietterich TG, Tresp V (eds). MIT Press: Cambridge, MA, 2001, pp. 946–952.

[8] Jackson T, *et al.* Distributed health monitoring for aero-engines on the grid: DAME. *Proceedings of the IEEE Aerospace Conference*. Big Sky, MT, 2005; pp. 3738–3747.

[9] Merrington GL. Fault diagnosis in gas turbines using a model-based technique. *ASME Journal of Engineering for Gas Turbines and Power* 1994 **116**:374–380.

[10] Li YG. Performance analysis based gas turbine diagnostics: a review. *Proceedings of the IMechE, Part A: Journal of Power and Energy* 2002 **216**(5):363–377.

[11] Volponi AJ, DePold H, Ganguli R, Daguang C. The use of Kalman filter and neural network methodologies in gas turbine performance diagnostics: a comparative study. *ASME Journal of Engineering for Gas Turbines and Power* 2003 **125**(4): 917–924.

[12] Ren X, Ong M, Allan G, Kadirkamanathan V, Thompson HA, Fleming PJ. Service oriented architecture on the grid for integrated fault diagnostics. *Concurrency and Computation: Practical and Experience* 2007 **19**(2):223–234.

[13] Tauber T, Johnson D. Experience with the electric oil debris monitoring system of the General Electric GE-90 gas turbine engine on the Boeing 777 aircraft. *Proceedings of the JOAP International Condition Monitoring Conference*. Mobile, AL, 2002.

[14] Jaw LC. Recent advancements in aircraft engine health management (EHM) technologies and recommendations for the next step. *Proceedings of 50th ASME International Gas Turbine and Aeroengine Technical Conference*, GT2005-68625. Reno Tahoe, NV, June 2005.

[15] Pfoertner H. The information content of turbine engine data—a chance for recording-based life usage monitoring. *Proceedings of the IEEE Aerospace Conference*. Big Sky, MT, 2002; pp. 6–2975–6–2985.

[16] Mustapha F, Manson G, Pierce SG, Worden K. Structural health monitoring of an annular component using a statistical approach. *Strain* 2005 **41**(3):117–127.

[17] Giarratano JC, Riley GD. *Expert Systems: Principles and Programming*, *Fourth Edition*. PWS Publishing: Boston, MA, 2004.

[18] DePold H, Gass FD. The application of expert systems and neural networks to gas turbine prognostics and diagnostics. *ASME Journal of Engineering for Gas Turbines and Power* 1999 **121**(4):607–612.

[19] Kolodner J. *Case-Based Reasoning*. Morgan Kaufmann: Cambridge, MA, 1993.

[20] Ong M, Ren X, Allan G, Kadirkamanathan V, Thompson HA, Fleming PJ. Decision support system on the grid. *International Journal of Knowledge-based and Intelligent Engineering Systems* 2005 **9**:315–326.

[21] Bishop CM. *Pattern Recognition and Machine Learning*. Springer: London, 2006.

[22] Fabri SG, Kadirkamanathan V. *Functional Adaptive Control: An Intelligent Systems Approach*. Springer: London, 2001.

[23] Patel VC, Kadirkamanathan V, Thompson HA. A novel self-learning fault detection system for gas turbine. *Proceedings of the UKACC International Conference on Control*. Exeter, 1996; pp. 867–872.

[24] Ganguli R. Application of fuzzy logic for fault isolation of jet engines. *ASME Transactions: Journal of Engineering for Gas Turbines and Power* 2003 **125**(3):617–623.

[25] Gustafsson F. *Adaptive Filtering and Change Detection*. John Wiley & Sons: ISBN 0-471-49287-6 UK, 2000.

[26] Li P, Kadirkamanathan V. Fault detection and isolation in nonlinear stochastic systems—a combined adaptive Monte Carlo filtering and likelihood ratio approach. *International Journal of Control* 2004 **77**(2):1101–1114.

[27] Spina PR. Reliability in the determination of gas turbine operating state. *Proceedings of the 39th IEEE Conference on Decision and Control*. Sydney, 2000; pp. 2639–2644.

[28] Kobayashi T, Simon DL. Application of a bank of Kalman filters for aircraft engine fault diagnostics. *Proceedings of ASME Turbo Expo*, GT2003-38550. Atlanta, GA, 2003.

[29] Polycarpou MM. An on-line approximation approach to fault monitoring, diagnosis and accommodation. *SAE Transactions: Journal of Aerospace* 1994 **103**(1):371–380.

[30] Spencer Jr BF, Ruiz-Sandoval ME, Kurata N. Smart sensing technology: opportunities and challenges. *Structural Control and Health Monitoring* 2004 **11**(4):349–368.

[31] Chapelle O, Scholkopf B, Zien A (eds). *Semi-Supervised Learning*. MIT Press: Cambridge, MA, 2006.

[32] Volponi AJ, Brotherton T, Luppold R. Development of an information fusion system for engine diagnostics and health management. *AIAA 1st Intelligent Systems Technical Conference*, AIAA 2004–6461. Chicago, IL, 2004.

Chapter 114
Landing Gear

R. Kyle Schmidt[1] and Pia Sartor[2]

[1] *Messier-Dowty SA, Vélizy-Villacoublay, France*
[2] *Department of Mechanical Engineering, University of Sheffield, Sheffield, UK*

1 INTRODUCTION

The landing gear of an aircraft is a unique component. It must bear extreme and varying loads when an aircraft maneuvers on the ground or alights, yet it must be lightweight and compact because it is stowed and unused during most of an aircraft's flight. The landing gear is both a structure and a machine. In its simplest form, it is a structure with energy absorption capability. However, in its modern incarnation, it is a complex machine with controlled articulation, multiple axes of energy absorption, and the sole structure that supports the aircraft.

In most applications, landing gears have no structural redundancy. This, coupled with the conflicting design requirements of high load carrying capacity versus minimum weight and size, results in a landing gear being predominantly manufactured from very high-strength (but relatively low toughness) steel, aluminum, and titanium alloys. This is in contrast to most airframes and other aircraft structures, which are typically made from ductile aluminum alloys that can withstand relatively long cracks that grow over time.

The significant difference between the aircraft structure and landing gear is also reflected in the fact that aircraft design and approval methodologies are quite different between the airframe structure and the landing gear. For example, many airframe designs use "damage tolerant" design methodologies, which allow cracks of known sizes to exist in structural members, whereas landing gears use "safe life" design methods, which do not permit cracks. Therefore, many of the technologies relating to health monitoring of the airframe, such as measuring crack growth and assessing the location and size of cracks, are not useful when considering health monitoring of landing gear structures.

The landing gear also has movable elements that form part of the structure. These components must

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

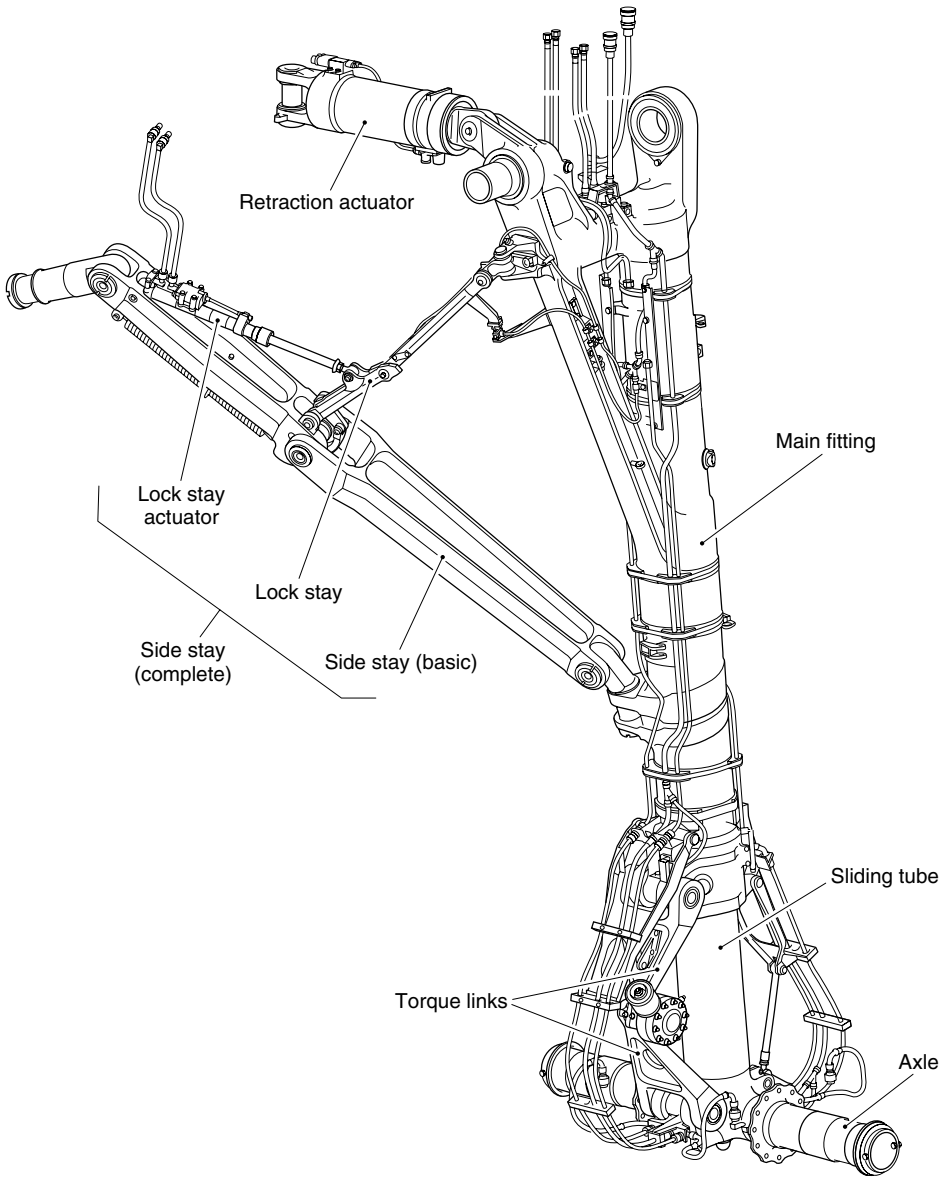


Figure 1. Typical main landing gear structure. [Reproduced with permission. © Messier-Dowty Ltd.]

be considered when taking into account monitoring of the entire landing gear structure.

The differences between landing gear systems and other aircraft systems mean that often, an alternative approach has to be taken with the structural health monitoring solutions of the landing gear. Historically, landing gears have only been fitted with the minimum number of sensors required to indicate the position of the structure or state of articulation. As aircraft system complexity has increased, so has the level of sensing on the landing gear. Several commercial aircraft presently fly with monitoring systems installed to sense tire pressure and brake temperature [1]. Newer aircraft are being fitted with systems to monitor the health of the shock strut element [1]. Direct measurements on the landing gear to determine the weight

of the aircraft have been attempted several times in the past [2]. This article discusses past and present approaches for monitoring the movable aspects of the landing gear, methods for transient overload detection, techniques for measuring the forces seen by the landing gear structure, and methods for determining the fatigue state of the landing gear structure. Figure 1 shows a typical medium-range commercial passenger aircraft main landing gear (MLG) structure. It is assumed that the reader has a prior knowledge of landing gear design and operation. For reference, the design and operation of landing gears are discussed extensively in Norman S. Currey's *Aircraft Landing Gear Design: Principles and Practices* [3].

2 LANDING GEAR TIRE, BRAKE, SHOCK STRUT, AND ACTUATION MONITORING

The movable portions of the landing gear demand utmost care and maintenance, and as such have lent themselves to the introduction of monitoring systems earlier than the purely structural portions. Typical of these systems is the tire pressure indication system (TPIS) available from Messier-Bugatti, which is now in its third generation of development [4]. Alternative systems for tire pressure monitoring are available from other vendors, such as Crane [5]. A tire pressure monitoring system eliminates the regular maintenance actions to inspect the pressure of the tire. This type of monitoring system increases the overall safety of the aircraft by helping to avoid incidents due to underpressurized tires, as well as reducing the overall maintenance burden.

Brake temperature monitoring systems detect the temperature of brake components and report the data to the pilots in the cockpit. The temperature information can also be used to directly control brake cooling fans. Monitoring of brake temperature permits the flight crew to make educated decisions regarding the dispatchability of the aircraft. Knowledge of brake temperature profile and braking performance can be used to determine the remaining brake life [6].

Shock strut servicing indicators, whether as a ground support aid or through the aircraft central maintenance computer, would help to ensure that the shock strut is performing as designed, thereby prolonging landing gear and aircraft life. Health monitoring of any oleo–pneumatic, single-state shock strut requires measurements of nitrogen gas pressure, hydraulic fluid volume, temperature, and shock strut position to determine the status of the landing gear shock strut.

Currently, to ensure that the landing gear is properly serviced, the ground crew measures the shock strut extension and nitrogen gas pressure and compares the measurements with the servicing chart on the shock strut. If the gas pressure is found to be too low, more nitrogen is added. However, low pressure could actually be due to a loss of hydraulic fluid rather than nitrogen. Therefore, this technique can lead to improper servicing. The only reliable way to know that the hydraulic fluid is at an acceptable level is to take the aircraft out of operation and perform full hydraulic fluid and nitrogen gas servicing or to perform the servicing check at two different aircraft weights, which is time consuming and expensive.

In the ideal situation, it would be possible to monitor the shock strut by measuring all four parameters with sensors. For example, Goodrich Corporation has developed a method for an oil volume sensor assembly [7] and NASA has developed a capacitive fluid level measurement system [8], both of which would be internal to the shock strut. In each of these cases, however, field repair would be impossible and an additional hole would have to be drilled in the shock strut wall for the extraction of measurement wires. Two ultrasonic approaches to fluid level measurement also exist—an external manual method [9] and an internal pulse echo method [10].

An alternative approach to measuring oil volume would be to include the element of time in the measurement so that the volume sensor is not included and two measurements are made at different shock strut positions. The simplest approach would be to make one of these measurements prior to landing and the second one when the shock strut has settled, potentially at the gate. This "two-point" method would allow the shock strut state to be determined reliably with conventionally available pressure, temperature, and position sensors.

Currently, in the case of temperature and pressure measurements, these can be obtained using a temperature–pressure sensor. For example, Messier-Bugatti has developed an oleo pressure monitoring system (OPMS), which is currently installed on

the A380 [11]. The OPMS takes shock strut pressure and temperature measurements and transmits the information to the cockpit, where the information is displayed on the multifunction displays. Any abnormal measurements are alerted to the flight and maintenance crew to ensure that maintenance is performed.

For shock strut position measurement, a rotary variable differential transformer (RVDT) could be used on both articulated and cantilevered gears. This piece of equipment is currently used extensively in landing gear systems. For an articulated landing gear, an RVDT could be used to measure the angle between the trailing arm and the main fitting and, using geometry, the amount of shock strut compression could be calculated. For cantilevered landing gear, an RVDT could be installed at a point that permits measurement of the torque link rotation. Although three pivot points are available for measurement, the upper torque link to the main fitting attachment point typically offers the greatest simplicity in terms of rigid mounting of the sensor and static harness routing. Alternatives for the position sensor include synchros and resolvers for rotation measurement. These devices are aerospace grade and their advantage is their larger rotary measurement range at the expense of a greater number of wires in the harness. The use of rotary potentiometers is also possible but not recommended because of the severe operational environment on the landing gear.

In any case, a rotary measurement method would be superior in most respects to a linear contrivance. The linear systems, which are currently used for flight test instrumentation, are too fragile for use on revenue aircraft since the exposed translating member could be easily damaged. If, however, the linear system is contained within the shock strut it may be protected from damage, but repair or replacement becomes onerous.

It should be noted that multistage shock struts offer some additional challenges, primarily on the computational side of determining the service state. At a minimum, an additional pressure and temperature sensor is required in the secondary chamber. It may also be required to know the position of the floating piston.

Currently, a more detailed study of shock strut servicing is being performed through a European Union (EU)-funded 6th Framework project, known as *TATEM* (*Technologies And Techniques for nEw Maintenance concepts*) [12].

Monitoring of the bearing interfaces on landing gear is not currently done, but several possibilities exist that could be exploited. The extension and retraction operations of the landing gear can be monitored using the conventionally installed proximity system. By trending the times for extensions and retractions, deviations from the norm could be identified, which may indicate excessive friction in the joints or problems with the actuation system. Wear of wheel bearings can lead to local overheating of the bearing lands and a subsequent softening of the material. Vibratory monitoring as employed on engines (*see* **Monitoring of Aircraft Engines**; **Gas Turbine Engines**) and other rotating machines (*see* **Large Rotating Machines**) could be employed to detect the early stages of bearing wear. Local temperature sensors—perhaps as an extension of the brake temperature monitoring system—could be employed as an alternative detection system.

3 LANDING GEAR TRANSIENT OVERLOAD DETECTION

One of the primary areas of interest in landing gear monitoring is the determination of transient structural overloads. Owing to the general high structural reliability of landing gears, these transient incidents are of utmost interest to aircraft operators. Aircraft landing gear overload can occur when any part of the landing gear exceeds the design/certification loads of its components. These overload events can include hard landings, overweight landings, off-runway events, runway/taxiway/apron incursions, or problems during towing operations. After any overload event, the operator needs to know whether the landing gear is suitable for safe dispatch. Presently, when an overload event occurs and a pilot makes a declaration, visual and nondestructive testing (NDT) inspections may be performed on the landing gear. In addition, information from the flight data recorder (FDR) may be downloaded from the aircraft and analyzed with the use of dynamic aircraft models to generate estimates of the loads seen during the event.

Given the amount of work required when an incident occurs, there has been significant interest in the

development of hard landing indicators. The potential solutions can be divided into three classes. The first is by the use of dynamic measurements, such as acceleration, velocity, and displacement indications. The second is by the use of purely mechanical devices to indicate if an overload event has occurred. The third method is by the use of force measurements, such as pressure and stress/strain.

Interest in monitoring the performance of the aircraft during the landing stage began as early as 1919. Dr Zahm developed an aircraft accelerometer, intended for research and not operational use, that consisted of a number of helical vertical springs with "styluses" attached to them. This device traced the sudden loads and shocks encountered during landing. However, it gave a discontinuous record of accelerations [13]. In 1934, the National Advisory Committee on Aeronautics (NACA) had developed and tested a seismographic "landing shock recorder" that gave a time–displacement record of movement in a given direction during the impact of an aircraft landing. Although the original intent of the device was to confirm the reliability of the accelerometers in use, this landing shock recorder was also able to make measurements to determine the accelerations under aircraft landing impact conditions [14].

This early work on the measurement of landing impacts is instructive as the vast majority of hard landing detection systems developed and employed use acceleration to determine whether any given landing was an overload event. Presently, acceleration data from digital FDRs is analyzed to provide information when a suspected overload event occurs. This analysis is time consuming and is hindered by the generally low recording rates of the parameters of interest to the landing gear versus the short time taken to perform a landing. This has led to the development of stand-alone hard landing detectors.

One intriguing example of a passive hard landing detector is a mechanical latching accelerometer developed in the 1960s. This indicator is composed of a pivotally mounted metallic mass/spring arrangement such that when a predetermined acceleration is exceeded, the mass pivots and is held in place by a magnet, giving an indication [15]. An electronic single-axis accelerometer approach is used on the Gulfstream IV. This aircraft has a gravity force "G" monitor system, which records the maximum vertical acceleration experienced by an aircraft during landing. The landing acceleration is indicated to the flight crew on a cockpit display. Any suspected hard landing can be checked by comparing the peak vertical acceleration against a plot of allowable accelerations versus aircraft mass. The disadvantages of this system are that it records only the peak acceleration in the vertical axis, and only near the aircraft's center of gravity.

Messier-Dowty has developed a hard landing indication system that records high-rate accelerations and roll rates in three dimensions from several remote inertial measurement units (RIMUs), as seen in Figure 2. The data from the RIMUs are then fused with the data from the aircraft's data bus. A computer can compare the data to predetermined threshold parameters to indicate if the aircraft has experienced a hard landing [16]. The main advantages of this system are that it can easily be retrofit onto any aircraft; it has several measurement units that can be placed close to the landing gears (to avoid the influence of the aircraft's structural dynamics) and it provides a go/no-go indication following each landing.

NASA has developed a bolt-on/bolt-off "Vehicle Health Monitoring Architecture", which could also be used to indicate overload events. This design is made up of a Remote Data Acquisition Unit (RDAU), Command and Control Unit (CCU), and Terminal Collection Unit (TCU). One or more RDAUs are located on the aircraft landing gear, with up to eight sensors being used. In the testing performed, these

Figure 2. Messier-Dowty remote inertial measurement unit (RIMU). [Reproduced with permission. © Messier-Dowty Inc.]

sensors included accelerometers. The CCU is located within the aircraft and all RDAU data are sent to the CCU. The TCU is located at the fleet terminal, and at the end of a flight, the aircraft's CCU analysis is forwarded to the TCU. Here the TCU can determine whether there are any common anomalies within the fleet and forward any relevant information to the maintenance personnel [17].

The major disadvantage of using kinematics to determine landing gear overload is that accelerations, velocities, displacements, roll angle, and roll rates are not direct measurements of the loads experienced by the structure. Ultimately a model needs to be employed to correlate the measured kinematic parameters with the landing gear loads.

Efforts to avoid the measurement of kinematic parameters and the use of models have centered on purely mechanical devices for the indication of an overload. Anecdotal evidence suggests that some Russian aircraft shock absorbers were fitted with lead targets opposite a hard indenter. If the shock strut was compressed beyond the normal range, an indelible mark would be left in the lead target. Another approach to this overtravel indication is a small pin that is permanently deformed upon exceeding the limit. Either example is easily implemented and detected using a visual inspection. The downside is that they provide an indication of vertical overload only—and only if the shock strut is properly serviced at the time of the incident.

An alternative approach to the shock strut overtravel system is to design a structural element that acts as a fuse: it fails in a noticeable way when overloaded. The difficulty in landing gear design is that there are typically no redundancies that would permit a full failure of any component. This has been approached by the development of links that take a permanent set when a limit is exceeded (but has sufficient strength margin to carry the full ultimate load). Another approach developed by Messier-Dowty is the pin-within-a-pin approach. A thin shelled pin is designed to fail at the overload limit, whereas a backup pin within the fuse pin carries loads up to the ultimate load [18]. A dye or other indicator can be used that escapes once the outer fuse pin ruptures, providing a visual indication of the overload [19].

These mechanical solutions are potentially good solutions when knowledge of only one load is required. They are also able to give a simple visual indication that a load has been exceeded, with no data processing. Unfortunately, landing and ground maneuvering loads are multidimensional in nature, and since these mechanical solutions are only able to provide indication in one axis, they provide a small amount of information. Historically, they have also been difficult to implement because of the need for careful calibration to set the correct trigger levels.

The technically superior option is to directly measure the forces acting on the landing gear in all axes. These forces could be measured using pressure or stress/strain. Recording of force data at sufficiently high rates to capture the landing event would provide a definitive answer in overload cases.

4 LANDING GEAR FORCE MEASUREMENT

In the 1950s, the use of accelerometers and strain gauges to directly measure the loads on landing gear was implemented for aircraft test programs. In 1954, NACA developed a method for measuring drag loads on small landing gear using angular and linear accelerometers [20]. In 1958, a large research aircraft was instrumented with strain gauges on the axle, linear, and angular accelerometers, as well as other instruments to measure vertical velocity, angular velocity, drift angle, and mean tire deflections for each wheel. The vertical, drag, and side forces imposed on the MLG under actual loading conditions were then measured [21]. At that time, however, there was concern that strain-gauge instrumentation applied to the landing gear would be subject to the effects of hysteresis and that the results would require correction for inertial effects occurring from the elastic response of the structure [20].

Subsequent work with strain gauges on the landing gear axle to measure landing loads was also the basis for developing alternative weight and balance systems for aircraft. Here, weight refers to the mass of the aircraft, fuel, occupants, and cargo, and balance refers to the position of the center of gravity of the aircraft. Both of these values are calculated before each flight to ensure that the aircraft is within its weight and balance limits.

The benchmark weighing system, a stationary scale onto which the aircraft is rolled to measure weight through the landing gear, is an impractical method for regular use since the weighing procedure usually takes a significant amount of time and scales cannot be easily transported [22]. Therefore, in an attempt to measure the loads of an aircraft through the landing gear, onboard aircraft systems were developed beginning in the late 1960s.

In 1969, BLH Electronics Inc. placed strain-gauge-based transducers in the hollow axles of the landing gear to measure the wheel axle shear forces caused by the vertical loading forces. These vertical loading forces were then used to calculate the gross weight and center of gravity of the aircraft [23]. In 1973, Electro Development Corporation proposed using force transducers, originally intended for an aircraft weight and balance system, on a landing gear axle to measure the shear force upon ground contact [24]. Recently, Boeing presented the idea of placing strain-gauge transducers on the critical components of the landing gear so that forces could be measured [25]. The idea of using strain gauges continually on an aircraft landing gear, and not just in flight testing, has its limitations. For example, strain gauges suffer from a lack of longevity owing to their reliance on adhesive bonding of the gauge to the area of interest. Furthermore, there are issues with corrosion where the electrical leads are terminated to the strain gauges because these terminations exist in a harsh environment. For strain gauges to survive in the landing gear area, they need to be completely encapsulated and protected from the environment.

Recently, Messier-Dowty has been working on the development of load-measuring pins to replace conventional pins in the landing gear. These instrumented pins contain strain gauges that are protected from the environment [26]. The drawbacks of this approach are that the load pins would have to be manufactured for each specific type of landing gear and the cost of retrofitting these load pins would be high. Messier-Bugatti has placed strain gauges into a cell, which could then be fit into the bore of a pin. This would allow for the cells to be easily installed on any type of landing gear, without requiring the cell to be dimensioned specifically for each application [27].

A weight and balance system certified in 1993 for all Airbus A330/A340 type aircraft used four variable reluctance sensors per landing gear to measure the shear deflection of the axle by directly measuring the displacement of the landing gear axle (which is proportional to the aircraft's weight) [2]. These variable reluctance sensors [28] were mounted on lugs, which were milled out of the landing gear axle or bogie beam [29, 30]. The system was expensive owing to the need to machine the lugs on a part that would normally have been turned on a lathe and was also difficult to calibrate and use [22].

Optical displacement methods have also been proposed as a means of measuring the structural deflection and loads on a landing gear axle. Both Airbus [31] and Messier-Dowty [22] have proposed devices using a laser-beam-emitting device connected at one end of the axle and a light position sensing device at the other end of the axle, which are able to determine the position of the laser beam and thereby calculate the load acting through the axle.

There are a variety of other potential methods to measure the loads directly on a portion of the landing gear. The use of fiber Bragg grating sensors (*see* **Fiber-optic Sensor Principles**; **Fiber Bragg Grating Sensors**)—housed in a suitable container—could provide an alternative to strain gauges. Another approach is the use of fixed Barkhausen noise sensors for the measurement of stress in the material [32]. This approach to loads measurement does not require the sensor to be mechanically strained with the loaded component, simplifying installation and increasing robustness.

Pressure can be used to determine landing gear loads. Segerdahl and Greene [33], Lindberg and Thomas [34], and most recently Nance [35, 36] had been developing Elfenbein and Mueller's 1970s approach to determining the weight over a landing gear, by measuring the fluid pressure in the shock strut [37]. Although Elfenbein and Mueller's approach did not compensate for friction inside the shock strut, methods developed by Segerdahl, Lindberg, and Nance did account for this friction. These approaches typically require complicating the shock strut of the landing gear with various valves, tubes, and actuators that by their existence reduce the reliability of the landing gear [22]. However, much of this work on weight and balance systems became the basis for measuring landing gear shock strut fluid levels, monitoring landing gear shock strut health, and

determining overload events, which can be seen in work done by Airbus [38].

5 LANDING GEAR FATIGUE MONITORING

For landing gear structural health monitoring, it is important to know the fatigue state of the landing gear at any point in its life cycle. While the benefits of landing gear servicing determination and event recording are direct and measurable to the landing gear operator, the determination of the landing gear fatigue state is an area where the benefits are not seen immediately, but are seen over the life cycle of the landing gear. For example, with knowledge of the landing gear's fatigue state during its life cycle, it is possible to remove landing gear components prior to failure, resulting in a potential reduction of landing gear–related in-service incidents and therefore increased safety. With these benefits in mind, one might determine the fatigue state of aircraft landing gear by "asking the material", using a sacrificial component or using one of the loads monitoring approaches.

5.1 Ask the material

The ideal approach to fatigue state determination would be to "ask the material" since the actual loading for any component is unique. The ability to ask the material would have to be nondestructive, as it is important to know the fatigue state at any point in the landing gear's life cycle. In addition, there may not be a single solution to determine the fatigue state in metals of different types (steel, aluminum, titanium, etc.) or for different metal chemistries (300M, 4340, etc.).

Given the foregoing, there are some measurement approaches that offer hope of determining fatigue state in steels. Both known techniques correlate the fatigue damage with magnetic properties of the metal. These techniques are the measurement of magnetic permeability, using MWM-array (*see* **Eddy-current Methods**; **Eddy-current *in situ* Sensors for SHM**) sensors offered by JENTEK Sensors Inc., shown in Figure 3, or using the MAPS system offered by ESR Technology, and the Barkhausen noise measurement,

Figure 3. Jentek MWM-array sensors fixed to the landing gear structure. [Reproduced with permission. © Messier-Dowty Inc.]

using sensors provided by a number of companies. These approaches, while having some success demonstrated in the laboratory and in specific in-service instances, still require significant development and correlation effort before they reach maturity.

5.2 Sacrificial component

Another approach to fatigue monitoring is to use a sacrificial component that is subjected to the same loading as the component of interest. Ideally this sacrificial component would have electrical, optical, or other properties that changed as its fatigue life was consumed. Thus, a simple measurement could be made, which could be correlated against the consumed fatigue life of the part of interest.

Such a component has been proposed and developed at a research level. A recent paper by Zhi reports on a modified strain gauge with consumable life [39]. This demonstration builds on the work of D. R. Harting in developing an S/N fatigue life gauge [40]. This type of modified strain gauge offers the promise of a simple approach to fatigue life determination. However, it does have a number of downsides. Extensive testing and correlation work is required before the part goes into service to relate the readings from the gauge to the actual fatigue life consumed in the part. The gauges are for single-axis systems and it may be difficult to locate or correlate on landing gear

components that are loaded multiaxially. The gauge gives no indication of the magnitude of the loads applied and this may be an issue in situations where a large but not excessive load early in the part life is the root cause of a premature fatigue failure.

Another alternative in this category is the integral strain gauge (ISG) offered by ISG Preventive Technologies Inc. of Canada. Essentially a sacrificial foil gauge, the foil material undergoes a visible reaction to increasing fatigue usage. Although unaided visual inspection can provide a rapid check on the fatigue state, microscopy and numerical analysis methods can reveal a more precise determination of the fatigue life consumed [41].

5.3 Loads monitoring

The last approach to fatigue life determination is based on loads monitoring (*see* **Loads Monitoring in Aerospace Structures**; **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**). There are many approaches within this subset as it represents the most established approach to fatigue life determination. Very broadly, there are two subcategories within this group: direct approaches and model-based approaches.

5.3.1 Direct approaches

The two direct approaches are both comparison approaches: applied load comparison and energy comparison.

Applied load comparison
In the applied load comparison approach, the loads into the landing gear are measured or otherwise determined and compared against the design loads, which are specified by Federal Aviation Regulations (FAR) or Joint Aviation Regulations (JAR). Any excessive load is then flagged for analysis.

This idea is expressed in an invention by Delest *et al.* [42] where, in a preliminary step, design loads, which are representative of the loads that the aircraft will be subjected to, are defined. At the moment of landing, input parameters such as the mass of the aircraft, accelerations, roll angle, roll rates, and load data are gathered from sensors. A computer receives these input parameters and load data, and a comparison algorithm evaluates the real values measured against the calculated allowable value of the design loads.

However, this applied load comparison approach gives, at best, only a rough estimation since the input loads can vary in so many ways from the design loads, whether in magnitudes, combinations, or repetitions. Therefore, a comparison in a real life situation would be difficult, if not impossible, to make.

Energy comparison
The energy comparison approach measures the loads input into the landing gear and the loads from the landing gear into the aircraft structure. When these loads are compared on an energy basis, any difference represents energy absorbed by the landing gear and therefore material damage.

The energy comparison approach represents a method for determining landing gear damage without requiring extensive time histories to be recorded. However, there are some significant challenges. The first is that the landing gear contains an energy absorbing device, the shock strut. This device, when operated nominally converts the energy from landing into heat. Any energy comparison device would have to exclude the shock absorbing element from its analysis, or else monitor the shock strut to quantify its energy conversion. The other challenge is the sheer number of measurements required to be made: landing gear wheel or axle input loads, landing gear attachment loads, and shock strut energy conversion. Beyond these complexities, the concept is very clever—the difference between energy in and energy out, when corrected for shock strut effects (by employing a numerical model of the dynamic response of the oleo–pneumatic strut), is energy that went into damaging the material of the landing gear.

5.3.2 Model-based approaches

The model-based approaches represent the most conventional approaches to determining part damage. Essentially, the landing gear wheel or axle input loads are determined by directly measuring them or by inferring them from aircraft reference frame measurements. These loads are saved as a time history for later analysis. A plain analysis of the input

loads is equivalent to the applied load comparison approach described in the section titled "Applied load comparison". However, if the input loads are converted into landing gear structural loads by developing a structural analysis model of the gear, such as a beam model, time histories of the applied load for each structural element of the landing gear can be developed. A fatigue analysis of each element could then be conducted to determine the amount of damage in each part.

Messier-Dowty has developed a strain logger concept that has been certified for installation by Transport Canada on the Bombardier Regional Jet landing gear. The technology, developed as a 4-channel or a 12-channel device, combines pulsed-power excitation to the strain gauges with onboard data analysis and a combined battery–memory module to facilitate in-field removal and replacement of the memory unit [26]. The four-channel strain logger can be seen in Figure 4.

The strain logger process of determining the life consumed in a landing gear part is as follows: measurement of in-service loads, conversion of these loads to input ground loads, conversion of these loads to critical section load histories by means of the landing gear beam model, rainflow counting of critical section loads, and computation of the damage in the part considering the section and material properties.

The above-mentioned process involves a significant amount of computation. To ultimately perform onboard computation of part damage, it is proposed to precalculate a matrix of damage coefficients for a given section. Each damage coefficient would relate to the damage input by one cycle of a particular range–mean pair of loads. The coefficients would be calculated to match the range–mean pairs generated by the rainflow counting of loads for a specific section. Thus, the calculation of damage would become a multiplication and accumulation calculation. The dot product of the rainflow count matrix and the damage matrix could be taken. Then a summation of all resulting values could be taken to arrive at the damage number. This approach would reduce the onboard computation difficulties to the calculation of the input ground loads (algebraic computations) and of running of the beam model.

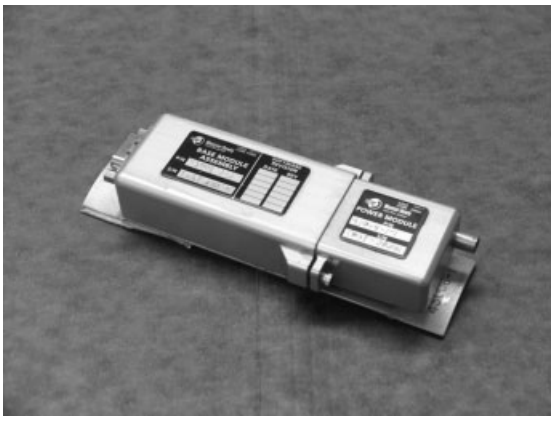


Figure 4. Messier-Dowty four-channel strain logger. [Reproduced with permission. © Messier-Dowty Inc.]

There are some significant challenges with this admittedly brute force approach. Data retention over long periods of time would be necessary. Extensive mathematical analysis capability would be required such that each landing gear component in the fleet would possess its own fatigue spectrum that would have to be updated regularly. Data must be conserved across a significant number of aircraft; an approach must be developed to deal with missing or incorrect data; and the infrastructure must exist to track damage data with a part-through operation and maintenance cycles. Lastly, the model-based approach still suffers from the limited understanding of the mechanisms of fatigue, such as scatter factor, a value used to account for statistical variation in fatigue.

Even given the limitations of the understanding of fatigue, the concept of measuring landing gear loads and recording them for later analysis provides significant advantages. Some of these advantages include utilizing the data by performing "one-off" analyses of the loads for incident investigations, performing a statistical analysis of ground loads so that fleet behavior could be tracked (usage outside of the norms could be identified), and collecting data that could be used to refine the design values for future aircraft.

6 CONCLUSION

Aircraft landing gear systems continue to develop. Structural and performance monitoring is being added to landing gears in operations where an economic or safety need exists. At the time of writing this article, most of the large aircraft landing gear systems

include monitoring of the brakes and tires, both to increase safety and to increase aircraft utilization. New aircraft are incorporating monitoring systems for the shock absorber in order to permit reduced maintenance effort.

Several technologies exist to permit more direct monitoring of the landing gear structure and to permit the material condition to be directly measured. Work continues in the development of robust sensors and effective algorithms to allow the commercial fielding of structural health monitoring for landing gear that offers reduced total cost of ownership along with improved awareness of the landing gear's condition.

As aircraft structures, and indeed landing gear systems, continue on their inexorable march toward lower weight, the ability of structural monitoring to track component life and identify potential failures before they occur will become indispensable.

ACKNOWLEDGMENTS

The authors would like to thank the following friends, colleagues, and fellow employees (past and present) of Messier-Dowty for their invaluable assistance in the development of landing gear structural monitoring systems and the drafting of this article: Mr Cosimo Commisso, Mr Lukasz Hul, Mr Stuart Lacy, Mr Ian McCluskey, and Ms Julia Payne.

REFERENCES

[1] Decourselle F, Messier-Bugatti, General Press Kit, *A Global Player in Aircraft Braking*, 16 August 2007, http://www.messier-bugattiusa.com/IMG/pdf/press file2007en.pdf.

[2] Kehlenbeck U. Airbus A340 weight and balance system. *58th Annual Conference of Society of Allied Weight Engineers*. San Jose, CA, 24–26 May 1999.

[3] Currey NS. *Aircraft Landing Gear Design: Principles and Practices*. American Institute of Aeronautics and Astronautics: Washington, DC, 1988.

[4] Pradier J-C, Gautier A. *Installation for Measuring Pressure of at Least One Airplane Wheel Tire*, US Patent 6,959,596, 1 November 2005.

[5] Products, Sensing & Utility Systems Solutions, Crane Aerospace & Electronics, 6 September 2007, http://www.craneae.com/Solutions/Sensing%20& %20Utility%20Systems/SmartStem.htm.

[6] Miller RJ, Marshall RJ, Bailey DA, Griffin NC. *Brake Condition Monitoring*, US Patent 2007/0007 088, 11 January 2007.

[7] Luce WE. *Aircraft Shock Strut having a Fluid Level Monitor*, Canadian Patent 2 500 458, 8 April 2004.

[8] Woodward SE. *A Wireless Fluid-Level Measurement Technique*, NASA Technical Memorandum 214320, 2006, p. 34.

[9] Allison SG. *Ultrasonic Measurement of Aircraft Strut Hydraulic Fluid Level*, NASA Langley Research Center 02WAC-1931, August 2007, http:// techreports.larc.nasa.gov/ltrs/PDF/2002/mtg/NASA-2002-wacd-sga.pdf.

[10] Seror C. *Method of Measuring the Compression of a Shock Absorber, and an Airplane Undercarriage Constituting an Application Thereof*, US Patent 2005/0230200, 20 October 2005.

[11] Monitoring Systems, OPMS, Messier-Bugatti, SAFRAN Group, 6 September 2007, http://www.messier-bugatti.com/rubrique.php3?id_rubrique = 61 &lang = en.

[12] TATEM, Technologies and techniques for new maintenance concepts. *6th European Research Framework Programme*, 17 August 2007, http://www.tatem project.com/.

[13] Zahm AF. Development of an airplane shock recorder. *Journal of the Franklin Institute* 1919 **88**(2):237–244.

[14] Brevoort M. *Landing-Shock Recorder*, National Advisory Committee for Aeronautics Technical Note NACA-TN-501, 1934.

[15] Finance R. *Detecting Hard Landings of Aircraft*, UK Patent 2,014,731, 30 August 1979.

[16] Schmidt RK. *System and Method for Determining Aircraft Hard Landing Events from Inertial and Aircraft Reference Frame Data*, International Patent WO 2006/130984, 14 December 2006.

[17] Woodward SE, Coffey NC, Gonzalez GA, Taylor BD, Brett RR, Woodman KL, Weathered BW, Rollins CH. Development and flight testing of an adaptable vehicle health monitoring architecture. *Journal of Aircraft* 2004 **41**(3):531–539.

[18] Messier-Dowty, *Shear Pin*, UK Patent Application Number 0718296.7, 19 September 2007.

[19] Messier-Dowty, *Overload Detection*, UK Patent Application 0718297.5, 19 September 2007.

[20] Theisen JG, Edge PM. *An Evaluation of an Accelerometer Method for Obtaining Landing-Gear Loads*, NACA Technical Report 3247, 1954.

[21] Hall AW, Sawyer RH, McKay JM. *Study of Ground-Reaction Forces Measured During Landing Impacts of a Large Airplane*, NACA Technical Report 4247, 1958.

[22] Schmidt RK, El-Samid NA. *Structural Deflection and Load Measuring Device*, International Patent WO 2006/024146, 8 March 2006.

[23] Kadlec C. *Aircraft Weight Measurements*, US Patent 3,426,586, 11 February 1969.

[24] Harris CL, Rama LC, Soward DV. *Aircraft Hard Landing Indicator*, US Patent 3,712,122, 23 January 1973.

[25] Cowan SJ, Cox RL, Slusher HW, Jinadasa S. *Airplane Hard Landing Indication System*, US Patent 6,676,075, 13 January 2004.

[26] Schmidt RK. *Monitoring Parameters in Structural Members*, UK Patent 2,387,912, 29 October 2003.

[27] Dellac S, Lafaye E. *Force-Measurement Cell and a Connection Pin Fitted with such A Cell*, US Patent 2006/0266561, 30 November 2006.

[28] Nelson HK, Kleingartner CA, Vetsch LE. *Strain/Deflection Sensitive Variable Reluctance Transducer Assembly*, US Patent 4,269,070, 26 May 1981.

[29] Patzig H-N, Schult K. *Arrangement of Sensors on the Landing Gear of an Aircraft for Measuring the Weight and Position of Center of Gravity of the Aircraft*, US Patent 5,257,756, 2 November 1993.

[30] Patzig H. *Method for Calibrating Sensors Arranged in Pairs on Loaded Structural Parts*, US Patent 5,239,137, 24 August 1993.

[31] Giazotto, ARB, *Optically Measuring the Dispacing or Load for an Aircraft Component, Landing Gear, Braking Control*, US Patent 2007/0006662, 11 January 2007.

[32] Kehlenbeck U, Vengrinovich V, Denkevich Y, Tsukerman V. Onboard aircraft weighing system using barkhausen noise sensors. *7th European Conference on Non-Destructive Testing*. Copenhagen, 26–29 May 1998.

[33] Segerdahl RR, Greene SI. *Method for Reducing Frictional Error in Determining the Weight of an Object Supported by a Pneumatic or Hydraulic Device*, US Patent 3,581,836, 1 June 1971.

[34] Lindberg GR, Thomas HO. *On-board Aircraft Weighing and Center of Gravity Determining Apparatus and Method*, US Patent 5,521,827, 28 May 1996.

[35] Nance K. *Aircraft Weight and Center of Gravity Indicator*, US Patent 5,548,517, 20 August 1996.

[36] Nance K. *Method of Determining Status of Aircraft Landing Gear*, US Patent 6,293,141, 1 September 2001.

[37] Elfenbein JA, Mueller MC. *Aircraft Weight and Center of Gravity Computer*, US Patent 3,513,300, 19 May 1970.

[38] Yates MS, Keen P. *Landing Load Monitor for Aircraft Landing Gear*, International Patent WO 2007/023280, 1 March 2007.

[39] Zhi Z, Duan Z, Jia Z, Ou J. New kind of structural fatigue life prediction smart sensor. *Proceedings—SPIE the International Society for Optical Engineering, Smart Structures and Materials 2004: Smart Sensor Technology and Measurement Systems*, March 2004; Vol. 5384, pp. 324–331.

[40] Harting DR. The S-N fatigue life gage: a direct means of measuring cumulative fatigue damage. *Experimental Mechanics* 1966 **6**:19–24.

[41] Oudovikine A. *Sensing Method Predicts the Fatigue Life of Materials*. ISG Preventative Technology, 17 August 2007, www.preventativetechnology.com.

[42] Delest T, Regis O, Schuster P. *Method and Device for Detecting that the Design Loads of an Aircraft have been Exceeded*, US Patent 5,511,430, 30 April 1996.

Chapter 117

Health Monitoring, Diagnostics, and Prognostics of Avionic Systems

Michael Pecht[1] and Yan-Cheong Chan[2]

[1] *Center for Advanced Life Cycle Engineering (CALCE), University of Maryland, College Park, MD, USA*
[2] *Electrical Engineering Department, City University, Kowloon, Hong Kong, China*

1 INTRODUCTION

Flight management and engine control of today's airplanes are highly dependent on electronics. However, despite the importance of electronics to all modern aircraft, avionics have a relatively low-volume market, and have many special requirements, especially in terms of emphasis on mission success, safety, and long-term reliability. Nevertheless, affordability of leading-edge technologies mandates that the avionics industry is largely dependent on commercial-off-the-shelf (COTS) electronic components, modules, and assemblies not originally designed for avionics (but for the consumer, computers, and telecom industry). The result is that COTS electronics reliability and prognostics (the forecasting of failure and prediction of remaining useful life) is a major concern of the aerospace community [1]. This concern is compounded by other concerns, such as part obsolescence, high maintenance costs, and the aerospace's response to the potential for a worldwide legislative ban on lead and other hazardous materials in all electronic devices.

The first efforts in health monitoring of avionics involved the use of built-in test (BIT), which is defined as an onboard hardware–software diagnostic means to identify and locate faults. A BIT can consist of error detection and correction circuits, totally self-checking circuits, and self-verification circuits. Two types of BIT concepts have been employed in avionic systems: interruptive BIT (I-BIT) and continuous BIT (C-BIT). The concept behind I-BIT is that normal equipment operation is suspended during BIT operation. The concept behind C-BIT is that equipment is monitored continuously and automatically without affecting normal operation.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Studies [2, 3] conducted on the use of BIT for fault identification and diagnostics showed that BIT can be prone to false alarms and can result in unnecessary costly replacement, requalification, delayed shipping, and loss of system availability. However, there is also reason to believe that many of the "false alarms" were "real" failures, but intermittent in nature [4]. In any case, BIT has generally not been designed to provide prognostics or remaining useful life. Rather, it has served primarily as a static diagnostic tool.

With increasing functional complexity of onboard autonomous avionic systems, there is an increasing demand for system-level health assessment, fault diagnostics, and prognostics, especially to meet critical safety and mission requirements. This is of special importance for soft faults and intermittent failures, which are some of the most common failure modes in today's avionics.[a] As a result, there is a need for effective and efficient reliability prognostics for avionic systems using algorithms that can fuse sensor data, discriminate transient (intermittent), and false alarms from actual failures, correlate faults with relevant system events (e.g., aircraft mode changes), and reduce redundant processing elements that are subject to common mode failures. Finally, there is the need to assess remaining useful life to aid in forecasting maintenance and system life.

Prognostics and health monitoring (PHM) techniques combine sensing, recording, and interpretation of environmental, operational, and performance-related parameters to assess a system's health and health trends (often in terms of remaining life). Product health monitoring can be implemented through the use of various techniques to sense and interpret the parameters indicative of performance degradation, such as deviation and trending of operating parameters from their expected values; changes in physical or electrical degradation, such as material cracking, corrosion, interfacial delamination, electrical resistance, or threshold voltage; and changes in a life-cycle environment, such as usage duration and frequency, ambient temperature and humidity, vibration, and shock.

2 APPROACH TO AVIONICS HEALTH MONITORING, DIAGNOSTICS, AND PROGNOSTICS

A heath monitoring and prognostics framework for avionics is shown in Figure 1. The first step involves a virtual life assessment, where design data, expected life-cycle conditions, failure modes, mechanisms, and effects analysis (FMMEA), and physics-of-failure (PoF) models, which were used in the product design for reliability (DfR), are the inputs to obtain a reliability (virtual life) assessment. On the basis of the virtual life assessment, it is possible to prioritize the critical failure modes and failure mechanisms. The existing sensor data, bus monitor data, and

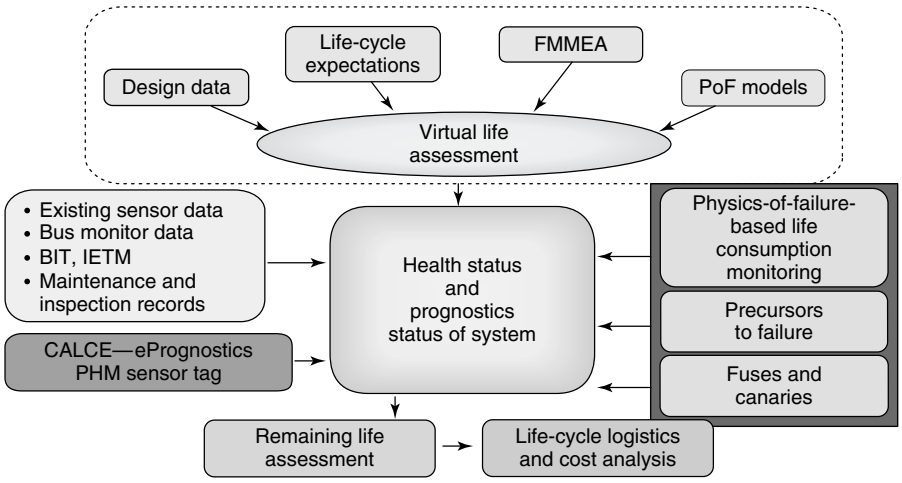


Figure 1. Prognostics and health management methodology for avionics systems.

maintenance and inspection records can also be used to identify potential weak items, as well as potential abnormal conditions and parameters. On the basis of this information, the monitoring parameters and sensor locations for PHM can be determined.

The input for the fault detection (diagnostics) and prognostics is the real-time system-level sensor data of the operating and environmental conditions. Operating condition data can also comprise the output of canary circuits (or fuse networks), as well as soft variables such as system loads and other performance metrics.

On the basis of the sensor and other input data, three approaches, including combinations of these approaches, to prognostics for avionics can be implemented. These are (i) the use of fuses and canary devices; (ii) monitoring and reasoning of failure precursors; and (iii) monitoring environmental and usage loads for damage modeling. It is then possible to assess the product life-cycle usage profile, health, and conduct failure prognostics.

2.1 Health monitoring and extraction of anomalies

Health management systems are programs that respond in a preemptive and opportunistic manner to the anticipation of failures. The health state of an avionics system is characterized by extracting features from the real-time data, taking into account varying operational (parameter) ranges and profiles that constitute a "healthy system".

Generally, a healthy system is defined from a training model. The training model captures "known healthy" system data and uses feature extraction to establish conditions of acceptable health. These conditions are then used to assess whether a monitored system can be classified as healthy.

Feature extraction data is also fed into fault algorithms that generate classifications and probabilities of "unhealthy" or fault states as well as define the scales appropriate for fault detection (i.e., some faults are detected on a nanosecond timescale, while other degradation mechanisms can have failure timescales that are equivalent to the system life). An anomaly can be further classified by analyzing it through fault algorithms, using statistical approaches for trend analysis. When the fault algorithms show positive trending or meet a fault classification, then the next step is to conduct fault and/or failure prognostics.

2.2 Parameter and fault isolation

The objective of the fault prognostic algorithms is to assess faults (parameter and fault isolation) and then predict failures within a reasonable time and prevent false alarms as well as missed alarms. Missed alarms are prevented by the fault prognostic algorithms independently accessing the health monitoring data and analyzing the probability of transition from a health state to a fault state. Evaluation of the transition probability, as well as the fault classification, is used to identify potential precursor(s) to failure. A no-fault-found (NFF) event may not have a well-defined (classified) fault; however, it can still change the fault prognostics by changing the transition probabilities of the states in the data-driven approaches and thus can be captured. Then, by combining data-driven fault isolation with PoF, the fault can be isolated and the root cause for failure identified.

Damage can be calculated from the PoF models to obtain the remaining life. Then PHM information can be used for maintenance forecasting and decisions that minimize life-cycle costs and maximize availability of some other utility function.

The output from fault prognostic algorithms can be stored for off-line analysis. Off-line models derived from real field use data offer more effective and efficient maintenance planning and can also influence mission scheduling, future designs, as well as line replaceable unit (LRU) selection and sparing. Another advantage of these off-line models is that maintenance algorithms can be tailored to look at much longer or finer windows of data that may be computationally intensive. For example, printed circuit board (PCB) faults are usually intermittent in nature and may not fully reveal their nature during the on-line model analysis. However, these intermittents can be tracked and correlated to assess features including excessive delay, electromagnetic interference, and cross talk [4]. Using PoF techniques and off-line models, failure signatures can be developed and further abstracted into failure states that can then be implemented into the on-line framework.

3 RELIABILITY REQUIREMENTS FOR ANALYSIS

Three key requirements that the avionics community considers to be critical in the implementation of diagnostic and prognostic techniques and algorithms are as follows: (i) robust on-line data fusion and state determination to discriminate transient and false alarms from actual failures, (ii) data analysis algorithms to correlate "soft" avionics faults with other relevant events (e.g., aircraft mode changes), and (iii) prognostic algorithms to track degradation and predict remaining useful life for avionics systems. The following sections discuss these requirements in terms of algorithm features and dependencies and how they relate to system reliability.

3.1 Robust on-line data fusion and state determination algorithms to discriminate transient and false alarms from actual failures

Failures in complex electronic systems can be extremely difficult to isolate and identify [5]. An electronic system that has been observed to fail in the field often functions correctly during subsequent fault-finding activities. Consequently, fully functional units are often replaced, which leads to ineffective cost management, or worse, to maintenance practices that permit returning a potentially faulty unit to the field, which can be a safety hazard [2].

Various terms such as "cannot duplicate (CND)", "retest OK (RTOK)", "no fault indicated (NFI)", "NFF" and "no trouble found (NTF)" are used to describe this phenomenon. These are also defined as *soft* faults. Typical causes of soft faults include transient failures due to particle radiation, power supply fluctuations, intermittently occurring faults due to loose connections, partially defective or deteriorating components, and poor hardware design. A soft fault also arises due to the inability of the laboratory environment to duplicate the field load conditions exactly and due to the self-healing of failures such as solder joint cracks during testing. In certain cases, the occurrence of soft faults indicates the use of inappropriate diagnostic procedures or that the total fault spectrum is larger than the fault coverage.

Intermittent failure causes of NFF can be grouped into five categories: PCBs, components, interconnects, connectors, and software. Figure 2 is a fishbone diagram for intermittent failures at an assembly (LRU) level.

In situ diagnostics is necessary to catch most soft failures. On-line data fusion refers to an algorithm's ability to analyze performance variables on all scales of actual measurement *in situ* to operation. At the same time, diagnostic and prognostic algorithms must be able to discriminate between events related to faults and common mode operational process outliers. For example, outliers can be present in the data due to natural changes in a process, due to intrusive changes, or simply due to noise, instrumentation, and human error.

One approach has been to center-normalize the data to remove differences in scale. Variants of this approach may be necessary to enable customization of scales and reduce noise dependent on the measurement metrics. Reduction and filtering techniques (affine projection using principal component analysis (PCA), Sequential Bayesian filtering, and regressive moving average) are ways to separate noise and extremes in data.

State determination in PHM refers to an algorithm's ability to make decisions pertaining to the health trends of a system. Data trending methods can be used to detect potential anomalies and, in some cases, to discriminate transient and false alarms from actual failures, but these methods only examine changes from a predefined normal (healthy) state. Unless there exists a history of failure data (degradation patterns in which the system degrades and ultimately fails), data trending is not sufficient to predict failures. In such cases, the association of an anomaly and trend to a failure mechanism and failure mode requires a PoF approach.

3.2 Data analysis algorithms must correlate "soft" avionics faults with other relevant events (e.g., aircraft mode changes)

To correlate "soft" avionics faults with other relevant events, especially intermittent failures and false errors, a correlation scheme of the system parameters linked to the related system characteristic or features

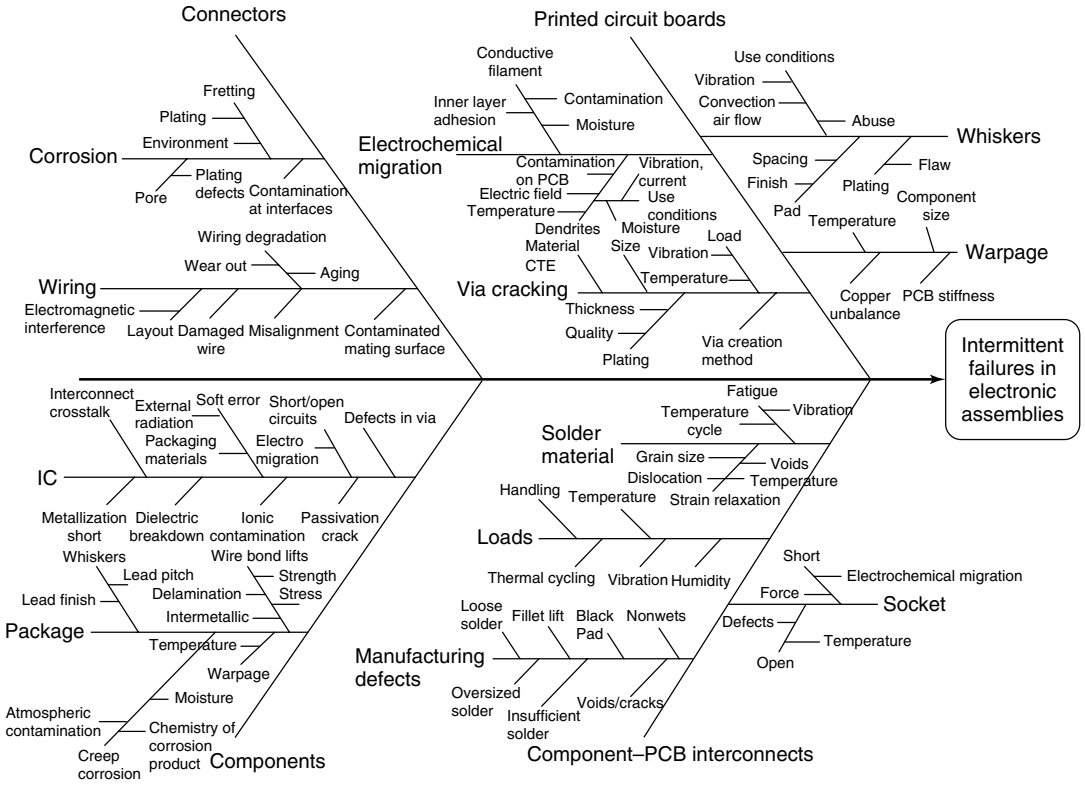


Figure 2. Cause-and-effect diagram for intermittent (soft) failures in electronic line replaceable units (LRU). [Courtesy of CALCE—University of Maryland.]

is a preferred approach. The critical task, however, is to select criteria to uncover the associations to the event of interest (e.g., aircraft mode changes). Again, this requires understanding of failure modes and mechanisms, which are the links to system-level performance. One approach is to use FMMEA to aid in the analysis.

FMMEA is based on understanding the relationships between product requirements and the physical characteristics of the product (and their variations in the production process), the interactions of product materials with loads (stresses at application conditions), and their influence on the product's susceptibility to failure with respect to the use conditions. FMMEA combines life-cycle, environmental, and operating conditions and the duration of the intended application with knowledge of the active stresses and potential failure mechanisms.

FMMEA prioritizes the failure mechanisms based on their occurrence and severity to provide guidelines for determining the major operational stresses and environmental and operational parameters that must be accounted for in the design or be controlled. The life-cycle profile is used to evaluate failure susceptibility. If certain environmental and operating conditions are nonexistent or generate a very low-level stress, the failure mechanisms that are exclusively dependent on those environmental and operating conditions are assigned low occurrence. Severity ratings are obtained from the failure modes associated with the mechanism and there can be more than one failure site mode for the same mechanism. The high-priority failure mechanisms identified through combination of occurrence and severity are the critical mechanisms. Each critical failure mechanism has one or more associated sites, modes, and causes in an FMMEA result.

Failure mechanisms are the physical, chemical, thermodynamic, or other processes that result in failure. Failure mechanisms are categorized as either

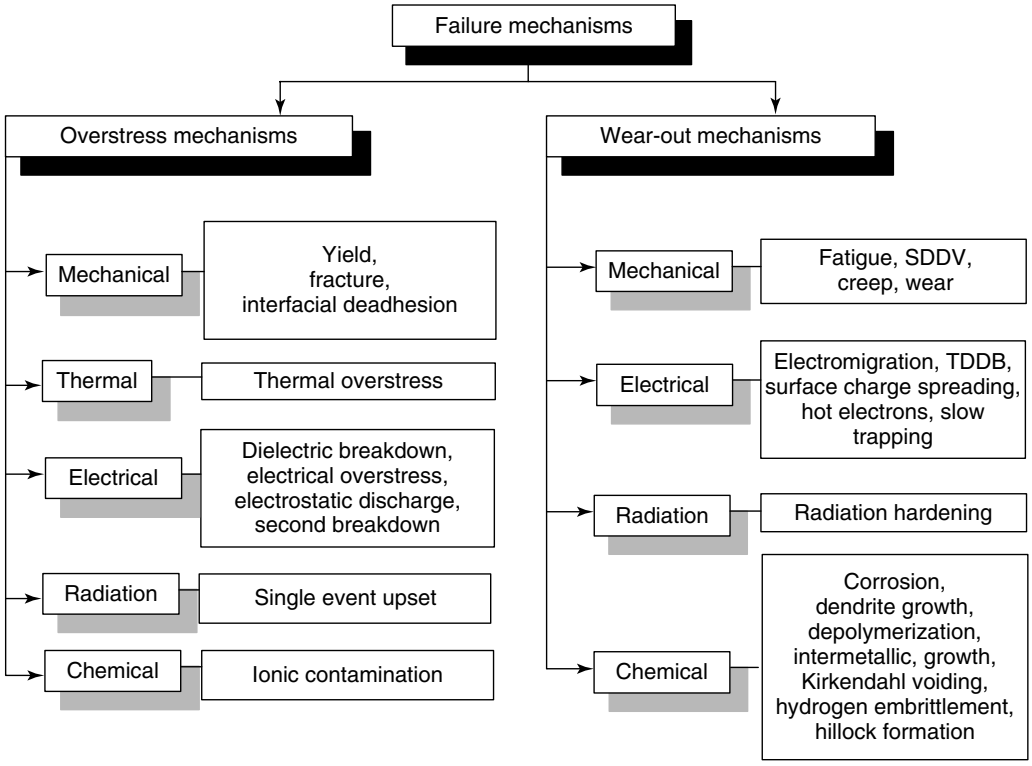


Figure 3. Example of failure mechanism breakdown into overstress and wear out for electronic products.

overstress or wear-out mechanisms (see Figure 3). Overstress failure arises as a result of a single load (stress) condition, which exceeds a fundamental material strength. Wear-out failure arises as a result of cumulative damage due to loads (stresses) applied over an extended time or number of cycles. Within current technology, prognostics can only be applied to the wear-out failure mechanisms [6].

3.3 Prognostic algorithms must track degradation and predict remaining useful life

Prognostic algorithms take time-dependent data history, extract a trend, and estimate the value of the function at some future time. The objective function for prognostics is the health state of the system. Time series, neural networks, support vector machines (SVMs), Markov chains, and parameter estimation are some popular techniques for forecasting avionic systems. However, PoF prediction is most often required to set the failure criterion and make the determination of the remaining useful life.

Mathew *et al.* [7, 8] applied a PoF-based PHM methodology to conduct a prognostic remaining-life assessment of circuit cards inside the space shuttle solid rocket booster (SRB). Vibration time history recorded on the SRB from the prelaunch stage to splashdown was used in conjunction with physics-based models to assess damage. Using the entire life-cycle loading profile of the SRBs, the remaining life of the components and structures on the circuit cards was predicted. It was determined that an electrical failure was not expected within another 40 missions.

Tuchband *et al.* [9] presented the use of prognostics for a military avionics LRU based on their life-cycle loads. The study was part of an effort funded by the Office of the Secretary of Defense to develop an interactive supply-chain system for the US military. The objective was to integrate prognostics, wireless communication, and databases through

a web portal to enable cost-effective maintenance and replacement of electronics. The study showed that prognostics-based maintenance scheduling could be implemented into military electronic systems. The approach involved an integration of embedded sensors on the LRU, wireless communication for data transmission, a PoF-based algorithm for data simplification and damage estimation, and a method for uploading this information to the Internet. It was shown that the use of prognostics for electronic military systems could enable failure avoidance, high availability, and reduction of life-cycle costs.

4 PROGNOSTICS

The identification of the variation trend in the parameter values of a system can provide information on drift and/or degradation in performance. Time series analysis techniques are efficient tools for capturing such trends. Other methods being considered by the avionics community include autoregressive moving average (ARMA), autoregressive with exogenous inputs (ARX), autoregressive moving average with exogenous inputs (ARMAX), and autoregressive integrated moving average (ARIMA). Symbolic time series analysis (STSA) is an extension of these traditional time series analysis techniques; it discretizes the continuous dynamic system in time and space, and represents the trajectories of observed variables/parameters as a sequence of symbols. Although time series methods perform well in parametric degradation forecasting, they are not well suited for failure forecasting, which requires prediction algorithms (usually PoF) capable of working with abrupt and quantized data.

SVMs are also being used for forecasting the system health [10, 11]. SVMs minimize an upper bound of the generalization error, through the use of kernel functions and the structural risk minimization principle, and generally achieve higher generalization performance than traditional neural networks in time series forecasting. However, they also suffer from difficulties in predicting failure occurrence without the use of PoF models.

Precursors to failure are the characteristic features that will help enable predictions of the system health. However, the metrics to assess the precursors is of critical importance in this step. For example, in an experiment conducted on intermittent failures of LRUs, it was shown that data trending prognostic methods provided no value. However, by embedding PoF relationships to directly modify scales of measurement, tremendous insight can be obtained and subsequent predictions made [12].

5 CONCLUSIONS

The avionics community has begun to implement fault event detection techniques using PoF-based methods for successful diagnostics. These include algorithms that extract cyclic ranges and means, ramp rates, dwell times, and dwell loads, and then correlate all these load parameters with PoF models to enable fault isolation and prognostics. In some instances, statistical projection pursuit, feature extraction, and pattern-recognition algorithms have also been used to detect and isolate faulty parameters, but mostly in experimental studies (products subject to accelerated aging conditions). In this case, PCA techniques have been utilized to build the projection models, cluster analysis techniques of multivariate data sets were used for anomaly detection, and multivariate state estimation technique (MSET), sequential probability ratio test (SPRT), and PoF-based hybrid methods were used to conduct the prognostics.

The next step for the avionics prognostics community is to extend these studies and provide

- efficient and cost-effective root-cause evaluation tools (software), which couple the PoF and time series PHM approaches to discriminate false alarms from actual failures;
- rapid fault identification and prognostics (software), which use SVM and STSA coupled with feature extraction, pattern recognition, and PoF analysis to predict system degradation, malfunctions, and remaining useful life of critical avionic systems;
- validated prognostic methods (algorithms) to detect and predict degradations, malfunctions, and failures on an LRU/line replaceable module (LRM) (including intermittent failures, soft faults, and random false alarm events); and
- demonstrated methods and algorithms to determine return on investment benefits.

END NOTES

[a.] An intermittent failure, which can lead to diagnostic conclusions of "CND", "RTOK", "NFI", "NFF"), and "NTF", is the loss of function for a limited period of time, and subsequent recovery of the function. This "failure" may not be easily predicted, nor is it necessarily repeatable. However, an intermittent failure can be, and often is, recurrent.

REFERENCES

[1] Vichare N, Zhao P, Das D, Pecht M. Electronic hardware reliability, avionics development and implementation. *Section 1–4: Development, Digital Avionics Handbook*, Second Edition, CRC Press, 2007, 4-1–4-22.

[2] Pecht M, Dube M, Natishan M, Knowles I. An evaluation of built-in test. *IEEE Transactions on Aerospace and Electronic Systems* 2001 **37**(1): 266–272.

[3] Johnson D. Review of fault management techniques used in safety critical avionic systems. *Progress in Aerospace Science* 1996 **32**(5):415–431.

[4] Williams R, Banner J, Knowles I, Natishan M, Pecht M. An investigation of 'cannot duplicate' failure. *Quality and Reliability Engineering International* 1998 **14**:331–337.

[5] Pecht M, Ramappan V. Are components still the major problem: a review of electronic system and device field failure returns. *IEEE Transactions on Components, Hybrids, Manufacturing Technology* 1992 **CHMT-15**:1160–1164.

[6] Pecht M, Dasgupta A. Physics-of-failure: an approach to reliable product development. *Journal of the Institute of Environmental Sciences* 1995 **38**:30–34.

[7] Mathew S, Das D, Osterman M, Pecht M, Ferebee R. Prognostic assessment of aluminum support structure on a printed circuit board. *International Journal of Performability Engineering* 2006 **2**(4):383–395.

[8] Mathew S, Das D, Osterman M, Pecht M, Ferebee R, Clayton J. Virtual remaining life assessment of electronic hardware subjected to shock and random vibration life cycle loads. *Journal of the IEST* 2007 **50**(1):86–97.

[9] Tuchband B, Pecht M. The use of prognostics in military electronic systems. *Proceedings of the 32nd GOMAC Tech Conference*. Lake Buena Vista, FL, 19–22 March, pp. 157–160, 2007.

[10] Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. *Least Squares Support Vector Machines*. World Scientific Publishing: London, Singapore, 2002.

[11] Muller KR, Smola AJ, Ratsch G, Scholkopf B, Kohlmorgen J. Using support vector machines for time series prediction. In *Advances in Kernel Methods—Support Vector Learning*, Scholkopf B, Burges CJC, Smola AJ (eds). MIT Press: Cambridge, 1999; pp. 243–254.

[12] Vichare N, Pecht M. Prognostics and health management of electronics. *IEEE Transactions on Components and Packaging Technologies* 2006 **29**(1): 222–229.

Chapter 110

Design Benefits in Aeronautics Resulting from SHM

Hans-Juergen Schmidt and Bianka Schmidt-Brandecker

AeroStruc—Aeronautical Engineering, Buxtehude, Germany

1 INTRODUCTION

The primary objective of the aerospace industry is to offer products that meet improved goals in terms of payload and range, resulting in a significant reduction of the direct operating cost (DOC) for the airlines. These goals require an integrated approach considering advanced materials, new design principles and improved inspection procedures, e.g., structural health monitoring (SHM).

One of the primary focuses of SHM applications has been on the reduction of the weight as proposed by Schmidt and Schmidt-Brandecker at Airbus [1] for metallic structures and by Goggin *et al.* at Boeing [2] for structure made of carbon fiber reinforced plastics (CFRPs).

Modern aircraft have to meet the so-called damage tolerance (DT) regulations according to the Paragraph FAR (Federal Aviation Regulations) 25.571-Amendment 96 independent of the material, design principles, and manufacturing methods. DT is the attribute of the structure to withstand fatigue, corrosion, manufacturing defects, or accidental damage without catastrophic failure throughout the operational life of the aircraft. This can only be assured with an adequate inspection program based on DT evaluation determining crack growth and residual strength, i.e., the number of flights between a detectable and critical damage (crack) as shown in Figure 1.

DT is the dimensioning criterion for the majority of the fuselage and wing panels; i.e., this criterion is responsible for the weight of these panels. The fuselage and wing panels are stiffened panels, which consist of the skin and internal stringers as well as frames or ribs, respectively. With respect to the difficult access to the internal structure, the traditional inspection concept consists of repetitive inspections of the external skin and less inspections of the internal structure. Consequently, the DT evaluations have to consider a "skin crack above a broken internal member", which is the most severe scenario and leads

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

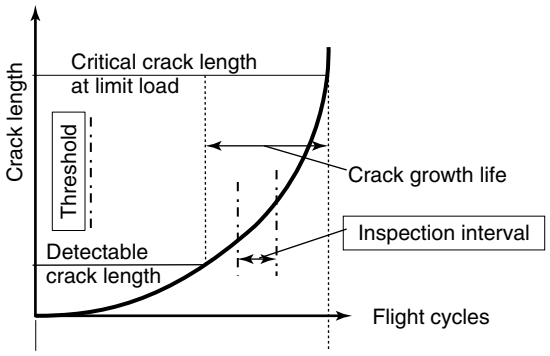


Figure 1. Primary objective of damage tolerance evaluation.

to the shortest crack growth period and the lowest allowable stresses for design.

The knowledge about the status of the internal members of in-service aircraft can be significantly improved by the application of SHM. In this case the DT analysis may consider less severe crack scenarios, e.g., "skin crack above an intact internal member" and/or "skin crack between intact members". Both scenarios lead to longer crack growth periods and higher allowable stresses. The advantage gained by the modified inspection procedure, i.e., external visual and SHM, can be used in different ways as shown in Figure 2:

- increased inspection interval while maintaining the allowable stress level and
- increased allowable stress level while maintaining the inspection interval.

Both improvements are of interest for the manufacturer and airlines. However, the design benefits seem

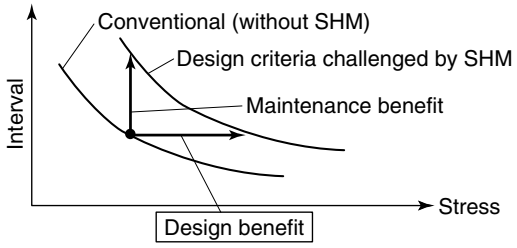


Figure 2. Principal maintenance and design benefits due to the application of SHM.

to be more interesting, because a significant weight saving can be realized to reduce the DOC.

In general the metallic fuselage structure is dimensioned for several design criteria. SHM provides significant weight or maintenance benefits to areas mainly dimensioned by DT, i.e., crack growth and residual strength. For other areas, dimensioned mainly by, e.g., static strength, deformation, and/or durability, SHM may provide maintenance benefits only. In the latter case, the allowable stresses and the weight cannot be improved.

2 FUNDAMENTALS

The DT evaluation for stiffened fuselage panels, which are biaxial loaded, considers three crack scenarios (I, II, and III) in circumferential direction and three crack scenarios (IV, V, and VI) in longitudinal direction as shown in Figure 3. In case of SHM application the most severe scenarios (I and IV) can be excluded, which leads to the benefits described in further detail below.

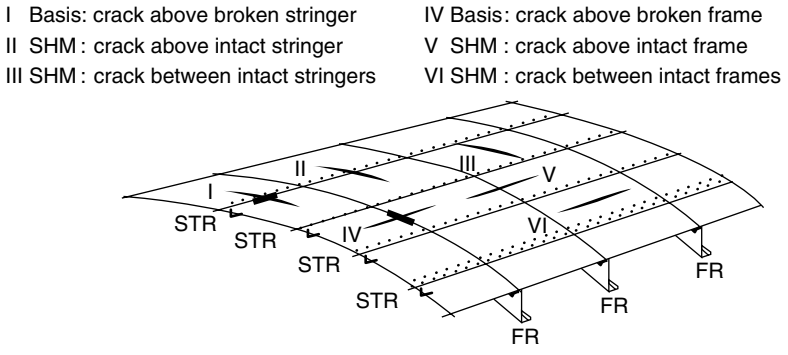


Figure 3. Crack configurations for DT evaluation with and without SHM.

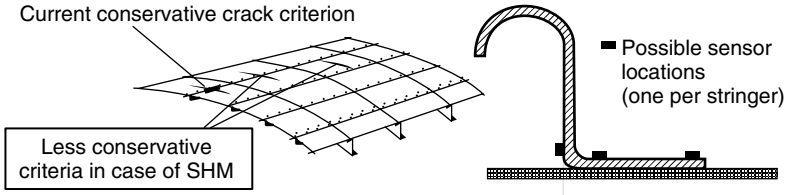


Figure 4. Location of sensor for crack detection at stringers.

The exclusion of scenario I for circumferential cracks requires a repetitive monitoring of the stringers for cracking or failure. Since cracks may originate from fastener holes in the stringer foot, this location is the favorable area for SHM application, see Figure 4. The SHM sensor should allow the detection of the complete failure of the stringer foot, which has a width of about 20 mm.

For longitudinal skin cracks, the structural behavior depends on the status of the frames and/or the skin. As principally shown in Figure 5, there are two possibilities for SHM application to allow higher stress levels resulting in significant weight savings, and these include

- skin monitoring for longitudinal cracks that will allow all scenarios IV, V, and VI to be eliminated. The circumferential crack sensors can be applied at the inside of the skin with a distance of approximately one frame bay;
- an alternative possibility in periodic measurement of the stresses at the outer frame flange considering the redistribution of skin loads into the adjacent frames in case of longitudinal cracks in the skin. The most severe scenarios IV and VI could be eliminated in case of SHM application. Since a significant longitudinal skin crack increases especially the stresses in the outer frame flange, this location is the optimum for the SHM sensor on the frame.

The weight and/or maintenance benefits due to SHM application can be determined by crack growth and residual strength analysis using less severe crack scenarios. A conventional linear analysis procedure is generally applied to fuselage structures using Forman, Paris or Walker equations [3]. Most of the computer codes are valid for the analysis of stiffened panels made of metal under uniaxial external loading. Comparisons of analyses using these codes and representative test results from curved stiffened panels tested under internal pressure and external loads revealed that these analyses are very conservative for the failure mode "crack above a broken stiffener", i.e., for scenarios I and IV [4]. The reason for these differences is the biaxial loading of the test panels, which is not represented in the analyses. Therefore, the modified analyses used in this article consider the following features:

- flexibility of skin-to-frame joint due to fastener flexibility and clip connection;
- bulging of the fronts of longitudinal cracks due to internal pressure causing out-of-plane bending;
- load (stress) distribution between skin and stringers due to internal pressure; and
- nonconstant stress distribution between frames.

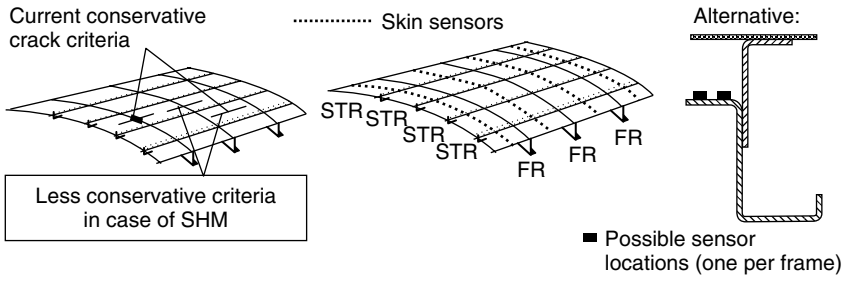


Figure 5. Location of sensors for stress measurement at frames and crack detection at skin.

Reference 4 contains the details of the modified crack growth analysis program providing excellent results, which are in line with the experience from representative tests with large curved panels with similar design and loaded by internal pressure as well as external longitudinal and circumferential loads.

3 QUANTIFICATION OF WEIGHT REDUCTION FOR TYPICAL FUSELAGE STRUCTURE

A correct application of SHM allows higher design stresses and therefore a reduction of the design weight. This can be realized in the fuselage areas where DT is the dimensioning load case. No weight benefits can be achieved in the areas dimensioned by static strength, e.g., in the forward bilge area of the rear fuselage. However, the weight benefits for panels with SHM located in different fuselage areas are not the same owing to different geometric configurations, stresses, and local load spectra. Therefore, all relevant panels have to be analyzed to determine the overall weight reduction.

For demonstration of the complexity of the issue, this article reports the analysis of the weight reduction for four fuselage locations of a modern wide-body aircraft used for medium range. These examples consider typical fuselage structure, i.e., the stiffened panels containing skin, stringers, frames, and frame-to-skin connections by clips. All design elements are made from 2024 material (AlCu alloy). Figure 6 shows the four locations selected and the ratio between the steady stresses in circumferential and longitudinal direction during the flight phase "cruise". The steady stresses consider the effect of the cabin differential pressure and the 1-g condition.

The DT analyses, i.e., crack growth and residual strength calculations, consider steady stresses as well as incremental stresses, e.g., due to gusts, maneuvers, and ground loads. Figure 7 shows the crack growth results for circumferential cracks in the aft upper panel of the rear fuselage, where $\sigma_{\text{skin circ}}/\sigma_{\text{skin long}} \approx 1.0$. The crack growth periods are given from an initial crack length of 76.2 mm to critical crack length. Scenario I is the dimensioning configuration, which can be excluded in case of stringer monitoring for cracks. With the stringer monitoring, the stresses can be increased by 15%, which results in a weight reduction of 13%.

Figure 8 shows the crack growth results for longitudinal cracks in the same panel. The given percentages are related to the basic crack growth period for scenario I in Figure 7.

In case of the described skin monitoring by SHM, the crack scenarios IV, V and VI will be detected before reaching a critical size. Therefore, the stated percentages (70, 170, and 60) for 15% higher stresses are not relevant, i.e., the weight reduction of the panel is defined by the results of the circumferential cracks.

All weight reductions for the four analysis points are presented in Figure 9. The weight reductions of the stiffened panels in the described locations are between 13 and 20%. These benefits can be achieved by SHM monitoring of either locating the sensors on the stringers, or the skin, or on both. These sensor locations are the result of the ratios of the steady stresses and the effect of the incremental load spectra.

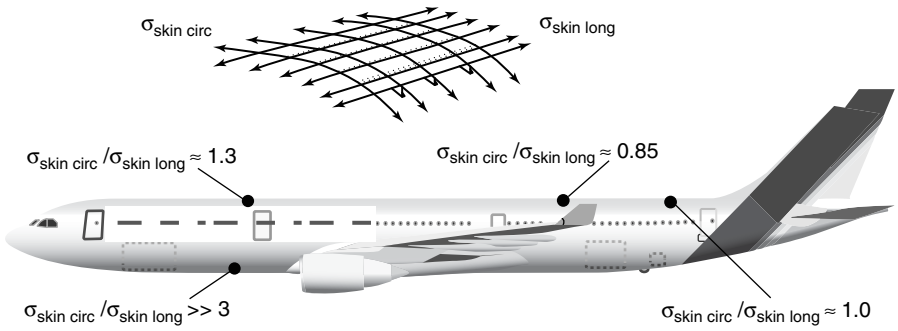


Figure 6. Fuselage analysis areas and ratio of steady stresses during cruise.

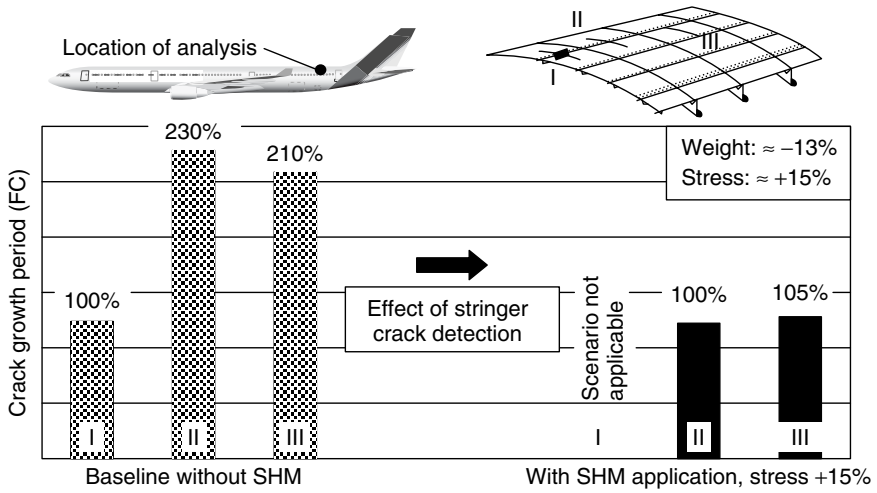


Figure 7. Crack growth results in the aft upper panel of the rear fuselage—circumferential cracks.

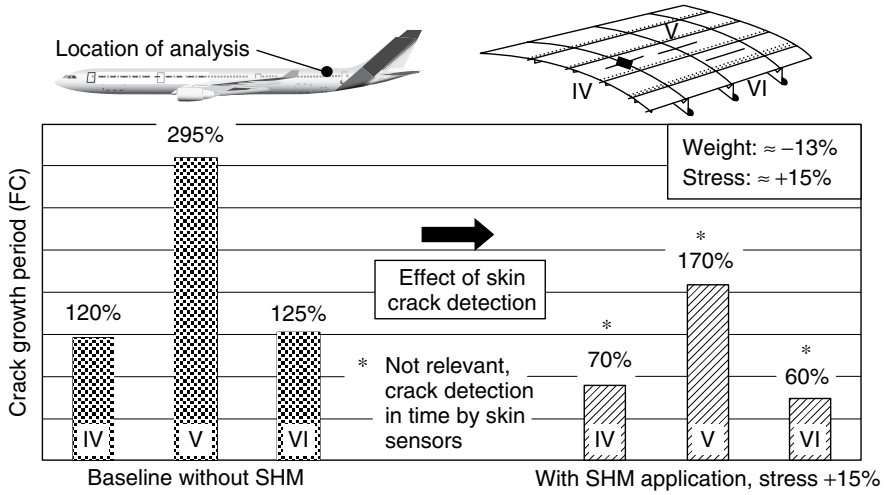


Figure 8. Crack growth results in the aft upper panel of the rear fuselage—longitudinal cracks.

In addition, the weight benefits for circumferential cracks are influenced by the level of the steady stresses. Lower stress levels lead to higher benefits given in percentage when scenario I (without SHM) is replaced by scenarios II and III (both with SHM). As an example Figure 10 shows a generalized chart about possible stress or weight benefits for the upper panels of the rear fuselage. The lower curve applies to panels without SHM, where crack scenario I is dimensioning; the upper curve applies to panels with SHM, i.e., scenarios II and III are to be considered. However, the external surface of the skin has to be inspected for circumferential cracks during heavy maintenance checks besides the SHM monitoring of the stringers.

The proposed skin monitoring shown in Figure 11 will allow the detection of all crack scenarios (IV, V, and VI) in time. Therefore the crack growth of longitudinal cracks according to scenarios IV, V, and VI is no longer a design criterion. In this case the critical crack length $2a_{\text{critical}}$ for limit load condition must be longer than the distance b between adjacent sensors.

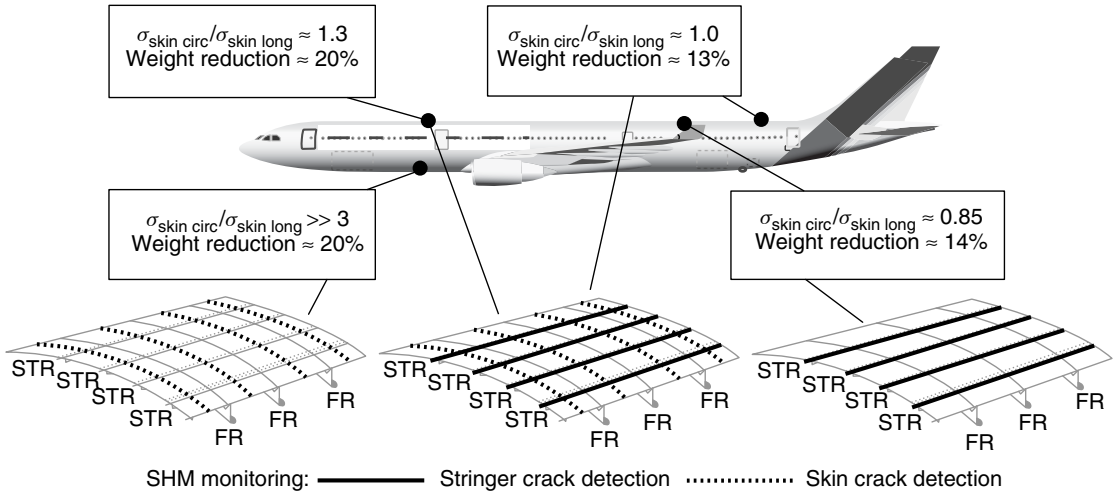


Figure 9. Weight reductions and location of SHM sensors for all analyses locations.

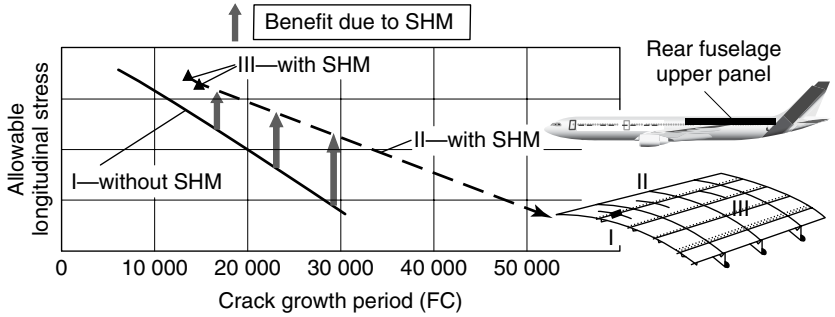


Figure 10. SHM benefits, allowable longitudinal stress for the upper panels of the rear fuselage.

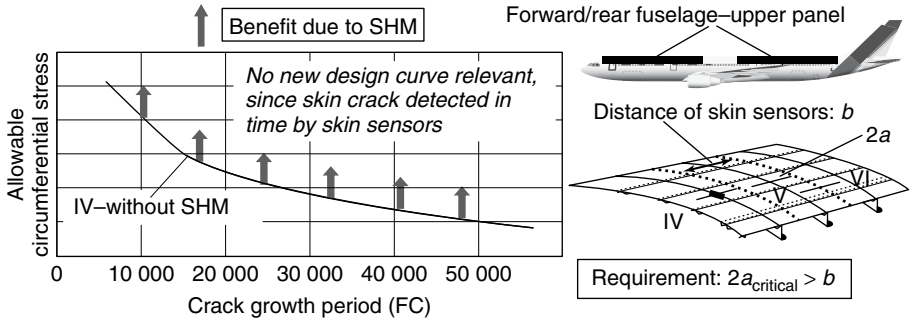


Figure 11. SHM benefits, allowable circumferential stress for the upper panels of the forward and the rear fuselage.

After application of the circumferential sensors no inspection of the external skin is necessary to detect longitudinal fatigue cracks. Therefore this criterion is not relevant anymore (see Figure 11). The alternative strain measurements at frames, as shown in Figure 5, lead to similar weight savings. However, the inspection of the external skin has to be maintained to detect longitudinal fatigue cracks, which is a

disadvantage compared with the skin monitoring proposed above.

4 CONCLUSIONS

The SHM application during the development and design of new metallic aircraft structure offers significant advantages for areas, where DT is the dimensioning criterion. In these areas the stringers or the skin or both elements have to be monitored for cracks by SHM. Alternative to the skin crack monitoring, the frames may be monitored for increased stresses. However, the benefits can only be realized if the SHM system is reliable during the entire life of the aircraft.

The possible weight reduction for four different panels of a typical metallic fuselage, analyzed as examples, is between 13 and 20%. The differences in these percentages are mainly the result of the different stress environments. The optimum definition of the SHM application requires similar DT analyses for all panels as performed for the examples presented in this article.

REFERENCES

[1] Schmidt H-J, Schmidt-Brandecker B. Structure design and maintenance benefits from health monitoring systems. *Proceedings 3rd International Workshop on Structural Health Monitoring: The Demands and Challenges*, September 2001. CRC Press: Boca Raton, FL, 2001; pp. 80–101.

[2] Goggin P, Huang J, White E, Haugse E. Challenges for SHM transition to future aerospace systems. *Proceedings 4th International Workshop on Structural Health Monitoring: from Diagnostics and Prognostics to Structural Health Management*, September 2003. DEStech Publications: Lancaster, PA, 2003; pp. 30–41.

[3] N.N. *Damage Tolerance Assessment Handbook*. FAA Technical Center: Atlantic City, NJ, October 1993.

[4] Schmidt H-J. Damage tolerance technology for current and future aircraft structure, plantema memorial lecture. *Proceedings 23rd ICAF Symposium of the International Committee on Aeronautical Fatigue: Structural Integrity of Advanced Aircraft and Life Extension for Current Fleets—Lessons Learned in 50 Years after the Comet Accidents*, Hamburg, June 8–10 2005, DGLR Report 2005-03, Vol. 1, pp. 1–41.

Chapter 17

Lamb Wave-based SHM for Laminated Composite Structures

Constantinos Soutis

Aerospace Engineering, University of Sheffield, Sheffield, UK

1 BACKGROUND

The development of nondestructive systems to detect and monitor the extent of damage in carbon fiber reinforced plastics (CFRP) during service life is a key challenge in many practical applications, especially in the aircraft industry. The lack of such a technique has severely limited the extensive use of composite materials [1]. Owing to concerns about the integrity of damaged composite structures, engineers are forced to overcompensate during the design process such that the advantages originally expected from composite materials are not fully achieved. An improved understanding of the behavior of sensor and actuator devices and how measured signals relate to damage will contribute to the development of an active system capable of continuously evaluating the condition of a structure. This technology could lead to improvements in composite design, the development of more rapid and appropriate repair strategies, and the withdrawal of expensive periodic maintenance inspections. Thus, the ability to evaluate the integrity of a structure without removing the individual structural components has become an important technology challenge. Several nondestructive evaluation (NDE) methods exist and are used in composite structures. Visual inspection, radiography, ultrasonics, shearography, and thermography are among the most common NDE methods in use. Despite their wide use and improvement in the last decades, the majority of NDE methods are not suitable for implementation in smart structures (self-diagnostic systems). Applications that require a movable probe to obtain data and scan a large area are disregarded as the basis for the development of self-diagnostic systems since they need the direct intervention of humans or a robot to perform the inspection. In principle, a smart in-service health monitoring system would imitate a biological system, where attached or built-in sensors continuously interrogate the structural

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

integrity throughout the component's life. Therefore, techniques that can operate from fixed locations in the structure while inspecting large areas are prime candidates for the development of a structural integrity monitoring system (SIMS). Furthermore, by fixing (attaching or embedding) the transducers, many variables affecting the reliability and repeatability of measurements are removed allowing the precise assessment of minute changes in structural behavior that permit the early detection of damage occurrence. Smart structure-based methods of large-area inspection using continuously sampled data would ideally supplement the use of existing NDE inspections that are performed off-line.

An attractive technique for the development of a SIMS is the use of Lamb waves. Their application has long been acknowledged as a potential solution for large-area nondestructive inspection because they are able to travel relatively long distances allowing the material between transmitter and receiver to be interrogated [1]. Hence, a line scan is achieved with each pulse rather than the comparatively slower point-scanning performance of conventional ultrasonic techniques. Fundamentally, this method involves the analysis of the transmitted and/or reflected waves after interacting with the test part at boundaries or discontinuities. The presence of damage is identified from changes in the response signal of subsequent tests when compared to the reference response of the undamaged configuration taken earlier in the structure's life.

Lamb waves can be excited and detected by a variety of methods, such as the use of interdigital transducers (IDTs) [2, 3], fine point contact transducers [4], air-coupled ultrasonic transducers [5], laser-generation methods [6], and the widely employed angled perspex wedge [7]. However, among these methods, only IDTs appear suitable for implementation in applications to smart structures, where a small and lightweight, permanently attached transducer system design is required. Still, IDTs present some limitations in the Lamb wave inspection of thick sections that are commonly employed in practical structures. When used in ultrasonic applications, piezoelectric materials are normally operated at their thickness-mode (d_{33}) resonant frequency, which is determined by the thickness of the element and the longitudinal wave velocity in the material. The thickness of piezoelectric elements for practical use varies from a few microns to a few millimeters; thus, the frequency range of piezoelectric transducers extends from the low megahertz (0.5) for thick elements to a few hundred megahertz for very thin films. The lower operational frequency of a transducer imposes an upper limit on the thickness of the plate that can be inspected at values under the cutoff frequency–thickness (fh) product of high-order Lamb modes. For instance, for frequencies less than 250 and 100 kHz, respectively, it would be necessary to have only the fundamental Lamb modes (A_o and S_o) propagating in typical composite aerospace laminates (multilayered plates), which have thicknesses ranging from 6 to 15 mm [8]. To generate the higher modes you need to operate at a higher frequency-thickness value, see Figure 2. Of course the advantage of the proposed method operating at low frequency and generating the basic modes is the simplicity of the signal analysis with the ability to detect relatively small defects in multi-layered composite structures. Likewise, if a 12-mm-thick steel plate is to be tested, frequencies below 136 kHz [9] must be utilized.

This work discusses the generation of Lamb waves for the NDE of composite laminates using surface-mounted piezoelectric elements in narrow strips, operated in the longitudinal mode (d_{31}). When voltage is applied to a bonded piezoelectric patch, it expands and contracts parallel to the surface inducing a bending moment in the structure. If the voltage applied is a sinusoid with a few cycles, the piezoelectric elements produce a transient flexural wave whose transmission, propagation, and subsequent reflections at the specimen's boundaries can be analyzed and used to identify the size and location of damage. The advantage of using longitudinal or radial modes (related to width and length, or diameter) of the piezoelectric element rather than the thickness mode is that the former can be excited at much lower frequencies, which allows the inspection of thicker laminates while keeping the fh product low, thus generating only fundamental Lamb modes. This condition is important in the use of Lamb waves for NDE applications since the excitation of a single Lamb mode favors signal interpretation. The study also examines the modal response of narrow beam specimens for wave velocity measurements using continuous wave excitation [9, 10].

2 EXPERIMENTAL PROCEDURE AND MEASUREMENTS

2.1 Materials and instrumentation

Tests are performed on narrow beam specimens made of aluminum and composite material. The composite beam is cut from a 24 ply $[\pm 45^{\circ}/0^{\circ}/90^{\circ}]_{3s}$ carbon/epoxy laminate of size 660 mm × 570 mm. The laminate is fabricated using T300-924C prepreg tapes. Individual test specimens 629 mm × 25 mm and 2.7 mm thick are cut from the laminate using a diamond-wheel saw. The elastic properties of the unidirectional ply are $E_{11} = 162$ GPa, $E_{22} = 11$ GPa, $\nu_{12} = 0.34$, $G_{12} = 5.7$ GPa, and the density is $\rho = 1536\,\mathrm{kg\,m^{-3}}$ [11]. The dimensions of the aluminum specimen are 814 mm × 16 mm and 3.3-mm thick. The beams are instrumented with two piezoelectric patches, 20 mm × 5 mm, made of commercial brass-backed piezoceramic resonators, which are used as an actuator and a sensor, respectively, and bonded near the beam's end, as shown in Figure 1. National Instruments' LabVIEW® signal processing and an analog-to-digital card (PCI-MIO-16E-1) are used in conjunction with a personal computer to implement the data transmission/acquisition under an automated framework and to perform the sensor response analyses.

2.2 Phase velocity measurement

The wave velocity is the fundamental characteristic of a Lamb wave because wave propagation may be analyzed by the variation in its velocity as a function of the *fh* product for each Lamb mode, as shown in Figure 2. These relations (dispersion curves) are found by numerical solution of the Rayleigh–Lamb relation for wave propagation in isotropic plates, described comprehensively by Viktorov [12]. They can also be determined experimentally using the amplitude spectrum method [13] and the phase spectrum method [14].

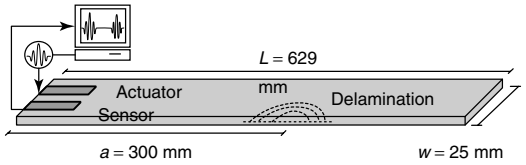


Figure 1. Experimental setup of the composite beam specimen.

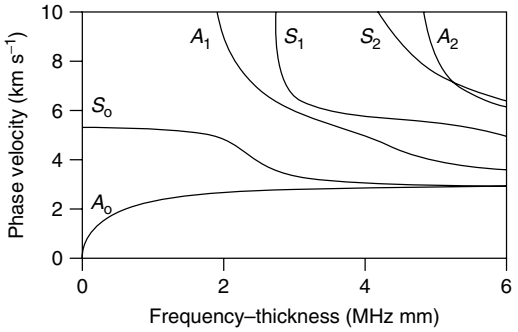


Figure 2. Lamb wave phase velocity dispersion curves for aluminum. [Reused with permission from Sergio H. Diaz Valdes and Costas Soutis, The Journal of the Acoustical Society of America, 111, 2026 (2002). © Acoustical Society of America, 2002.]

In this work, the dispersion curve of the A_0 Lamb mode in the low *fh* product range is determined experimentally from the mechanical resonant response of the beam specimens [15]. In a narrow beam of length L, the particle velocity at resonance consists of a series of standing waves whose frequency correspond to the condition that an integral number of half wavelengths fits in the sample. Thus, resonance exists when $2L = \lambda n$, where n is the harmonic integer and λ is the wavelength. Since the phase velocity is given by

$$c_{\mathrm{p}} = \lambda f \tag{1}$$

the nth mechanical resonance frequency can be expressed as

$$f_n = \frac{nc_{\mathrm{p}}}{2L} \quad \text{or} \quad \omega_n = \frac{2\pi nc_{\mathrm{p}}}{2L} \tag{2}$$

The resonance spectrum of the aluminum specimen is obtained using forced mechanical vibration. The actuator is excited with a 10-V sine-sweep signal [16], varying from 0.1 to 50 kHz in 0.2 s. The structural response captured with the sensor is sampled at a rate of 0.8 MHz and is then Fourier transformed to obtain the resonant frequencies as shown in Figure 3.

The well-defined peaks shown in Figure 3 correspond to the flexural modes of the beam, whereas the smaller peaks appearing above 20 kHz correspond to axial modes. Below 1 kHz, the resonance maxima are only poorly excited so that the modal number of each peak cannot be indexed reliably. Therefore, the appropriate value of n is estimated by comparison with theoretical values of the natural frequencies of

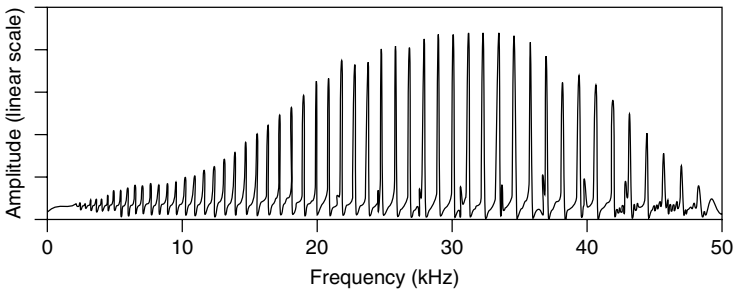


Figure 3. Resonance spectrum of the response of the aluminum beam to sine-sweep excitation.

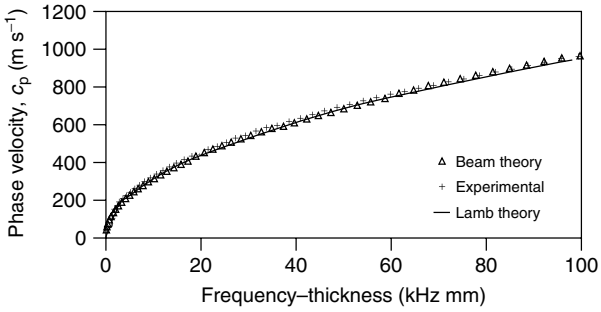


Figure 4. Phase velocity in a 3.3-mm-thick aluminum beam measured using the mechanical resonance method. Beam and Lamb wave theory estimates are also presented.

a uniform beam in transverse flexural vibration using the following expression [17],

$$f_i = \frac{\lambda_i^2}{2\pi L^2}\sqrt{\frac{EI}{m}}; \quad i = 1, 2, \ldots \tag{3}$$

where m is the mass per unit length of the beam, E is the modulus of elasticity, I is the second moment of area, and λ_i is the solution of the characteristic equation for the imposed boundary conditions (free–free).

Figure 4 shows the dispersion curve for the A_o mode in aluminum obtained using the resonance frequency values from Figure 3 that are substituted into equation (2). Also, the beam theory results derived using equations (2) and (3) are presented in Figure 4 along with the Lamb theory curve for aluminum and show excellent correlation with the experimental results.

The phase velocity of the A_o Lamb mode in the composite specimen is also measured using the resonance spectrum method. The sine signal used to excite the actuator is swept from 0.1 to 25 kHz in 0.2 s, and the structural response is sampled at a rate of 0.5 MHz. Figure 5 shows the experimental curve for the laminated beam along with the Lamb theory curve calculated using the average elastic properties of the orthotropic plate [18]. The experimental data shown in Figure 5 have the characteristics of the dispersion curve of the A_o Lamb mode, although the agreement between Lamb theory and experimental results is not quite as good as for the isotropic case, especially as the *fh* product increases (decreasing wavelength). The difference can be attributed to the fact that transverse shear deformation effects, which are neglected in the formulation of the classical plate theory, are significant in the case of laminated plates due to the relatively low transverse shear modulus [19]. It can be seen from Figure 4 that Lamb wave and beam theory can be used to successfully predict the dispersion curves of an isotropic material. However, for composite laminates, no obvious analytical solutions and numerical techniques are required to model such systems. In this study, dispersion curves are measured experimentally using the resonance spectrum method, which is a reliable procedure for measuring low-frequency, long wavelength, flexural wave phase velocity.

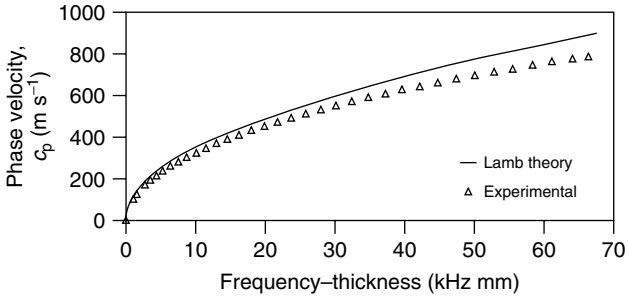


Figure 5. Phase velocity in a 2.7-mm-thick $[\pm 45^{\circ}/0^{\circ}/90^{\circ}]_{3s}$ composite beam measured using the mechanical resonance method. Lamb theory estimates are also presented.

2.3 Delamination detection

Owing to the relatively low interlaminar strength of composite laminates, damage in the form of delaminations can be easily introduced from low-velocity impacts during service, with the subsequent degradation of the mechanical properties of the laminate that can lead to the premature failure of structural components. In practice, composite laminates are designed to tolerate certain degrees of damage and it is often only necessary to find relatively large defects such as 10–20-mm-diameter delaminations [20]. For instance, the typical size of a critical defect in a composite structure for a Harrier aircraft is approximately 20–25 mm [8].

To study the ability of Lamb waves for delamination detection at low *fh* values, the composite beam is examined at different damage scenarios. The beam specimen is excited with short sinusoidal pulses rather than with continuous wave excitation, as was previously done for phase velocity measurements. The undamaged beam is tested and its response history is captured and kept as a baseline. Then a sharp and thin scalpel blade is inserted into the beam's midplane, 300 mm from the left end, as shown in Figure 1. The blade is forced into the material initially producing a small delamination, whose dimension is increased each time the blade is forced into the midplane. In this manner, the delamination area is gradually extended from a small incision located at one edge of the beam until it almost reaches a full-width delamination. This type of artificially induced delamination is thought to better represent damage patterns observed in fatigue loading [10, 15] than beams with a full-width delamination of fixed length, as is commonly found in the literature. The damage area, A_d, is measured at every stage by conventional ultrasonic C-scan, as shown in Figure 6.

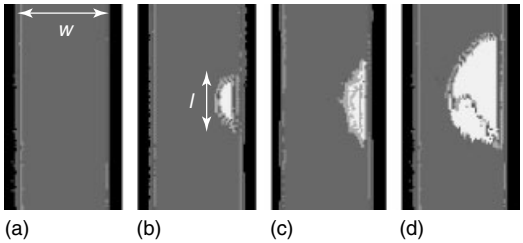


Figure 6. Ultrasonic C-scans of the composite beam: (a) no delamination, (b) $A_d = 22\,\text{mm}^2$, (c) $A_d = 47\,\text{mm}^2$, and (d) $A_d = 220\,\text{mm}^2$.

Figure 7(a) shows the response of the undamaged composite beam when the actuator is excited with a 15 kHz sinusoidal pulse of 5.5 cycles modulated by a Hanning window. The response history shows the input pulse followed by another large wavelet, which is the first reflection from the opposite end of the beam. The time delay between the transmitted and reflected signals corresponds to the propagation distance of the wave, which is twice the length of the beam. The second and third reflections are also identified using longer acquisition times, suggesting a propagation range well over 2 m. However, the response signals shown in Figure 7 only contain the first reflection. It can also be observed that the shape of the wave changes as it propagates along the beam due to the dispersive nature of the A_o mode at this *fh* value.

The same test is performed after damage has been induced, thus providing the specimen response at a different stage of delamination growth. In comparison to Figure 7(a), Figure 7(b–d) shows an extra reflection between the input pulse and the first reflection from the end of the laminate. It can also

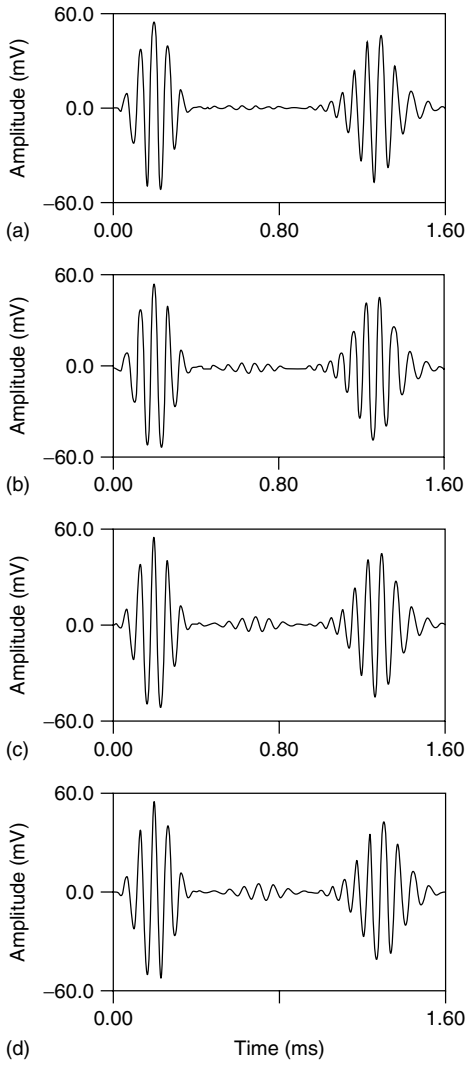


Figure 7. Measured response of the composite beam at different stages of damage: (a) no defects, (b) $A_d = 22\,mm^2$, (c) $A_d = 47\,mm^2$, and (d) $A_d = 220\,mm^2$.

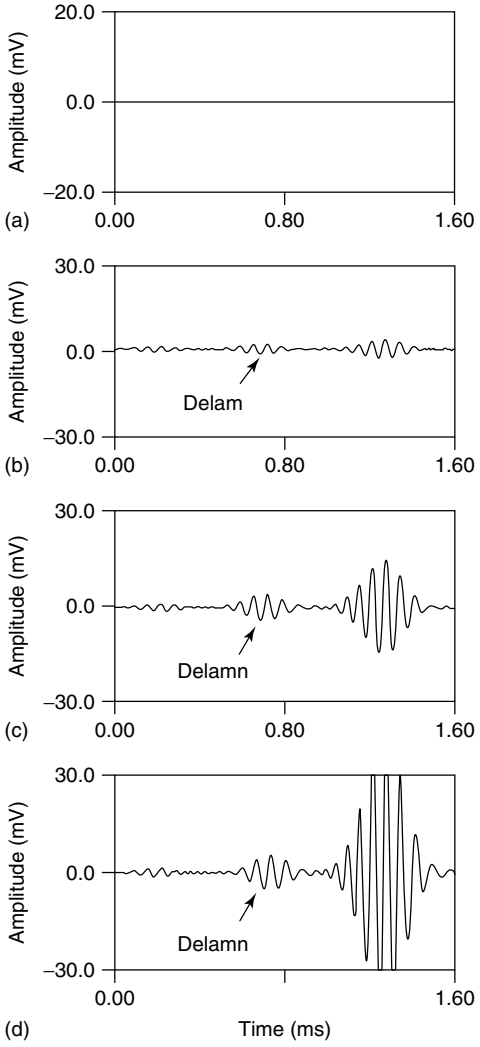


Figure 8. The arithmetic difference between the response of the undamaged and the response at different stages of damage: (a) no defects, (b) $A_d = 22\,mm^2$, (c) $A_d = 47\,mm^2$, and (d) $A_d = 220\,mm^2$.

be observed that the amplitude of the first reflection is affected (reduced) due to the presence of damage. These effects can be better appreciated in Figure 8, which represents the arithmetic difference between Figure 7(a) and each one of Figure 7(a–d), respectively. This basic operation allows each time history to be compared to the undamaged configuration, thereby eliminating the large amplitude sections of the response signals and making it possible to identify small changes in the propagating wave, such as variations in the amplitude and phase shifts. This procedure can be implemented in a health monitoring system, where damage occurrence would be identified by the appearance of ripples or wavelets in an otherwise straight-line signal representative of the undamaged structure.

The sensitivity of the propagating wave to even the smallest delamination can be explained in terms of the relative dimension between wavelength and defect size. In general, the sensitivity of a given

Lamb mode to defects will increase as its wavelength decreases [7]. The wavelength of the excitation signal can be calculated from equation (1). The pulse center frequency used is 15 kHz and the plate thickness is 2.7 mm, which yield a frequency–thickness product of 40.5 kHz mm. At this *fh* value, the phase velocity of the A_o mode obtained from Figure 5 is close to 640 m s^{-1}, from which the wavelength is found to be $\lambda \approx 42$ mm. From the C-scans shown in Figure 6, the l dimension (parallel to the length of the beam specimen) of the smallest delamination tested is found to be 10 mm indicating that even though the delamination size is only 25% of the wavelength, the system is capable of detecting it.

2.4 Delamination location

The position on the timescale of the reflection generated at the damage site and the reflection from the end of the laminate, shown in Figure 7, can be used to estimate the location of the defect along the beam span. The time difference (Δt_L) between the maximum peak of the input pulse and the maximum peak of the reflected signal from the end of the beam is about 1060 μs, which corresponds to twice the length of the specimen ($2L = 1258$ mm). Similarly, the time difference (Δt_{dam}) between the maximum peak of the input pulse and the maximum peak of the reflected signal from the damage site is about 490 μs, which corresponds to a round trip of the wave between the receiver and the delamination ($2a$). Substituting these values into equation (4), the location of damage (a) is found to be at 290 mm, which is a fair estimate of the actual location ($a = 300$ mm) of the artificially induced delamination shown in Figure 1. This measurement is only approximate since the shape of the wave packet does not remain the same during its propagation along the beam due to the dispersive nature of the A_o Lamb mode at this *fh* value.

$$a = \frac{L\ \Delta t_{\text{dam}}}{\Delta t_L} \tag{4}$$

These experiments demonstrate the potential use of Lamb waves at very low *fh* values for NDE applications. Along with a simple and effective signal-processing method, Lamb waves can be used for the development of an in-service, health monitoring technique capable of detecting delaminations in composite components. However, the specimens used were essentially narrow beam elements where lateral wave spreading effects are minimized. In the following section, the wave propagation on wide plates is studied using the finite element (FE) method [21]. Different actuation configurations are examined including the use of a linear array of actuators for the damage evaluation of large surfaces.

3 FINITE ELEMENT ANALYSIS

3.1 Wave propagation in wide plates

FE analysis is carried out in parallel with an experimental approach to study Lamb wave generation and propagation in quasi-isotropic beams and plates. Several models are used to qualitatively investigate the interaction of elastic wave at boundaries to predict the response of a surface-bonded sensor. A mesh in the xy plane is implemented to model the laminate using quadrilateral shell elements, where the nodes are defined on the midthickness of the shell and each node has both translational and rotational degrees of freedom. The x coordinate is aligned along the length direction and the z coordinate is aligned normal to the surface of the plate. The plate thickness is given as a geometrical parameter when the mesh is generated. The excitation of the plate using piezoelectric patches is modeled with uniformly distributed moments of opposite sign applied along two short parallel nodal lines (actuator width), which are separated by a distance equivalent to the length of the piezoelectric actuator as indicated with bold lines in Figure 9.

The excitation signal employed in all of the simulations is a 5.5 cycle, 20 kHz sinusoidal wave modulated by a Hanning window. It is defined in the FE code as a time-history variation of the amplitude

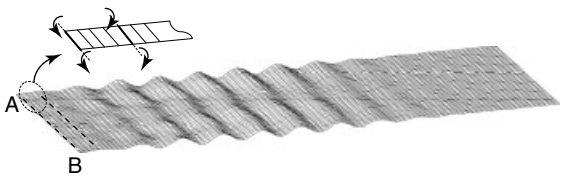


Figure 9. The predicted response of a 2-mm-thick laminate to a sinusoidal load applied at the locations indicated by bold lines along the AB boundary.

of the applied load. Then an explicit central difference scheme is employed in the *MARCH* module of the *FE77*, which carried out a step-by-step time marching integration to solve the wave propagation simulation.

Figure 9 shows a snapshot of the predicted response of the specimen at a certain time step after being excited by the described load case. It shows the propagation of the A_o mode across the surface of a 500 mm × 200 mm plate excited with a linear array of actuators distributed along its left edge (AB boundary). This arrangement produces a fairly uniform wave front across the width of the plate that is reflected by the right edge with minor lateral wave spreading. Such response favors the interpretation of the signal produced by a bonded sensor, because the input pulse and subsequent reflections are similar to those observed in the narrow beam and can be easily identified in the signal time history as shown in Figure 10. It presents two well-defined sinusoids and a fairly flat signal between them, similar to the observation for the narrow beam element in Figure 7(a).

3.2 Damage detection

The advantages of using a linear array of actuators for the inspection of large areas can be greatly enhanced if the actuators comprising the linear array of transducers are also used as sensors in a pitch–catch mode, thus exploiting the dual capabilities of piezoelectric materials as a receiver and a transmitter. This approach has been examined using a uniformly distributed sinusoidal load applied along the left edge

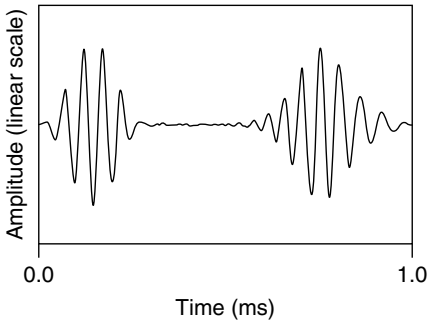


Figure 10. The predicted displacement history of a node located at the center of the AB boundary. The plate was excited with a sinusoidal load applied to the AB edge.

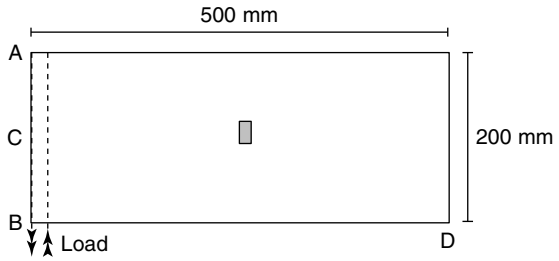


Figure 11. Sketch of the plate used in the FE calculations, showing a central area with reduced material properties.

of the plate model, the AB boundary on Figure 11, while monitoring the in-plane nodal displacement along that same edge. Both undamaged and damaged configurations are simulated and their responses are compared to identify the effects of damage on the model behavior. Damage is simulated by degrading the material properties of a selected number of elements within the mesh. This reduction of the mechanical properties varies from zero stiffness to 50% of the elastic properties of the undamaged material, thus representing the cases of an open hole and damage due to low-velocity impact, respectively. Figure 12 shows the response of the undamaged plate presented in a 3-D plot of amplitude of in-plane displacement, time, and plate width. Similar to the experimental results, the first part of the response shows the input pulse followed by another large wavelet, which is the first reflection from the opposite end of the plate. Figure 13(a) presents the response history of the plate containing damage in the form of a square (10 mm × 10 mm) cutout. This displacement history is, at first glance, similar to that of the undamaged model; however, evident differences between the two time histories are revealed in Figure 13(b), which represents the arithmetic difference between Figures 12 and 13(a). Qualitatively, similar results to those shown in Figure 13 are obtained in the case where the elastic properties in the damaged area are reduced by 50%, although the amplitude of the wave reflection generated at the damage site is lower than in the open-hole case. The presence of structural discontinuities can be easily inferred from Figure 13(b). In addition, the location of damage can be estimated from visual inspection of these figures or by correlating the position on the timescale of the wave reflection generated at the damage site with the laminate length as was done in the experimental section for the

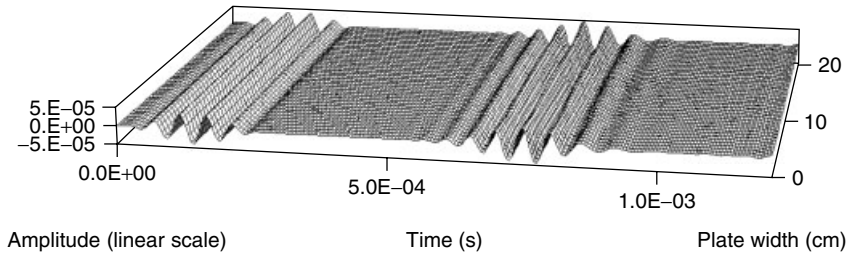


Figure 12. The predicted in-plane displacement history of the nodes located along the AB boundary of the undamaged mesh.

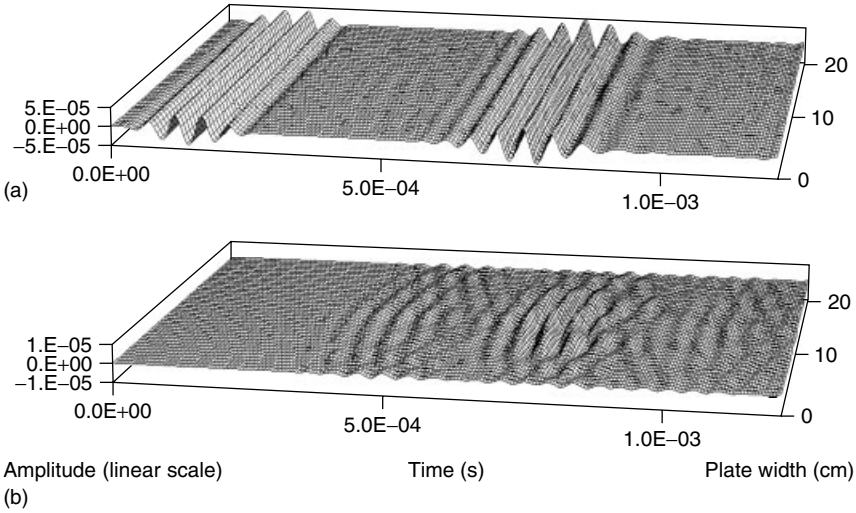


Figure 13. The predicted in-plane displacement history of the nodes located along the AB boundary of the damaged mesh. (a) Time history for the 10 mm × 10 mm open-hole case, (b) arithmetic difference between undamaged and damaged response histories.

narrow beam. Furthermore, the severity of damage can also be estimated because the amplitude of the waves appearing in Figure 13(b) directly relates to the damage size. These results strongly suggest that damage can be found in large-area specimens using a permanently bonded linear array of piezoelectric sensor/actuator elements.

3.3 Damage characterization

Defect detection has been demonstrated using experimental measurements and FE estimates. However, the determination of delamination size and position through the thickness of the laminate is still a challenge in the use of Lamb waves for NDE purposes. This subsection presents a methodology to determine these two parameters from changes in the time of arrival of a tone burst when it travels through a region with delaminations. The time of flight (TOF) of the input pulse, defined as the time it takes for the input pulse to complete a round trip along the length of the specimen, is altered in the presence of damage. In general, the wave velocity changes in the delaminated area because the waves must travel through a region with an effective thickness smaller than the undamaged laminate thickness. As a consequence, the *fh* value decreases and the phase velocity of the wave varies according to the dispersion curves of the material (Figure 2). The shape of the dispersion curve of the A_o Lamb wave in the low *fh* value region is such that a reduction of the

fh value produces a reduction of the phase and group velocities. Therefore, the TOF of an A_o Lamb wave traveling in a specimen with delaminations is greater than that of a wave traveling in the same undamaged specimen. This difference, ΔTOF, can be estimated assuming that through the delamination, the wave propagates via two sheets of material and that the thickest section is the fastest wave path, which is the one used for calculating the *TOF*. Therefore, the *TOF* of a wave propagating along an undamaged and damaged specimen are given, respectively, as

$$TOF_u = \frac{2L}{c_g} \tag{5}$$

$$\text{and} \quad TOF_d = \frac{2(L-l)}{c_g} + \frac{2l}{c'_g} \tag{6}$$

from which the ΔTOF is obtained, i.e.,

$$\Delta TOF = 2l\left(\frac{1}{c'_g} - \frac{1}{c_g}\right) \tag{7}$$

where L is the length of the specimen, l is delamination size, and c_g and c'_g are the group velocities in the undamaged and damaged regions, respectively. An inspection of equation (7) shows that ΔTOF is a function of delamination size (l) and its position through the thickness (d) because c_g is a function of d. Hence, for a given ΔTOF there are several combinations of l and d that satisfy equation (7).

To find the characteristics (l and d) of a given delamination, it is necessary to perform at least two tests of the ΔTOF at two different frequencies (different *fh* values) so that a unique combination of l and d would satisfy both ΔTOF values measured. To illustrate the method, consider a tone burst excited at 10 and 20 kHz, traveling along a 1000-mm-long (L) and 10-mm-thick (h) aluminum plate with a 20-mm-long (l) and 2-mm-deep (d) defect. Using dispersion curves for aluminum, the group velocity of a 20-kHz tone burst traveling through the undamaged ($h = 10$ mm) and damaged regions ($h - d = 8$ mm) are found to be 2284 and 2125 m s^{-1}, respectively. Substituting these values into equation (7) gives a ΔTOF of 1.310 μs. Likewise, a 10-kHz tone burst travels at 1775 and 1618 m s^{-1} through the undamaged and delaminated regions, respectively. Substituting these values into equation (7) gives a ΔTOF of 2.186 μs. Figure 14(a,b) shows ΔTOF as a function of damage size (l) and lines of constant depth (d) for a pulse generated at 20 and 10 kHz, respectively. The possible combinations of l and d for each ΔTOF previously calculated are those values where a horizontal line at 1.310 and 2.186 μs on Figure 14(a,b), respectively, intersects the lines of constant depth. These values are listed in Table 1 along with results for similar tests performed at 40 and 30 kHz. It can be observed from this table that the defect size that remains almost constant for all frequencies is that of $l = 20$ mm in the $d = 2$-mm-depth column (second column), demonstrating the use of this methodology.

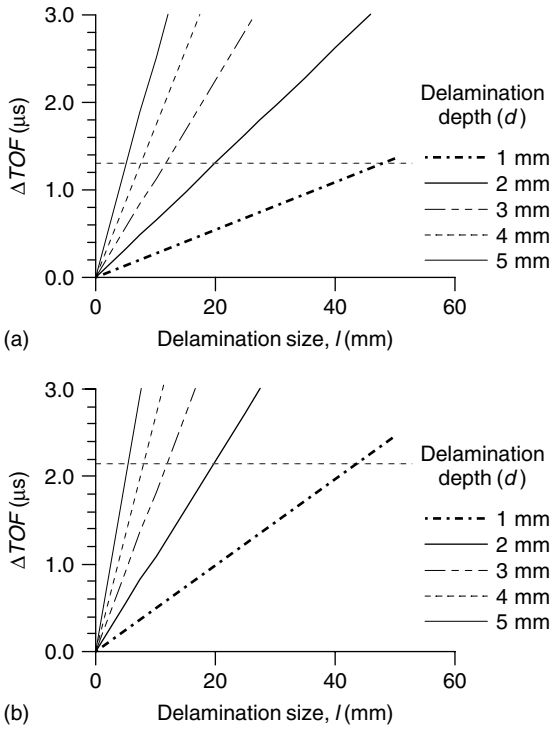


Figure 14. ΔTOF plots calculated for (a) 20-kHz and (b) 10-kHz pulse propagating in a 1000-mm-long, 10-mm-thick aluminum plate.

4 CONCLUDING REMARKS

The generation of low-frequency Lamb waves using small and thin piezoceramic patches has been presented for the development of an on-line, structural

Table 1. Values of defect size and depth for given ΔTOF calculated at different frequencies

Defect (mm)		1.0	2.0	3.0	4.0	5.0	ΔTOF (μs)	f (kHz)
	Depth							
	Size	44.88	19.99	11.73	7.63	5.04	0.792	40
		44.51	19.99	11.83	7.76	5.32	1.022	30
		47.78	19.99	11.63	7.60	5.22	1.310	20
		44.05	20.06	12.05	8.02	5.59	2.186	10

integrity assessment system capable of detecting delaminations in composite laminates. The study offers a solution to some of the problems encountered with current Lamb wave generation techniques where the use of standard ultrasonic probes makes them unsuitable for the development of a built-in, health monitoring system due to their considerable size and shape. The inspection of CFRP laminates has been performed producing promising results at excitation frequencies in the low ultrasonic range (<100 kHz) where IDTs have certain limitations in operations that restrict their use for the damage inspection of thick laminates.

Simple and effective signal-processing techniques have been employed for the detection of small changes in the structural response. Their minimal computational demand makes these techniques appropriate for real-time continuous damage monitoring. Also, the good sensitivity to delaminations of relatively small size (1 cm^2), considerable propagation distances (over 2 m), and the very low requirement of electronic hardware offers a cost-effective solution for structural monitoring in terms of implementation and subsequent operation.

FE analysis has shown the potential use of a linear array of transducers to achieve large-area scanning from fixed locations. The measurement of in-plane displacement response to a given excitation has been demonstrated to be an appropriate procedure to detect, locate, and quantify damage. In addition, a methodology to determine damage characteristics (size and depth) was investigated, which is based on the velocity reduction of a wave passing through regions with defects that produce a measurable delay of its TOF. These analyses can be easily implemented in an array of transducers, providing additional capabilities toward the development of a fully integrated, continuous health monitoring system for large composite structures.

Recent work by the author and coworkers [22–25] involves the use of an arrays of sensors, whose optimum number and spatial distribution were determined from experimental and FE analyses, to detect damage in the form of resin (matrix) cracking and delamination in structural configurations such as composite sandwich panels, stiffened panels, and patch-repaired composite laminates. In these studies, it has been shown that inspection can be realized in large complex composite structures while economizing on hardware and processing complexity. The piezoelectric elements were individually bonded to the structure; however, the concept has been conceived to consist of a multielement array of ultrasonic transducers bonded to a thin flexible strip that can be bonded to new and existing structures. The possibility of embedding these devices into the laminate structure during the fabrication process also needs to be explored. Ultimately, to achieve the objective of enhanced structural reliability, research is required that will involve active sensor development, structure/sensor network integrated manufacturing, signal processing and interpretation, and system integration. This will lead to (i) revolutionary design and manufacturing concepts and analyses for the design of advanced ultrareliable structures for the new century and (ii) theoretical and computational models for predicting sensor performance, structural integrity, and damage-detection capabilities. A significant challenge here is to transition microscale material and structural behavior through mesoscale and macroscale behavior into full-scale structural system performance.

RELATED ARTICLES

Monitoring Marine Structures

Wind Turbines

SHM and Lifetime Management of Industrial Piping Systems

REFERENCES

[1] Percival WJ, Birt EA. A study of Lamb wave propagation in carbon-fibre composites. *Insight* 1997 **39**(10):728–735.

[2] Monkhouse RSC, Wilcox PD, Cawley P. Flexible interdigital PVDF lamb wave transducers for the development of smart structures. *Ultrasonics* 1997 **35**(7):489–498.

[3] Wilcox PD, Cawley P, Lowe MJS. Acoustic fields from PVDF interdigital transducers. *IEEE Proceedings-Science Measurement and Technology* 1998 **145**(5):250–259.

[4] Degertekin FL, Khuri-Yakub BT. Hertzian contact transducers for nondestructive evaluation. *The Journal of the Acoustical Society of America* 1996 **99**(1):299–308.

[5] Farlow R, Hayward G. Real-time ultrasonic techniques suitable for implementing non-contact NDT systems employing piezoceramic composite transducers. *Insight* 1994 **36**(12):926–935.

[6] Pierce SG, Culshaw B, Philp WR, Lecuyer F, Farlow R. Broadband Lamb wave measurements in aluminum and carbon/glass fibre reinforced composite materials using non-contacting laser generation and detection. *Ultrasonics* 1997 **35**: 105–114.

[7] Alleyne D, Cawley P. Optimization of Lamb wave inspection techniques. *Ndt & E International* 1992 **25**(1):11–22.

[8] Birt A. Damage detection on carbon-fibre composites using ultrasonic Lamb waves. *Insight* 1998 **40**(5):335–339.

[9] Díaz Valdés SH, Soutis C. Application of the rapid frequency sweep technique for delamination detection in composite laminates. *Advanced Composites Letters* 1999 **8**(1):19–23.

[10] Díaz Valdés SH, Soutis C. Delamination detection in composite laminates from variations of their modal characteristics. *Journal of Sound and Vibration* 1999 **228**(1):1–9.

[11] Mackinley, CP, *Compressive Failure of CFRP Laminates Containing Pin-Loaded Holes*, Ph.D. thesis, Imperial College, Department of Aeronautics, 2000.

[12] Viktorov IA. *Rayleigh and Lamb Waves-Physical Theory and Applications*. Plenum press: New York, 1967.

[13] Pialucha T, Guyott CCH, Cawley P. An amplitude spectrum method for the measurement of phase velocity. *Ultrasonics* 1989 **27**:270–279.

[14] Sachse W, Pao YH. On determination of phase and group velocities of dispersive waves in solids. *Journal of Applied Physics* 1978 **49**(8):4320–4327.

[15] Díaz Valdés SH, Soutis C. Real-time nondestructive evaluation of fibre composite laminates using low-frequency Lamb waves. *The Journal of the Acoustical Society of America* 2002 **11**(5):2026–2033.

[16] White RG, Pinnington RJ. Practical application of the rapid frequency sweep technique for structural frequency response measurement. *Aeronautical Journal* 1982 **86**:179–199.

[17] Blevins RD. *Formulas for Natural Frequency and Mode Shape*. Krieger Publishing Co.: 1984.

[18] Datoo MH. *Mechanics of Fibrous Composites*. Elsevier Science: 1991.

[19] Tang B, EG Henneke II, Stiffler RC. Low frequency flexural wave propagation in laminated composite plates. In *Proceedings of Acousto-Ultrasonics: Theory and Application*, Duke JC Jr (ed). Plenum press: New York, 1988, pp. 45–65.

[20] Guo N, Cawley P. Lamb wave reflection for quick nondestructive evaluation of large composite laminates. *Materials Evaluation* 1994 **52**(3):404–411.

[21] Hitchings D. *Finite Element Package FE77*. Imperial College, Department of Aeronautics, 1997.

[22] Diamanti K, Hodgkinson JM, Soutis C. Detection of low-velocity impact damage in composite plates using Lamb waves. *Structural Health Monitoring Journal* 2004 **3**(1):33–41.

[23] Diamanti K, Soutis C, Hodgkinson JM. Lamb waves for the non-destructive inspection of monolithic and sandwich composite beams. *Composites A* 2005 **36**(2):189–195.

[24] Diamanti K, Soutis C, Hodgkinson JM. Non-destructive inspection of sandwich and repaired composite laminated structures. *Composites Science and Technology* 2005 **65**(13):2059–2067.

[25] Diamanti K, Soutis C, Hodgkinson JM. Piezoelectric transducer arrangement for the inspection of large composite structures. *Composites A* 2007 **38**(4):1121–1130.

Chapter 111

Design, Analysis, and SHM of Bonded Composite Repair and Substructure

Constantinos Soutis[1] and Jeong-Beom Ihn[2]

[1] *Aerospace Engineering, University of Sheffield, Sheffield, UK*
[2] *The Boeing Company, Advanced Structures Technology, Phantom Works, Seattle, WA, USA*

1 BACKGROUND

The proportion of fiber composite materials being used in aerospace, industrial, automotive, and marine structures is increasing year on year. Continuously reinforced thermosets are currently the most popular composite systems, which are offering higher specific stiffness and strength compared with conventional engineering materials. A number of different resins and reinforcements have been "qualified" for use in a number of different market areas. These structures are, without doubt, outperforming their forerunners when structural performance parameters alone are considered. However, the inherently brittle nature of composites makes them susceptible to damage caused by low-velocity impact. Consequently, it has been necessary to develop repair methods so that costly components are not scrapped owing to in-service damage. Current repair concepts of composites include a wide range of approaches from highly refined and structurally efficient but expensive flush patch repairs to the external mechanically attached metal or composite patch [1–3]. In all these repair methods, the main concerns are the prediction of both strength and durability of the repaired laminate and also the structural health monitoring of the bonded repair.

In this study, the mechanical behavior of bonded external patch repairs is examined; the compressive loading mode is more severe than the tensile mode

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

due to instability of delaminated plies, instability of the patch, and skin strength reductions occurring under elevated temperatures and absorbed moisture conditions. Using a simple "shear-lag" analytical model and a three-dimensional finite element (FE) analysis, design guidelines are produced for the selection of patch size, shape, and membrane stiffness; strength measurements of repaired laminates are also presented. Following some of these design guidelines, an aluminum plate with a cracked hole is repaired with a unidirectional boron/epoxy patch that contains two layers of piezoelectric transducers, which are used to monitor crack growth under fatigue loading.

2 BONDED REPAIR METHODS

Adhesively bonded repairs are the most common type of repair carried out with composite materials [4, 5]. *Cosmetic repairs* refer to damage that is not structurally significant (e.g., scratches, dents, or missing surface plies). The repair is made to restore surface smoothness. In these repairs, a potting compound or a liquid adhesive is spread into the damaged area and formed to the component's contour. *Injection repairs* are another type of repair procedure, which is used for minor disbonds or delaminations. In this procedure, a number of holes are drilled to the depth of the damage. Filler resin is heated to decrease its viscosity and injected under pressure until the excess flows out of adjacent holes. Pressure can be applied to the repaired area to ensure mating of neighboring regions. If serious damage is encountered, then more rigorous repair techniques must be employed. In this case, there are two types of bonded patch that can be used to repair structural damage, namely, *flush scarf patches* and *external patches*.

Flush scarf-type bonded repairs are used where surface smoothness is essential. This approach provides the highest joint efficiency of any repair method. Scarf repairs are used on critical components where load concentration and eccentricities, especially for compressive loading, must be avoided. Thick monolithic structures lend themselves to such repair since an external patch would cause excessive out-of-mouldline thickness and unacceptably high bond-line peel and shear stresses. The flush repair procedure requires careful preparation of the damaged area to obtain the correct scarf angle and dimensional tolerances; taper ratios of 20:1 to 40:1 are typical. The laminate orientation of the patch must match that of the damaged section that has been removed and the scarf patch can be cocured (cured on the damaged plate) or precured (cured and then bonded onto the damage).

The external patch technique is simpler to apply and less critical in nature than a scarf approach. This type of repair is considered as temporary repair and aims to restore the mechanical strength required to permit aircraft operation until a permanent repair can be carried out. In this approach, the load is taken over and around the damaged area. Any bending strains due to the eccentric load path must be considered in the patch design. The patch must also be capable of withstanding the high peel and shear stresses, which develop at the edges of the overlap. Figure 1 illustrates the test specimen configuration studied in this article. The optimum patch size, thickness, shape, and lay up are identified by first analyzing a simple double-lap joint.

3 DOUPLE-LAP BONDED JOINT

The double-lap joint shown in Figure 2 is an idealized case of the bonding composite repairs with external patches on both sides. Hart-Smith [6, 7] showed that its strength is determined by the area under the shear stress−strain curve of the adhesive. Consequently, a stronger joint is obtained with a weaker, but ductile, adhesive if the area under its stress−strain curve is larger than that of a stronger, but brittle, adhesive. A further consequence is that although failure stress and strain will be altered by environmental effects, such as temperature change and moisture absorption, the joint strength will not be affected if the area under the stress−strain curve remains the same, and provided the integrity of adherend/adhesive interface is maintained. The simplest approach for solving stresses in a double-lap joint is the linear analysis due to Volkersen [8], the so-called shear-lag analysis; the shear deformation of the adhesive and the elongation of the adherends are the only factors considered.

The next section presents a theoretical method based on Hart-Smith's work [6, 7] for predicting the optimum patch geometry and the failure strength of a double-lap joint.

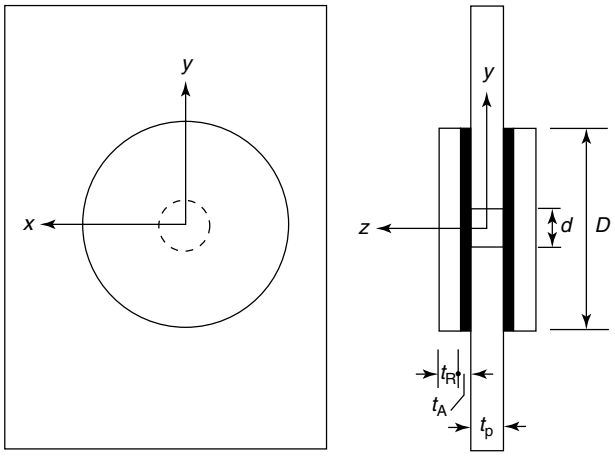


Figure 1. External patch repair.

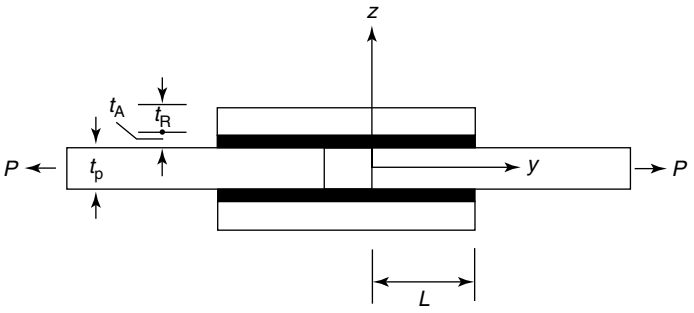


Figure 2. A schematic of a double-lap joint.

3.1 Adhesive shear model

Consider the double-lap joint shown in Figure 2 under in-plane loading P. The length of overlap is L. The thickness of the parent plate (adherend), the repair patch, and the adhesive layer are t_P, t_R, and t_A, respectively. The elastic moduli of the parent plates and the patch (repair) are E_P and E_R. The shear modulus, failure shear stress, and shear strain of the adhesive material are G_A, τ_s, and γ_s. The problem is simplified by making the following assumptions: (i) the parent plates and repair patches are predominantly subjected to direct stress along the x direction, (ii) the adhesive layer is mainly subjected to shear stress in the $x-z$ plane, (iii) the parent plates and the repair patches are elastic and the adhesive layers are elastic-perfectly plastic; ductile adhesives exhibit very high shear strains at failure, and for the purposes of strength prediction can be idealized as elastic-perfectly plastic, and (iv) the maximum shear–strain criterion is used to determine the failure of the adhesive material.

The detailed analytical solution and a Fortran program that estimates the stress distributions and the ultimate failure load of a double-lap joint are given in a previously published technical report [9]. Here, only the output stress results are presented; in particular, the following significant parameters that affect the stress distribution and joint strength are discussed:

1. Optimum overlap length

The overlap length has a significant effect on joint strength. The first question that the repair designer has to answer is how big the patch should be. Figure 3 shows the shear stress distribution in the adhesive layer of joints with short ($L = 5$ mm), medium

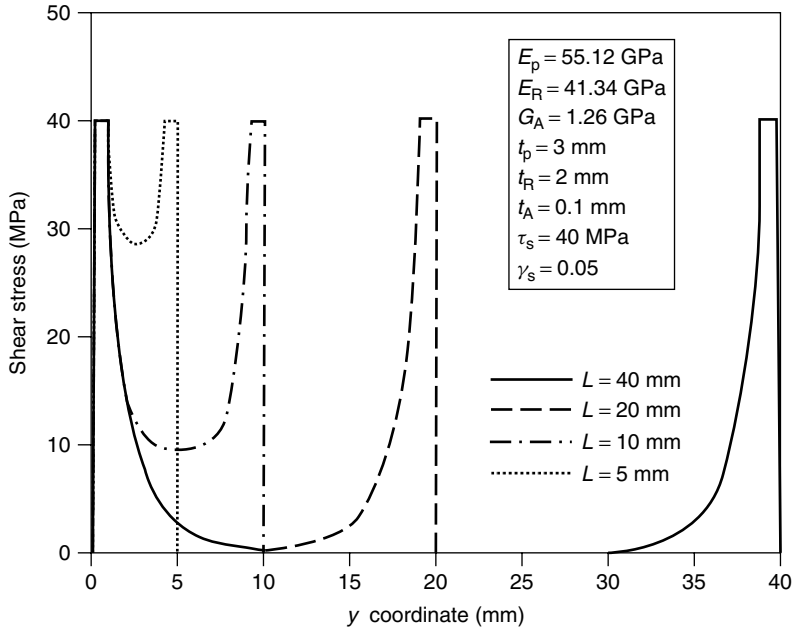


Figure 3. Shear stress distribution in the adhesive layer under ultimate load.

($L = 10$ mm), and long overlap length ($L \geq 20$ mm) under ultimate loading. With short overlap joints, all adhesive material is under high shear stress. For longer overlap joints, the great majority of the load is carried by the plastic adhesive zones separated by a lightly loaded elastic region (trough). Since the load transfer zones occur at the ends and eventually reach a constant length, no increase in joint strength is achieved once these zones are fully developed. The limiting value of joint strength, based on the failure criterion of total adhesive shear strain, is obtained in Figure 4. There is no point in increasing the overlap beyond this critical value, in this case, $\ell \approx 12$ mm, since no significant enhancement in strength, σ_{max}, results. However, considering various effects such as imperfect bonding, patch delamination, environmental effects, and a safety factor, the limiting overlap for current carbon fiber-epoxy systems is around $30t_R$, where t_R is the *repair patch thickness* defined in Figure 2; large flaws in the middle of such a joint would impose no loss of strength since there is no load being transferred there.

2. Optimum patch thickness

The influence of the membrane stiffness (i.e., the product of the elastic modulus and thickness) of patches is demonstrated in Figure 5. It is found that the optimum value of the total patch membrane stiffness ($2E_Rt_R$) is equal to the membrane stiffness of the parent laminate (E_pt_p). If the parent plate has the stiffness of Et, the optimum patch stiffness is $(Et)/2$. As expected, a joint with soft patches (patches are thin or with low elastic modulus) has low strength. However, using stiffer patches does not mean enhancing the strength of the joints. Figure 5 indicates that very thick or overstiff patches become harmful since they increase weight and reduce strength due to larger direct and shear stresses developed at the joint. As patch thickness increases, peel (through-thickness tensile) stresses become increasingly important and can become sufficiently large to limit joint strength and cause failure in an adherend with a low through-thickness tensile strength.

3. Influence of the adhesive

The strength properties of the adhesive certainly influence the strength of the joint, as shown in Figure 6. The present nonlinear analysis indicates that using stronger adhesive materials produces stronger joints; the ultimate adhesive shear strain affects the joint strength more than the ultimate adhesive shear stress does.

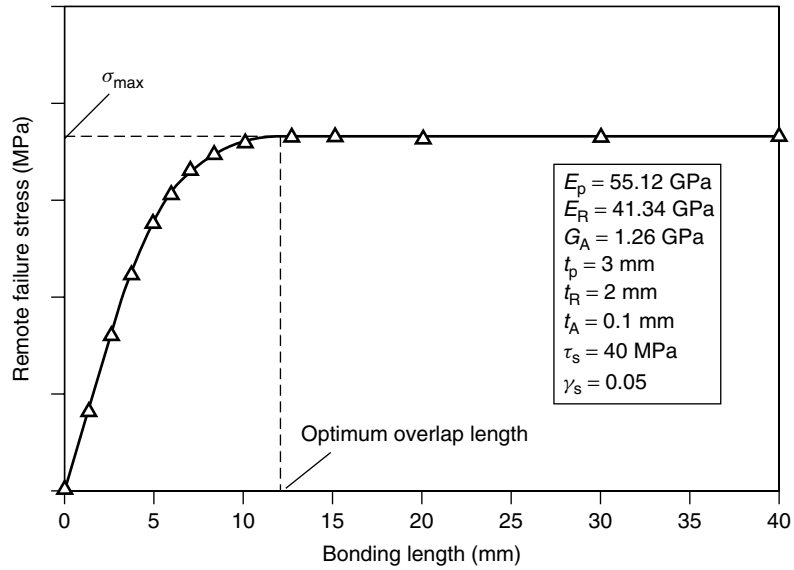


Figure 4. Remote failure stress of a double-lap joint as a function of overlap length.

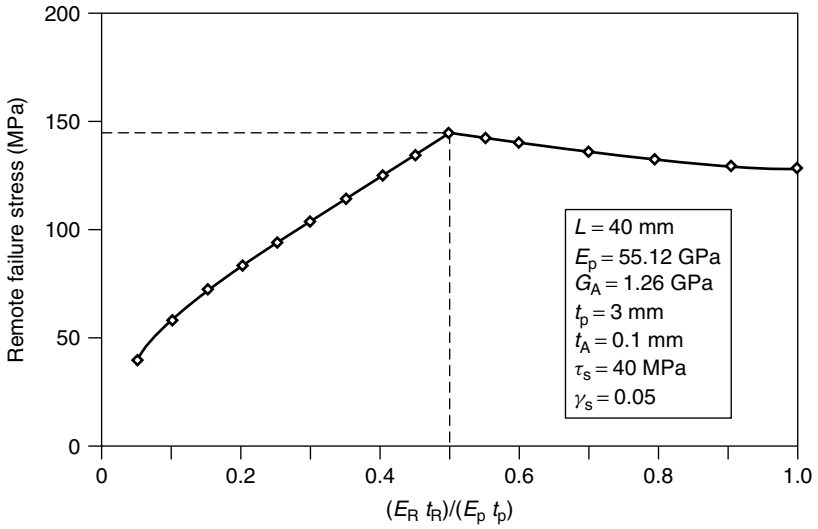


Figure 5. Remote failure stress of a double-lap joint with a large bonding length ($L = 40$ mm) against patch membrane stiffness.

Since high localized stresses are developed at the ends of the overlap, special care is required during the design process of the joint. Figure 7 indicates that the high shear strain can be markedly reduced by deliberately increasing the adhesive thickness at the edge of the overlap. A joint with patches tapered from inside (near the ends of the overlap) can create the local thickening of the bond and substantially reduce the stress concentration in the adhesive layer. However, it is important to remember that good adhesive bonds can be produced only in a small range of thicknesses (typically 0.125–0.25 mm) since thick bonds tend to be porous and weak while ultrathin bonds are too stiff and brittle.

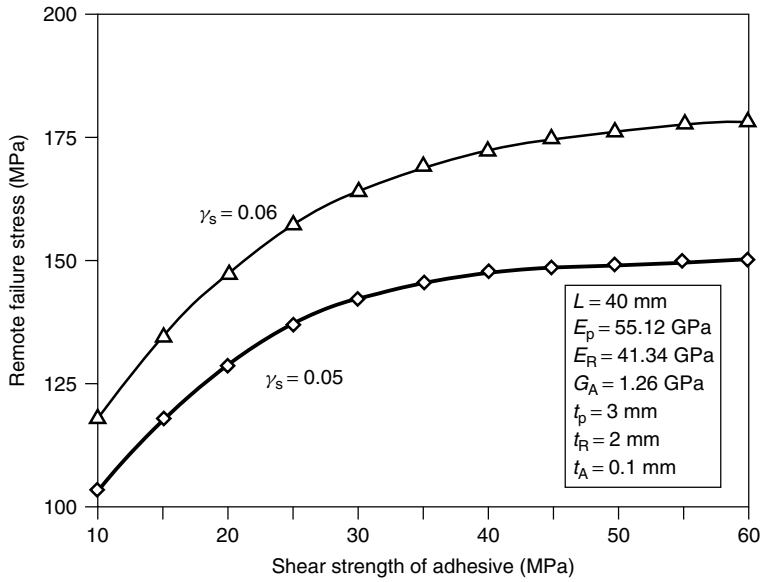


Figure 6. Influence of the ultimate adhesive shear strain on joint failure stress.

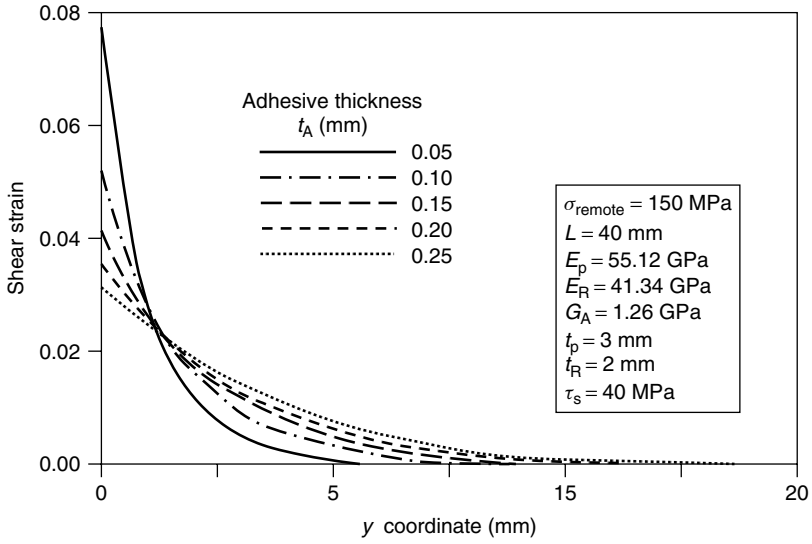


Figure 7. Influence of the adhesive thickness on the maximum shear strain at the end zone of the overlap.

4. Peel stress and optimum patch shape

Peel (through-thickness tensile) stresses can become significant in double-lap joints with thick adherends. High peel stresses are normally induced at the ends of the overlap. As the tensile allowable (through the thickness) for the adherends is generally less than that for the adhesive, failure occurs in the adherend [6, 7], as explained in Figure 8. The maximum peel stress induced in the adhesive and adjacent adherend at the end of the overlap is given by [10],

$$\sigma_{\text{peel}} = \tau_{\max} \left(\frac{3E_z(1-\upsilon_{xz}^2)t_R}{E_x t_A} \right)^{1/4} \tag{1}$$

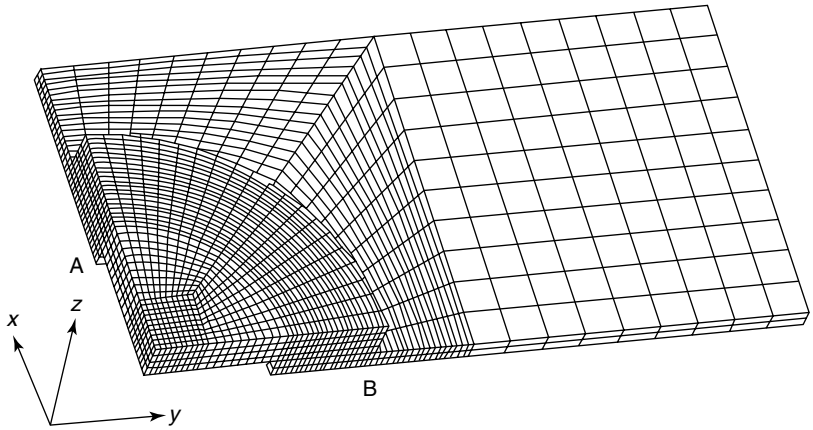


Figure 8. Typical finite element mesh used in the analysis.

where τ_{max} is the maximum adhesive shear stress, E_x and E_z are the elastic moduli of the adherend laminate in the loading direction and thickness direction, respectively, and ν_{xz} is the Poisson's ratio. Equation (1) indicates that the peel stress increases as the patch thickness t_R increases and decreases with increasing the adhesive thickness t_A. A reduction in the tip thickness of the repaired patch and a local increase in the adhesive layer thickness are beneficial since they reduce the peel stress significantly. As discussed in the previous section, this tapering is also advantageous in reducing the peak shear stress and strain in the adhesive. More detailed analysis of the optimum design for the double-lap joint examined here is presented in [11].

5. Maximum strength and failure modes

The failure load and modes of a double-lap joint are dependent on the adherend thickness. For sufficiently thin adherends, failure will always occur in the adherend outside the joint; adhesive strength is greater and the peel stresses are negligible. For slightly thicker adherends, adhesive failure will dominate and if the adherends are sufficiently thick, failure will always be by peeling apart the adhesive or the top layer of the composite laminate [11].

4 EXTERNAL PATCH REPAIR

The external patch repaired laminate, shown in Figure 1, is generally a three-dimensional (3-D) problem, which is oversimplified by the double-lap joint model described earlier. In a double-lap joint, all loads are transferred through the bonded patches. In fact, the parent plate can still carry load after losing the support of the patches (plate with an open hole). Therefore, the two-dimensional (2-D) analytical model underestimates the strength of the repaired laminate and a 3-D stress analysis should be performed. However, this is difficult to solve analytically and in the following section an FE method is used to calculate stresses developed in the patch, adhesive layer, and parent laminate.

4.1 Finite element model for a repaired laminate

Consider a symmetric laminate with a hole of diameter d, external patches bonded on both faces, and subjected to a uniaxial compressive loading, as shown in Figure 1. The $x-y$ plane of the Cartesian coordinate system lies in the mid plane of the laminate and the origin is at the center of the hole. This is an idealized case that simulates a situation where the impact-damaged laminate has been repaired by drilling a hole and then external patches are attached on both sides of the plate. An orthotropic lay up $[0/\pm\theta/90]_{ns}$ is selected; the total length of the carbon fiber-epoxy panel is 100-mm long by 50-mm wide, the hole diameter is 10 mm, and the laminate thickness is about 3 mm.

The FE77 FE package [12] is used and the analysis is based on displacement formulation employing

a curved isoparametric 20-node element. Owing to the symmetry of loading, hole location, and lay up, only one-quarter of the laminate is modeled; the FE mesh for the entire plate with circular patches is shown in Figure 8. The laminate and the patches are treated as homogeneous, elastic, and orthotropic materials with the following stiffness properties: $E_{xx} = E_{yy} = 55\,\text{GPa}$, $E_{zz} = 9.3\,\text{GPa}$, $G_{xy} = 20.95\,\text{GPa}$, $G_{xz} = G_{yz} = 4.4\,\text{GPa}$, $\nu_{xy} = 0.315$, and $\nu_{xz} = \nu_{yz} = 0.175$. The subscripts x, y, and z denote the loading, transverse, and thickness directions, respectively. The adhesive layer (Araldite 2005) of thickness $t_A = 0.1\,\text{mm}$ is modeled by using isotropic elements.

4.2 Stress results

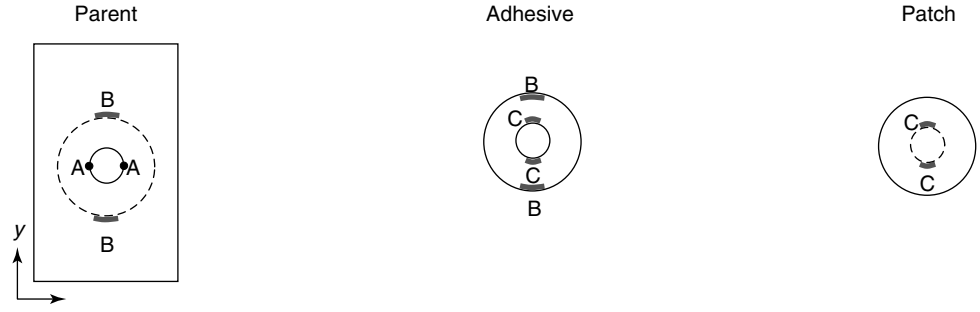
The stress concentration factors (SCF) in the parent, patch, and adhesive materials are summarized in Table 1 for a repaired plate with external round patches.

In the parent laminate, the normal stress σ_{xx} is the predominant component, which may cause failure. Two critical positions suffering of high stresses are identified; point A at the hole edge, along the y axis and point B at the edge of the overlap, along the x axis (the load direction) (see inset in Table 1). The stress magnitude at points A and B depends very much on the thickness of the patch. As the patch thickness increases, the SCF at point A is reduced but when it exceeds a certain value, $t_R > 1.5\,\text{mm}$ ($0.5\ t_p$), high stresses appear at point B.

The patch is subjected to relatively low stresses. The maximum value (339 MPa) of the dominant stress component, σ_{xx}, is less than the remote loading (350 MPa), when the patch thickness is 1.5 mm (half of the parent plate). Therefore, patches are generally safe if they are not too thin (here, greater than 30% thickness of the parent plate).

The adhesive/adherend interface is another location where failure may occur. The interlaminar (between plies) shear stress, τ_{xz}, is the main stress component

Table 1. Stress concentration factors, critical locations of round external patch repair, and the influence of patch thickness t_R



t_R (mm)	$\frac{(\sigma_{yy})_{max}}{\sigma_{remote}}$		$\frac{\tau_{max}}{\sigma_{remote}}$	$\frac{(\sigma_{zz})_{max}}{\sigma_{remote}}$	$\frac{(\sigma_{von}-\text{Mises})_{max}}{\sigma_{remote}}$	Locations	$\frac{(\sigma_{yy})_{max}}{\sigma_{remote}}$
	at "A"	at "B"					at "C"
0 (no patches)	3.19	—	—	—	—	—	—
0.5	2.03	1.13	0.250	0.119	0.461	"C"	1.62
0.8	1.73	1.19	0.294	0.174	0.539	"B"	1.34
1.0	1.57	1.22	0.314	0.204	0.575	"B"	1.21
1.3	1.40	1.27	0.338	0.240	0.616	"B"	1.05
1.5	1.31	1.29	0.350	0.261	0.637	"B"	0.968
1.7	1.24	1.32	0.362	0.279	0.656	"B"	0.901
2.0	1.16	1.34	0.377	0.302	0.681	"B"	0.821
2.5	1.06	1.38	0.396	0.332	0.714	"B"	0.727
3.0	0.99	1.40	0.411	0.354	0.738	"B"	0.666

to initiate failure. High shear stress concentration at the overlap edge may cause debonding of the repair patch. Only the narrow zones at the interface edges transfer load from the parent laminate to patches and most adhesive materials are in low stress state. High peel stresses, σ_{zz}, can also contribute to the final failure; they may cause the parent plate to delaminate because of its poor through-thickness strength.

The possible failure modes that can happen in the patch repaired configuration are identified in Figure 9, based on the stress results presented in Table 1 and a maximum stress failure criterion. For repairs with good bonding and thick patches ($t_R > t_p/2$), delamination may initiate at the edges of the bonding area (Figure 9a) due to high peel stress and fracture finally occurs across the net section after the loss of patch support. In repair configurations with strong bonding and thin patches, fracture initiates at the hole edges (point A, $SCF > 2$) and propagates across the plate width (Figure 9b). Finally, Figure 9(c) illustrates the case where the patch partially debonds resulting in large local stresses at the hole edge, point A, which then initiate fiber breakage (fiber microbuckling when loaded in compression), and final failure along a line almost perpendicular to the loading axis (Figure 9d and e).

In conclusion, the stress results presented in Table 1 suggest that the optimum external patch thickness is 1.5 mm, which, in this case, is half of the parent plate thickness; this is similar to the result obtained for the double-lap joint (Figure 5). Special attention should be paid to the bond shear and peel stresses, which markedly increase as the repair patch becomes thicker. Oversized patches are harmful because they increase structure weight and induce high stress concentrations in the repaired region.

5 EXPERIMENTAL EVIDENCE

In an earlier work by Soutis and Hu [13], repaired carbon fiber-epoxy laminates with specimen configuration, similar to that shown in Figure 1, were tested in uniaxial compression. Two different patch geometries were examined: (i) small overlap length, $\ell =$ 5 mm (10-mm hole, covered with 20-mm patches) and (ii) large overlap length, $\ell = 12.5$ mm (10-mm hole

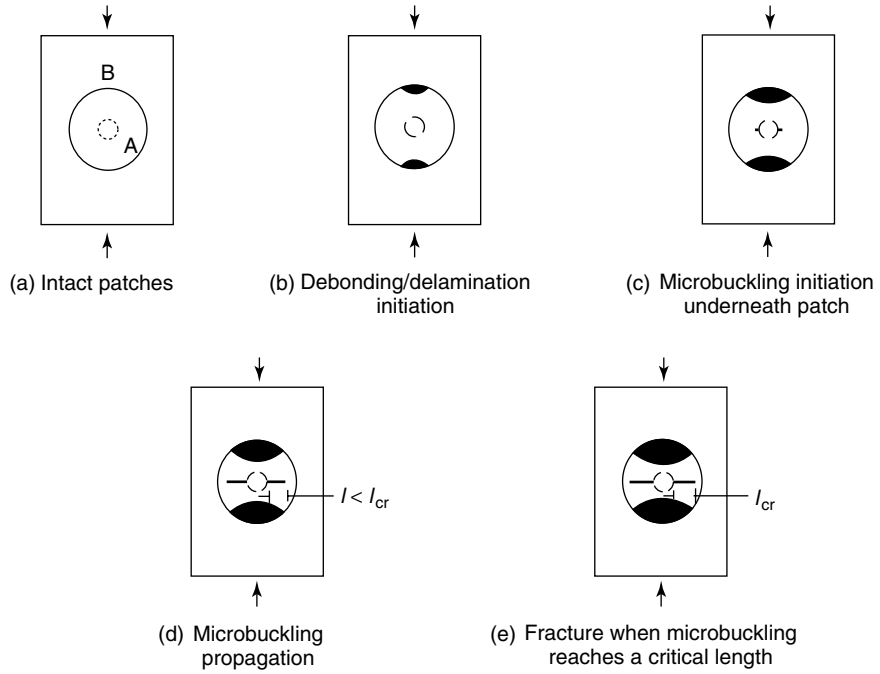


Figure 9. Possible failure mechanisms in a patch repaired laminate under compression.

Table 2. Compressive strength of repaired laminates

Overlap length (mm)	Square patch (MPa)	Round patch (MPa)
5	446.7	461.3
12.5	560	562

Gauge length = 100 mm, plate width = 50 mm, hole diameter = 10 mm unnotched strength = 680 MPa, notched strength = 437 MPa.

with 35-mm patch). The measured residual strengths of repaired laminates with round and square patches are summarized in Table 2.

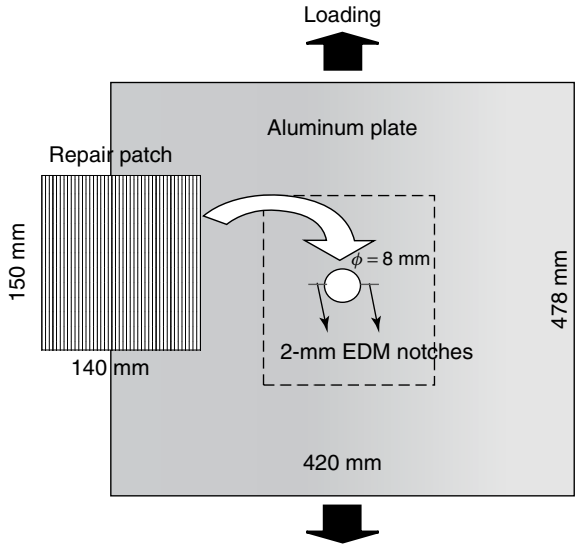
The round patch geometry performs slightly better than the square patch owing to smaller stresses developed in the repair region. The small overlap length recovers 65–68% of the undamaged strength compared to more than 80% for the larger patch. Notice that the 12.5-mm overlap is that obtained from the double-lap joint analysis. If the composite were perfectly brittle, failure would occur when the maximum stress in the structure equals the unnotched compressive strength of the material,

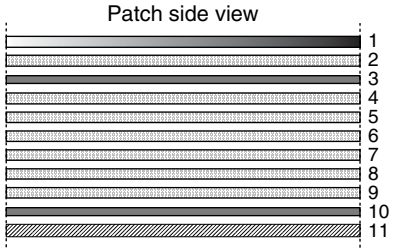
$$\sigma_{\mathrm{R}} = \frac{1}{K_t}\sigma_{\mathrm{un}} \tag{2}$$

where K_t is the stress concentration factor at point A and σ_{un} is the unnotched compressive strength. In this case, $K_{\mathrm{t}} = 1.31$ (Table 1) and $\sigma_{\mathrm{un}} = 680$ MPa, predicting a residual strength of 519 MPa, which is only 8% less than the measured value (560 MPa). Some stress redistribution is possible owing to local damage in the form of resin cracking and fiber/matrix splitting that is not considered in the present analysis. It should be noted that the elastic stress concentration factor for the orthotropic plate with an open hole is 3.19 and can be substantially reduced by selecting an appropriate patch ($K_t = 1.31$) and accounting for the nonlinear response of the adhesive. Further experimental work is reported in [13–17], where local damage initiating from the patches is monitored using X-ray radiography and scanning electron microscopy. Analytical and numerical models that incorporate the load redistribution in the failure stress calculations can be found in these publications [13–17]. The following section discusses how a rectangular aluminum plate with a cracked hole is repaired with a unidirectional boron/epoxy patch that contains two layers of piezoelectric transducers to monitor crack growth under fatigue loading.

6 DAMAGE DIAGNOSIS OF A BONDED PATCH REPAIR

A rectangular aluminum plate (420 mm × 478 mm × 3.175 mm) with 2-mm-long EDM (electric discharge machining) notches at an 8-mm-diameter hole was repaired with a unidirectional boron/epoxy patch. The thickness of the boron/epoxy patch and aluminum plate was chosen on the basis of the stiffness match design criterion, $E_{\mathrm{R}}t_{\mathrm{R}} \approx E_{\mathrm{p}}t_{\mathrm{p}}$ (where $E_{\mathrm{R}} = 2.4E_{\mathrm{p}}$). The repair configuration and repair materials used are listed in Figure 10. It can be seen that the size of the patch is proportionately larger that the patch used in Section 5 to account for environmental effects, manufacturing defects, and fatigue loading, parameters that can affect the patch efficiency and were not considered in the static analysis presented in the earlier sections. The stiffness matrix of the boron/epoxy laminated patch was taken as $C_{11} = 210.1$ GPa, $C_{12} = C_{13} = 5.64$ GPa, $C_{22} = C_{33} = 26.3$ GPa, and $C_{23} = 4.5$ GPa, with shear stiffness $C_{44} = 10.3$ GPa and $C_{55} = C_{66} = 7.2$ GPa. A Young's modulus of 2.07 GPa and a Poisson's ratio of 0.34 were used for the adhesive layer. Two smart layers were fabricated [18, 19] with an embedded network of piezoelectric actuators/sensors and inserted into the patch at different ply locations (Figure 10). The lower smart layer (layer 10) was inserted right on the interface as close to the neutral axis as possible to excite a more symmetric (Lamb wave) mode [20]. Conversely, the upper smart layer (layer 3) was inserted near the top layer to excite a more antisymmetrical mode. Thus, the lower smart layer targets for the crack growth on the aluminum plate and the upper smart layer targets for the possible patch disbond from the aluminum plate. The locations of the piezoelectric actuator and sensors with damage diagnostic paths on the smart layers are shown in Figure 11. The repair layers and smart layers were stacked as seen in Figure 10 and cocured with aluminum plate, with a vacuum bag, under a constant temperature of 121 °C for 90 min. A smart suitcase [19] (a portable active diagnostic instrument designed to interface with piezoelectric transducers) was implemented to generate diagnostic signals and





Materials

Legend	Material	Supplier	Thickness (mm)
	5521 Boron/epoxy prepreg	Textron	0.127
	FM 73 adhesive	Cytec	0.254
	SMART layer	Acellent	0.1016

Patch lay ups

No.	Lay up (top to bottom)	Length (mm)	Width (mm)
1	Boron/epoxy	150	140
2	Boron/epoxy	150	140
3	SMART layer #2	150	140
4	Boron/epoxy	150	140
5	Boron/epoxy	150	140
6	Boron/epoxy	150	140
7	Boron/epoxy	150	140
8	Boron/epoxy	150	140
9	Boron/epoxy	150	140
10	SMART layer #1	150	140
11	Adhesive	150	140

Figure 10. Bonded patch specimen assembly.

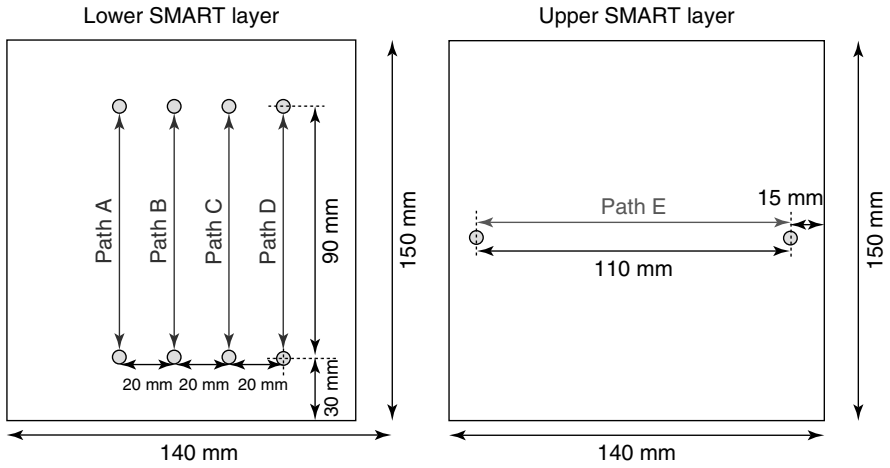


Figure 11. smart layers with damage diagnostic paths.

record measurements from the smart layers. Using a built-in waveform generator, a windowed sine burst wave was generated as an input signal over a wide frequency range of 100–600 kHz.

The repaired specimens were loaded in tension–tension fatigue and the lower smart layer (layer 10, Figure 10) was used for monitoring crack growth. Sensor measurements from the lower smart layer

were taken at each specified loading cycle while the specimen was unloaded. The corresponding crack lengths were visually measured.

7 LAMB MODE IDENTIFICATION

7.1 Fiber direction

The boron/epoxy patch-adhesive layer-aluminum plate specimen was modeled numerically using the disperse code [21, 22] and the group velocity dispersion curves in various modes were generated (Figures 12 and 13). Owing to the directional properties of the patch, the group velocity dispersion curves can be obtained at various wave propagation angles using the code. Figure 12 illustrates the numerically calculated group velocity dispersion curve of the asymmetric and symmetric Lamb wave modes (a_0 and s_0) when the wave propagation direction is parallel to the boron fibers ($\theta = 0$) and compared to experimental measurements by the lower smart layer (that is used for crack detection and the diagnostic

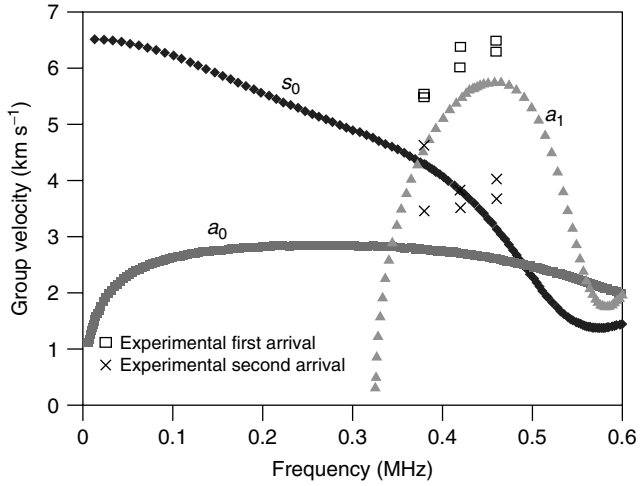


Figure 12. Group velocity dispersion curve of boron/epoxy-adhesive-aluminum (fiber direction, $\theta = 0$).

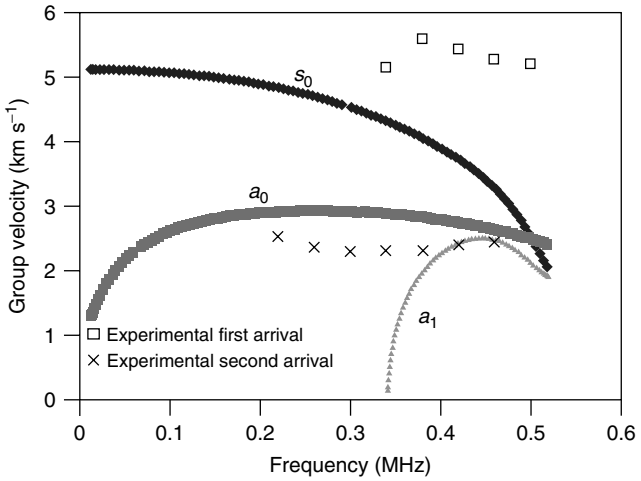


Figure 13. Group velocity dispersion curve of boron/epoxy-adhesive-aluminum ($\theta = 90^{\circ}$).

paths are parallel to the boron fibers). Similarly, in Figure 13 the equivalent results are presented when $\theta = 90$, i.e., wave propagation is perpendicular to the fiber direction; the experimental data in Figure 13 were from the upper smart layer that was designed to detect the patch disbond. A reasonably good agreement between analysis and experiment is observed. Sensor measurements were taken from the path E (Figure 11) in the upper smart layer and Figure 14 shows the amplitude spectrums of both the a_0 and s_0 modes before and after the debond damage where the amplitude was normalized by the maximum value of the a_0 amplitude spectrum. It appears that the fundamental antisymmetric mode a_0 shows higher sensitivity to the disbond damage.

8 DAMAGE INDEX RESULTS

8.1 Damage index

The damage index (DI) is defined as the relative ratio of the scatter energy contained in a Lamb mode wave packet to the baseline energy contained in the same Lamb mode wave packet. This ratio can be obtained by the time integration of the power scatter spectral density within a specified time window at a specified frequency, which can then be nondimensionalized by baseline information such that

$$\text{Damage index (DI)} \equiv \left(\frac{\int_{t_i}^{t_f} |S_{\text{sc}}(\omega_0, t)|^2 \, \mathrm{d}t}{\int_{t_i}^{t_f} |S_{\text{b}}(\omega_0, t)|^2 \, \mathrm{d}t} \right)^{1/2} \quad (3)$$

where S_{sc} denotes the time varying spectral amplitude of scatter signal, S_{b} is the time varying spectral amplitude of baseline signal, ω_0 is the selected driving frequency, and t_f and t_i denote the upper bound and lower bound of the selected Lamb mode in time domain, respectively. A time varying spectral amplitude $S(\omega, t)$ of a given signal $s(t)$ is obtained by the short-time Fourier transform:

$$S(\omega, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega\tau} s(\tau) h(\tau - t) \, \mathrm{d}\tau \quad (4)$$

where $h(t)$ is the Hanning window function. Typically, the DI selects the symmetric mode (s_0) for crack detection and the antisymmetric mode (a_0) for disbond-type damage.

8.2 Upper smart layer

The DI can be used for detecting both crack and debond damage by choosing the s_0 mode and a_0

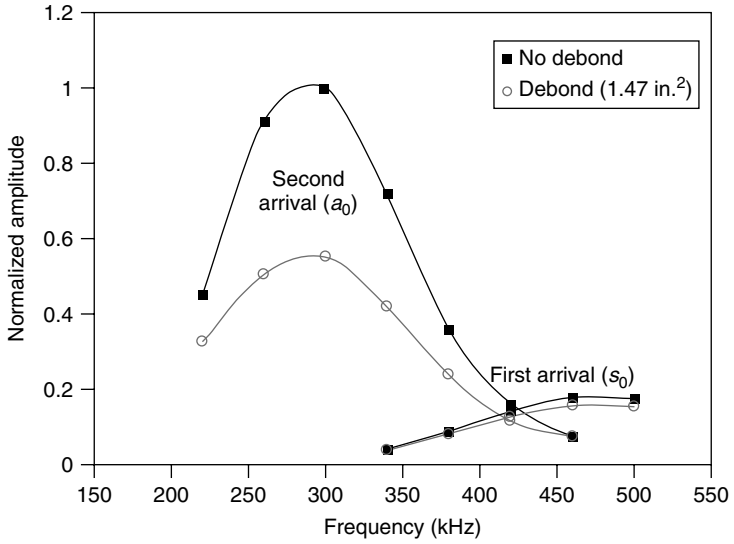


Figure 14. Normalized amplitude responses of the upper smart layer (path E) to debond damage ($1\text{in.}^2 = 645.16$).

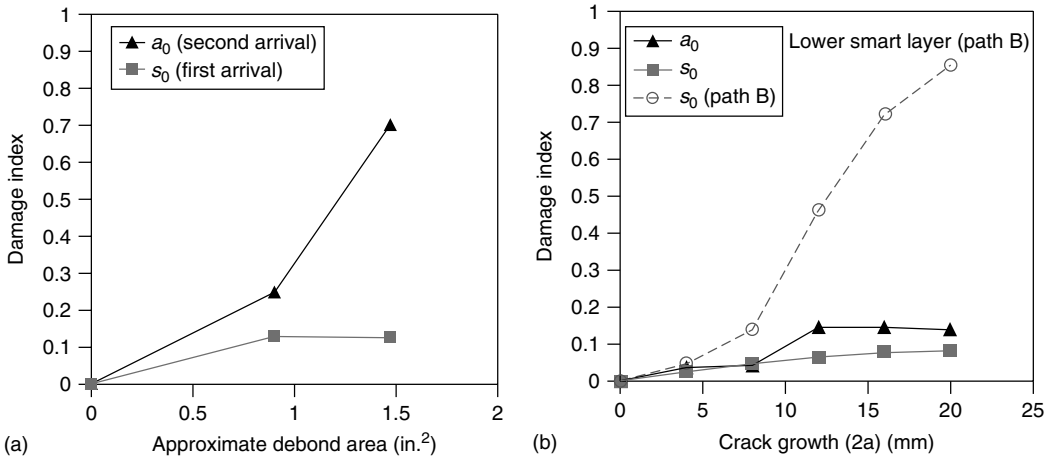


Figure 15. The response of s_0 and a_0 modes to (a) patch debond and (b) crack growth.

mode, respectively. Figure 15 shows the sensitivity of the s_0 and the a_0 modes generated by the upper smart layer to the debond damage and the crack growth. As previously observed in the amplitude spectrum plot (Figure 14), the fundamental antisymmetric mode (a_0) is more sensitive to debond damage than the s_0 mode. Both the a_0 and the s_0 modes generated by the upper smart layer were not sensitive to the crack growth when compared with the s_0 mode of the lower smart layer. The relatively low sensitivity of the s_0 mode of the upper smart layer (path E) is due to its wave propagation direction that is parallel to the crack propagation direction.

8.3 Lower smart layer

The sensor data from the 420-kHz input signal was used for the analysis as this data has the highest signal-to-noise ratio. The DI was evaluated on the basis of the time of arrival of the second wave packet. The new baseline was taken after the specimens were cycled until an initial crack growth of 2 mm had occurred. Further crack growth was visually identified and at the same time monitored by the DI, which was evaluated at different diagnostic paths as shown in Figure 16(a). The results are plotted in Figure 16(b), which indicates the clear crack growth detection capability of the DI. It also shows the higher sensitivity of the DI in a consistent manner, as the diagnostic paths (paths B and C) are closer to the initial crack-tip location.

8.4 Overall variations of damage index

As shown in Figure 17, various mechanical tests were conducted on the bonded patch specimen at different times, which are denoted as "serial events". The specimen was first tested only for the patch debond (that was introduced statically by pushing down the patch from the back face using a cylindrical rod) and then fatigued under cyclic loading for the fatigue crack growth. The initial sensor measurement at "event 0" or "pristine condition" was kept as a baseline for the entire evaluations to see the overall variations of the DI under different loading and time history. It can be observed in Figure 17 that the DI by the upper smart layer shows good sensitivity to the debond damage and remains with small variation afterward over time. On the other hand, the DI by the lower smart layer shows good sensitivity only to the crack growth. Also the DIs from both smart layers show no or minimal changes when there is no damage occurring during the cyclic loading.

9 CONCLUDING REMARKS

The analytical/numerical approaches presented in this work can improve the design of external patch repairs and estimate their strength with some success. For the cases examined, the optimum overlap length of the patch is approximately 12–15 mm. The optimum patch membrane stiffness (i.e., the product of elastic

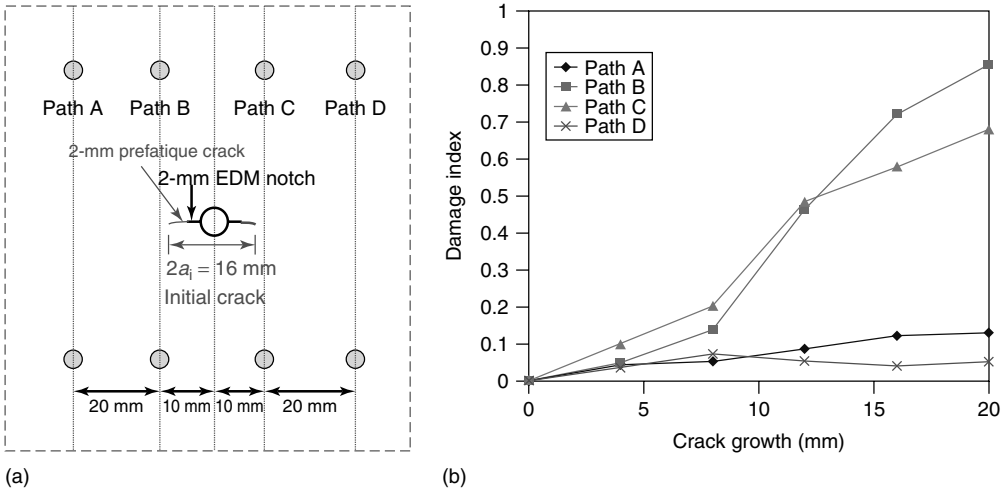
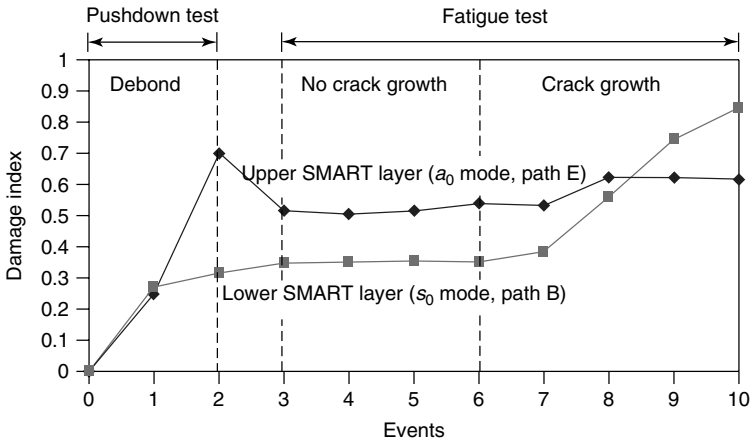


Figure 16. Diagnostic paths and damage index of lower smart layer for crack detection. (a) Diagnostic paths and (b) damage index versus crack growth.



Serial events	0	1	2	3	4	5	6	7	8	9	10
Debond (in.^{2})	0	<1	1.47	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.54
Cycle (x1000)				65	80	123	429	616	852	902	948
Crack growth 2a (mm)	0	0	0	0	0	0	8	12	16	20	24

Figure 17. Variation of damage index in the series of debonding and crack growth tests.

modulus and thickness) is half of the parent stiffness. Overstiff or too-thick patches are harmful because they produce higher peel and shear stress concentrations and increase the weight of the structure. Tapering the patch edges and increasing the local adhesive thickness reduces the local stresses substantially. Using the optimum patch configuration, bonded repairs can recover up to 80% of the undamaged laminate strength. The through-thickness stresses can be analyzed by performing ply-by-ply

FE analysis and damage initiation predictions can be made using simple stress-based failure criteria. However, fracture mechanics needs to be applied to model the process of damage growth. Environmental effects (hot/wet) and durability of the repairs are also to be considered in the repair design, since moisture and temperature do alter the properties of the resin and affect the load-carrying capability of the repaired configuration. In addition, it is demonstrated that smart sensors could be embedded in the patch to monitor the structural health of the repaired configuration. A diagnostic system based on the active piezo sensor network was instrumented with an aluminum plate repaired by a boron/epoxy composite repair patch. The DI results from the two active-sensing layers inserted at different ply locations clearly indicate that crack growth underneath the patch and debond damage, although they are colocated, can be separately monitored. More details of the embedded active-sensing approach and the DI results can be found in the literature [23–26]. Of course, other damage inspection methods could be used to monitor the structural health of the repaired configuration. Radiography, ultrasonics, shearography, thermography, embedded optical fibers [27], shape memory alloy (SMA) wires [28], and fiber Bragg grating (FBG) sensors [29] are among the most commonly used, but it is demonstrated in this article that Lamb waves offer an attractive way for large-area nondestructive inspection, as the waves are able to travel relatively long distances allowing the material between transmitter and receiver to be successfully interrogated. The Lamb wave based approach could be applied to the health monitoring of wind turbines (*see* **Chapter 147**) and marine structures (*see* **Chapter 142**).

REFERENCES

[1] Baker AA, Jones R (eds). *Bonded Repair of Aircraft Structures*. Martinus Nijhoff Publishers, 1988.

[2] Armstrong KB. British airways experience with composite repairs. In *Composite Materials in Aircraft Structures*, Middleton DN (ed). Longman Scientific & Technical, Longman Group, 1990, 368–378.

[3] Myhre SH, Labor JD. Repair of advanced composite structures. *Journal of Aircraft* 1981 **18**(7):546–552.

[4] Matthews FL (ed). *Joining Fibre-reinforced Plastics*. Elsevier Applied Science, 1987.

[5] Matthews FL, Rawlings RD. *Composite Materials: Engineering and Science*. Chapman & Hall, 1994.

[6] Hart-Smith LJ. Further developments in the design and analysis of adhesive-bonded structural joints. *Symposium on Joining of Composite Materials*. American Society for Testing and Materials, (STP 749), Minneapolis, MN, 16 April 1980.

[7] Hart-Smith LJ. An engineer's viewpoint on design and analysis of aircraft structural joints. *Proceedings of the Institution of Mechanical Engineers Part G-Journal of Aerospace Engineering* 1995 **209**:105–129.

[8] Volkersen O. Die Nietktraftverteilung in zugbeanspruchten Nietverbindungen mit konstanten Laschenquerschritten. *Luftfahrtforschung* 1938 **15**:41.

[9] Hu FZ, Soutis C. *Analysis and Optimisation of Bonded Patch Repairs in Composite Structures*, Imperial College Technical report, GR/K54892. EPSRC, March 1996, p. 70.

[10] Jones JS, Graves SR. *Repair Techniques for Celion/LARC-160 Graphite/Polyimide Composite Structures*, NASA-CR-3794. NASA, 1984.

[11] Hitchings D. *Finite Element Package FE77 User's Manual*. Imperial College, 1995.

[12] Vlattas C, Soutis C. Composite repair: compressive behaviour of CFRP plates with reinforced holes. *7th European Conference on Composite Materials*, London, 14–16 May 1996; pp. 87–92.

[13] Soutis C, Hu FZ. Design and performance of bonded patch repairs of composite structures. *Journal of Aerospace Engineering* 1997 **211**:263–271.

[14] Soutis C, Duan D-M, Goutas P. Compressive behaviour of CFRP laminates repaired with adhesively bonded external patches. *Composite Structures* 1999 **45**(4):289–301.

[15] Hu FZ, Soutis C. Strength prediction of patch repaired CFRP laminates loaded in Compression. *Journal of Composites Science and Technology* 2000 **60**(7):1103–1114.

[16] Soutis C, Hu FZ. Failure analysis of scarf-patch-repaired composite laminates loaded in compression. *AIAA Journal* 2000 **38**(4):734–740.

[17] Soutis C. Fibre reinforced composites in aircraft construction. *Progress in Aerospace Sciences* 2005 **41**(2):143–151.

[18] Lin M, Chang F-K. *Manufacturing of Composite Structures with a Built-in Network of Piezoceramics*, Ph.D. Dissertation. Department of Mechanical Engineering, Stanford University: Stanford, CA, 1998.

[19] Lin M, Qing X, Kumar A, Beard S. SMART layer and SMART suitcase for structural health monitoring applications. *Proceedings of SPIE on Smart Structures and Material Systems*. Newport Beach, CA, 2001; Vol. 4332, pp. 98–106.

[20] Diamanti K, Soutis C, Hodgkinson JM. Non-destructive inspection of sandwich and repaired composite laminated structures. *Composites Science and Technology* 2005 **65**(13):2059–2067.

[21] Pavlakovic B, Lowe M. *A System for Generating Dispersion Curves—User's Manual for Disperse v.2.0*, June 2000.

[22] Lowe MJS, Pavlakovic BN, Cawley P. Guided wave NDT of structures: a general purpose computer model fro calculating waveguide properties. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Stanford, CA, 2001; pp. 880–888.

[23] Ihn J-B, Chang F-K. Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: part I. Diagnostics. *Smart Materials and Structures* 2004 **13**:609–620.

[24] Ihn J-B, Chang F-K. Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: part II. Validation through riveted joints and repair patches. *Smart Materials and Structures* 2004 **13**:621–630.

[25] Ihn J-B, Chang F-K. A smart patch for monitoring crack growth in metallic structures underneath bonded composite repair patches. *Proceedings of the American Society for Composites 17th Technical Conference*. Purdue University, Lafayette, IN, 2002.

[26] Ihn J-B, Chang F-K. Built-in diagnostics for monitoring crack growth in aircraft structures. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Stanford, CA, 2001; pp. 284–295.

[27] Jones R, Galea S. Health monitoring of composite repairs and joints using optical fibres. *Composite Structures* 2002 **58**:397–403.

[28] Qiu Z-X, Yao X-T, Yuan J, Soutis C. Experimental research on strain monitoring in composite plates using embedded SMA wires. *Smart Materials and Structures* 2006 **15**(4):1047–1053.

[29] Takeda S, Yamamoto T, Okabe Y, Takeda N. Debonding monitoring of composite repair using embedded small diameter FBG sensors. *Smart Materials and Structures* 2007 **16**:763–770.

Chapter 104

Thermal Protection System Monitoring of Space Structures

William H. Prosser[1], Eric I. Madaras[1], George F. Studor[2] and Michael R. Gorman[3]

[1] *NASA Langley Research Center, Hampton, VA, USA*
[2] *NASA Johnson Space Center, Houston, TX, USA*
[3] *Digital Wave Corporation, Englewood, CO, USA*

1 INTRODUCTION

Damage caused by the impact of foam insulation shed from the external tank of the space shuttle shortly after launch was suspected as a leading candidate for the cause of the loss of the space shuttle Columbia during reentry on February 1, 2003. As a result, an experimental test program was initiated during the accident investigation to reproduce this impact event and estimate the resulting damage to the thermal protective systems (TPSs) on representative shuttle wing structures. In addition to reproducing the impact and resulting damage that led to the accident, NASA had the foresight to use these impact tests to develop and demonstrate acoustic sensor technology to detect impact damage on future shuttle flights. Previous testing [1, 2] had already demonstrated that such sensors might be used to detect and locate micrometeoroid and orbital debris (MMOD) impact events on spacecraft. Although ascent debris damage was the focus of the Columbia investigation, MMOD had also been identified as a significant potential danger to both the shuttle and the space station [3]. Both low-frequency accelerometer and high-frequency ultrasonic acoustic emission (AE) sensors were evaluated for this purpose during the accident investigation.

Testing during the investigation successfully validated the capability of these sensors for detecting major impact damage. However, additional testing was necessary to develop this sensing approach for application to the remaining shuttle fleet. These

 ISBN: 978-0-470-05822-0.

tests have included the determination of sensor response to a range of energies of foam impact events on TPS materials including those that are near to and below the threshold of damage. Additionally, impact tests have been performed with a number of other potential impact materials that can damage the shuttle during ascent including ice, ablator, and metal. Also, since it is desirable to have the impact sensing system detect not only ascent debris impacts but also those of MMOD during orbit, testing has been performed to measure sensor response to hypervelocity impacts on TPS materials. In addition to impact testing on structural test articles, testing was performed on the shuttle Endeavor to study wave propagation effects and evaluate differences in structural configuration between Columbia test articles and the remaining shuttle fleet. An overview of these test results is presented, along with a discussion of results obtained from wing leading-edge impact-detection system (WLEIDS) sensors deployed on the space shuttles during recent missions.

2 COLUMBIA ACCIDENT INVESTIGATION FOAM IMPACT TESTING

At the onset of the Columbia accident investigation, it was not known exactly where the foam debris impacted the shuttle wing. Video images showed that it struck somewhere on the lower surface of the left wing. However, the views and resolution available did not indicate whether it struck the leading edge, which consists of reinforced carbon–carbon (RCC), or the lower wing surface, which has thermal protection consisting of tile. Thus, a variety of test specimens were fabricated to investigate the damage caused by foam impacts on these structures. In addition, preliminary testing to calibrate the foam mpact gun performance as well as test instrumentation configuration was performed using aluminum plate targets. Accelerometers and AE sensors were included on all of these tests and successfully detected the impacts in all cases.

As the investigation progressed, sensor data from Columbia and forensics of debris provided indications that the damage had occurred on the leading edge, specifically on RCC panel 8. The focus of the impact testing turned toward foam impacts on leading-edge panels mounted on a leading-edge support structure (LESS) as shown in Figure 1. This test article consisted of a section of leading edge spar using the honeycomb structural configuration from Columbia, to which leading-edge panels 5–10 were attached. Because of the enormous expense and limited availability of RCC panels, initial testing was performed using fiberglass replicas of the leading-edge panels, with final testing performed on flight RCC panels. An array of eight AE sensors (Digital Wave Corporation model B225-5) was attached on the interior side of the spar. The bandwidth of these transducers was specified by the manufacturer to be 30–300 kHz. However, responses well below 10 kHz were measured. Initial testing with the sensors arrayed close to the impact

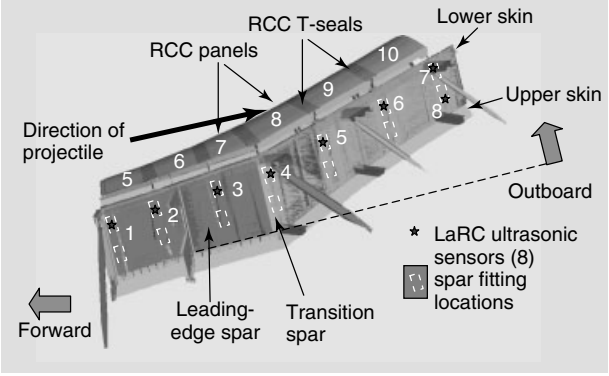


Figure 1. Leading-edge support structure with RCC panels 5–10 and T-seals shown. The locations of the AE sensors 1 through 8 are indicated with black stars.

point demonstrated that signals of significant amplitude were produced and that these signals propagated through the attach fittings into the spar. For later testing, the sensors were arrayed along the length of the spar as shown in Figure 1 to determine how well the signals propagated along the spar, and thus how remote the sensors could be located and detect the impact. As the foam impacts and the attenuation of the complicated structure resulted in very low AE frequency signal content, the AE data was acquired at a sampling frequency of only 500 kHz with a total of 32 K points acquired for each sensor.

For the defining test of the investigation, a foam block of 0.757 kg mass (1.67 lbs weight) was launched at a velocity of 237 m s^{-1}, striking panel 8 as indicated in Figure 1. This impact produced a significant hole in the RCC panel providing conclusive evidence for the Columbia Accident Investigation Board in determining the cause of the accident [4]. The AE signals that were detected from this foam impact event are shown in Figure 2. Only 6 dB of gain was applied to the signals from the AE sensors. As would be expected, the largest signal, arriving earliest in time, was that from sensor 5, which was nearest the impact site. Quantitatively, decreasing arrival times and amplitudes of signals from sensors located further away from the impact point were observed. Although not noticeable in this figure as all signals are plotted on the same scale, signals were detected all the way down to the location of sensor 1, suggesting that impact events can be detected by sensors mounted several RCC panels away, a distance of more than 1 m. Examination of the arrival times for signals from sensors 7 and 8 showed that the impact site could be localized with respect to the upper and lower surface of the leading edge.

3 RETURN TO FLIGHT TESTING

At the completion of the accident investigation, a number of questions remained regarding the capability of acoustic impact sensing on the shuttle. These included the detectability of much smaller foam impacts including those near or below the threshold of damage, the characteristics of signals caused by other potential impact source materials including ice, ablator, and metal at ascent velocities, as well as hypervelocity impacts to simulate orbital impacts. These effects needed to be assessed for impacts on both the leading edge as well as on the tile protected lower wing surface including the main landing gear door. Another issue was that the construction of the wing spar on the remaining shuttle fleet varied considerably from that of Columbia and the effects of this difference on acoustic wave propagation had to be investigated. Thus, a comprehensive test program was initiated to address these questions. As it is impossible, as well as expensive, to test all possible combinations of impact parameters, a simultaneous modeling effort was initiated to develop capabilities to model impact events on shuttle wing structures. One key experimental piece of data required for these models was the measurement of the transfer function of the acoustic signals from the RCC leading edge to the spar where sensors are located. Additional experiments were performed to acquire this critical data.

3.1 Launch debris impact testing

Additional foam impact tests were performed on RCC panels over a range of projectile sizes and impact velocities. These impact tests were performed on different panels on the LESS test article, as

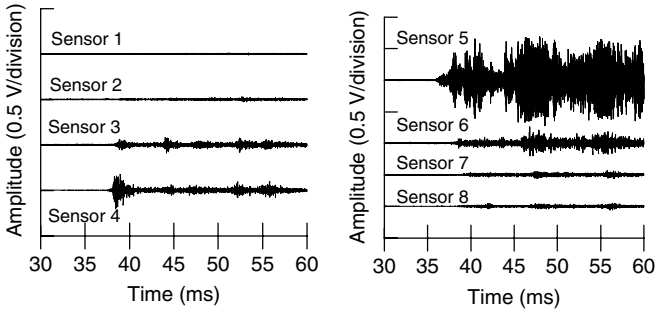


Figure 2. AE signals from foam impact on shuttle RCC wing leading-edge (LESS) tests.

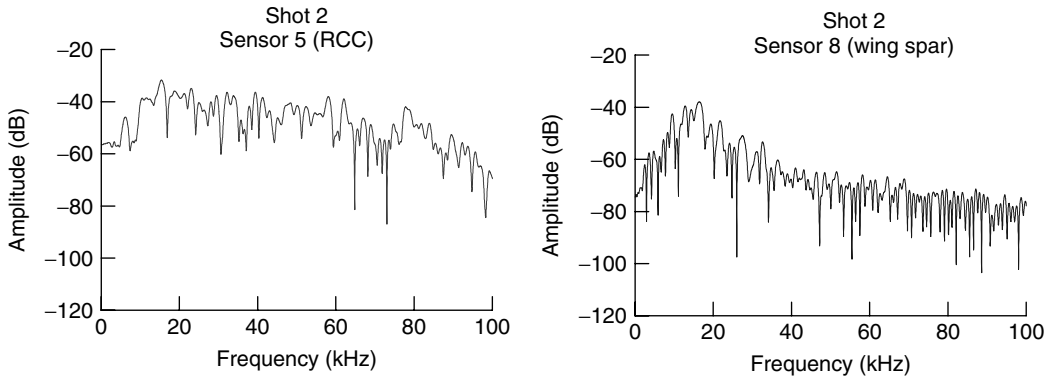


Figure 3. Frequency content of foam impact signals for sensors on RCC panel and wing spar.

well as on the T-35 test article, which represented a more outboard section of the wing. This test article allowed impact tests on panels 16 and 17 and further provided the opportunity to evaluate the effect of differing impact locations on measured signals. Signals from small projectiles and/or low-impact velocities producing impact energies below the threshold of damage were still readily detected. Variations in the signal amplitude correlated with the impact energy for a given type of impact material. However, different impact materials such as foam and ice exhibited different amplitude impact energy relationships. In addition to sensors on the spar, sensors were also placed on the RCC panel of the T-35 test article to measure the transfer function response from the RCC panel to the spar. The frequency response plots in Figure 3 show the significant loss in high-frequency signal content that occurs as the signal propagates from the RCC to the wing spar of Columbia construction. Preliminary testing on test articles with the wing spar construction of the remaining shuttle fleet suggests that this high-frequency attenuation might not be as severe.

Foam impact tests were also performed on lower wing specimens representative of regions on which the thermal protection material is tile. Specimens from this region of the wing also included a main landing gear door. Representative damage for a wing specimen impacted by foam at approximately $290\,\mathrm{m\,s^{-1}}$ is shown in Figure 4 in which a hole formed by a tile that was broken away by the impact can be observed. Again, AE transducers readily detected signals for all impact conditions studied.

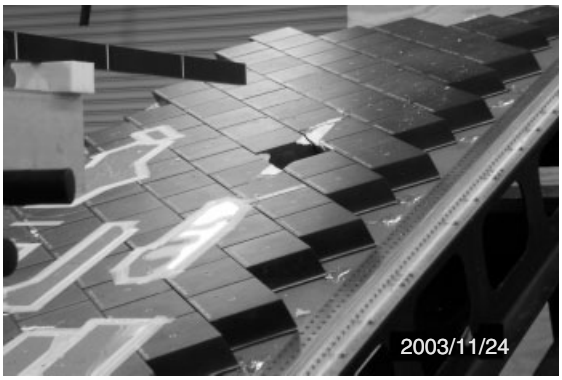


Figure 4. A wing acreage tile test article showing the resulting tile loss due to a foam impact.

Although the signals were very complex due to the complicated nature of the source and the complex structural geometry of the tile and wing specimen, source location could be determined using appropriate frequency filtering to selectively analyze the flexural mode of propagation.

Impact testing on RCC and tile specimens was also performed using other types of potential launch debris. These materials included ice, ablator, and metal. Again, the impact velocity and energy was varied over a range from below the damage threshold to that causing substantial damage. AE and accelerometer sensor data were obtained for all tests. Preliminary analysis shows that all impacts were successfully detected with both accelerometers and AE sensors, and that again there was a correlation between signal amplitude and energy of impact for a given impact material.

3.2 Hypervelocity impact testing

Hypervelocity impact tests were performed to simulate MMOD damage that can occur once the shuttle is in orbit. Initial tests were performed on flat metal and fiberglass plates to develop a database to support modeling efforts as well as to determine appropriate instrumentation settings. Figure 5 shows typical damage resulting from two hypervelocity impact events at 6.8 km s^{-1} in a fiberglass plate. A 2-mm-diameter aluminum projectile created the smaller impact while the larger was created by a 6-mm aluminum projectile, which fully penetrated the plate. Figure 6(a) shows the signals from a hypervelocity impact, while for comparison, a lead break simulated AE signal near the impact site is shown in Figure 6(b). Curiously, the flexural mode amplitude was generally smaller than the extensional mode, especially at the higher energy shots. This is interesting since low-velocity impact usually produces a large amplitude flexural mode due to the source motion perpendicular to the plate target. In the present case, the attenuators played a role in filtering the low frequencies that generally confirm the presence of a flexural wave. However, the source function for hypervelocity impact is quite a bit different than ball drop at low velocity or a lead break. It is also interesting to note in comparing these signals that there was 64 dB of attenuation applied to the signals from the hypervelocity impact as compared to 47 dB of gain for the lead break signal. There is a tremendous amount of energy in the hypervelocity impacts. Figure 7 shows the raw signal amplitude, after adjustment for the attenuation, from a series of hypervelocity impacts on a fiberglass plate as a function of impact energy. As shown in this figure, the raw signal amplitude increases with corresponding impact energy until it peaks at nearly 80 V for an impact energy of nearly 100 J. In the fiberglass plates, for impacts exceeding 100 J, the projectile penetrates the plate and a decrease in AE signal amplitude was observed. However, for actual RCC leading-edge specimens, a decrease in AE was not observed after impact penetration.

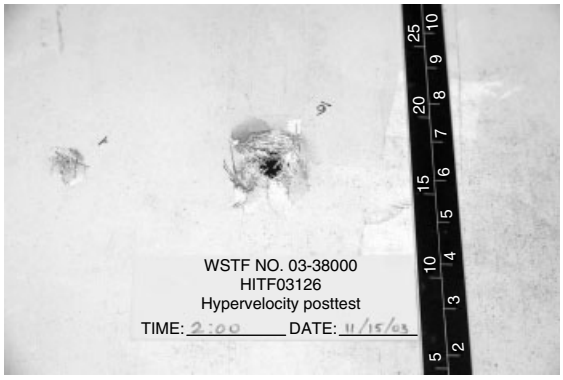


Figure 5. Fiberglass panel showing damage from two hypervelocity impacts.

Propagation effects on AE signals from the impacted material through attachment mechanisms to likely sensor locations on the spar were also investigated. Initial testing for this consisted of multiple plates connected by threaded rods, followed by testing on a realistic shuttle wing spar test article. Again, because of the expense of RCC panels, testing included hypervelocity impacts on a number of fiberglass replicas of a leading-edge panel, followed by shots on an

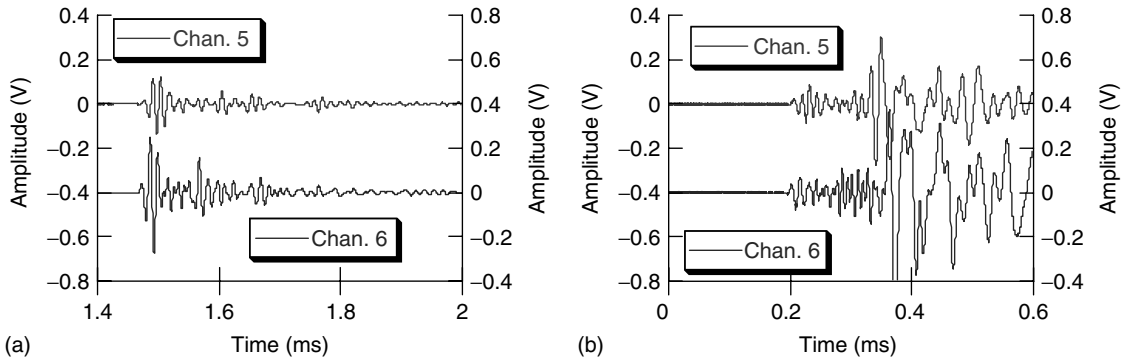


Figure 6. AE signals produced by (a) hypervelocity impact and (b) pencil lead break.

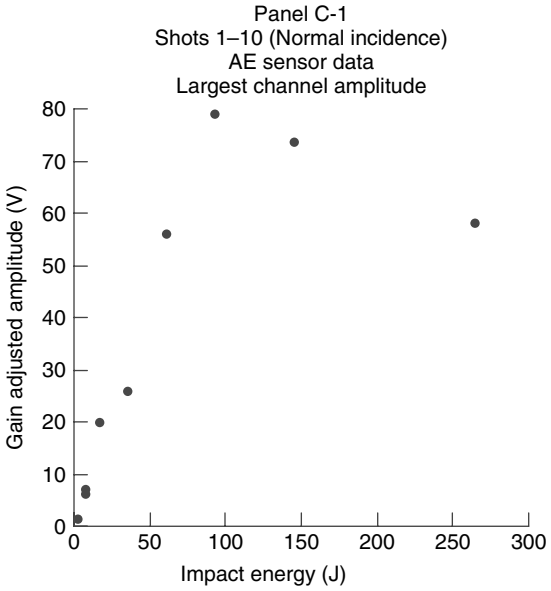


Figure 7. AE signal amplitude versus impact energy on fiberglass targets.

actual RCC panel. These tests demonstrated that the much higher frequency hypervelocity impact signals are much more heavily attenuated than was observed for the lower frequency foam impact signals. Further analysis was performed to determine the transfer function from the RCC to the spar where the sensors are located on the flight vehicle.

3.3 Impact hammer testing

Impact hammer and pulsed pitch–catch ultrasonic measurements were made on the wing spar of the shuttle Endeavor to investigate the effects on wave propagation due to differences in wing spar construction. As noted previously, the LESS and T-35 test articles represented the Columbia wing spar construction, which is different from the remainder of the fleet. Transducers were attached to the leading edge of the shuttle's wing, as indicated in Figure 8. At various locations, ultrasonic signals between 10 and 150 kHz were introduced and recorded on the fixed transducers. In addition, a series of low-energy instrumented hammer impacts (9.07, 27.22, 68.04, 113.40 kg) were performed on the wing's leading edge. Similar experiments were performed on the LESS and T-35 test articles to develop a correlation between the different structures. Figure 9 shows the frequency response of AE sensors to a hammer impact on the shuttle Endeavor wing spar as well as on the LESS test article. Although the overall peak amplitudes of the time-domain signal are similar, the frequency response shows that the peak amplitude is at a much lower frequency with much higher frequency attenuation for the LESS as compared to the shuttle. These differences are significant in that they indicate that higher frequency signal components may propagate from impacts on the shuttle to and along the spar. Such

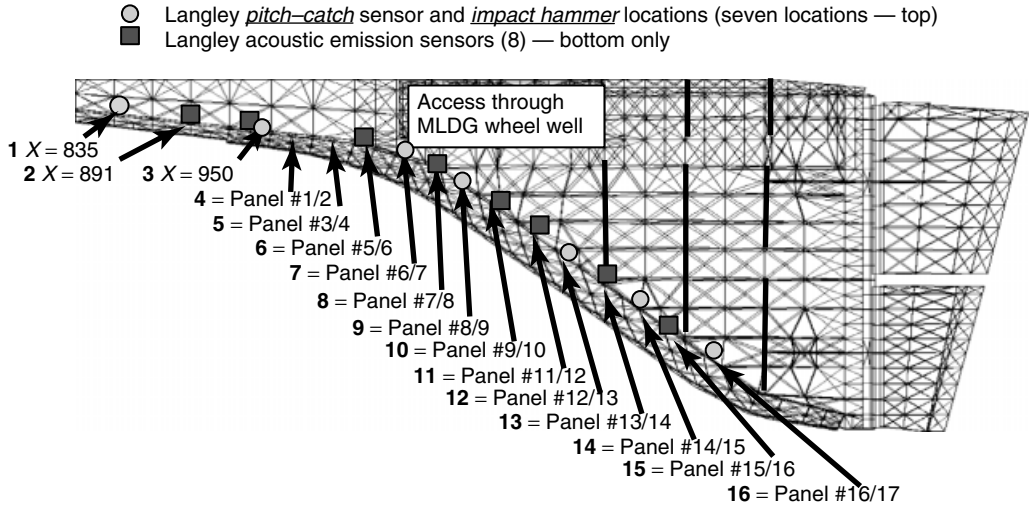


Figure 8. Layout of transducer locations inside the shuttle Endeavor's wing.

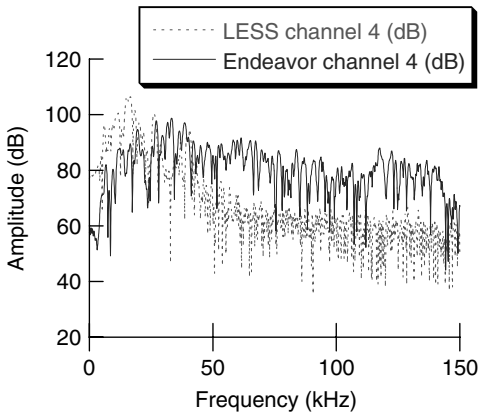


Figure 9. Frequency response for 68 kg (150-lb) hammer impact on shuttle Endeavor wing spar and leading-edge structural system (LESS) test article.

higher frequencies may enable improved signal to noise for detection as the background noise is expected to decrease with increasing frequencies. However, no database exists for measurements of the background noise for ultrasonic frequencies on the shuttle.

4 SHUTTLE IMPACT-DETECTION IMPLEMENTATION

Although both accelerometers and ultrasonic AE sensors were demonstrated to be successful at detecting impacts on space shuttle structures, accelerometers were chosen for the implementation of the WLEIDS because of the availability of existing flight-qualified sensors and instrumentation. Arrays of 66 accelerometers have been deployed on each wing leading-edge spar of all space shuttles. The data from these sensors are recorded by arrays of 22 battery-powered data-acquisition/wireless transmission units mounted in each wing cavity. Each data-acquisition unit records the output from three accelerometers as well as one temperature sensor. The system records data from all sensors continuously during launch and ascent to orbit, digitizing the signals at a sampling frequency of 20 kHz. Then, to conserve battery life, the system is switched into on-orbit monitoring mode during which smaller sets of sensors are monitored to record any triggering MMOD impacts. Data is transmitted wirelessly to a laptop computer in the crew compartment and then downlinked to mission control

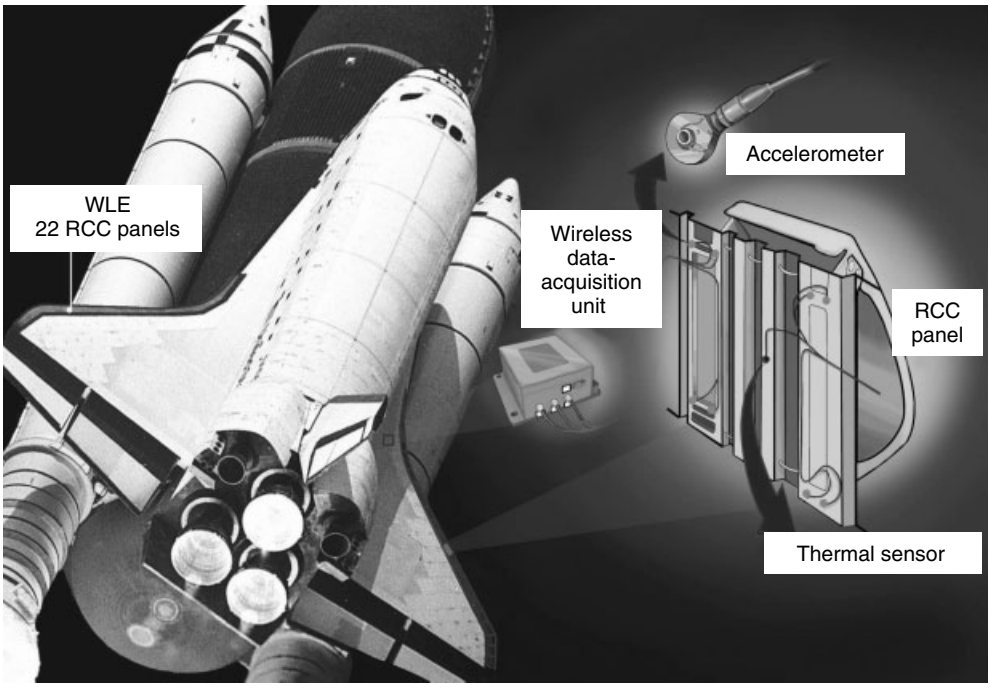


Figure 10. Key components of the shuttle wing leading-edge impact-detection system.

at the Johnson Space Center for analysis. Figure 10 shows the key components of the WLEIDS system.

Because of the limited data-acquisition unit battery life and telemetry bandwidth, the complete time-history data from all sensors cannot be transmitted to mission control for analysis during the flight. Preprocessing routines in the data-acquisition units calculate Grms (rms value of the g forces recorded by the sensors) values for the sensor units and then create summary files of the largest Grms peaks. These summary files are then downlinked for preliminary analysis. Grms peaks that occur globally across the wing are discounted as impacts and most often correlated with mission-specific events such as main engine ignition, solid rocket booster (SRB) ignition, maximum dynamic pressure, and tank and SRB separation. Local peaks are analyzed as potential impacts by downlinking and evaluating short intervals of the time-history response for multiple sensors near a suspected impact location. Additionally, suspected impact events are correlated with other data sources such as video and radar recordings of the vehicle during launch and ascent.

For all flights since Columbia, the WLEIDS has performed exceptionally well. All sensor data-acquisition units have successfully triggered at launch and data has been recorded from all sensors. The summary files have been successfully downlinked and a number of small probable impact events identified on each flight. None of these probable impact events have been of amplitude consistent with critical damage to the RCC leading edge and in-flight inspection at the suspected impact locations using the orbital boom sensing system have not revealed damage. The complete time-history data from all sensors is retrieved from the vehicle after each flight and is analyzed. The focus of this postflight analysis is to determine if any potential impacts were missed during the analysis of the summary files during the mission, and to develop and evaluate improved algorithms for impact signal identification during future flights.

5 CONCLUSIONS

AE sensors and accelerometers were used to monitor foam impact tests on shuttle test articles as part of the Columbia accident investigation. These tests demonstrated that acoustic sensing could be used to detect and locate impact events on the shuttle wing leading edge. Follow-on testing has demonstrated this capability for a wide range of impact conditions on both the leading edge as well as the lower wing surface. These tests have included much smaller impact energies at and below the threshold of damage, different impact materials, and hypervelocity impact conditions designed to simulate MMOD damage. Additional testing has analyzed the effects of different wing spar constructions on the propagation of impact generated acoustic waves along the spar.

As a result of this testing, a shuttle WLEIDS was developed and successfully deployed. Accelerometers are used in this system due to the availability of previously flight-qualified sensors and wireless data-acquisition units that could be easily integrated into the shuttle wing spar. The system has performed as designed detecting only a small number of probable impact events that have been of a magnitude small enough to have not caused damage. Postflight analysis of the complete data from each mission is performed to develop improved impact-detection methodologies for future shuttle flights.

RELATED ARTICLES

Acoustic Emission

Applications of Acoustic Emission for SHM: A Review

Lamb Wave-based SHM for Laminated Composite Structures

Wireless Sensor Network Platforms

Commercial Fixed-wing Aircraft

REFERENCES

[1] Prosser WH, Gorman MR, Humes DH. Acoustic emission signals in thin plates produced by impact

damage. *Journal of Acoustic Emission* 1999 **17**(1–2): 29–36.

[2] Nelson JM, Lempriere BM. *Space Station Integrated Wall Design and Penetration Damage Control*, Final Report for NASA Contract NAS8-36426, 1987.

[3] GAO. *Space Program: Space Debris is a Potential Threat to Space Station and Shuttle*, GAO Report # GAO/IMTEC-90-18, 1990.

[4] Columbia Accident Investigation Board Report. Government Printing Office, 2003.

Chapter 112

Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control

Stephen C. Galea[1], Tony Trueman[1], Len Davidson[1], Peter Trathen[1], Bruce Hinton[1], Alan Wilson[1], Tim Muster[2], Ivan Cole[2], Penny Corrigan[2] and Don Price[3]

[1] *Air Vehicles Division, Defence Science and Technology Organisation (DSTO), Melbourne, VIC, Australia*
[2] *Commonwealth Scientific and Industrial Research Organisation (CSIRO) Materials Science and Engineering, Clayton, VIC, Australia*
[3] *Commonwealth Scientific and Industrial Research Organisation (CSIRO) Materials Science and Engineering, Lindfield, NSW, Australia*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 Commonwealth of Australia. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

The development of corrosion in aircraft structural components over their operating life is largely due to the gradual deterioration and breakdown of protective paint coatings, anodized coatings, conversion coatings, metallic coatings, and sealants (*see* **Environmental Monitoring of Aircraft**; **Loads and Temperature Effects on a Bridge**). Unfortunately, most of the high-strength aluminum alloys and steels used for aircraft structural components are susceptible to corrosion (*see* **Design Principles for Aerospace Structures**). The frequency and time taken to detect, repair, replace, and repaint corroded components increases with aircraft age, and the corresponding increase in the overall costs are important factors in determining the eventual life of the aircraft.

In military aircraft, corrosion is responsible for a large proportion of maintenance time and expenditure. Studies by the United States Air Force (USAF) have revealed that the cost of corrosion was US$800 million in 1998 [1]. Corrosion prevention measures

accounted for 29% of the expenditure, while repair-related costs made up 71%. The United States Navy (USN) spends more than US$1.5 billion annually on aircraft systems on repair and maintenance due to corrosion [2]. In fact, a significant proportion of the corrosion effort arises from the cost of looking for corrosion. For example, the US Naval Air Systems Command noted that about 50% of the corrosion items processed were scheduled corrosion inspections, whereas corrosion repair accounted for about 20% of the items processed [3]. For support equipment, the figures showed that about 83% of man-hours were spent on inspecting, and only 2% were spent on repairing. Clearly, there are significant incentives to reduce the number of inspections. While these maintenance costs may be readily identified, it is the costs associated with not having the aircraft capability available for operational purposes that are harder to quantify, and which may be orders of magnitude greater. If prolonged corrosion-associated maintenance periods cause an aircraft to be unavailable for service, then the cost of that corrosion should also include the cost of hiring a replacement aircraft for that period.

Inspections for corrosion damage are usually time consuming and complex, require aircraft disassembly, and, quite often, reveal no corrosion damage. These inspections may also introduce additional costs resulting from incidental damage to the structure or increased likelihood of corrosion due to the replacement of factory seals by possibly inferior, less-durable seals. In an effort to reduce maintenance costs and increase aircraft availability, aircraft maintainers are considering condition-based maintenance centered around a structural health monitoring (SHM) system (with a prognostic capability) for aircraft structures, rather than inspections based on elapsed flying hours.

The philosophy of SHM of structures prone to corrosion damage is based on the concept of continually monitoring a structure to (i) identify when coatings have broken down, (ii) identify when corrosion has commenced, and (iii) characterize the nature of the environment in a particular area of the structure. The system is based on the concept of locating corrosion sensors and environment monitors in areas of aircraft structure that are generally difficult to inspect and/or are only inspected at major servicings, i.e., at three- to five-year periods for military aircraft. The basic tools for effective SHM are (i) the application of *in situ* corrosion sensors and environment monitors in a robust network to provide accurate, reliable local microclimate (corrosivity) and corrosion information, (ii) corrosion-prediction models (CPMs), and (iii) maintenance decision making approaches. An extension of an SHM system is a prognostic structural health monitoring system that should be capable not only of indicating the state of degradation (i.e., corrosion) and its effect on the structural integrity of the aircraft but also of predicting: (i) when protective coatings will fail; (ii) when corrosion is likely to occur in remote areas that are not monitored and difficult to inspect; and (iii) the rate of corrosion growth. The system should also be able to indicate how these predictions will change depending on the location of the aircraft's base and its mission type. With this information, components could be replaced before failure due to corrosion or fracture resulting from section loss, thus maintaining structural safety, increasing aircraft availability, and reducing recurring maintenance costs.

Subsequent sections of this article discuss the key elements required for a corrosion SHM system (*see* **Principles of Structural Degradation Monitoring**; **Design Principles for Aerospace Structures** and **Maintenance Principles for Civil Structures**). The main elements discussed here include

1. sensors to monitor the environment, state of the protective coating, and to detect and monitor corrosion in order to measure local microclimate environments and corrosion;
2. robust system architectures for a sensor network to enable the provision of accurate, reliable environmental and corrosion data;
3. novel parametric and physics-based CPMs, which use the sensor data to accurately predict the current state and future trends of corrosion in all locations of the aircraft, including areas with and without sensors; and
4. the appropriate maintenance decision for the operator or maintainer, such as whether the corroded components may be repaired, replaced, or treated and left in place until a more convenient inspection period (*see* **Design Principles for Aerospace Structures** and **Environmental Monitoring of Aircraft**).

In discussing the above elements, this article focuses on two slightly different corrosion diagnostic

and prognostic approaches. One approach led by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), with the Boeing Company and the Australian Defence Science and Technology Organisation (DSTO), involves a novel parametric-based CPM and uses an agent-based approach to the sensor network system architecture [4]. The other approach, led by DSTO, involves the use of a physics-based CPM with a more conventional sensor network system architecture. The article also outlines current laboratory demonstrators of both approaches. A description of some current in-flight demonstrators within the Australian Defence Force (ADF) have also been included. These in-flight demonstrators consist of incorporating some corrosion and environmental monitors within military aircraft to produce interim deliverables, which will assist in developing smarter maintenance corrosion management practices for military aircraft and to facilitate acceptance of the SHM approach to corrosion management by the operator and maintainer (*see* **Principles of Structural Degradation Monitoring**).

2 BACKGROUND

2.1 Knowledge of corrosion problem areas

Information on corrosion-prone areas is readily available where a fleet of aircraft has been in operation for many years. For a new aircraft, it is difficult to know where corrosion may occur. However a knowledge of aircraft construction, experience in how environments develop within particular areas of the structure, and information about local stresses, the nature of structural joints, types of alloys and fasteners at these locations, and the type of protective coating system, all assist in making judgments on where corrosion may develop. Often, the onset of corrosion may be associated with human factors—either workmanship errors or issues that arise during service. While it is not possible to determine if workmanship errors will occur, it is possible to identify areas where workmanship errors could, if they occurred, raise the corrosion risk. Thus it is possible to assess the propensity of corrosion in a component by considering factors such as

- proximity to heat sources or heat sinks;
- proximity to sources of external air;
- orientation/location of components;
- features of components that may provide crevices (fasteners, lap joints);
- nature of material and coating;
- microclimate and factors that may influence microclimate (limited air flow, etc.).

2.2 Operating environments

It is generally accepted that most corrosion in aircraft structures occurs while the aircraft is on the ground or flying at low altitude. At high altitudes, temperatures are usually too low throughout most of the structure for the corrosion reaction to proceed. In the vicinity of engines or other onboard heat-emitting sources, those low temperatures will not be reached, and corrosion may take place. A detailed knowledge of what the environment is in the vicinity of particular components, e.g., how often the area becomes wet, what contaminants are present, and how these factors vary with mission type and base location is probably too difficult to obtain. However, the installation of sensors and monitors in the general area of interest can provide this information.

2.3 A conceptual model of SHM sensor and monitor location

SHM is most likely to be useful in areas of structure that are not accessible or easily inspected. Corrosion on exterior surfaces at rivets and joint lines, and within open areas such as landing gear bays and weapons bays, may be relatively easy to detect, and are not the main focus of an SHM system.

The schematic drawing in Figure 1 represents the model of an internal space in an aircraft structure.

Some joints and fastener holes are associated with the external surfaces of the space, e.g., wing plank to spar attachments contained within the space, and other joints within the space not associated with external surfaces, e.g., wing spar webs. Aircraft generally experience an external environment, which can be represented broadly as follows. On the ground, it may consist of rain, hail, snow, or high humidity and regular washing. In addition, if the temperature falls below the dew point and the air contains sufficient water vapor, condensation occurs. The external

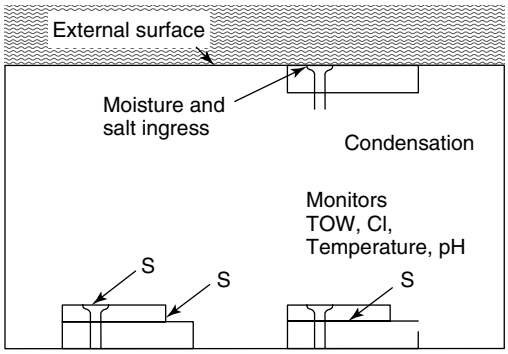


Figure 1. Schematic of the air space within an aircraft structure, where S indicates possible locations of sensors. The sensors indicated are to measure time of wetness (TOW), temperature (Temp), chloride concentration (Cl), and pH..

surfaces of the aircraft collect airborne contaminants such as sea spray during flights at low altitude over the ocean or sea spray carried inland by onshore winds from the ocean. Studies have shown that salt contamination on these surfaces may vary from 40 to 700 mg m^{-2} [5]. Other airborne contaminants such as oxides of nitrogen, sulfates, etc., may also be present. If the relative humidity (RH) is above the deliquescent RH for a particular salt contaminant, the salt wets out, providing a moisture film on the surface.

Moist air may be drawn into internal airspaces, but moisture also enters via internal joints adjacent to the skin through working fasteners and joints where cracks in sealants and paint coatings have developed. Because of the closed nature of many internal spaces, while the aircraft is on the ground condensation may occur because of the daily heating and cooling cycles and the "hot box" effect. When an aircraft descends from high altitudes, most internal surfaces are cold (e.g., $-40\,^{\circ}$C is not uncommon) and when ambient air is drawn into the space, condensation occurs. It is therefore clear that, once the coating has failed, all the necessary ingredients for corrosion to occur can penetrate, and corrosion can develop within the aircraft structure.

It is not possible or practical to locate corrosion and environment sensors and monitors everywhere within an internal air space. Under the SHM methodology, it is proposed to locate these at key points within areas of interest. The discussion that follows describes various methodologies to account for the other non-sensed regions.

The types of environment sensors vary according to the inputs required for the CPM. However, time of wetness (TOW), chloride level, temperature, RH, and, possibly, pH are considered as the main parameters to be monitored. Pressure is not essential; however, because it is related to altitude, it does provide a method for checking other monitors such as TOW and temperature. In general, the role of strategically placed corrosion sensors is to provide onboard validation of the output of the CPM. When the corrosion sensors are located beneath a coating within joints or fasteners holes, they may also provide an indication of when coatings have failed and corrosion has occurred.

2.4 Paint coating/sealant failure

Corrosion does not generally occur until a protective coating fails. This failure may be induced by mechanical damage in the course of maintenance, but it is more usual for a coating on aircraft components to mechanically fail because of stress and environmental factors. Stresses arise from operational loads such as those around working joints and fasteners, thermal cycles (e.g., from -30 to $+80\,^{\circ}$C would not be uncommon for military combat aircraft, with less extreme upper limits for commercial and military transport aircraft), the freezing of moisture absorbed into a paint coating, and blistering resulting from moisture accumulating at the paint–metal interface. The environmental factors most relevant to an SHM system, which impact on paint failure and corrosion, are those that develop inside the aircraft. They are condensation resulting from the thermal cycle, the "hot box" effect associated with enclosed spaces, or prolonged exposure to high temperatures if near an engine or just beneath the outside skin (which may be heated by frictional stress of air flow). Following paint failure, attack on the exposed metal generally follows when the inhibiting pigment included in the paint coating is depleted by leaching.

Currently, models that predict coating failure under the combined action of all these variables do not exist. There have been studies to identify the rate of depletion of inhibitors from primers once the coating has been breached; however, these results are empirical and have usually focused on a particular test environment such as the neutral salt spray (NSS)

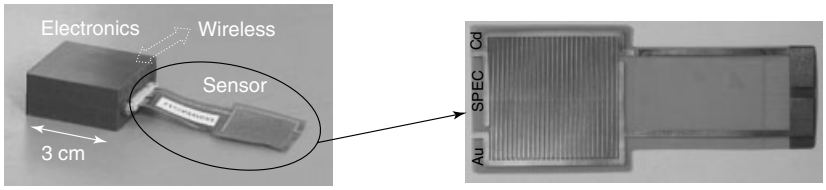


Figure 2. US Navy SPEC galvanic sensor. (Dimension of sensor is typically 28 mm square).

test [6, 7]. It is generally accepted that the NSS test is quite extreme and is neither typical nor representative of the environment around an aircraft component, which is more likely to be subjected to condensing droplets of moisture and periods of drying.

While paint coating failure models are necessary, it is also important to have some sensors strategically located onboard for detecting coating failure, and to allow for onboard validation of these models. One approach for an SHM system would be to locate sensors under coatings in areas of interest, for example, in joints or fastener holes, under the lavatory and galley areas, at doors and door edges, etc., which indicate when corrosion on the sensor elements occurs, thus indicating not only when deterioration of the coating has occurred but also when inhibitors in the coating have been depleted.

3 ENVIRONMENTAL AND CORROSION SENSORS

In this article, various types of sensors and monitors that may be considered for use in an airspace, similar to those described in Section 2.3, are discussed (*also see* **Nanoengineering of Sensory Materials**; **Miniaturized Sensors Employing Micro- and Nanotechnologies** and **Environmental Monitoring of Aircraft**).

3.1 Corrosion and galvanic sensors

The SPEC Inc. galvanic (TOW) sensor system developed by the USN is based on the galvanic interaction of interdigitated electrodes of two dissimilar metals, gold and cadmium, deposited on a thin polymer substrate. The electrodes are connected to a miniature data logger about the size of a matchbox. The unit and the sensing element are shown in Figure 2. The current flow between the two dissimilar metals when covered by moisture is measured as a function of time—thus the sensor indicates the time at which the component is wet. The logger is self-contained and battery powered, and the data may be downloaded via a wireless link. The sensor system has been tested on USN P-3C and Seahawk aircraft.

An earlier version of this type of sensor was developed by the DSTO in 1998 [8]. It is much larger in size than the USN version, and relies on a galvanic couple of copper and tin to provide a current output and to measure the TOW. The DSTO sensor has been flown for many years in the tail section of a Royal Australian Air Force (RAAF) P-3C Orion and the forward equipment bay of an RAAF F-111.

Analatom supply the linear polarization resistance (LPR) corrosion sensor as shown in Figure 3 (*see* **Miniaturized Sensors Employing Micro- and Nanotechnologies**). This sensing system was developed by the DSTO [9] and licensed to Analatom in the USA [10]. The system uses the electrochemical technique of LPR, and involves measuring a current response following an application of a small potential to microelectrodes. The sensing elements are made from the alloy of interest using either micromachining or lithography techniques. The LPR sensor not only provides a measure of the corrosion rate for the alloy from which the sensor element is made but is also a TOW sensor. Analatom AA7075 aluminum alloy LPR sensors have been flown in the forward cargo hold of a Delta Airlines Boeing 737 [11].

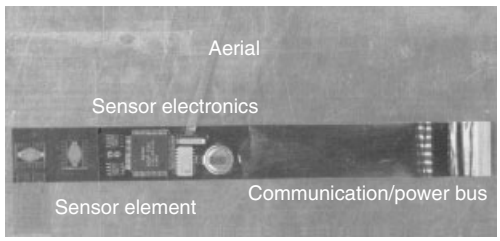


Figure 3. Analatom sensor system. (The LPR sensor is approximately 10 mm square).

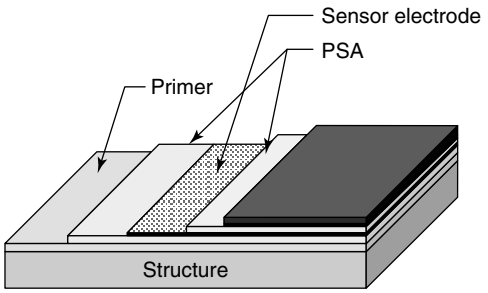


Figure 4. DACCO smart appliqué [12].

DACCO SCI Inc. has developed a smart-appliqué corrosion sensor. Smart appliqués are peel-and-stick fluoropolymer films containing an embedded sensor electrode as depicted in Figure 4 [12]. The appliqué is applied over primer coatings in lieu of a topcoat. The sensor is interrogated using electrochemical impedance spectroscopy (EIS) with the embedded sensor as one electrode and the aircraft structure as another. Currently, it requires the use of external potentiostat/impedance instrumentation when the aircraft returns to base. This type of sensor would enable detection of damage of the coating system, including the appliqué, as well as detection of the start of corrosion of the metallic substrate. At this time, no miniaturized electronics package is provided to interrogate the electrodes *in situ*. Further work is also being undertaken to apply the concept of this sensor to typical aircraft coating schemes.

The BAE Systems Advanced Technology Centre (BAES ATC) galvanic sensors use a galvanic couple of gold and aluminum or gold and tungsten and is shown in Figure 5(a). The interdigitated tracks of the two different metals (the thickness of the deposited tracks is approximately 1–2 μm) are vapor deposited onto a silicon wafer. The potential generated between the two metals when moisture is present is measured using a high impedance voltmeter. The sensor acts as a TOW sensor and the measurement of the potential as opposed to current provides a longer lifetime for the sensor electrodes.

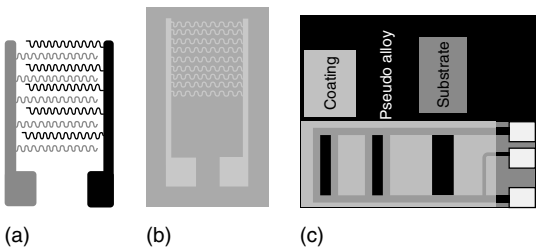


Figure 5. BAE Systems (a) galvanic sensor, light = Au, black = Al or W, (b) resistance sensor, and (c) sentinel sensor.

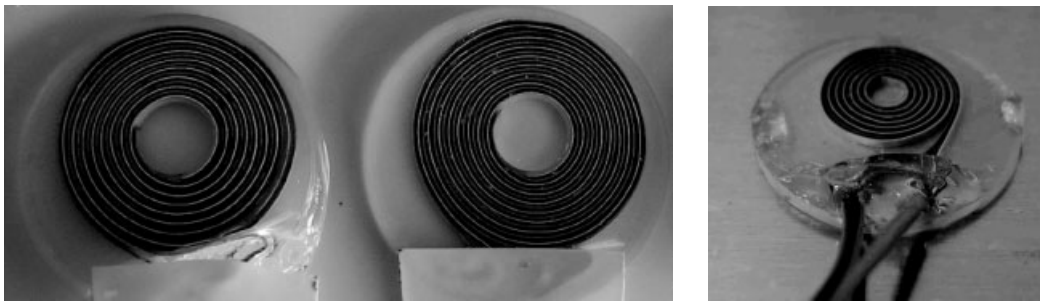
The BAES ATC resistive sensors consist of nine tracks of an aluminum alloy, e.g., Al–Si–Cu, vapor deposited onto a silicon wafer (Figure 5b). Pitting or general corrosion of the tracks results in sectional losses of the tracks causing an increase in the track resistance. The resistance changes are logged, and they provide an indication of corrosion development and possibly corrosion rate.

The BAES ATC sentinel sensor is approximately 25 mm × 25 mm and is made from material similar to the BAES ATC resistance sensor (Figure 5c). This sensor has three electrodes, and each electrode is masked with different size slots. The sensor is primed and topcoated, it is then bonded to the airframe, and the slots are then unmasked. The different-sized slots simulate different sizes of damage in the coating. It is claimed that the sensor resistance changes as the ability of the corrosion inhibitor in the various slots in the primer is exhausted [13].

The galvanic corrosion sensors, developed by CSIRO (Figure 6), consist of coils of dissimilar metals made using 0.1-mm-thick foils of copper and aluminum, both with a purity of 99.9% [14]. The dissimilar metals generate a current flow when exposed to corrosive conditions because of the differing solution potentials that exist at each electrode. The nature of the dissimilar metals may be changed depending on the alloy substrate of interest. For example, laboratory studies using aluminum coupons and aluminum–copper coil sensors allow the correlation of pitting damage in the aluminum coupon with the sensor output. Like the LPR sensor, outputs from this sensor not only provide a measure of the corrosion rate for the alloy from which the sensor element is made but also provide an indication of the TOW.

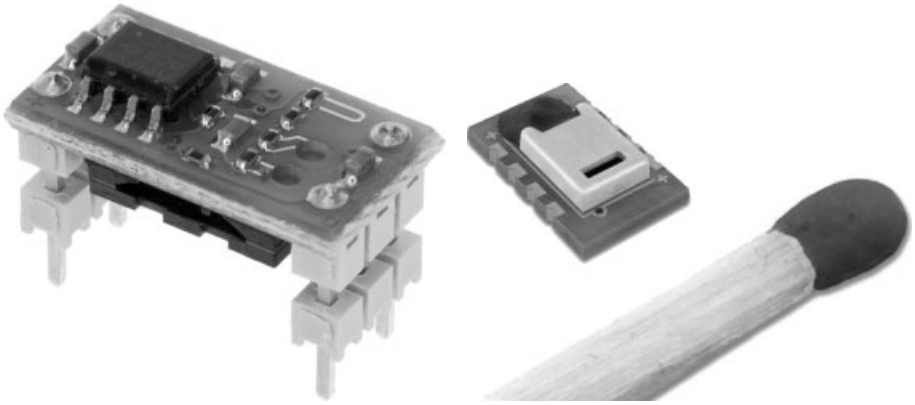
3.2 Environment sensors and monitors

Several commercial sensor systems are available to monitor RH, temperature, and pressure. Some typical solid-state temperature and humidity sensors are produced by Humirel Ltd and Sensirion (Figure 7a



(a) (b) (c)

Figure 6. Galvanic coil sensors. (a) Zn/Cu, (b) Al/Cu Type A, and (c) Al/Cu Type B.



(a) (b)

Figure 7. Humidity and temperature sensors: (a) Humirel HTF3130 and (b) Sensirion SHT1x.

and b, respectively). Both units are very small modules (Humirel and Sensirion units are about 18 mm × 9 mm and 8 mm × 5 mm, respectively), designed to have high resistance to shock and vibration, with a fast response time. Both devices are based on a capacitive polymer sensing element to measure RH. It is claimed that they have a high resistance to a range of chemicals and that they are unaffected by water. These are desirable properties for sensors to be fitted to aircraft.

CSIRO has also produced thin-film sensor arrays which monitor both the environment and corrosivity [15]. As illustrated in Figure 8, the arrays are produced on printed-circuit boards using traditional techniques for etching copper tracks and physical vapor deposition to produce thin films of aluminum. The design incorporates an aluminum–copper galvanic sensor, where the surface area of the copper electrode is minimized to increase the lifetime of the thin film of aluminum (which is oxidized upon being placed in corrosive environments). A thin-film aluminum electrical resistance sensor enables

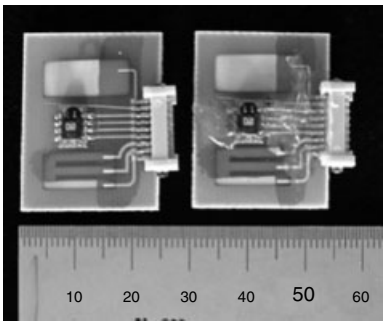


Figure 8. Thin-film sensor arrays support electrical resistance and galvanic corrosion sensor elements and a surface-mounted Sensirion RH/temperature sensor.

a second measurement of corrosivity. The printed-circuit-board design enables the incorporation of a surface-mounted RH/temperature sensor (Sensirion, SHT1x). In addition, a 50-nm-thick evaporated layer of alumina is deposited over the general area of the sensor to encourage the metal elements to wet when the sensor surface is in contact with moisture. This allows the sensor surface to replicate the improved wetting properties of aging metals.

4 SYSTEM ARCHITECTURE

The initial objective of the SHM system was to monitor corrosion damage in a number of corrosion-prone areas within an aircraft, such as in joints or fastener holes, under the lavatory and galley areas, at doors and door edges (as discussed in Section 2.3). The development of the system involves the nontrivial issue of *in situ* sensor conditioning, signal processing, data storage, and system networking. The systems need to achieve sufficient sensitivity, reliability, and robustness to be able to attain data measurements with reasonable confidence (*see* **Sensor Network Paradigms**). The system employs the following principles:

1. Clusters of sensors in small local regions measure local microclimatic parameters, including temperature, humidity, surface wetness, pH, and conductivity of surface moisture. The sensor cluster would also include sensors for measuring corrosion rate such as the Al–Cu galvanic corrosion sensor or the aluminum alloy LPR sensor.
2. Corrosion damage information is obtained from the sensor readings either through CPM or through validated relationships between corrosion damage on aluminum alloy test specimens and sensor readings, established in environmental chamber tests (Section 5).
3. The system has the capability to learn about damage progression from "similar" regions of the structure in which the damage is further advanced and, ultimately, from other aircraft within a fleet.

4.1 Intelligent agents (CSIRO approach)

The approach used here to make the system robust, to allow it to be more readily extendable (scalable), and to minimize communications traffic is to adopt a system architecture based on the use of hardware "agents" [4]. Each agent consists of a cluster of sensors that make measurements in a local region, along with data acquisition, processing, and communications hardware: it is an autonomous sensing unit. The electronic hardware of the agent should be located close to the sensors to minimize effects of cable impedance and noise pick up. The aim for each agent is not only to acquire data from its cluster of sensors but also to carry out any processing of the data that depends only on the local data. Thus, each local agent can analyze the data time series from its sensors, and perform a local diagnosis—thus enabling local checking of the consistency of sensor data, to provide a level of confidence that the sensor measurements are accurate.

The agents communicate with a central processor, which contains global knowledge of the structure, and of each individual agent's location within the structure. A schematic diagram of this system architecture is shown in Figure 9.

The central processor, also known as the *information technology* (IT) platform, provides any required location-related information to an agent, possibly including learnt information from similar locations, which may relate to the agent's local data processing. The agent communicates reduced information to the central processor, rather than raw sensor data.

One of the features of this system is that it is able to predict corrosion damage at some points that do not have sensors, as long as their materials, geometry, and microclimate are sufficiently similar to those at sensed points that they can be reliably deduced from the characteristics of the sensed points. This is done explicitly by the use of "virtual agents" located within the IT platform, which generate diagnostic information from the information obtained from the most appropriate real agents. Clearly, the virtual agents cannot predict the occurrence of specific incidents that may contribute to corrosion (e.g., a water spill), but they may be useful in locations in which unpredictable incidents are unlikely, or in which the probability of such random incidents is similar to that in a sensed location.

Information regarding the airframe structure, and situational information for the individual agents, is maintained within the IT platform in software structures known as *intelligent objects* [16]. Typically, an intelligent object defines a structural component

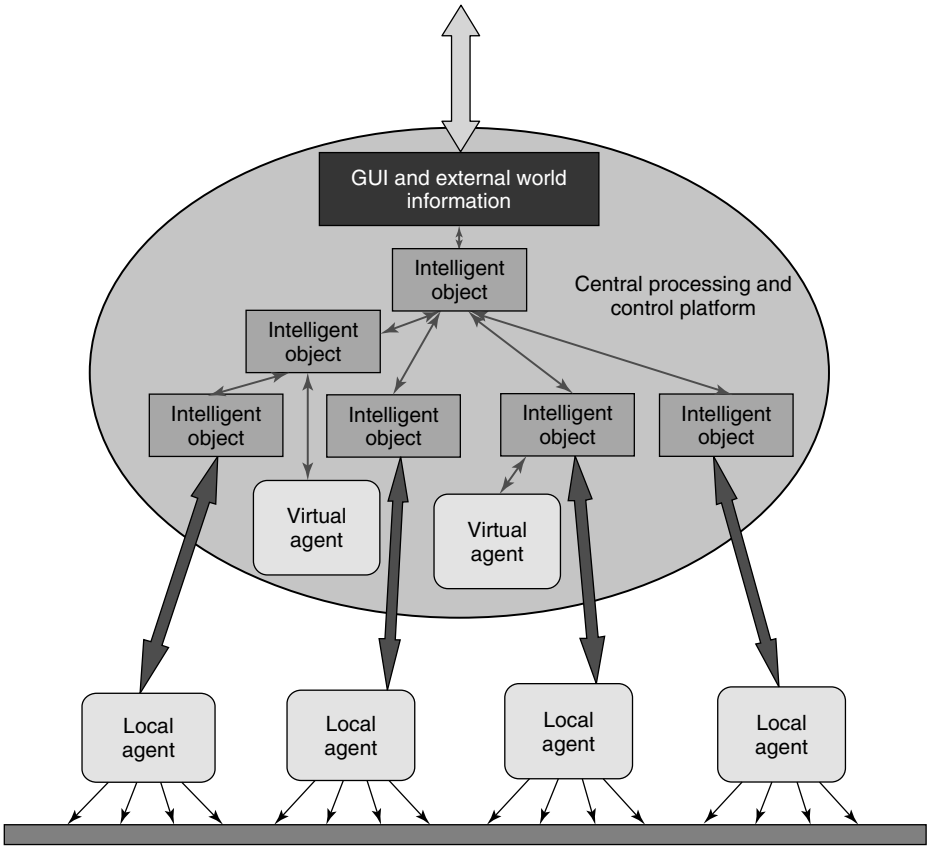


Figure 9. Schematic diagram of one possible system architecture, including (1) local agents that consist of a group of sensors and processing hardware, communicating with a central processor and (2) the IT platform which consists of virtual agents, the intelligent objects and the graphical user interface (GUI). The agents communicate diagnostic information to the intelligent objects.

or a substructure: it contains information about the geometry and materials of the object, and defines the relationships and connections with other objects to form the complete structure. Intelligent objects may be aggregated to form objects representing larger components of the structure. One or more local agents and/or virtual agents may relate to a single intelligent object. Figure 9 contains schematic details of this IT platform architecture, which also contains the interface to the outside world.

A laboratory demonstrator containing 11 sensing agents based on an under-floor region of a large commercial aircraft was implemented in an environmental chamber and has been running for more than 18 months. Single-agent systems, operating purely as data loggers, are operating on two operational aircraft (one commercial and one military). These are described in Section 7.1.

4.2 Sensor nodes (DSTO approach)

A combined sensor interface and networking technology is under development with the following key features [17]:

- low power for long endurance when battery powered;
- packaging of the sensor signal conditioning, network interface, and sensors in one small lightweight package;
- remote software upgrade capability;
- ability to handle large networks of sensors;

- software and hardware adaptability to accommodate new sensors;
- a software interface that is flexible enough to accommodate single or multiple sensors;
- reduced network bus traffic by using local data storage and processing, and local timing for data logging.

Critical to achieving these features is the use of low-power, microcontroller technology that incorporates a wake-on-demand protocol. Using this approach, a sensor node may be in one of three states, namely, disabled, enabled (with only the sensor node network communications active), and active (where the sensor node is fully operational).

The sensor node switches from the disabled to the enabled or active state on receipt of a special serial communications byte or an internally generated interrupt such as a timing signal. The sensor node moves to the disabled state either when it is deselected by the network controller or under internal program control.

Both the hardware and software are modular, consisting of electronics and applications for specific sensors interfaced to an invariant core communications and processing module (Figure 10). The software division is taken to the extent that the memory stack for sensor applications is separate from the CPU stack. A call-back methodology is used for sensor applications to access routines in the core module. The software modules for specific sensors are associated with user-defined command names, known as a *tag*, and both of these are externally downloaded into ROM under control of the core processes. Tag lists that consist of a list of tag names, like a list of function calls, can also be downloaded. Addressing of individual sensors is achieved by assigning them different tag names. The downloadable tag and tag list approach facilitate debugging and development of sensor-specific code. Core processes control all input and output, memory management, the sensor state etc., and also include a number of utility functions to manage the tags. The core routines only occupy 18% of the available ROM, leaving significant space for sensor interfacing code and any desired local data processing. This hardware and software modularization allows for the rapid incorporation of new sensors.

A number of nodes have been developed, including the following:

- high impedance, two-electrode electrochemical sensors;
- inter-integrated circuit (I^2C) compatible temperature and humidity sensors;
- metal foil strain gauge rosettes;
- microelectromechanical system (MEMS) accelerometers (three axis);
- single frequency impedance (amplitude and phase) measurement;
- any sensor that requires a 3–5 V excitation and produces a voltage between 0 and 5 V;
- any sensor that has an open/closed circuit response.

To date, small sensor networks have been implemented in the laboratory, monitored by desktop and embedded PCs. A larger project is underway to install a sensor network consisting of roughly 30 nodes and 100 individual sensors on a medium-sized boat.

5 CORROSION-PREDICTION MODELS

Once coating failure occurs, and corrosion-inhibiting pigments have been depleted to a level where protection is no longer effective (available from laboratory studies), it is reasonable to assume that corrosion will quickly occur. Corrosion models are required that will predict the depth and distribution of corrosion damage with time, whether it is pitting or intergranular corrosion, since it is such damage that may act as fatigue crack initiators and jeopardize aircraft structural integrity.

5.1 Functional approach (CSIRO)

To enable accurate corrosion predictions, a sound understanding of the relationship between the actual damage occurring on structures and the output from corrosion sensors must be established. The corrosion sensors measure the rate of accumulation of damage to the sensors themselves, not to the underlying material of the aircraft frame. To date, the damage occurring on several materials of relevance to aircraft structures, under accelerated corrosion conditions, has been evaluated and related to the output of the corrosion sensors [4]. The materials studied include

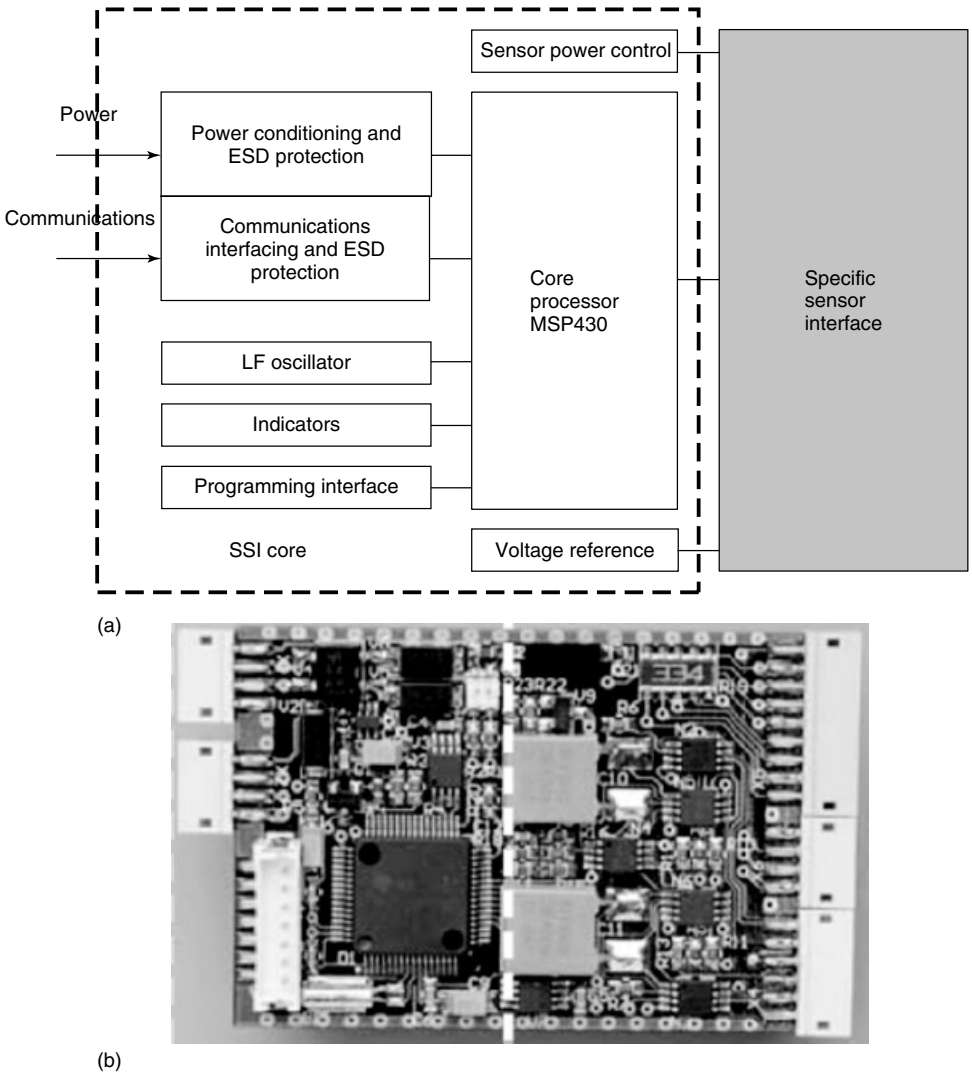


Figure 10. Sensor interface is modular with both (a) the software and (b) the hardware divided into a core communications and processing module (left-hand side of a and b) and software and hardware specific to the sensor (right-hand side of a and b).

- AA2024 (major alloying additions Cu and Mg)
- AA7075 (major alloying additions Zn, Mg, and Cu)
- painted AA7075 with chromate conversion coating and epoxy-polyamide primer.

Corrosion sensors and panels of the aluminum alloys have been exposed to accelerated corrosion environments. The test cycle parameters were defined by General Motors Standard GM9540P [18], with some modifications to make the environment more aggressive.

The degradation of the uncoated aluminum panels was evaluated by determinations of mass loss, pit depth, pit area, and pit density using standard techniques. The most reproducible parameters for aluminum alloy damage assessment were found to be mass loss, the average maximum pit depth, and the average maximum pit dimension. In this case, the average pit dimension is defined as the average pit

width × the average pit depth. Pit-size characterization is usually perceived as being a more important and realistic damage parameter for aluminum alloys, particularly where fatigue crack initiation is of concern.

The degradation of the coating on the painted panels has also been evaluated using EIS and Raman spectroscopy to monitor the depletion of the chromate inhibitor. With EIS, the impedance modulus at 0.1 Hz was adopted as a measure of the barrier properties of the primer as a function of time. This measure is based upon the work of Potvin *et al.* [19]. Using Raman spectroscopy, maps of epoxy, chromate, and titania species were established on the basis of the baseline-corrected intensities of absorption peaks, which occur at known wave numbers. Good correlation was achieved between the two methods of evaluating the degradation of the coating (Figure 11).

Using these analysis techniques, relationships were developed for the degradation of the materials with respect to the number of cycles of the environmental test. Having determined the failure modes of the materials, these needed to be related to the output from the corrosion sensor (Figure 12). During accelerated testing, the temporal data output from 10 galvanic corrosion sensors was collected to enable correlation with the state of the simultaneously exposed aluminum alloy panels.

To predict damage, relationships are sought to link the damage (corrosion) sensor outputs with the causes of corrosion, as measured by an array of microclimate sensors. Using an approach that combines aspects of both physics-based modeling and statistical analysis, these relationships are established and updated on

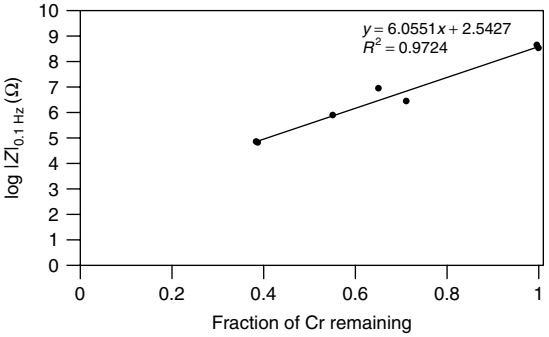


Figure 11. Correlation of the two measures of coating deterioration, EIS, and fraction of Cr remaining (determined from Raman spectroscopy).

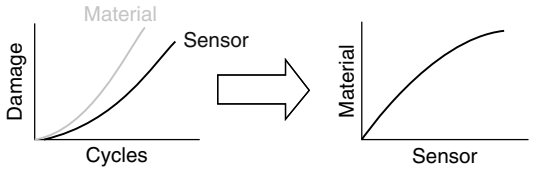


Figure 12. The material degradation and sensor outputs are initially related to the number of cycles of environmental testing. A relationship between the material damage and the sensor output is then derived.

an ongoing basis as more data is received by the agents. The system does not attempt to investigate the underlying factors controlling the microclimate or damage; rather the relationships between sensed damage and sensed surface response are determined using mathematical manipulations of the raw data.

Predictions of future corrosion behavior are made by using the relationships established on current data with a predicted future microclimate scenario.

5.2 Physics of failure approach (DSTO)

When surfaces are contaminated by NaCl and/or sea salts, and wet due to deliquescence, discrete moisture droplets are formed [20]. The number and size of these are dependent upon the contaminant density and the ambient RH. For low contaminant densities, discrete droplets form, but as the contaminant density increases, the droplets may merge to form larger droplets. These droplets constitute the local corrosion cells upon the surface and can be considered to be independent of each other. To predict the number and size of pits formed, the number, size, and composition of these droplets needs to be determined. This is achieved by using atmospheric science combined with atmospheric parameters that are practical to measure, using sensors.

The environmental model has been formulated with a requirement to have only simple atmospheric variables as inputs. In this way, the sensor requirements are reduced, and the use of simple commercial off the shelf sensors can be realized. To this end, a sensor suite proposed for the DSTO corrosion SHM system consists of a combined atmospheric temperature and RH sensor, a surface temperature sensor, and a TOW sensor. However, these sensors are required to be monitored continuously. The sensor suite may also be expanded to introduce redundancy

in the measurements and by the use of more exotic sensors, for example, paint degradation or pH sensors. Hence this approach could be made more universal, allowing its use where more complex contaminants occur. The sensors used in this approach consisted of the USN TOW, Sensirion RH/temperature, and National Semiconductor LM35 surface temperature sensor. These were selected after numerous trials of potential sensors as they were found to give accurate responses to rapid environmental changes. The results of testing of these sensors in an environmental chamber are shown in Figures 13 and 14 [21].

From Figure 13 it can be seen that the RH sensor was able to follow the programmed chamber RH closely with a maximum error of approximately 5%. The temperature of the chamber as measured by the Sensirion sensor (Figure 14) was an identical match to the programmed temperature. The temperature as measured by the surface sensor (LM35) responded more slowly to the changing temperature due the thermal mass of the surface to which it was attached (Figure 14). The difference in the two temperature measurements is important, as it will affect the condensation and evaporation of moisture from the surface.

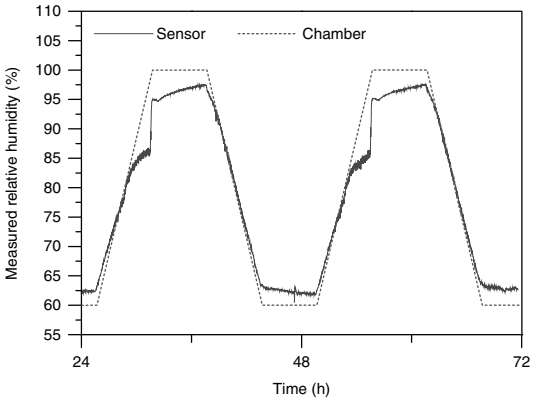


Figure 13. Sensirion RH sensor output versus programmed environmental chamber condition [21].

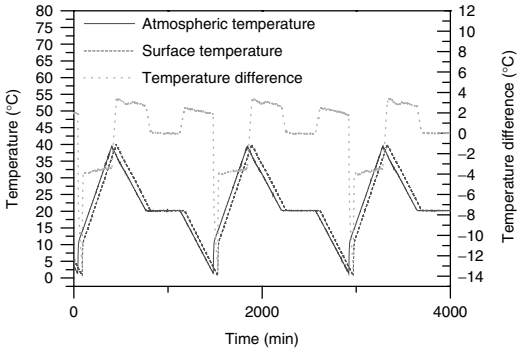


Figure 14. Atmospheric and plate temperature sensor outputs plotted with temperature differences [21].

The results from the TOW sensor showed how the measured current is scalable with wetness. Figure 15 shows the response with a high salt density on the surface and Figure 16 is the response of the TOW sensor with a low salt density.

5.2.1 Environmental model

The function of the local environment model is to take the raw sensor data and determine those parameters that are required to predict corrosion levels. The inputs and outputs from this model act as inputs into the atmospheric corrosion model that are listed in Table 1.

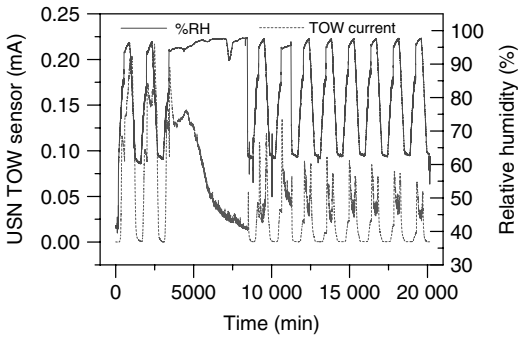


Figure 15. Response of USN TOW sensor to varying RH when doped with 5 g m^{-2} NaCl [21].

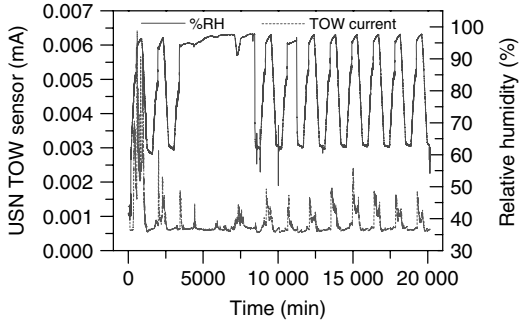


Figure 16. Response of USN TOW sensor to varying RH when doped with 0.1 g m^{-2} NaCl [21].

Table 1. Inputs for and outputs from the local environment model

Local environment modeling inputs	Local environment modeling outputs
Time of wetness	Periods of wetness (minutes)
Relative humidity	Extent of wetness (percentage wet)
Atmospheric temperature:	Surface salt coverage (milligrams per square meter)
Plate (aircraft structure) temperature	Number of moisture droplets
	Distribution of droplets areas
	Pollutant composition
	Droplet concentration (moles per cubic decimeter)

The outputs from the local environment model were calculated using atmospheric physics [22–24]. For example, the surface moisture concentration was calculated using the measurement of RH and the relationship shown in Figure 17. The interactions between multiple sensor outputs can also be used; for example, the salt crystal density can be calculated using the extent of wetness, from a scalable TOW sensor, combined with the RH, as shown in Figure 18. The period of wetness is obtained directly from TOW readings and the droplet size distribution is calculated using crystal salt densities and RH measurements in combination with probabilistic mathematics.

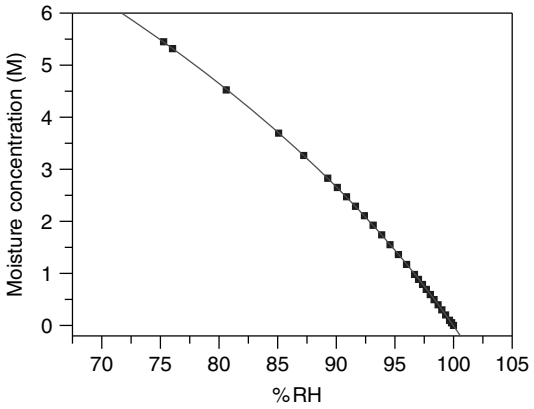


Figure 17. The dependence of surface moisture concentration on RH for NaCl [21].

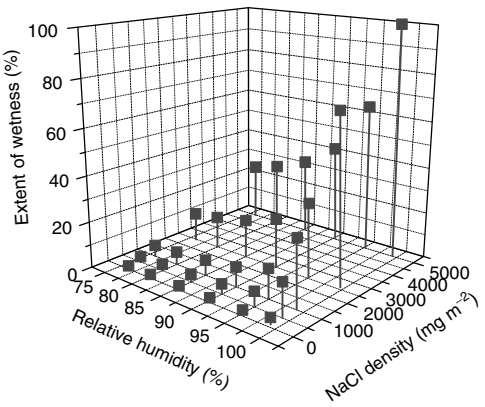


Figure 18. The dependence of salt crystal density on relative humidity and the extent of wetness for NaCl [21].

5.2.2 Corrosion modeling

As detailed above (Section 5.2), the atmospheric pitting corrosion of aluminum alloys occurs in packets when the surface is wet. To model the corrosion, the environmental sensor data is first sorted into wetness periods and the environmental parameters are calculated for each time interval. The time-resolved environmental parameters are then used to build corrosion models. The current corrosion model determines the number and size of corrosion pits; however, further models are being designed to predict intergranular and exfoliation corrosion. The pitting corrosion model has two components: a pit-creator algorithm and a pit-growth algorithm. The pit-creator algorithm employs probabilities of pitting calculated electrochemically and the pit-growth algorithm employs electrochemically measured kinetic parameters. These two algorithms work synergistically to determine the corrosion damage.

Pit-initiation algorithm

The pit-initiation algorithm employs stable pit probabilities calculated from electrochemical measurements. Pit probabilities are calculated using the method explained elsewhere [25], but will be described briefly here. The technique electrochemically determines the distribution of the largest metastable pit sizes occurring on the alloy of interest in the expected droplet solution—a typical experimental curve is displayed in Figure 19. The current spikes in the potentiostatic trace indicate metastable pitting events occurring on the aluminum alloy electrode.

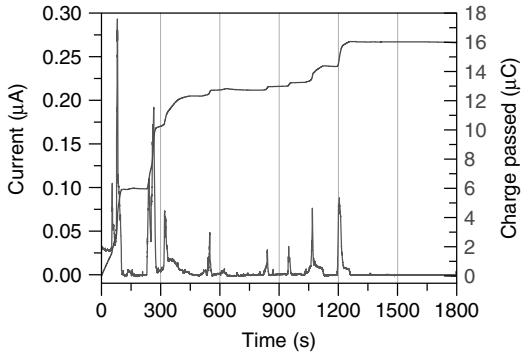


Figure 19. Potentiostatic current response of AA2024-T3 in NaCl solution [25].

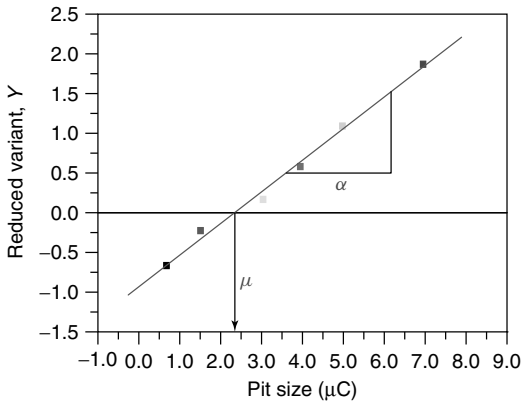


Figure 20. Maximum pit sizes for each plotted versus reduced variant, $-\ln(-\ln(1-(i/(n+1))))$ for n data points where i = 1, n [25].

The integrated area under the current spikes may be converted to pit depth using Faraday's law. The largest transients (current spikes)—metastable pits—measured are used in an extreme value statistics procedure to determine their distribution (see Figure 20, [25]). This distribution is then used to calculate the probability for a metastable pit larger than that required for the pit to become stable, "stable pit criteria" will occur, as shown in Figure 21. Multiple probability determinations are undertaken to ensure all possible droplet electrolyte compositions are covered. The matrix of probabilities is then used to generate a function that describes the distribution of probabilities that is used in the pit-initiation algorithm. In this manner, the electrolyte composition predicted in the environmental model is an input into the function to determine the initiation of a corrosion

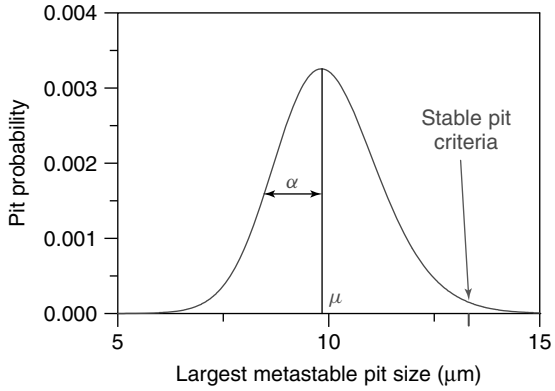


Figure 21. Distribution of maximum pit sizes predicted from extreme value statistics showing cut of size for stable pits [25].

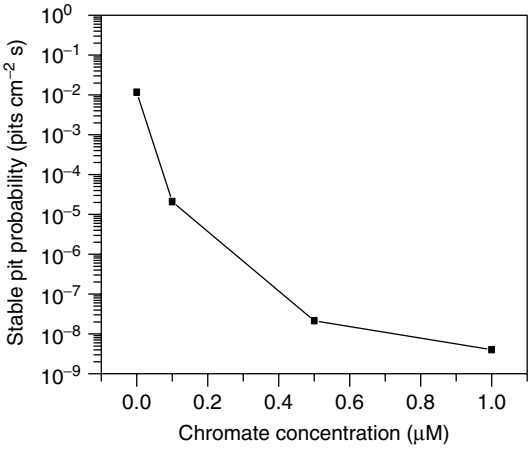


Figure 22. Stable pit probabilities calculated for AA2024-T3 in 3.5% NaCl at different chromate levels [25].

pit. As an example, the calculated probabilities for AA2024-T3 in 3.5% NaCl in the presence of varying concentrations of chromate are shown in Figure 22.

The method described above is time consuming and needs an experienced electrochemist to undertake the analysis. Hence current activities involve the development of a simpler less time consuming analysis to predict stable pit probabilities.

Pit-growth algorithm

The pit-initiation algorithm determines if and when a pit is created, by comparing the pit probability to a random number using a Monte Carlo simulation. If a pit is created, it is then grown by the pit-growth

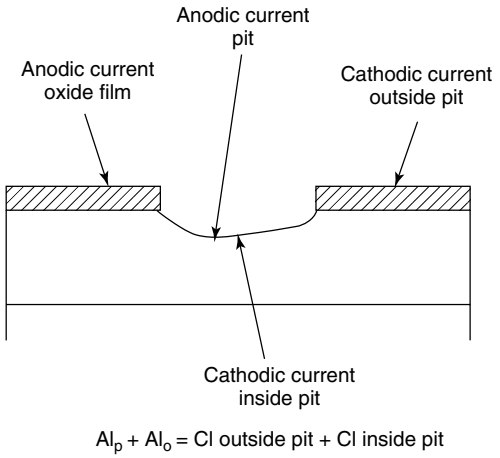


Figure 23. Schematic of a growing pit.

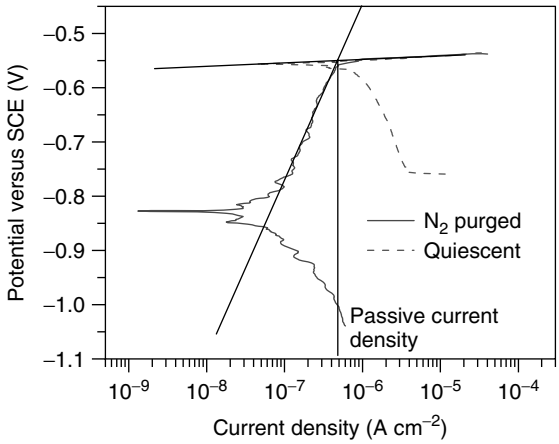


Figure 24. Potentiodynamic scans of AA2024-T3 in 1 M NaCl with and without nitrogen purging [25].

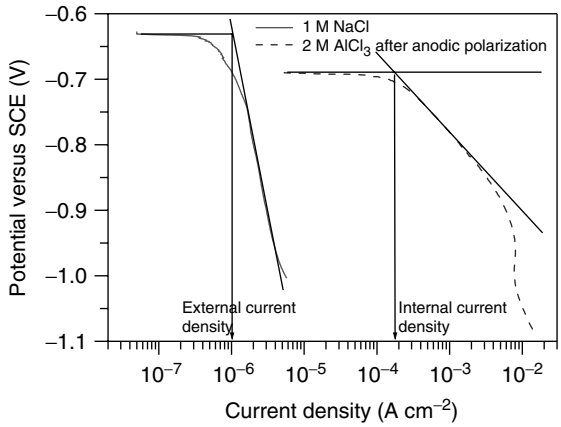


Figure 25. Cathodic polarization curves for AA2024-T3 in 1 M NaCl and 2 M $AlCl_3$ after anodic polarization [25].

algorithm. This algorithm performs an electrochemical kinetic current balance of the cathodic and anodic reactions taking place within a droplet. If we consider the schematic in Figure 23 for a growing pit on an aluminum surface, the sum of the anodic current producing the oxide film (AI_o) and the anodic current feeding the growth of the pit (AI_p) is equal to the cathodic current associated with the reduction of oxygen outside the pit (CI_o) plus the current associated with the evolution of hydrogen within the pit (CI_i). For various alloys, polarization data have been produced to give values for AI_o, CI_i, and CI_o. Of these four currents, only the anodic pitting current is unknown, and it can be calculated as a pit grows by performing a simple balance between the anodic and cathodic currents. However, in the early periods of pit growth, the pit does not grow at this rate, but is limited by the transport of metal ions out of the pit and counter ions into the pit, and not by the supporting cathodic reaction rates. This period of time, where the pit is not large enough to consume the available cathodic current, is when the pit grows at the fastest rate. It is assumed all pits grow at this rate.

The rates of the anodic and cathodic reactions are measured electrochemically using potentiodynamic polarization, as shown in Figures 24 and 25. In cases where the pitting potential was coincident with the open circuit potential (the case for many aircraft aluminum alloys in chloride solutions), a nitrogen-purged potentiodynamic scan was also conducted to enable the passive current density in an aerated solution to be estimated (Figure 24).

To determine the internal-pit cathodic current density, potentiodynamic polarization was conducted in 2.0 M $AlCl_3$ after removal of the surface oxides by the initial application of 1 V versus saturated calomel electrode (SCE) for 5 min. The 2.0 M $AlCl_3$ solution closely represents the chemistry of the solution within pits. The internal cathodic current density in the model environment was significantly higher than the external cathodic current density as shown in Figure 25. When the total area of pitting i.e., sum of the area of individual pits (assuming each pit is a hemispherical) times 1 mA cm^{-2} exceeds the cathodic

current outside the pits plus the cathodic current inside the pits, minus the current to grow the oxide film, the growth of a pit becomes cathodic current limited, and the total current available is shared by all pits. Thus the size of a pit and density of pits at any time may be calculated.

6 MAINTENANCE DECISIONS

Traditional corrosion inspections are undertaken at specified intervals (often based on flying hours) and are set by safety and economic considerations. In many cases, these inspections involve pulling apart seals (sometimes factory seals) and may cause more harm than good. In fact, a significant proportion of the corrosion effort arises from the cost of looking for corrosion. Clearly, there are significant incentives to reduce the number of inspections. It is one of the aims of corrosion SHM programs (as shown in the flow diagram of Figure 26) to ultimately decrease the frequency of traditional inspections with a system of sensors and models that inform the maintainer about the structural condition of the aircraft. Figure 26 is a simple representation of the way a corrosion SHM system fits together, and the types of decisions to be expected from it, when complete. There are, however, significant technical and regulatory hurdles that need to be overcome before the structural condition information inferred from a corrosion SHM system could be used in preference to condition information acquired through traditional inspections.

In order that a corrosion SHM system be used in lieu of traditional visual inspections, it will need to be capable of predicting corrosion in a joint in a remote area. This will require that the system be capable of diagnostically determining when the protective coating has failed, both physically via cracking, and chemically via the leaching and removal of corrosion inhibitors from the area. While the BAE Systems' *Sentinel* sensor [13] goes some way to this end (see Section 3.1), in that it provides information regarding the depletion of the corrosion inhibitors given physical defects in the sealant, no present day corrosion SHM system is capable of determining when the coating has failed both physically and chemically. However, efforts within DSTO and the ADF are designed to produce interim capabilities that will assist in the current management practices of ADF aircraft and will facilitate the acceptance of the SHM approach to corrosion management by the operator and maintainer. Examples of such interim capabilities are shown in Figure 27: The roadmap for DSTO's corrosion SHM technologies. This figure is a schematic illustration of how improvements in the corrosion SHM system capability result in increased aircraft availability. For

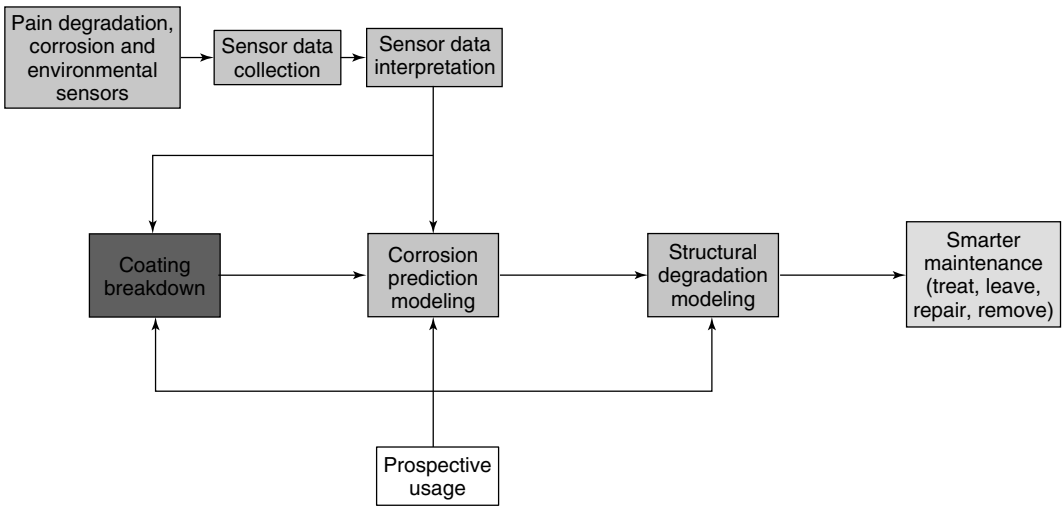


Figure 26. Schematic of the components of the DSTO SHM system for corrosion maintenance.

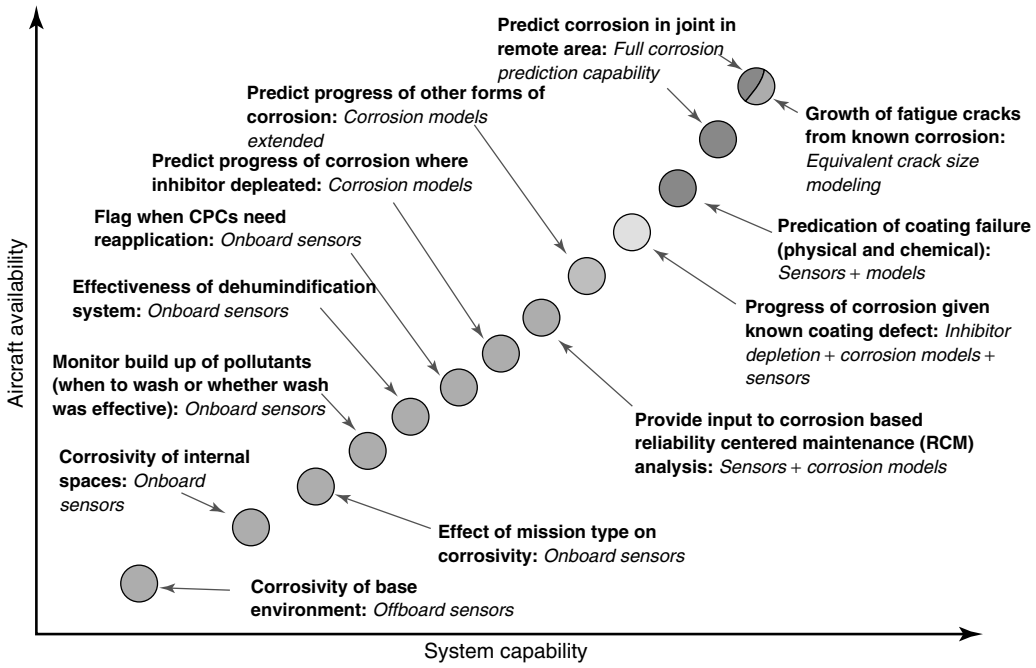


Figure 27. DSTO's corrosion SHM technology roadmap. The enabling capability (or technology) is in italics, while the resulting information made available to the maintainers is shown in bold. The color of the point indicates the level of technical maturity, where green (light grey) corresponds to existing capabilities and red (dark grey) corresponds to capabilities that have not yet been achieved.

each point on the roadmap, the enabling capability (or technology) is italicized, while the resulting information (outputs) made available to the maintainers is shown in bold. For example, the first point on the roadmap illustrates that a system of *off-board sensors* can provide information regarding the **corrosivity of the base environment**. Each point is also colored according to the level of technical maturity, green (light grey) corresponding to existing capabilities and red (dark grey) corresponding to capabilities that have not yet been achieved. What is clear from the roadmap is that there is a wealth of information that the existing system can provide to the operator/maintainers including

1. the corrosivity of the base environment (including hangars, shelters, or the flight line);
2. the corrosivity of the actual internal spaces within the aircraft;
3. the effect of mission type on the corrosivity within the aircraft;
4. the build-up and concentration of pollutants (salts etc.) on the aircraft, which can indicate when or how often the aircraft needs to be washed as well as how effective the wash is at removing pollutants;
5. the effectiveness of dehumidification systems, especially their efficacy in hard to access locations or locations containing high-value equipment;
6. when corrosion preventative compounds no longer provide protection and therefore need to be replaced;
7. information regarding the progress of corrosion where the inhibitor is depleted;
8. input to corrosion based reliability centered maintenance analyses.

In short, the existing system can advise the maintenance process and enable the forward prediction of maintenance requirements, including preventative maintenance, with the result that aircraft availability will increase and maintenance costs will decrease.

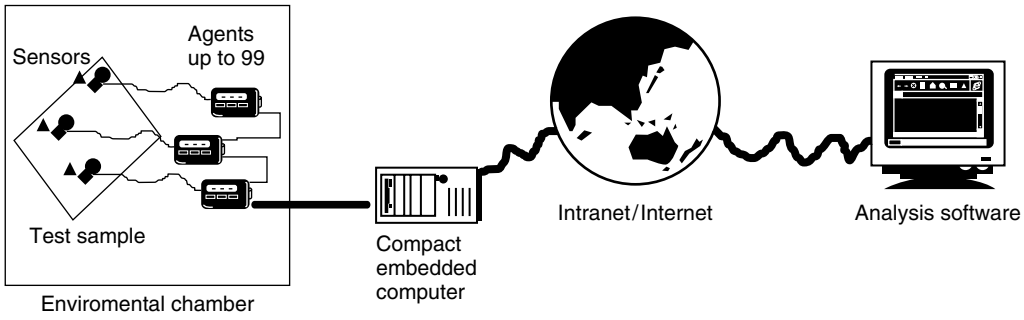


Figure 28. A schematic representation of the laboratory demonstrator.

7 LABORATORY AND FLIGHT DEMONSTRATORS

7.1 Functional approach

A laboratory demonstrator system has been set up to evaluate and demonstrate the methodology of this approach. A subsection of an aircraft (the subfloor region of the lavatory/galley area of a Boeing 747-400) has been set up with 11 sets of sensors and agents (Figure 28) [4]. The whole system has been placed in an environmental chamber and subjected to a continuing cycle of temperature and humidity fluctuations. Some variations in environment have been provided by undercooling of some areas (using a Peltier device) and sprays of contaminating liquids likely to be found in the lavatory/galley area. The demonstrator has been running for 18 months and has provided data for testing the analysis algorithms.

A set of sensors with a data logger has also been placed on a Boeing 747 aircraft regularly flying across the Pacific and the data collected from these trials has also been useful in validating the analytical approach. A similar set of sensors and data logger have been installed in an operational military aircraft more recently.

Data from the demonstrator and flight trials have been analyzed using the proposed technique and the results are illustrated in Figures 29 and 30. Figure 29 is based on the sensor outputs from one of the agents on the laboratory demonstrator. It shows the corrosion sensor (damage) output and predictions derived from different subsets of the data. In each case, the data

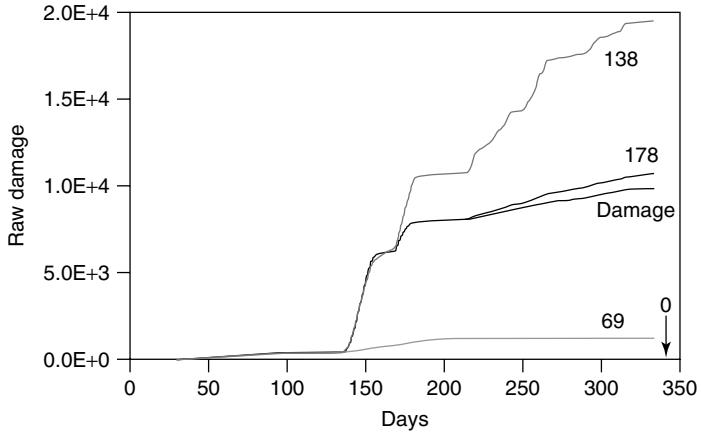


Figure 29. Fit of models with measured damage (cumulative corrosion current in microamperes) for galvanic corrosion sensor at one position on the laboratory demonstrator. The models are generated from nine days of data starting at day 0, 69, 138, and 178. The line 0 runs coincident with the x axis and is not visible.

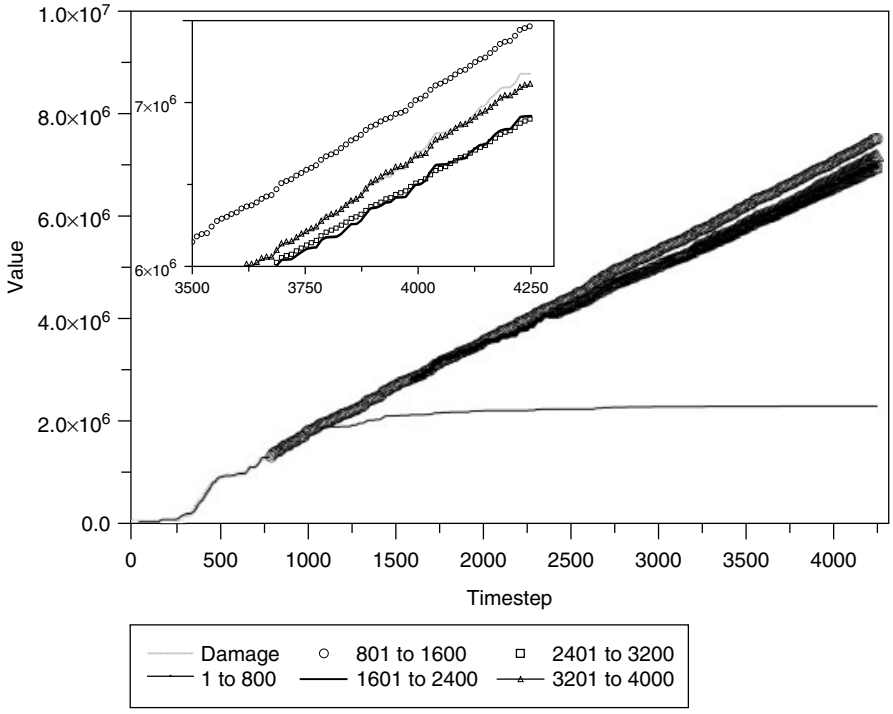


Figure 30. Analysis of real in-flight data using moving windows to generate models.

model is calculated from about nine days of data and then applied to predict the data for the rest of the time sequence using the measured microclimate parameters. Thus the lines marked 0, 69, 138, and 178 are predictions derived from data 0–9, 69–78, 138–147, and 178 to 187 days respectively, and then applied into the future. This analysis demonstrates that the models can closely fit the damage stream from the microclimate data and can successfully predict damage into the near future. It also illustrates how the damage/microclimate relationship changes as damage develops, since early models do not remain valid. Thus the method of continuously regenerating the model that is embedded in the software will ensure that the prediction will remain accurate.

The analysis of in-flight data is shown in Figure 30. Again the models are trained on different portions of the data and used to predict into the future using the microclimate data. This shows that predictions can be made with good accuracy into the immediate future given that sufficient data is available to establish the modeling relationships. It also illustrates that the technique is applicable to real in-flight data.

In these illustrations, the raw output from the corrosion sensor has been used, and the relationships developed to map this onto actual damage of aluminum materials (discussed in Section 5.1) have not been integrated into the calculations.

7.2 DSTO model validation

The physics-based CPM developed to date predicts the occurrence of pitting corrosion at damaged sites in paint coatings on aircraft. The CPM has been validated in the laboratory with this use in mind by exposing the sensor suite and corrosion coupons in an environmental chamber programmed with two varying RH and temperature cycles. The first cycle was designed to induce few and small corrosion pits and was called the *benign cycle*, and the latter was designed to produce many and larger pits and was denoted the *severe cycle*. These two cycles are displayed in Figures 31 and 32. The benign cycle has a maximum RH of 90% and no rapid condensation

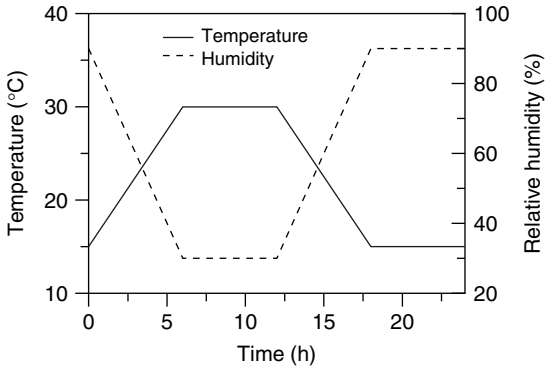


Figure 31. Benign environmental cycle.

period, whereas the severe cycle has a maximum RH of 100% and a rapid condensation period.

Coupons of aluminum alloy AA7050 from the rolled surface of a 100-mm plate were finished to 1200 grit. The coupons were left unpainted to mimic areas of exposed aluminum on aircraft after the paint has degraded. The coupons had an area of 1 cm^2, to give a representative group of pits, but the model could have easily used smaller exposed areas of bare metal, such as those experienced on aircraft. The corrosion coupons and sensor suite were doped with NaCl to two different salt densities before exposure. The first salt density was 5 $g\,m^{-2}$, a level designed to result in full coverage of the surface with condensed moisture, and the second salt density was 0.1 $g\,m^{-2}$, a salt density typically found on ADF aircraft [5]. The sensor suite was bonded to aluminum specimen (see Figure 33a) and the sensors were interrogated using a DSTO designed data logger system. The sensor system has also been miniaturized for deployment onboard aircraft and is shown in Figure 33(b).

The corrosion model predicts, using Faraday's equation, the number and size of pits occurring on aluminum surfaces from the measured coulombs of charge. The volume of the pits on the corrosion coupons was estimated from the measured depth, width, and length using the equation for the volume of a spherical segment. This volume was then normalized into an equivalent pit size for a hemispherical pit of the same volume. In this manner, the predicted and measured pit sizes were compared.

The modeled pit diameter distributions generated from both the one-week and two-weeks sensor data in the severe cycle doped with 5 $g\,m^{-2}$ NaCl were very similar to the measured pit diameter distributions (Figures 34 and 35). The largest difference between the measured and modeled results, for both periods, was for the lowest pit diameter bin where the modeled number of pits was greater than the measured number of pits. However, experimental verification is difficult for small pit sizes and due to time constraints and the very large numbers of these small pits on these coupons, a cutoff size (dependent upon the number and size of pits present) was imposed, where pits smaller than this size were not measured. However, they were predicted by the model.

The results displayed in Figures 36 and 37 are for exposure to the severe cycle with 0.1 $g\,m^{-2}$ NaCl doped surfaces. In this test, the modeled and measured pit diameters were similar in number, maximum size, and also in size distribution. Again

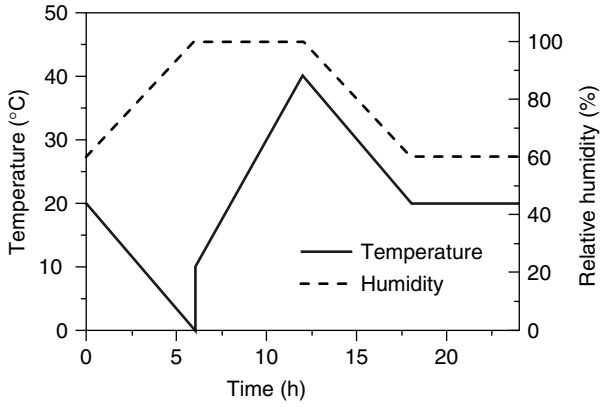


Figure 32. Severe environmental cycle.

(a) (b)

Figure 33. This figure shows (a) validation test apparatus and specimen, and (b) miniaturized sensor system (with dimensions 54 mm long by 38 mm wide).

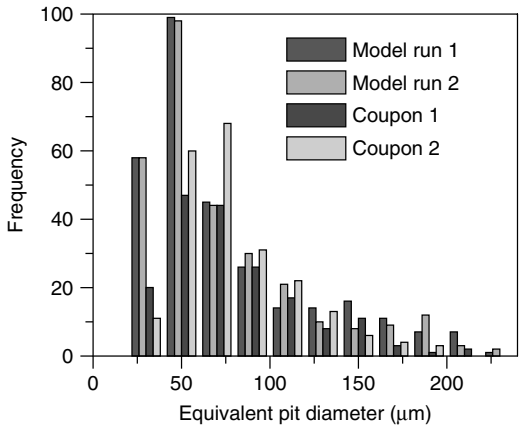


Figure 34. Pit sizes for AA7050 exposed for one week to the severe cycle after being doped with 5 g m^{-2} NaCl.

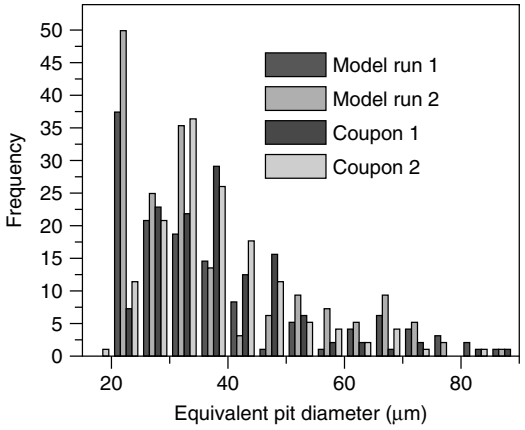


Figure 36. Pit sizes for AA7050 exposed for one week to the severe cycle after being doped with 0.1 g m^{-2} NaCl.

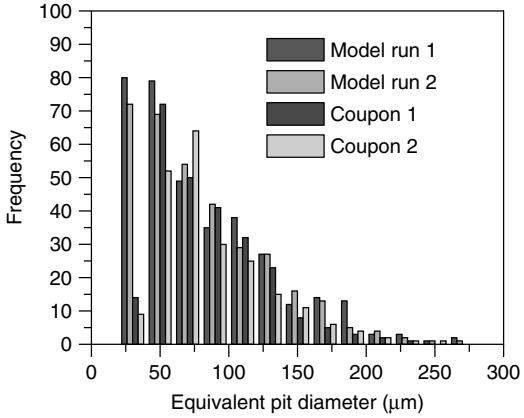


Figure 35. Pit sizes for AA7050 exposed for two weeks to the severe cycle after being doped with 5 g m^{-2} NaCl.

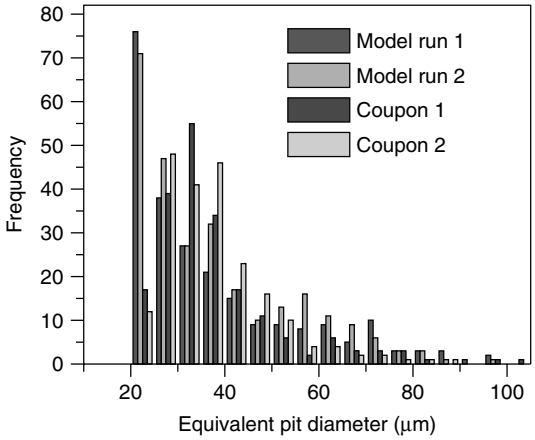


Figure 37. Pit sizes for AA7050 exposed for two weeks to the severe cycle after being doped with 0.1 g m^{-2} NaCl.

the pit numbers for the smallest bin were higher for the modeled distribution than the measured distribution due to the measurement cutoff at the small sizes.

For the test conducted in the benign cycle with 5.0 g m^{-2} NaCl doped surfaces, the modeled pit diameters were similar to the measured pit diameters in this test with respect to numbers of pits and maximum pit sizes. However, there were some differences in the shapes of the distributions (Figures 38 and 39). The reasons for this are variations in the pit probabilities at higher chloride concentration that have not yet been incorporated into the modeling. That is, the benign cycle did not include a condensing period and the maximum RH was 90%, resulting in chloride concentrations above 2.5 M at all times. In concentrated chloride solutions, the probability of pitting was reduced dramatically because of the lower solubility of oxygen.

The pit diameters modeled and those measured were in approximate agreement for maximum pit size and pit numbers, but again the distribution was dissimilar for the test conducted in the benign cycle with 0.1 g m^{-2} NaCl doped surfaces (Figures 40 and 41). The reason for this is the higher NaCl concentrations expected at the lower RH as discussed above.

The modeled results predicted using environmental sensor inputs were in good agreement with those

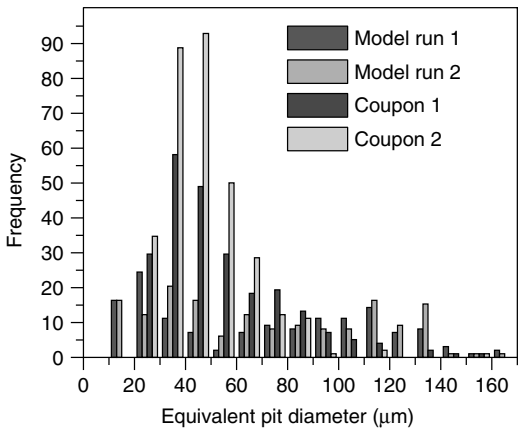


Figure 38. Pit sizes for AA7050 exposed for one week to the benign cycle after being doped with 5 g m^{-2} NaCl.

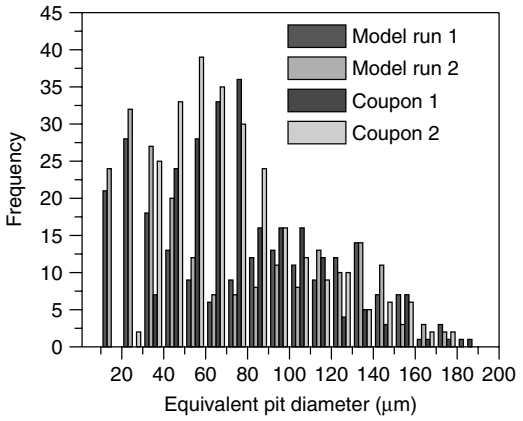


Figure 39. Pit sizes for AA7050 exposed for two weeks to the benign cycle after being doped with 5 g m^{-2} NaCl.

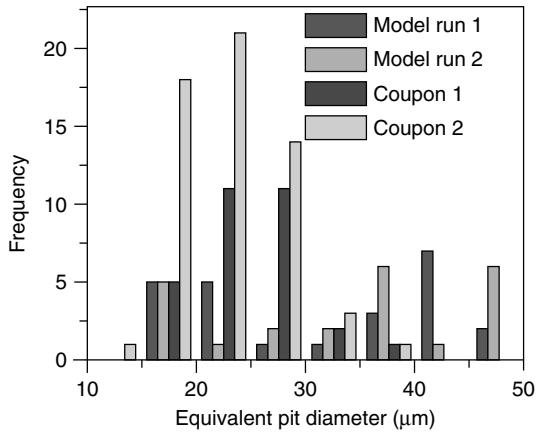


Figure 40. Pit sizes for AA7050 exposed for one week to the benign cycle after being doped with 0.1 g m^{-2} NaCl.

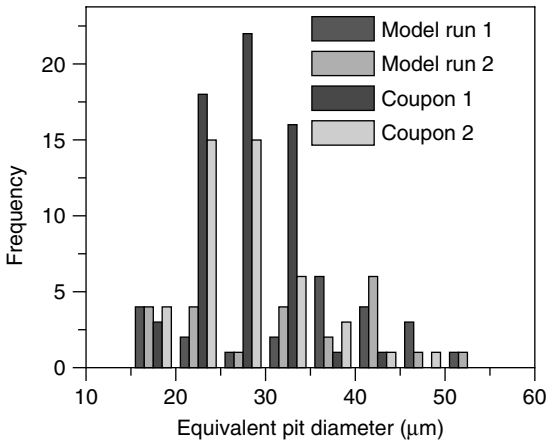


Figure 41. Pit sizes for AA7050 exposed for two weeks to the benign cycle after being doped with 0.1 g m^{-2} NaCl.

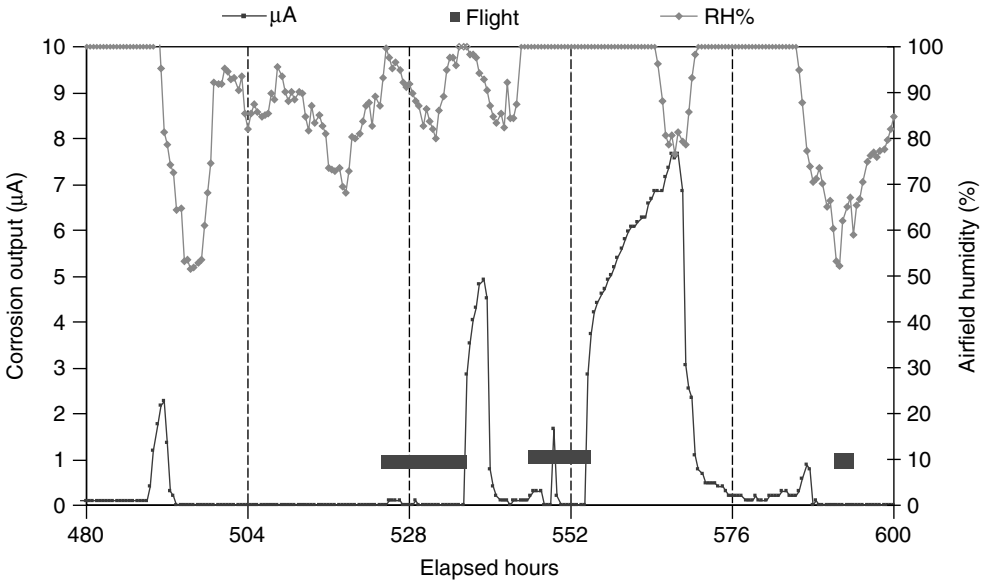


Figure 42. Typical output from a corrosion monitor located in the tail section of an RAAF P-3C maritime patrol aircraft.

measured from corrosion coupons exposed adjacent to and treated the same way as the sensors. This was especially true for the test panels exposed to the severe environment where the pit numbers, pit sizes, and pit-size distributions predicted by the corrosion model were in close agreement with those measured. The pit sizes and pit numbers predicted for panels exposed to the benign environment were also in close agreement, but the pit-size distributions were different. The reasons for these differences are the reduction in pit probabilities in concentrated NaCl solutions.

Overall, the use of corrosion/environment sensors with mechanistic-based modeling has been demonstrated to be capable of predicting, with some accuracy, the different levels of pitting corrosion observed on AA7050 exposed to a variety of simulated aircraft exposures of different severities.

7.3 In-flight demonstrators

A DSTO-developed TOW monitor was first mounted in the tail section of an RAAF P-3C Orion maritime patrol aircraft in 1998 [8]. This section is unpressurized, and the internal structure is exposed to the atmosphere via various vent and actuator holes. Some of the corrosion activity from the monitor could be attributed to the ambient conditions at the base while the aircraft was between flights. Corrosion activity was present during flight on many occasions, but only on flights lasting more than 1 h. There was no significant correlation of mission type or flight duration with corrosion activity. However, the data analysis revealed that a very significant amount of activity occurred shortly after the aircraft landed. This "postflight" activity was attributed to condensation when ambient air entered the structure, which had cooled below the dew point during flight. The amount and duration of this activity varied with the ambient humidity. If the RH was above 80%, nearly half the flights had condensation after landing, while if the RH was below 50%, only about a tenth of the flights experienced condensation. Typical output from the monitor is shown in Figure 42. The thin black data is the output from the monitor in microamperes (μA), and the gray data is the RH at the airfield. The solid black lines represent the time the aircraft is in flight. Some corrosion activity occurs while the aircraft is on the ground, the high humidity (100%) indicating rain. Significant peaks have also occurred immediately after landing, with the output being higher, and of a longer duration, when the airfield humidity is high [8].

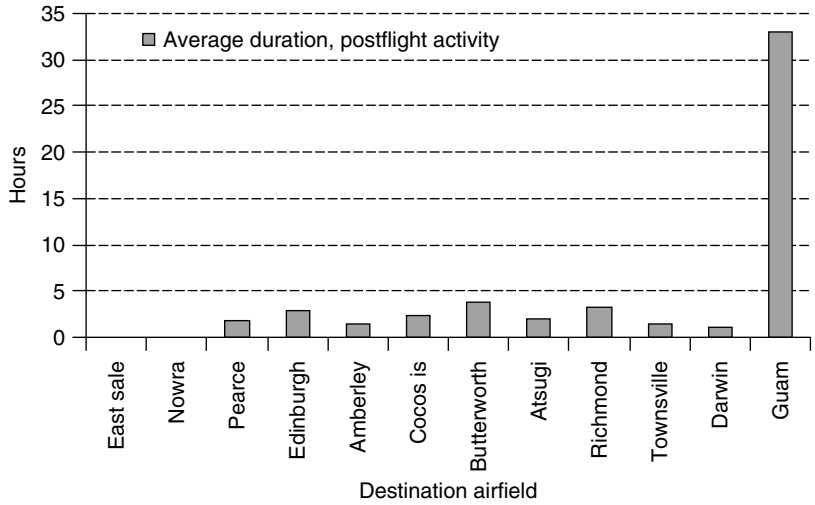


Figure 43. Duration of postflight corrosion activity at various airfields.

In the F-111, the monitor was located in the forward section of the weapons bay, also unpressurized (P. Trathen, private communication, 2007). In contrast to the P-3C, no condensation occurred after landing, in the F-111. This difference in behavior is attributed to the temperatures reached during flight. The P-3C, being a propeller-driven aircraft is significantly slower than the F-111. It is estimated that the P-3C would cool to $-25\,^{\circ}C$ in high-level flight. The F-111, a fast jet, would rarely go below ambient temperature, and, in fact, the skin temperatures could heat to $100\,^{\circ}C$ or higher during a typical mission owing to frictional heating from airflow. The significance for aircraft maintainers is that flights on P-3C may be a significant contributor to corrosion damage, due to the postflight activity. The F-111 corrosion is likely to depend entirely on ambient conditions at the base and be unrelated to flight. The importance of geographic location was demonstrated by the P-3C flights to Guam, where all flights to this small island base produced sustained corrosion activity, as shown in Figure 43.

The USN SPEC galvanic sensor system has been fitted to two Royal Australian Navy (RAN) Seahawk helicopters. Initial results from this trial have shown significant activity from sensors fitted externally, but less from those mounted inside the structure. Data is being collected from both land-based and ship-based aircraft.

8 SUMMARY

SHM systems that provide diagnostic and prognostic information on corrosion-related damage will enable maintainers and operators of aircraft to manage the prevention and control of corrosion in structural aircraft components on a condition basis rather than on the basis of an elapsed number of (flying) hours. The purpose of this article has been to outline various concepts, methodologies, and technologies involved in a corrosion SHM system that could provide corrosion diagnostic and prognostic capability. SHM systems for predicting corrosion damage are still under development; however, the results presented above show that the development of various corrosion-prediction models, using physics of failure and statistical approaches, is progressing rapidly. The fine-tuning of these models will depend on model testing and validation using laboratory and onboard environment monitoring systems. While further improvements in sensing capability would be desirable, major deficiencies in current systems at this time are (i) the absence of models that predict the mechanical and chemical failure of paint coatings and (ii) the lack of capability to predict the corrosion state in unsensed regions.

Current programs in place are designed to produce interim deliverables, which will assist in the current management practices of aging aircraft fleets. For example, the outputs from environment monitors

on ADF aircraft have identified mission types and base locations that are associated with high corrosion activity. This type of information may lead to more efficient and effective corrosion preventative maintenance.

ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance of Peter Vincent and Ian Powlesland with their inputs on the DSTO sensor system, and Wayne Ganther for his contribution to the CSIRO testing program.

REFERENCES

[1] Cooke G, Cooke G Jr, Kawanishi G. *A Study to Determine the Annual Cost of Corrosion Maintenance for Weapons Systems in the USAF*. Prepared for AFRL/MLS-OL by NCI Information Systems, Inc, Contract No. #F09603-95-D-0053, February, 1998.

[2] Agarwala V, Ahmad S. Corrosion detection and monitoring—a review. *Proceedings of Corrosion 2000*. NACE International: Houston, 2000; Paper 00271.

[3] Conroy W LCDR. *Brief on Naval Air Systems Command Aircraft/Engine/Support Equipment Corrosion Data*. Naval Air Systems Command, Air Vehicle Integration, 21 April 2004.

[4] Cole I, Corrigan P, Ganther W, Muster T, Paterson D, Price D, Galea S, Hinton B. A novel system for corrosion monitoring, diagnosis and prognosis in aircraft structures. *Proceedings of the 6th International Workshop on Structural Health Monitoring*. Stanford University, Stanford, CA, September 2007.

[5] McAdam G, Russo S, Trathen PN, Trueman A, Galassi A, Duxbury E. *Salt Contamination on Some ADF Aircraft*, DSTO Corrosion Control Report No. 8/00, Defence Science and Technology Organisation (DSTO), June 2000.

[6] Hughes AE. *Run Off From Slotted Panels*, BAE Systems Australia CPM Report No. 1269/029, BAE Systems, Australia, 2000.

[7] Nikpour T, Curtis PR, Hughes AE. *Cr-Leaching From Inhibited Primers During Immersion and NSS Exposure II*, BAE Systems Australia CPM Report No. 1269/075, BAE Systems, Australia, 2000.

[8] Trathen PN, Hinton BRW. Corrosion monitoring on defence force aircraft. *Proceedings of Australian Corrosion Association Conference*. Adelaide, 2002.

[9] Lai PK, Trathen PN, Hinton BRW. An atmospheric corrosion sensor for use in aircraft structure. *Proceedings of Corrosion and Prevention'98, No7, Australian Corrosion Association Conference*. Hobart, 1998.

[10] Niblock TGE, Surangalikar HS, Morse J, Laskowski BC, Castro-Cedeno MH, Wilson AR. Development of a commercial micro corrosion monitoring system. In *Proceedings of SPIE's International Symposium on Smart Materials, Nano-, and Micro-Smart Systems, Smart Materials II Conference*, Wilson AR, Varadan VV (eds), SPIE: Bellingham, WA, 2002 Paper 4934-25; pp. 179–189.

[11] Trego A. Installation of the autonomous structural integrity monitoring system. *Proceedings of the 4th International Workshop on Structural Health Monitoring*. Stanford University, Stanford, CA, 15–17 September 2003; pp. 863–870.

[12] Davis GD, Vargo TG, Dalgleish AW, Deason D. Corrosion protection and health monitoring using smart appliqués. *Materials Performance* 2004 **43**(8):32.

[13] Stonham A. JSF corrosion sensor overview. Presented at the *2007 Ageing Aircraft Users' Forum*. Brisbane, (http://www.boeing.com/global/Australia/AgeingAircraft2007/index.html) 2007.

[14] Ganther WD, Muster TH. Coiled galvanic corrosion sensors for atmospheric corrosion monitoring. *Proceedings of Corrosion and Prevention 2005*. Gold Coast, 20–23 November 2005, Paper 032.

[15] Muster TH, Sexton BA, Smith F, O'Halloran R, Davis T. Thin film sensors for atmospheric corrosion and structure monitoring. *Proceedings of Corrosion and Prevention 2005*. Gold Coast, 20–23 November 2005, Paper 082.

[16] Hayes-Roth B. An architecture for adaptive intelligent systems. *Artificial Intelligence: Special Issue on Agents and Interactivity* 1995 **72**:329–365.

[17] Vincent PS, McMahon PJ, Muscat RF, Zeve L, Wilson AR. A small low-power networked and versatile sensor interface. *Proceedings of SPIE International Symposium on Smart Materials, Nano-, and Micro-Smart Systems 2006: Smart Structures, Devices and Systems III*. Adelaide, December 2006; Paper 6414-39.

[18] GM9540P General Motors Engineering Standard. Accelerated Corrosion Test, December, 1997.

[19] Potvin E, Brossard L, Larochelle G. Corrosion protective performances of commercial low-VOC epoxy/polyurethane coatings on hot-rolled 1010 mild steel. *Progress in Organic Coatings* 1997 **31**(4):363–373.

[20] Leygraf C, Graedel TE. Atmospheric corrosion. *Electrochemcial Society Series*. John Wiley & Sons: New York, 2000.

[21] Trueman AR, Begbie K, Hinton B. Modeling the atmospheric pitting corrosion of aluminium alloys. *Proceedings of Australasian Corrosion Association*. Hobart, 2006.

[22] Montrith JL, Unsworth MH. *Principles of Environmental Physics*, *Second Edition*. Edward Arnold: London, 1990.

[23] Ansari AS, Pandis SN. Prediction of multicomponent inorganic atmospheric aerosol behaviour. *Atmospheric Environment* 1999 **33**:745–757.

[24] Topping DO, McFiggans GB, Coe H. A curved multicomponent aerosol hygroscopicity model framework: part 1—inorganic compounds. *Atmospheric Chemistry and Physics* 2005 **5**(5):1205–1222.

[25] Trueman AR. Determining the probability of stable pit initiation on aluminium alloys using potentiostatic electrochemical measurements. *Corrosion Science* 2005 **47**(9):2240–2256.

Chapter 103

Experience with Health and Usage Monitoring Systems in Helicopters

Dy Dinh Le

Federal Aviation Administration, Air Traffic Organization, William J. Hughes Technical Center, Atlantic City International Airport, NJ, USA

1 INTRODUCTION

The health and usage monitoring systems (HUMSs) in the late 1980s comprised one or more stand-alone and basic analog monitoring units to monitor vibrations and exceedances in the engine, gearbox, drive system, or bearings (*see* **Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance**). Early systems were used to provide warnings of impending failures of rotorcraft components or subsystems. Current HUMS technologies include analog and full digital systems to provide health and some limited maintenance credits such as rotorcraft track and balance. However, the HUMSs along with a ground-based station have not been certified by the Federal Aviation Administration (FAA) to provide usage or condition-based monitoring for maintenance credits. "Maintenance credit" means to give approval to a HUMS application that adds to, replaces, or intervenes in industry-accepted maintenance practices or flight operations.

In the 1990s, HUMS technologies had been advanced significantly in many areas, including flight regime recognition, usage monitoring, diagnostics and prognostics, data fusion, and sensors. The advancement of HUMS has begun to allow the civil as well as military operators of rotorcraft to discuss HUMS potentials in approaching a true condition-based maintenance or in determining maintenance credits.

Currently, the HUMSs along with the ground-based station (GBS) have not been certified by the FAA to provide usage or condition-based monitoring for maintenance credits. Certification of a HUMS for maintenance credit purposes is considered to be a complex endeavor. The Advisory Circular (AC) 29-2C, Section MG-15, hereafter referred to as the *HUMS AC*, is the only FAA document providing guidance for HUMS airworthiness approval [1]. The HUMS AC provides guidance for transport category rotorcraft to achieve airworthiness approval for installation, credit validation,

ISBN: 978-0-470-05822-0.

and instructions for continued airworthiness for a full range of HUMS applications. The HUMS AC establishes an acceptable means, but not the only means, of certifying a rotorcraft HUMS regardless of complexity or intended usage to modify maintenance and/or operational actions.

Section MG-15 of AC 29-2C was developed by the Rotorcraft Health Usage Monitoring System Advisory Group and first released in 1999. This advisory group consisted of representatives from the FAA Aircraft Certification Services, FAA Aircraft Evaluation Group, US Industry, European Industry, and Joint Airworthiness Authority. The HUMS AC was written in a generic manner such that it provides the basic requirements and guidance for certification of HUMS. Since no HUMS has yet to be certified in accordance with this AC on a maintenance credit basis, the FAA is conducting research to assist in the substantiation of the certification guidance. Additionally, HUMS research efforts are to develop, validate, and/or demonstrate HUMS operational requirements; applicable technologies including processes, methodologies, and algorithms, and other required information including data to guide the certification of HUMS. The results of these Research and Development (R&D) efforts will also allow the FAA to incorporate any lessons learned into the certification process.

2 APPLICATIONS OF HUMS

2.1 HUMS benefits

HUMS technologies for health monitoring have been widely used in the past 20 years to detect impending failures of rotorcraft mechanical systems (*see* **Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance**). There have been numerous studies to quantify the impacts and benefits of using HUMS. One study, conducted by the Civil Aviation Authority (CAA), involved HUMS-equipped helicopters [2] flown between 1992 and 1996, and was based on 500 000 h of rotorcraft operation in the North Sea. According to the CAA data presented by Shell Aircraft International [3], HUMS can successfully detect 69% of mechanical defects in critical rotating parts before catastrophic failures (Figure 1). In the United Kingdom, the Civil Aviation Publication 693

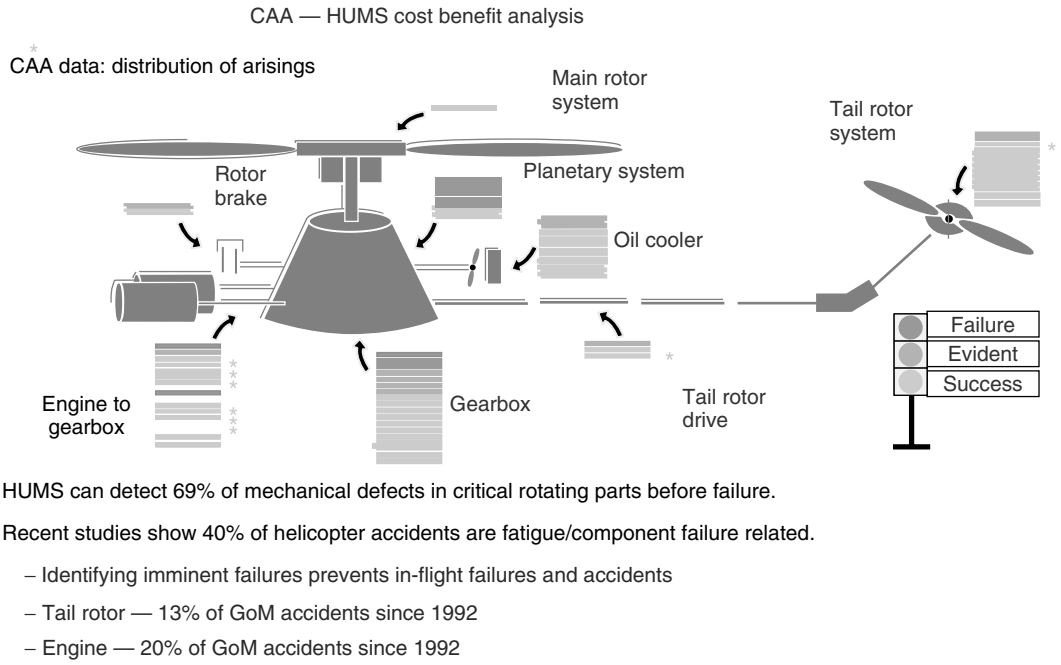


Figure 1. HUMS detection rate. [Reproduced with permission from Ref. 3. © Sheffield, B.]

entitled "Acceptable Means of Compliance Helicopter Health Monitoring" issued in May 1999 [4] states "the first generation of HUMS ... has already demonstrated the ability to identify potentially hazardous and catastrophic failure modes, and has already reduced fatal accident statistics."

2.2 HUMS technologies

2.2.1 Certification issues

A low degree of qualification is required for installing a basic HUMS that monitors rotorcraft parameters, identifies exceedances, and provides advisory data to maintenance personnel. A much higher degree of qualification is required to certify a HUMS to provide onboard warnings to the pilot or for usage or condition-based maintenance credits. So far, the FAA has only certified a few HUMS for health monitoring and one HUMS for limited maintenance credits. The lessons learned from the initial certification of HUMS provided valuable insights on how the future HUMS are to be certified for full usage and maintenance credits. Current certified HUMS maintenance credits are restricted to providing rotor track and balance solutions. Usage data and recorded exceedances are only used for maintenance advisories. Future HUMS will potentially take advantage of the state-of-the-art technologies to alter maintenance requirements and schedules using actual usages and the state of aircraft health monitored by HUMS.

Currently, no rotorcraft HUMS with a GBS, which is typically a commercial off-the-shelf (COTS) system, has been certified to provide usage or condition-based maintenance credits. Certification of such a COTS HUMS is considered to be very difficult and faces many challenges because it has not been certified to the same level of criticality used for airborne fixed-wing systems (*see* **Agile Military Aircraft**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**). Many issues still need to be resolved. For example, the use of a floppy drive, flash memory card, local area network, or the Internet to install software has the potential to introduce viruses and worms into the system. The processed data stored in the GBS will ultimately be used to make decisions on whether actions are required to address the safety or continued airworthiness of rotorcraft. Since the GBS is one of the important links of HUMS, it must have high integrity and accuracy.

Technologies for usage credits

The next generation of HUMS for usage credits will become more integrated and equipped with advanced technologies such as fault or damage detection, flight regime recognition, direct measurement, and/or data fusion to accurately capture the actual usage (how the aircraft is being flown) or to detect a crack in rotorcraft drive train systems before it reaches its critical length. Advanced tools used in HUMS for usage credits cover a wide spectrum of technologies, which require extensive knowledge of numerous areas.

The physics of proposed credits also needs to be fully understood and validated. Technologies used to substantiate the physics of proposed credits may include damage tolerance (DT), probabilistics, prognostics, and other HUMS-related methodologies. The next generation of HUMS can also be used to address the DT requirement. HUMS-DT technologies will, therefore, be inevitably combined to optimize safety benefits. A probabilistic approach and statistical tools are also critical in managing the uncertainty that may be inherent in HUMS data and information. These technologies can be used to develop statistical confidence measures to demonstrate the reliability and robustness of the technique used in deriving characteristics of collected HUMS data.

2.3 HUMS for enhancing safety and continued airworthiness

2.3.1 Rotorcraft safety concerns

The use of HUMS for usage monitoring is also becoming critical in an effort to reduce the rotorcraft accident rate. According to a summary report of the International Rotorcraft Safety Symposium [5] held in September 2005 in Montreal, Canada, the US civil rotorcraft accident rate per 100 000 flight hours is 8.09 (the fatal accident rate is 1.48). These accident rates were based on the 2004 US rate of an estimated 2 225 000 total flight hours. The US commercial transport aircraft carrier (14 CFR Part 121) accident rate, however, is 0.159 (the fatal accident rate is 0.011). By comparison, the rotorcraft accident rate is 51 times higher than the commercial transport aircraft

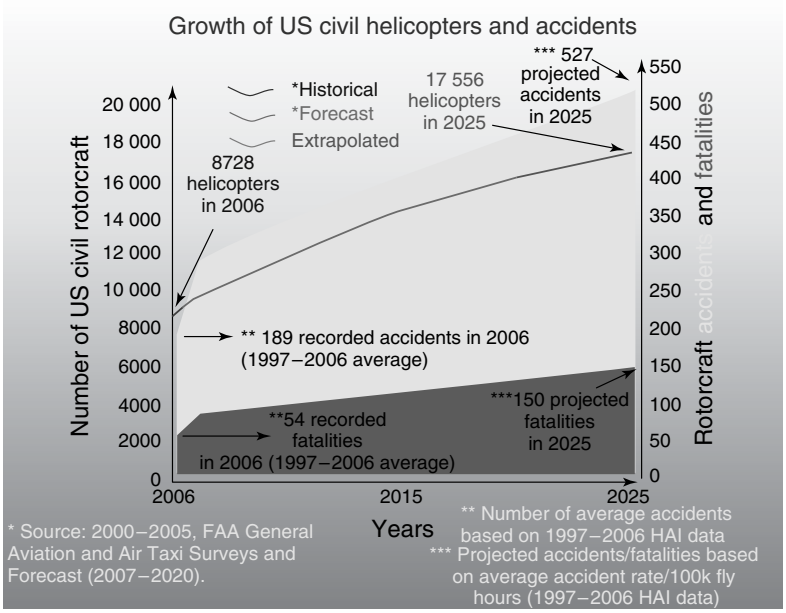


Figure 2. Forecast of the number of US civil rotorcraft and accidents.

accident rate. In 2003, the US Congress enacted legislation requiring the FAA to conduct the research and development to achieve five different goals including the reduction of the rotorcraft accident rate [6]. Additionally, the US Congress also mandated the FAA to develop a plan to implement and support the next generation air transportation system by 2025 [7].

Figure 2 shows the forecast of growth of the number of US civil helicopters by 2025. In 2006, according to the FAA General Aviation and Air Taxi Surveys [8], the US civil rotorcraft fleet has 8728 helicopters. It is projected that the total number of the US civil helicopters may increase to 17 566 by 2025. The figure also shows the historical rotorcraft accidents and projected occurrences in 2025. According to the Helicopter Association International data collected from 1997 to 2006 [9], on average, there were 189 rotorcraft accidents occurred per year. It is projected that the rotorcraft accidents may increase to 527 in 2025. Similarly, the fatalities may increase from an average of 54 deaths in 2006 to 150 in 2025. These projections are based on the estimated flight hours per year for rotorcraft. If 80% reduction in accidents can be achieved, the total number of accidents and fatalities may drop to 105 and 30, respectively. The 80% reduction goal was adopted and pursued by the industry during the International Helicopter Safety Symposium in Montreal, Canada, in September 2005 [5].

2.3.2 Rotorcraft usage monitoring

The US military released several thousands of its surplus helicopters into the commercial sector and these were designed, manufactured, and assembled prior to the early 1960s. Hence many helicopters flying in commercial and restricted category service today are from this pool. A large number of these helicopters are currently being used under contract by the US Forest Service for firefighting and logging in US national parks. 'The original military requirements that were reckoned in designing and developing these helicopters are much different from those governing their current usage. The fatigue design for these surplus rotorcrafts was based on, at that time, the perceived utility mission. Archived data for the loads and fatigue calculations are sparse and not available for the more recent missions being flown. In fact, there is no known existing database that has recorded the repeated heavy lift missions, including logging and firefighting. Most of the flight load data that exist are for the utility mission. Scripted flights have not

been accomplished in these types of environments to determine the scatter data associated with how different maneuvers are flown in heavy usage situations.

Modern helicopters are also experiencing significantly more complex and variable loadings than in the past. Therefore, the spectrum data that are currently used by the FAA to certify designs and modifications have limited relevance to current operations. Since stress states that reflect actual operations are critical for achieving accurate lifetime predictions for both safe-life and DT approaches, an up-to-date and comprehensive mission spectra database is needed. As the rotorcraft DT philosophy becomes widely used among the US rotorcraft industry, obtaining sufficient data to determine accurate typical mission spectrum is even more critical for the success of DT applications.

The insertion of HUMS technologies, including direct load measurement and flight regime recognition to determine component loads and how helicopters are flown, is critical for usage monitoring. Using flight regime recognition, the actual helicopter usage can be obtained and compared with certification usage spectrum. An accurate determination and measurement of helicopter gross weight, center of gravity, and other key structural loads is also critical for establishing loads for rotorcraft components and determining their consumed fatigue life. Depending on how helicopters are actually flown, HUMS data can be used to enhance safety or reduce operating costs.

To assure the continued airworthiness of the civil rotorcraft, the FAA has begun to collect usage data using HUMS on selected rotorcrafts. Further efforts are being planned to cover more US civil rotorcraft operations usages. Scripted flights using HUMS and other equipment and accessories are being adapted to obtain statistically valid data on the percentages of time that given specified maneuvers are being executed in modern rotorcraft operations. These data will span the range of current and potential future missions, including long-line heavy lift, emergency medical service, offshore support, and corporate transport.

The FAA also plans to establish a database of usage spectrum and loads from helicopters that are flown in the restricted category environment. Additionally, HUMS-equipped helicopters with instrumented devices to record various strains in critical components can also be used for scripted flights. Using HUMS, data can be collected, transmitted to, and maintained by the FAA for analysis. The analyzed flight load data can then be compared with the original or extrapolated loads used for the design of the military surplus rotorcrafts to determine the need for safety actions. Using the measured flight loads and recorded usage of helicopters, fatigue life of life-limited components can be determined and compared with the original fatigue life of the military helicopters being used for civil operations.

2.4 FAA HUMS research efforts for future maintenance credits

2.4.1 Damage tolerance design of rotorcraft components using HUMS usage spectrum

In the past 15 years, the FAA has funded numerous research projects to collect usage data using HUMS. One of the efforts involved the use of helicopters flown in numerous missions [10]. Three usage spectrums, namely, Atlanta Short Haul Mission (ASHM), Gulf Coast Mission (GCM), and Utility Mission in Morgan City (UMMC), were obtained using HUMS. The ASHM usage spectrum was considered severe usage involving many short maneuvering flights to provide pickup and delivery services at the Atlanta Olympics. The GCM usage spectrum was considered mild usage involving long cruise flights. The UMMC involved cruise flights along with some shorter flights for pickup and delivery services. The UMMC usage spectrum was more severe than the GCM but less severe than the ASHM.

Using the certification and other usage spectrum collected by HUMS, an analysis was conducted to calculate the fatigue crack growth in selected critical dynamic components, which were designed based on the safe-life method. On the basis of the analysis results, the selected components were theoretically redesigned to successfully meet the DT requirement with acceptable inspection intervals. Two major conclusions can be drawn from this hypothetical study.

1. HUMS can be used to determine the true flight spectrum of modern helicopters, which leads to a full understanding of how they are being flown. The successful determination of the rotorcraft usage spectrum will help determine accurate

Table 1. FAA HUMS R&D roadmap

HUMS research areas	Research duration (years)									
	1	2	3	4	5	6	7	8	9	10
HUMS AC compliance and demonstration										
HUMS demonstration and equipped flight testing										
Usage monitoring and flight regime recognition										
Direct loads monitoring										
Maintenance credits validation (indirect load measurement)										
Maintenance credits validation (direct load measurement)										
Operational development of HUMS										
Hardware										
Sensor										
Airborne systems										
Ground station and accessories										
Software										
Data management										
Diagnostics and monitoring										
Maintenance management										
Commercial validation of HUMS										
Algorithm and methodologies										
Safety monitoring										
Structural usage monitoring and credit validation										
Diagnostics, health, and prognostics										

retirement times or help in the establishment of the effective inspection intervals of rotorcraft life-limited components. If the actual usage spectrum collected using HUMS is more severe than the certification usage spectrum, life-limited components may be removed before their retirement times are fully consumed. This will enhance the aircraft safety and potentially reduce the rotorcraft accident rate.

2. If the actual usage spectrum collected using HUMS is less severe than the certification usage spectrum, the retirement times of selected components may be extended, resulting in reduced operating costs.

2.4.2 Determination of maintenance credits using HUMS flight regime recognition and usage monitoring

Using the developed FAA HUMS R&D strategic plan [11] and roadmap (shown in Table 1), the FAA is currently supporting research efforts to develop

HUMS processes, methodologies, and validated data to substantiate proposed maintenance or usage credits. One of the programs is the demonstration of HUMS on a newly designed transport helicopter using existing usage monitoring and flight regime recognition technologies. In this demonstration program, a process for developing fleet-wide and component usage-based maintenance credits is being established. The program is collecting HUMS operational data including flight regime, gross weight, center of gravity, and other associated helicopter operational parameters. The collected data are then used to determine the retirement times of selected components. Usage-based maintenance credits are determined by comparing the calculated retirement times based on HUMS data with current component retirement times based on the existing life calculation methodology for the worst-case spectrum and maximum loads. From the collected information and analysis results, the FAA can establish and validate the credit validation, introduction into service, and instructions for continued airworthiness plans.

3 CONCLUSIONS

The current civil rotorcraft accident rate remains high. The FAA safety goal is to reduce the rotorcraft accident rate to a level equivalent to that of fixed-wing transport aircraft. Therefore, there is a need to deploy state-of-the-art equipment and advanced technologies to address the safety goal. HUMS, if installed, can detect impending failures of rotorcraft components or subsystems. HUMS with advanced functionalities and technologies including the regime recognition capability can be used for usage monitoring to enable the accurate determination of rotorcraft flight profile and, eventually, the precise evaluation of the load distribution in rotorcraft components. If the actual usage spectrum is more severe than the certification usage spectrum, life-limited components may be removed before their retirement times are fully consumed. This will enhance the aircraft safety and potentially reduce the rotorcraft accident rate. If the actual usage spectrum is less severe than the certification usage spectrum, the retirement times of selected components may be extended, resulting in reduced operating costs.

The FAA has been conducting HUMS research including flight testing to validate and demonstrate HUMS technologies for maintenance credits. The validated technical information including methodologies, processes, and data will be incorporated into the HUMS AC for use in certification.

REFERENCES

[1] Federal Aviation Administration (FAA). *AC 29-2C–Certification of Transportation Category Rotorcraft*, December 2, 2003.

[2] McColl J. *Overview of Transmissions HUM Performance in UK North Sea Helicopter Operations*, Institution of Mechanical Engineers Seminar S553, November 1997.

[3] Sheffield B. *Presentation on Helicopter Safety in the Oil and Gas Business*, Washington, DC, November 2004.

[4] CAA. *CAP 693–Acceptable Means of Compliance Helicopter Health Monitoring CAA AAD 001-05-99*, London, 1 May 1999.

[5] Flater R. *International Helicopter Safety Symposium 2005* 26–29 September 2005, Montreal, Quebec, Canada–Final Report. American Helicopter Society, October 17 2005.

[6] Vision 100-Century of Aviation Reauthorization Act (P.L. 108–176), Section 711, Rotorcraft Research and Development Initiative, 2003, http://www.nps.gov/legal/laws/108laws.htm.

[7] FAA. *Concept of Operations for the Next Generation Air Transportation System, Version 2.0*, 13 June 2007, http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/oep/nextgenvision/.

[8] FAA Aerospace Forecasts FY 2007–2020, Forecast Tables, General Aviation (Table 27–30), http://www.faa.gov/data_statistics/aviation/aerospace_forecasts/2007–2020/.

[9] U.S. Civil Helicopter Safety Statistics–Summary Report (1997–2006), Helicopter Association International, http://www.rotor.com/Default.aspx?tabid = 597.

[10] Michael J, Collingwood G, Augustine M, Cronkhite J. *Continued Evaluation and Spectrum Development of a Health and Usage Monitoring System*, FAA final report DOT/FAA/AR-04/6, May 2004.

[11] Le D, Cuevas E. United States federal aviation administration health and usage monitoring system R&D strategic plan and initiatives. *Proceedings of the 5th DSTO International Conference on Health and Usage Monitoring*, Melbourne, Australia, 2007.

Chapter 105

Validation of SHM Sensors in Airbus A380 Full-scale Fatigue Test

Christophe Paget[1], Holger Speckmann[2], Thomas Krichel[3] and Frank Eichelbaum[4]

[1] *Airbus, Filton, UK*
[2] *Airbus, Bremen, Germany*
[3] *Airbus, Hamburg, Germany*
[4] *IABG, Dresden, Germany*

1 INTRODUCTION

As part of the qualification process of any aircraft, it is mandatory to test the full-size (also known as *full-scale*) aircraft under static and fatigue loading. The full-scale fatigue test allows the fatigue behavior of the aircraft structure to be characterized, thus defining its repair scheme. This article focuses solely on the world's largest aircraft, the A380. Such an aircraft obviously takes longer to inspect. Moreover, the airline industry is keen on hastening the inspection time, while maintaining the current high level of aircraft safety. One solution to achieving the aforementioned is structural health monitoring (SHM) [1–9]. The potential of SHM is well understood in Airbus [10] and in the airline industry, and several research projects are aimed at such a technology. Various SHM sensors have been implemented on the A380 full-scale fatigue test specimen (also called *A380 EF*), as this specimen was available at the time of sensor implementation. There are, however, no plans to implement or retrofit SHM sensors on the in-service A380.

SHM technology can be considered as part of the nondestructive test (NDT) technology family. Both NDT and SHM are used or intended to be used on various structures; however, this article focuses only on aircraft structures.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Previous papers [2, 10] have emphasized the major benefits and potentials of SHM compared with the current NDT and visual inspections of aircraft structures. It is nevertheless expected that SHM will ultimately result in fundamentally changing the way an aircraft inspection is done. The current focus on the A380 EF case is on surface-bonded or sandwiched SHM sensors, rather than embedded ones. In the case of A380 EF alone, most interrogation units are installed near the specimen for the duration of the testing. However, data processing by the interrogation unit was carried out just prior to the NDT or visual inspections, that is, just as in any other inspection method. This work is obviously in addition to the existing NDT and visual inspections, but any findings from the SHM systems are, in any case, investigated for research and development purposes.

A full-scale structure like the A380 EF provides a realistic loading condition vital for evaluating the performance of the SHM technology. Moreover, it is easy to install and maintain the SHM systems and sensors on a full-scale specimen, such as the A380 EF, compared to installing them onto an aircraft that is already part of an operating fleet. Implementation on aircraft ready for flight test is, of course, the next step and is not discussed in this article.

The article deliberately does not describe SHM technologies as have other articles of this encyclopaedia. Instead, it focuses on the A380 EF specimen and its mechanical testing equipment [11, 12]. This article also describes the various requirements in the selection of a potential SHM technology for future use in structural tests only.

2 STRUCTURAL TESTS

Ultrasonic, eddy-current and X-ray techniques as well as visual inspection are currently the main methods of testing for structural damage. These well-known techniques are used to detect surface or hidden cracks, such as at rivet holes or countersinks. With the typical service conditions and because of the large inspection intervals, the aircraft design allows for the detection of damage from 2 mm in length with high-frequency eddy-current technique [13] and 4.5 mm in length with ultrasonic testing technique [14]. Although classical NDT techniques can detect much smaller damage, these tests are not carried out between the given inspection intervals. Since SHM sensors are permanently on board the aircraft, the effective inspection interval can be drastically reduced, resulting in our being able to detect smaller damage and, therefore, initiate a potentially smaller, lighter, quicker, and cheaper repair than with the current inspection techniques.

SHM can be used differently on ground-based structural test monitoring, compared with in-flight structural monitoring. Indeed, the required information on ground tests is “when did the crack initiate?” or “what was the initial flaw concept?” This information could then be used to identify more accurately in time the damage initiation, for further optimizing the in-service inspection programme of the monitored aircraft structure, resulting in potential increased aircraft availability.

An expected future requirement is to demonstrate, throughout the lifetime of the aircraft, that its structure remains free of multiple site damage (MSD), which may influence one another. This urges the need to develop new monitoring techniques, such as SHM technologies, to detect small cracks (down to 1 mm in length) with a satisfactory degree of overall reliability.

Another benefit of SHM is its continuous monitoring of structures, helping to determine the time or flight cycle number at which the damage was initiated. Today, on coupon specimens, for instance, the specimens have to be disassembled, checked for damage, and reassembled. This exercise has to be repeated several times to obtain the load cycle number at which the damage initiates. The frequency of inspection must be as high as possible to prevent missing the damage initiation, to further discovering a much larger damage, resulting in not rating the specimen being tested. Evidently, such a test is time consuming and the use of SHM would provide a more cost-effective technique to greatly improve such a process.

Similarly, on component (e.g., 5-m single-shell) and full-scale fatigue tests, the load cycle has also to be interrupted to determine the point in time of damage initiation.

A range of benefits can be derived using continuous SHM techniques on structural test specimens. These can help minimizing test interruptions, detecting initial flaws, reducing human error as well as monitoring hot-spot and wider areas. They can also

monitor the structures in hidden and difficult access areas. The other advantages of SHM are that all the above can be carried out simultaneously and instantly, reducing drastically the inspection time as well as avoiding the classical NDT or visual inspections on areas without damage, leading to a more directed maintenance concept.

However, the aforementioned benefits from SHM techniques must be validated on test benches, such as the A380 EF and future structural test specimens.

3 REQUIREMENTS

The SHM technology to be implemented for structural testing is not solely for that application but can also be a validation toward in-flight use. Therefore, it is expected that the selection criteria for the SHM technology for structural testing should also include requirements for airborne applications. If some SHM equipment does not fulfill airworthiness requirements, such as durability, it is understood that it might still be usable for structural testing.

As part of the requirement definition process, the preparation of specifications for qualifying SHM systems is to be taken into consideration. Contrary to the current procedures applicable to the NDT technology, the qualification of SHM systems requires new specifications. They are a prerequisite for qualification and thus certification.

The SHM system comprises a sensor and the interrogation unit. The sensor also comprises the sensing item, any additional conditioning unit (such as a preamplifier), the connector, the cable, the sensor/structure bondline, and any protective layers, such as sealant and topcoat. The interrogation unit usually includes a data-acquisition unit, a postprocessing and management data unit, and a data-storage unit. Some or all parts of the interrogation unit can be an existing part of aircraft avionics, such as remote data concentrator (RDC), or a stand-alone piece of equipment, for both aircraft retrofit and structural testing.

3.1 System

The general SHM system requirements include performance, reliability, and commercial aspects, such as availability, hotline support, and cost effectiveness.

The performance of the SHM system must be well above those of classical NDT techniques to have a chance of replacing them in the near future. Indeed, the overall sensitivity must be below 4.5 mm to compete with eddy-current technology, for instance. High accuracy and resolution are vital to monitor the damage size and its growth. Ideally, the system should also be able to monitor a wide range of damage types (crack, corrosion, delamination, debonding, etc.) and sizes, as well as various material types (metallic, composite, and sandwich structures) and thicknesses, with the same interrogation unit and sensors.

The reliability level [15] expected from SHM is relatively high for it to at least support structural inspection by replacing classical NDT, with neither positive- nor negative-false alarms. The reliability of the overall system, linked to the development assurance level (DAL), necessitates the use of redundancies in the sensor network, in the interrogation unit hardware as well as in the use of a few self-diagnostic tools.

Commercially, the SHM system (including spares) must be available worldwide. In addition, in-field and hotline supports are paramount in aircraft operation. The same applies to full-scale fatigue testing where any downtime can cost and compromise the programme certification whenever the SHM technology is used as a replacement of classical NDT or visual inspection, which is still not the case today.

Additional commercial requirements contain the element defining the business case. Indeed, an attractive business case imposes the use of low-cost SHM system, low-cost installation and maintenance of the system on the aircraft. The SHM system should also be lightweight to reduce the fuel consumption cost generated by the system weight. The SHM equipment must also be appealing to the user and therefore has to be quick and easy to use, to have a user-friendly human interface, and with only the off-board equipment being portable. The mean time between failures (MTBF [13]) of the system needs to be as long as the application lifetime to further reduce the need of maintenance and therefore improve the business case.

For structural tests mainly, low-cost systems could suggest the use of commercial, off-the-shelf (COTS) components. These components could also be used

for airborne application, as long as obsolescence of such components is carefully controlled.

3.2 Sensor

The main top-level sensor requirements are obviously the way the sensor shall be connected to the structure (surface bonded, sandwiched, or embedded), as well as its size with regard to any geometric limitations and the complexity related to its calibration.

Additional requirements are associated with the type of sensors, its physical principle with respect to the damage to be monitored and its environment, preferably wireless-data-transfer-enabled, costs of sensors, costs and time of sensor installation, weight of sensors/cables/fixtures, type of bonding process (cold, warm, etc.) for both sensor and its protective coating, data processing time, electromagnetic interference, durability of equipment within its working environment for 30 years, compliance to typical Airbus specifications, maintainability of sensor, and robustness of sensor design to reduce maintenance needs. The sensor should also survive external mechanical interference, such as loads, vibrations, impacts, pressure, and temperature changes.

It is also important to install the sensor on the existing paint system at the surface of the structure, since removing it may affect corrosion protection.

For structural test monitoring, it is recommended to use surface-bonded sensors for easing its installation. In the case or A380 EF, it was compulsory not to modify the structural integrity (therefore, no embedded sensors).

3.3 Interrogation unit

The sensors must be connected either to the onboard evaluation unit or via a node (connector, router, etc.) to an external connector of an evaluation unit. Depending on the technology used, cables (electrical) or hoses (vacuum technique) would be installed. The main top-level requirements for the interrogation unit are similar to that of the sensor, with some exceptions. The list of requirements is however not exhaustive, as it highly depends on the application. Nevertheless, the goal of such a list is to support the selection of an interrogation unit for structural tests, including full-scale specimen monitoring. The interrogation unit should be small, light, cheap, durable, easy to install, reliable with a high MTBF, and compliant with specific application requirements. The interrogation unit should support the sensor in offering high accuracy and resolution information. It also combines a self-diagnostic tool determining whether a system component is malfunctioning or not functioning or if the sensor partly or completely disbonded from the structure.

4 SHM TECHNOLOGY DEVELOPMENT

Extensive investigations have to be carried out to demonstrate that the requirements placed on an SHM system have been met. The Airbus list of requirements follows the NASA US DoD TRL concept, where TRL6 signifies the end of technology development. At TRL6, the system is considered as almost the finished product, and two requirements dominate such a TRL: reliability and durability.

To ascertain reliability, a combination of both probability of detection (PoD) and pyramidal testing is used. The PoD is similar to that of classical NDT, based on various techniques such as the 90/95 method. The PoD investigation also provides what is the minimum damage size that the SHM system is qualified for. The pyramidal testing, shown in Figure 1, emphasizes first the need to carry out a large number of coupon tests in order to evaluate the performance and capabilities of the SHM system under investigation. The system is then evaluated on several subcomponent tests, generally on specimens around $5\,\text{m} \times 3\,\text{m}$. The system is then validated on a full-scale specimen, such as, in this case, the A380 EF. Implementing the system on in-service or flight test aircraft during ground testing can also further complement the validation. This test is mainly to emulate some environments that were not captured during previous mechanical tests, such as influence of vibration caused by the aircraft engine onto the SHM system functioning and performance. In all cases of reliability evaluation and validation, the SHM system must provide clear evidence of neither positive- nor negative-false alarms when monitoring a given damage.

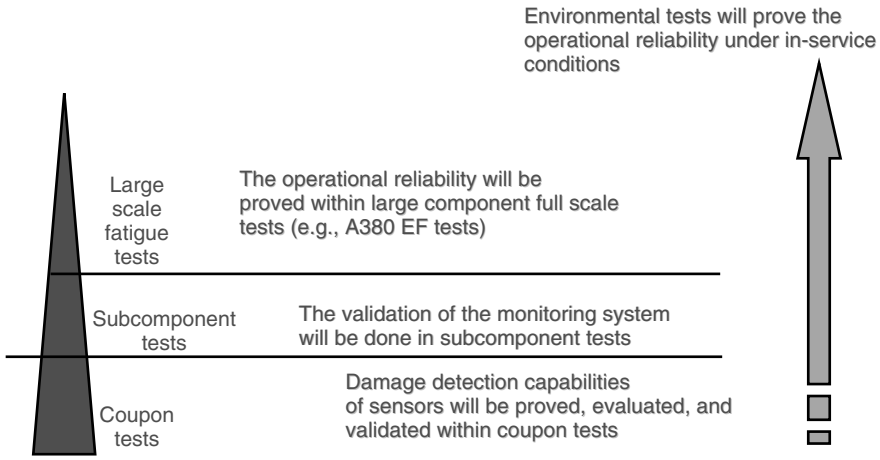


Figure 1. Technology development plan.

Parallel to the reliability testing, the durability of the SHM system must be characterized to ascertain whether the system (mainly the sensor in this case) is capable of coping with the harsh aircraft environments for a period of 30 years.

Universities and institutes are typically initiating such SHM systems, with maturing levels ranging from TRL3 to TRL5. Industries then adapt such technologies for their needs and requirements as well as commercialize the technology to reach TRL9. It is at that point where the user's requirements, in this case the aviation industry, come into play and enable the technology to be translated to the respective applications.

The A380 fatigue test specimen provides ideal conditions for testing the SHM sensors under realistic loading conditions on the ground. In parallel, the sensors are easy to access for data-acquisition and maintenance purposes. Today's objective for this kind of activity is not to monitor the A380 fatigue test with SHM (that is currently the role of NDT techniques), but the further development and gathering of experience for further optimization of the different sensor types applied to the specimen at this point in time.

Currently, the A380 specimen is equipped with a number SHM sensor types, each of them covering specific areas of the test structure (see Figure 2):

1. Comparative vacuum monitoring (CVM^{TM})
These sensors are mainly used in areas where quick access is possible. For the fatigue test, this means typically different locations along the fuselage, from both inside and outside the specimen.

2. Acoustic emission (AE) Vigilant
AE sensors have been applied to the wing sections of the specimen. The analysis of the AE data and the correlation of these data with fatigue finding along the specimen are currently under analysis by the experts.

3. Eddy-current foil sensors (ETFS)
ETF sensors are tested in the root joint area

4. Crack-wire sensors
Crack-wire sensors are well-known sensor types, which are used to follow crack growth of fatigue damage during the test performance phase. These crack wires are modified for airborne applications.

Some further SHM sensor types are to be installed during the remaining test performance phase. These are

5. Imaging ultrasonic (IU)

6. Acousto-ultrasonic (AU)

5 A380 EF DESCRIPTION

This section provides a general overview of the A380 fatigue test and a brief description of test requirements and features of the test setup.

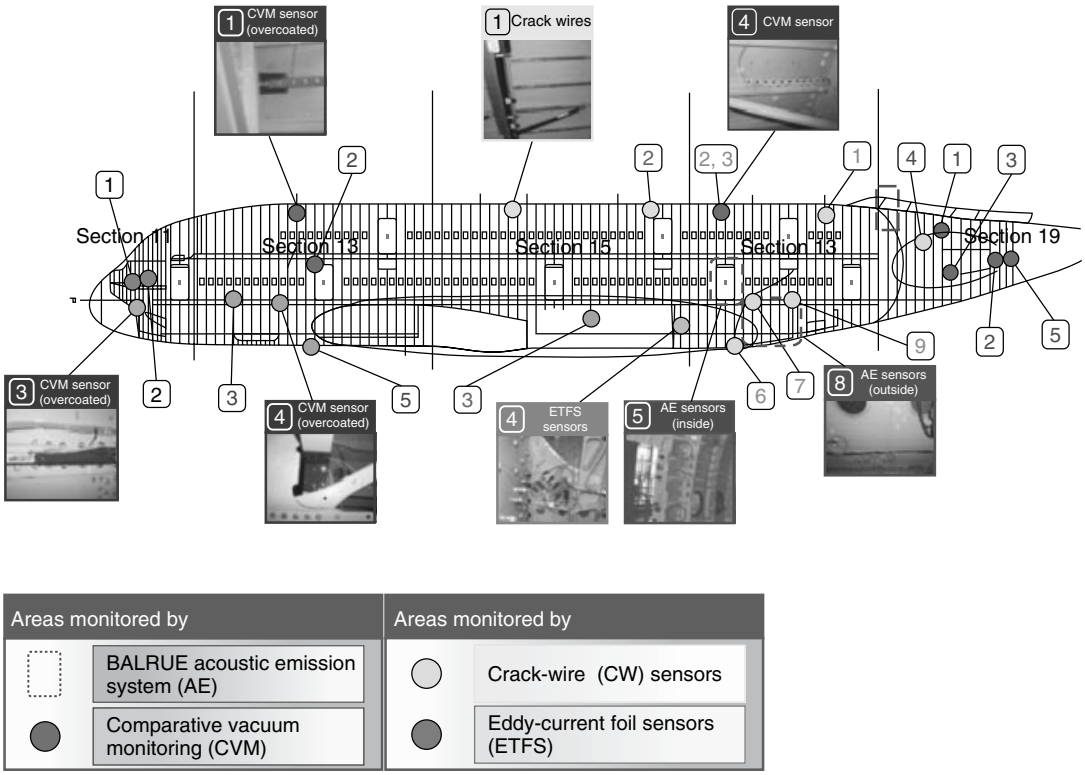


Figure 2. SHM sensor position on the A380 EF specimen.

For the type certification of any Airbus aircraft family, a large number of structural tests are carried out at various levels. On top of the variety of different component tests, Airbus is conducting three major structural tests:

- the static test of the entire aircraft structure (known as *A380 ES*);
- the full-scale fatigue test of the entire aircraft structure (the focus of this article, known as *A380 EF*); and
- a combined static and fatigue test of the rear fuselage and empennage (known as *A380 RET*).

The A380 EF ("Essai Fatigue"—the A380 full-scale fatigue test) is used as means of compliance to reach the so-called type certification of the A380 programme. The main objectives of this test are the validation of fatigue life and crack propagation of A380 primary metallic structure as well as the validation of structural repair manual (SRM) and maintenance programme. The A380 EF inspection results are also used for cross-checking and validating the simulations and structural design tools in terms of structural fatigue behavior.

The A380 fatigue test project first started in 2001 within Airbus. The test is currently performed by Airbus and its subcontractor IABG in Dresden, Germany, further shown in Figure 3. The A380 EF is used to prove the fatigue worthiness and the damage tolerance of the metallic aircraft primary structures. The design service life goal—the guaranteed service life—of the A380 family comprises either 25 years of operation or 19 000 flight cycles or 140 000 flight hours, whatever comes first. For a successful type certification, the A380 EF must prove that within this design service life goal no critical damage is caused by structural fatigue.

After successful accomplishment of the initial 5000 simulated flight cycles in December 2005, this contribution to the A380-type certification had been achieved.

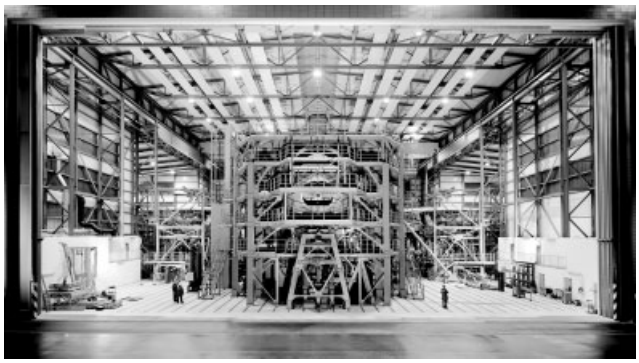


Figure 3. A380 EF test hangar and test setup.

The A380 EF test will be fatigued for a total of 47 500 simulated flight cycles for validating the structural design tools. These 47 500 simulated flight cycles represent 2.5 times the required 19 000 flight cycles of the design service life goal and thus allowing Airbus to analyze the initiation and propagation of natural cracks based on the fatigue behavior of the materials used. In addition to these natural damage, a number of artificial damages (cutting into shells, cutting of stringers or comparable stiffeners, removal of bolts, etc.) have been introduced to analyze the damage tolerance behavior of the dedicated structure.

The structural health of the A380 test specimen is inspected the same way as on in-service aircraft. Each individual inspection utilizes a range of techniques, such as visual and detailed visual inspection as well as NDI tools (e.g., eddy-current and ultrasonic systems, and in some specific cases X ray).

6 A380 EF TEST SETUP

The A380 EF test setup can be divided in different subsystems, such as the test specimen (representing all metallic primary structure), the dummy structures and the load introduction features (e.g., whiffle trees with load pads), test rigs, hydraulic loading system, pneumatic loading system, control and monitoring system (CMS), and data-acquisition system (DAS), each of them consisting of various subsystems.

The test specimen comprises all metallic primary structures for the entire fuselage and the wing as well as all other structural elements contributing to the global stiffness of the fuselage and wings, as shown in Figure 4. In some areas, secondary structure is also included (wing-fixed leading edge and trailing edge, as well as the belly fairing) as all those contribute to the overall stiffness. In the case of the empennage, dummy structures are representing the horizontal tail plane and tail cone. Additionally, a stub vertical tail plane torsion box is used as load introduction dummy to the empennage.

A number of dummy structures replace specific components of the aircraft for load introduction purposes. Namely, the engines and pylons, flap and slat tracks, and all landing gear legs are replaced by stiff steel structures allowing for hydraulic jack attachments and introducing the loads into the corresponding attachment fittings of the aircraft. The same is valid for the horizontal tail plane that is not a part of this test and that is represented by a dummy steel beam. All of these aircraft components represented by dummy structures are certified through separate component tests.

In addition to those load introduction structures representing aircraft components omitted for the A380 EF, further load introduction structures are used to distribute the loads to a large number of load introduction points along the fuselage and the wings. In most cases these additional load introduction structures are designed as whiffle trees, as shown in Figure 5.

A push–pull loading principle is used to minimize the required number of hydraulic jacks. Consequently, tension and compression loads are to be considered for the design of all whiffle trees.

The test rigs of the A380 EF test setup comprise in total more than 1.800 t of steel used for the fuselage, wing, and pylon rigs. The loading rigs are used for supporting the hydraulic actuators and their

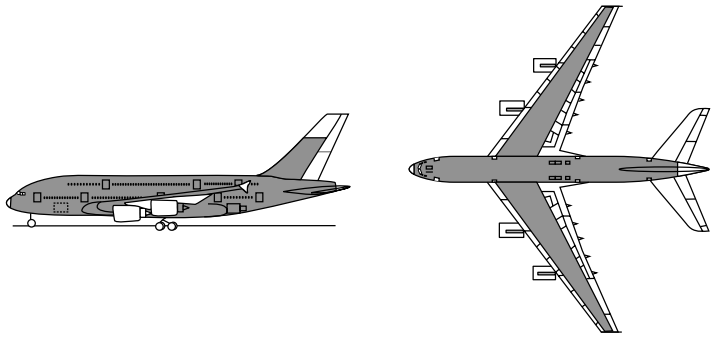


Figure 4. Overview A380 EF test specimen.

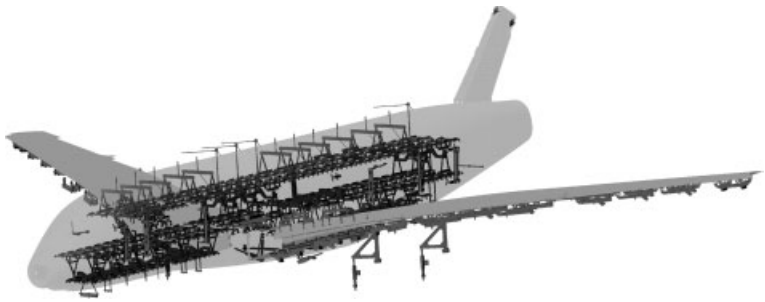


Figure 5. Specimen loading features.

counterweights as well as for transferring the reaction forces of these hydraulic actuators into the strong floor of the test hangar.

The inspection rigs provide access to the specimen for the inspection staff. The fuselage rig, however, is the only rig that combines the inspection and loading rigs, as shown in Figure 6.

The hydraulic loading system consists of the hydraulic power supply, the hydraulic piping, the hydraulic accumulators, and the hydraulic actuators, as shown in Figure 7.

Twelve hydraulic pump modules provide in total a hydraulic oil supply of up to 6000 l min^{-1}, to be controlled via the two main hydraulic valve manifolds and to be distributed via the hydraulic main piping system to the 182 hydraulic actuators mechanically loading the specimen. At specified positions near those hydraulic actuators with the longest stroke, a system of hydraulic accumulators has been introduced to provide extra power for most severe peak phases.

The hydraulic actuators provide a stroke of up to 6.5 m and loads of up to 150 t. They are equipped with double channel load cells and control valve blocks controlling the activated loading via two independent short-circuit control loops independent from the CMS for the A380 EF test setup.

The pneumatic loading system is powered by three compressor units feeding three pneumatic accumulators with an air pressure of 7.5 bars. For each individual flight, during the flight-by-flight simulation, the differential air pressure is provided through a pneumatic piping system including acoustic dampers via 16 dummy windows into the specimen, as shown in Figure 8. A pressure level of 1.67 bar is generated within the 2200 m^3 volume of all three decks (passenger upper and main deck as well as cargo deck) within 45 s.

The CMS is displaying the overall test information provided by the decentralized architecture based on a number of field cabinets being distributed all over the facilities. This CMS allows a high-speed test cycling at uncompromised loading accuracy.

These field cabinets, in addition, carry subsystems for the A380 EF DAS, which allows measurement campaigns of up to 7000 measurement channels, collecting data via data-bus lines.

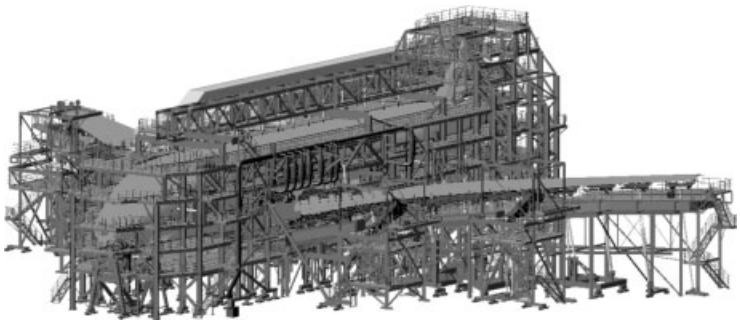


Figure 6. Test and inspection rig.

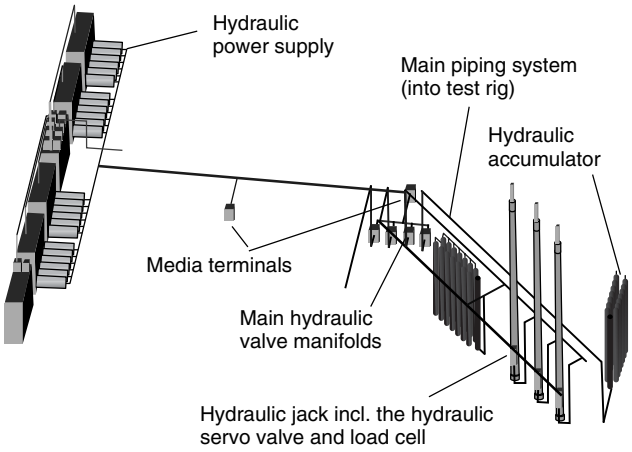


Figure 7. Hydraulic loading system.

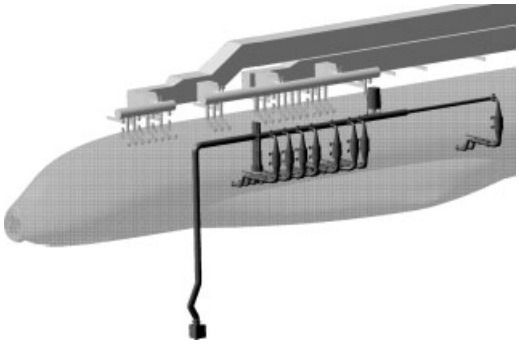


Figure 8. Overview pneumatic loading system.

Both static and dynamic measurement campaigns can be run automatically. In addition, a video system, controlled by the CMS, can be used for automatic visual inspection and recording for specific items, see Figure 9.

7 MECHANICAL TEST RESULTS

The A380 EF fatigue cycling had started on September 1, 2005. The contribution to the A380-type certification process had successfully been accomplished with the 5.000th simulated flight conducted on December 24, 2005.

Since that initial fatigue cycling phase, the remaining 42.500 simulated flights are being performed at a test speed of about 1000 simulated flights per week during test run phases. This is achieved in a three-shift 24/7 test run scheme. These test run phases are interrupted for carrying out scheduled inspections combined with specimen modifications, which are required either for keeping the test specimen at the latest serial production standard or for introducing structural technology innovations for test validation in preparation for future Airbus aircraft programmes.

Figure 9. Operator view of the A380 EF control and monitoring system.

8 SUMMARY

The article has presented and described the A380 EF specimen, with its extensive infrastructure to provide mechanical loading, fuselage pressure, maintenance, and structural inspection access. The article then discussed the benefits of the SHM technology for both in-service and structural test monitoring. The SHM technology requirements were provided to support possible technology down-selection for structural test monitoring as well as in-service applications. An SHM technology development was herein described. The SHM sensors were successfully implemented on the A380 EF and further results are expected at the completion of the A380 EF test campaign.

REFERENCES

[1] Boller C, Buderath M, Speckmann H. Measures for assessing structure-integrated damage monitoring systems in aircraft. *Proceedings of the SHM—First European Workshop*. Paris, 10–12 July 2002.

[2] Boller C, Staszewski W.J, Chang F.-K, Ihn J.-B, Speckmann H. Smart systems for in-service crack monitoring of aircraft components. *Proceedings of the Third International Workshop on Structural Health Monitoring*. Stanford, CA, 12–14 September 2001.

[3] Chang F.-K. Manufacturing and design of built-in diagnostics for composite structures. *52nd Meeting of the Society for Machinery failure prevention Technology*. Virginia Beach, VA, 30 March—3 April, 1998.

[4] Giurgiutiu V, Redmond J, Roach D, Rackow K. Active sensors for health monitoring of aging aerospace structures. *SPIE Conference on Smart Structures and Integrated Systems*. Newport Beach, CA, March 2000.

[5] Lemistre M, Gouyoun R, Kaczmarek H, Balageas B. Damage localization in composite plates using wavelet transform processing on lamb wave signals. *2nd International Workshop of Structural Health Monitoring*. Stanford University: San Francisco, CA, 8–10 September, 1999.

[6] Rock C. *Installation of Comparative Vacuum Monitoring (CVM) Technology Using the LI-50 Linear Indexer Kit*. Airbus Deutschland GmbH: Hamburg, Internal Memorandum – March 2002.

[7] Schmidt H.-J, Schmidt-Brandecker B. Structural design and maintenance benefits from health monitoring systems. *Proceedings of the Third International Workshop on Structural Health Monitoring*. Stanford, CA, 12–14 September 2001.

[8] Schmidt H.-J, Schmidt-Brandecker B. Management of aging civil aircraft—the challenge of the aerospace industry. *Proceedings of the Eighth International Fatigue Congress (Fatigue 2002)*. Stockholm, 2–7 June 2002.

[9] Ihn J.-B, Chang F.-K, Speckmann H. Built-in diagnostics for monitoring crack growth in aircraft structures. *Proceedings of the 4th International Conference on Damage Assessment of Structures (DAMAS)*, *Key Engineering Materials, Vols. 204-205*. Trans tech Publisher: Cardiff/Wales, 2001; pp. 299–308.

[10] Speckmann H. Structural health monitoring with smart sensors approach to a new NDI method. *Proceedings of the SPIE Conference on Smart Structures and Materials and NDE for Health Monitoring and Diagnostics*. San Diego, CA, 17–21 March 2002.

[11] Schwarberg F, Eichelbaum F. *An Efficient Load Introduction Concept for the A380 Full Scale Fatigue Test*. ICAF, 2005.

[12] Krichel T. A380 major fatigue test—program & progress. *3rd Dresden Airport Seminar Structural Health Monitoring*. Dresden, 07 Nov 2007.

[13] Airbus non-destructive testing manual NTM 51-10-08.

[14] Airbus non-destructive testing manual NTM 51-10-13.

[15] Military standard MIL-STD-781D, *Reliability Testing for Engineering Development, Qualification and Production*, DoD, Washington, DC, October 1986.

Chapter 99

Flight Demonstration of a SHM System on a USAF Fighter Airplane

Matthew C. Malkin
Phantom Works, Boeing, Seattle, WA, USA

1 INTRODUCTION

The F-16 station 341 fuselage bulkhead is susceptible to fatigue crack growth due to a maintenance-induced crack initiation site. The bulkhead is a single-piece machined structure, and replacement of a damaged bulkhead is time consuming and costly. Adhesively bonded repairs were developed by South West Research Institute and applied to the structure as an alternative to replacing damaged bulkheads. A health monitoring system was developed in order to assess the health of the repaired structure.

Health monitoring systems for composite bonded repairs have been applied to structures in laboratory environments [1–3]. Throughout 2004 and 2005, the United States Air Force Research Laboratory and Boeing have performed increasingly complex laboratory tests of structural health monitoring (SHM) systems for bonded repairs. Basic coupon testing has been performed, followed by component-level testing, building up knowledge and experience that led to the flight demonstration on an in-service F-16 airplane, and this is discussed here [4].

This article attempts to capture the lessons learned from the experience of overcoming the difficulties encountered in developing and installing a laboratory-level SHM system on a complex, in-service structure. It is hoped that this knowledge will smoothen technology transitions in the future.

2 STRUCTURAL HEALTH MONITORING SYSTEM OVERVIEW

Two structural failure modes of the repaired bulkhead, crack growth under the repair and disbonding of the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

repair from the structure, motivate the use of SHM for this application. The SHM system for this application responds to structural changes that occur due to either failure mode.

A piezoelectric-based SHM system was selected for this application. A network of piezoelectric transducers was permanently attached to the structure near the repair. To assess the health of the repaired structure, an electrical signal is sent to the piezoelectric transducers. The transducers convert this signal into an ultrasonic strain wave that travels through the structure and is detected by other piezoelectric transducers in the system. These transducers convert the signal into a voltage, which is read by the data-acquisition system.

A data-acquisition computer directs data collection. The computer is connected to the transducers with an electrical cable. The data-acquisition computer contains a signal generator, an actuation signal amplifier, the switching hardware, a sensor signal amplifier, a data-acquisition board, and the control software.

When damage is present in the structure, the transmitted strain waves change causing changes in the detected electrical signals. These changes are interpreted to give an indication of the damage.

3 SENSOR LAYER DESIGN AND MOCK-UP

One principal difference between SHM and conventional nondestructive inspection is that the sensing portion of the SHM system is permanently attached to the airplane. The transducers for this application are piezoelectric disks, each of a diameter of 6.35 mm and a thickness of 0.25 mm, permanently mounted on a polyimide substrate material. Wiring is printed on the substrate, and a connector is integrated for electrical access to the sensors.

Structure in the keel area of the F-16 station 341 bulkhead is complex, with material interfaces, fastened joints, and system penetrations. Figure 1 shows two views of the area in the main landing gear wheel well of an F-16 airplane.

Designing the sensor involved locating the transducers, routing the flat wiring, and properly positioning the data cable connector on the airplane. A paper mock-up of the sensor layer was created on an actual airplane. Mocking up the sensor on an actual airplane maximized chances that the installation fits properly and does not interfere with any airplane systems.

An iterative, design-analysis process was not followed for designing the transducer network. Layout of the transducers and wiring was performed by examining the structure and placing the transducers in locations that covered the repairs with sensor–actuator paths, while remaining within the constraints imposed by the geometry of the airplane structure. The number of transducers was limited by the number of channels available on the data-acquisition computer. A paper mock-up of the transducer layer was created. After the paper mock-up was complete, a sensor layer was fabricated to match the paper mock-up. This design process, as is any design process, was based on the experience and knowledge of the engineers performing the design. The analysis and simulation of SHM system designs would allow for iterative design and improved system performance.

4 INSTALLATION

The demonstration sensor was installed on aircraft tail number 87-0395 at Hill Air Force Base (Ogden,

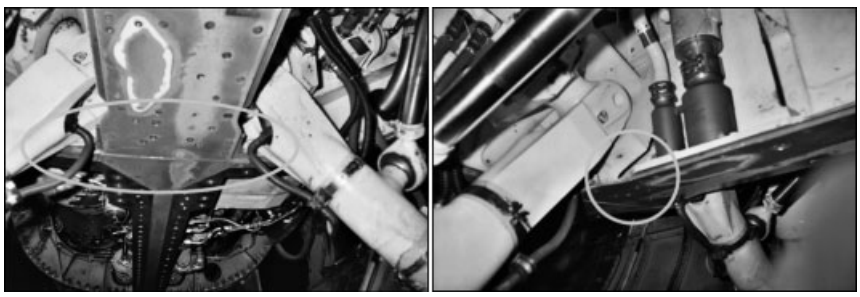


Figure 1. The main landing gear wheel well of an F-16 contains complex structure and multiple system installations.

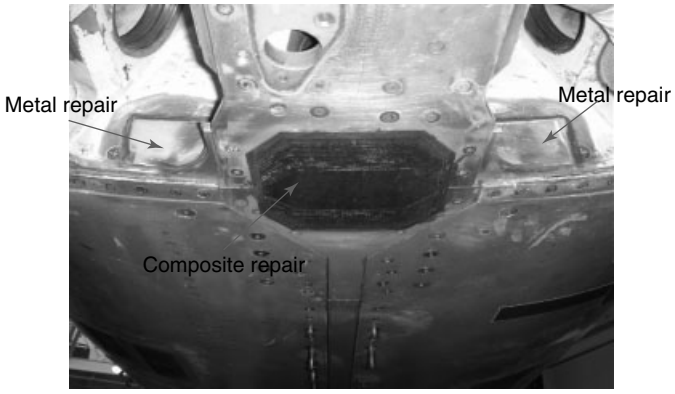


Figure 2. Three repairs, two metal and one composite, were installed in the keel area of the station 341 bulkhead.

Utah, USA) in February 2006. The repairs, pictured in Figure 2, were installed by Southwest Research Institute, and thermography was performed to validate the initial installation of the repairs.

After the repair installation was completed, the SHM sensor installation began. The first step in the sensor installation was to lightly abrade the airplane surface with sandpaper and clean the surface with isopropyl alcohol. Epoxy adhesive was mixed and applied to the sensor layer and structure. As shown in Figure 3, the sensor was adhered to the structure and fixed temporarily with clamps and tape.

Neither heat nor vacuum pressure was applied to the sensor during cure. After cure, the tape and clamps were removed, the region outside the sensor was masked off, and an overcoat of epoxy adhesive (Figure 4) was brushed on the exterior of the sensor layer. This is the first application of a sensor of this type to the exterior surface of a supersonic airplane.

Figure 3. The sensor was carefully wrapped around the repairs, positioned, and adhered to the structure.

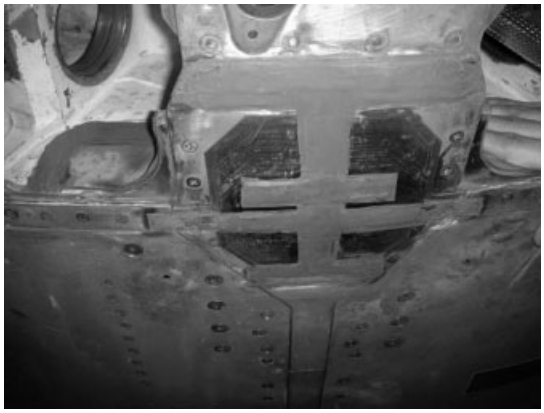


Figure 4. An overcoat of adhesive was applied over the sensor layer.

Even with a careful mock-up during sensor design, the clearance between the landing gear door and the portion of the sensor layer that wraps around the keel structure from the interior of the wheel well to the exterior of the airplane could not be verified. After the repair, the sensor and the landing gear were installed, and a gear swing was conducted as part of a normal maintenance. When the landing gear doors closed, there was approximately 3.18 mm (0.125 inches). clearance between the sensor layer and the inboard edge of the landing gear door.

The connection between the on-aircraft sensor and the off-aircraft data-acquisition system was carefully designed to ease access and ensure reliability. Easy access to the electrical connector for the sensor helps

to minimize the data collection time. A 90° backshell was installed on the connector on the data cable in order to work around clearance problems with wiring and adjacent structure in the wheel well. A self-locking protective cap fits over the airplane-mounted connector when the mating plug on the data cable is not in place. Wiring interfaces in SHM systems are critical and should be carefully engineered.

With the sensor layer installed, data collection began. Initial baseline data sets were taken after the adhesive cure cycle was complete. After the airplane completed a postmaintenance test flight, it was painted. Data was collected from the SHM sensor before and after paint was applied. The sensor was ready for service.

5 SERVICE EXPERIENCE AND THE DATA-ACQUISITION SYSTEM

After maintenance on airplane 87-0395 was completed at Hill Air Force Base, the airplane entered service at Luke Air Force Base outside of Phoenix, Arizona, USA. The SHM data-acquisition system was brought to Luke Air Force Base (AFB) for use by United States Air Force (USAF) personnel. The data-acquisition equipment is roughly the size of a small desktop computer. In order to make the data-acquisition system as transportable and easy to use as possible, a self-contained system was assembled. The system, shown in Figure 5, includes a battery backup unit, a monitor, a keyboard and mouse, a data-acquisition computer, and a data cable. All components were mounted inside a wheeled case with a retractable handle. The system can be rolled out to the airplane and data can be collected without a need for external power.

6 SENSOR DURABILITY

After approximately two months in service, it was discovered that two of the sensors had been torn off of the airplane. The reason for the departure of these sensors is not known. Possible causes for the loss of the sensors are poor initial surface preparation, strain at the adhesive bondline due to the sensor wiring

Figure 5. A self-contained data-acquisition system was developed for collecting data while the airplane is in service.

bridging the interface between two structural components, aerodynamic loads on the sensor, interference between the sensor and the center fuel tank pylon, or interference caused by other maintenance actions. These two sensors were repaired in the field on June 10, 2006.

The wheel well of an F-16 is a harsh environment for a fragile piezoelectric ceramic sensor and thin copper traces printed on a polyimide substrate that can easily tear. Testing is useful for determining if sensing systems can survive the service environment. Tests based on MIL-STD-810, a US government standard defining a variety of environmental tests for flight vehicle components, were conducted in 2006 to assess the capability of these sensors to withstand the wheel well environment. The testing showed that for test specimens with well-bonded sensor layers, the layers could adequately withstand the expected environment.

After several more months in service, the left side repaired sensor began showing signs of disbonding. It is possible that a good bond was not obtained during the repair because of the difficulties with in-service surface preparation, which explains the current disbonding of the sensor. As SHM sensing elements are mounted on structures, the sensing systems must be maintainable, and effective maintenance practices must be developed for the sensing systems.

As the system continued in use, additional anomalies were discovered. The right-side metal patch transducer wiring broke and is not repairable without removing the landing gear. It is not known how the wiring failed. One contributing factor may be that when the sensor was originally installed, some of the sensor wiring did not bond to the structure. For sensors on a flexible substrate, a good bond to the host structure is critical for system longevity.

The copper traces in the layer are slightly exposed where the sensor layer wiring wraps around the edge of the keel structure on the left side of the airplane. The effect of the exposed traces on transducer signals is unknown. An optimum configuration would allow for adequate clearance and protection to keep wiring traces covered.

This in-service wear of the system provides information about the effects of the environment in which SHM systems need to perform. By performing this demonstration, it has been shown that sensors can be flown, and that data can be collected by Air Force Personnel in the field. Areas have been found in which further development is needed before a production system is fieldable, at least for an environment like a landing gear enclosure.

7 TECHNOLOGY NEEDS

This demonstration unearthed a variety of technology needs before practical, wide-scale implementation of *in situ* sensors. To start, SHM system design methods need study and advancement. The system flying on the F-16 was not designed with any analysis or modeling. A design method for SHM systems that includes a requirement-based iterative design process with system-level simulation is needed.

Algorithm development is required to correlate signals received at transducers with the physical phenomena occurring within the structure. The objective of the damage evaluation portion of a SHM system is to localize and size damage. Localization means providing the position of damage, and sizing is to provide the extent (dimensions and shape) of damage. Damage evaluation must be possible with the structure in the environment (electromagnetic noise, dirt, temperature, and fluids) appropriate to that sensor installation. Algorithms to provide this capability are under development for *in situ* sensor-based SHM systems.

Miniaturization and strengthening of data-acquisition equipment are needed, and improvement of the user interface is required. For a system to operate in the field by personnel not familiar with SHM system principles, a lightweight, simple-to-operate system that meets field operation requirements is required. In the not-too-distant future, systems with energy harvesters, onboard data processing, and wireless data links to ground systems will replace bulky equipment and cumbersome connectors. Minimizing the skill level and difficulty involved in sensor installation, system maintenance, and system operation will minimize the recurring cost of production installations.

Further development of a SHM-based maintenance philosophy is needed. The US Air Force is gaining experience with *in situ* structural sensors on fixed wing aircraft. Data paths for SHM information need to be established, including feedback to allow modification of the usage of the airplane. The calculation of residual strength and remaining life could modify maintenance actions. Existing rotorcraft health and usage monitoring systems and health management methods of the systems are leading the way. SHM needs to fit into the US Air Force structural maintenance plan.

8 CONCLUSION

The installation of an SHM sensor system on a bonded repair of an F-16 airplane has been demonstrated. With the successful installation and operation of the sensor, experience with fielding a laboratory-level SHM system was gained. Technology needs related to SHM are great—improvements in design methods, algorithms, data-acquisition equipment, and the maintenance philosophy will all improve SHM-based maintenance practices in the future.

ACKNOWLEDGMENTS

The efforts of many people enabled this work to happen. The names of those who contributed are as follows: Dave Wieland, Bob Wallace, Eric Haugse, Ken Hunziker, Mark Derriso, Matt Leonard, Kevin Brown, Emilio Talipan, Nathan Taylor, David Day, Jean Rojas, Shone Topham, Peter Qing, and Shawn Beard; many more are not mentioned.

REFERENCES

[1] Dennis RP. Results from Fed-Ex pilot program to assess durability and health monitoring of bonded composite doubler repairs. *Presented to the 2005 Air Transport Association Nondestructive Testing Forum*, Orlando, FL, 2005.

[2] Baker AA, Galea SC, Powlesland IG. A smart patch approach for bonded composite repairs to primary airframe structures. *Proceedings of the 2nd Joint NASA/FAA/DoD Conference on Aging Aircraft*. Williamsburg, VA, August 31–September 3 1998; pp. 328–338.

[3] Chiu WK, Galea SC, Koss LL, Rajic N. Damage detection in bonded repairs using piezoceramics. *Smart Materials and Structures* 2000 **9**:466–475.

[4] Malkin M. Structural health monitoring for bonded repairs: complex structure testing. *Proceedings of the 5th Structural Health Monitoring International Workshop*. Stanford University, September 2005; pp. 470–477.

Chapter 108

Aerospace Applications of SMART Layer Technology

Xinlin P. Qing[1], Shawn J. Beard[1], Roy Ikegami[1], Fu-Kuo Chang[2] and Christian Boller[3]

[1] *Acellent Technologies, Inc., Sunnyvale, CA, USA*
[2] *Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA*
[3] *Saarland University & Fraunhofer Institute for Non-Destructive Testing, Saarbrücken, Germany (and formerly of The University of Sheffield, Sheffield, UK)*

1 INTRODUCTION

Structural health monitoring (SHM) technology is perceived as a revolutionary method of determining the integrity of structures involving the use of multidisciplinary fields including sensors, materials, signal processing, system integration, and signal interpretation. The aim of the technology is not simply to detect structural failure but also to provide an early indication of physical damage. The early warning provided by an SHM system can then be used to define remedial strategies before the structural damage leads to failure [1–3].

Recent advances in sensor technology, material processing, damage modeling, and system integration have enabled new developments in structural evaluation and inspection technologies to overcome the shortcomings of existing inspection systems. An SHM system consists of three major components: a sensor network, integrated hardware, and diagnostic software to monitor *in situ* the condition of in-service structures. Among several types of sensors that can be used for SHM, piezoelectric sensors are being widely used because they can be used as either active or passive sensors due to their piezoelectric effect (*see* **Piezoelectric Wafer Active Sensors**; **Piezoelectric Paint Sensors for Ultrasonics-based Damage Detection**).

Unique SHM technologies have been developed [4–6] through the use of built-in distributed piezoelectric sensor networks integrated with composite and metal structures. As shown in Figure 1, the basic idea of the technology is to use a network of

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

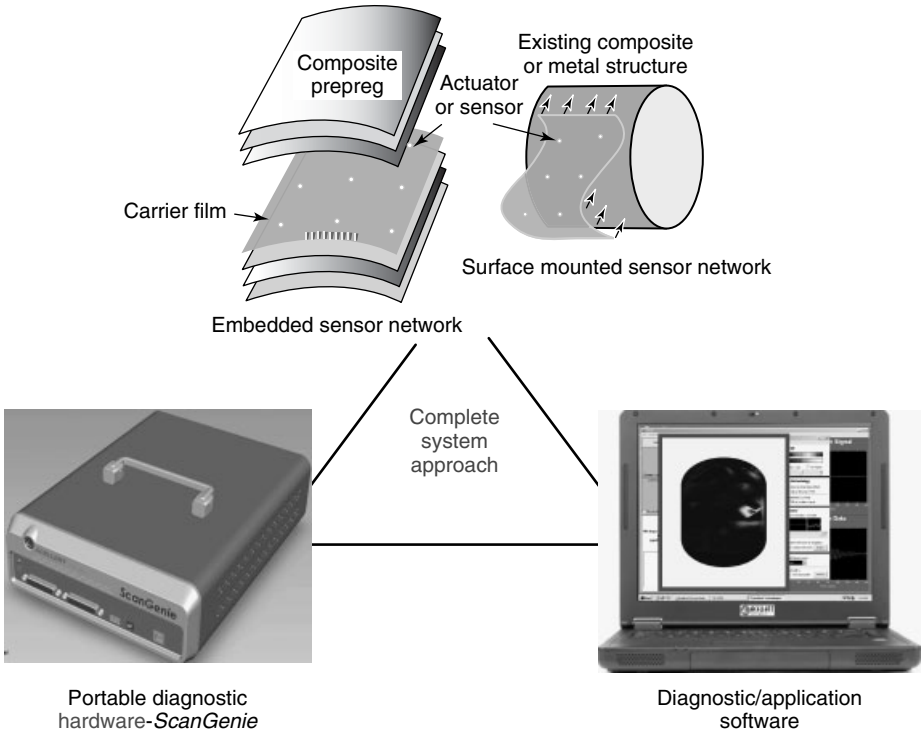


Figure 1. Complete system approach.

distributed piezoelectric sensors/actuators embedded in a thin dielectric carrier film called the *SMART Layer* to monitor and evaluate the integrity of a structure. A portable diagnostic hardware unit called *ScanGenie* is used to collect and process diagnostic signals obtained by the SMART Layer during the monitoring process. The signals obtained can then be analyzed to determine the integrity of the structure.

The SMART Layer technology has the unique ability to provide wide structural coverage for gathering diagnostic data with a network of transducers embedded in a layer, thus eliminating the need for each transducer to be installed individually (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**). There are two types of SHM systems defined by their functionalities: the active-sensing system and the passive-sensing system [7–10]. The SMART Layer can work with both the active-sensing system and the passive-sensing system.

In terms of aerospace applications, there are two operation modes of SHM systems: onboard and off-board modes. The onboard mode considers the system to record data during flight, while off-board mode only considers data to be taken when the aircraft is on the ground. The former is therefore the more challenging approach, since it requires the complete monitoring system (signal and power generator, sensors, signal acquisition, writing, and processing unit) to be placed onboard the aircraft, and, as a result, it requires certification of a very high degree of airworthiness, since the system is due to operate in flight. Off-board monitoring can be considered more as a substitution of conventional NDT methods where only the transducers and part of the wiring (or the antenna system in the case of wireless monitoring) have to stay on board the aircraft.

Both operation modes have advantages, since each of them very much depends on the application. A passive monitoring system, such as acoustic emission, can only be run on board. The same applies for impact detection or any loads to be monitored and occurring in service. Off-board monitoring is useful when a structure is allowed to tolerate damage, such

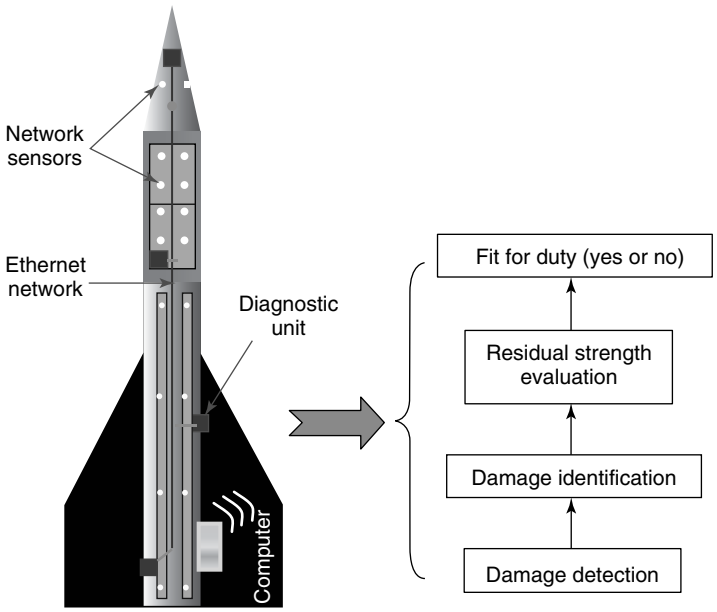


Figure 2. Overview of an onboard SHM system.

as when it is applied to a large number of metallic structures with regard to fatigue and fracture in a damage-tolerant design. An example for an onboard system is provided in Figure 2, while an off-board system is shown in Figure 3.

Within this article, a variety of applications are presented for both onboard and off-board usage to demonstrate the success of the systems in the real world. The applications include monitoring damage in large composite structures, cracks in the pipes of liquid rocket engines, multisite damage in riveted joints, bondlines in composite repairs, detection of impacts on the thermal protection systems (TPS) of space vehicles, and monitoring composite rocket motor cases. Further background regarding SMART Layer technology can be found in **Stanford Multi-actuator–Receiver Transduction (SMART) Layer Technology and Its Applications** and partially also in **Hybrid PZT/FBG Sensor System**.

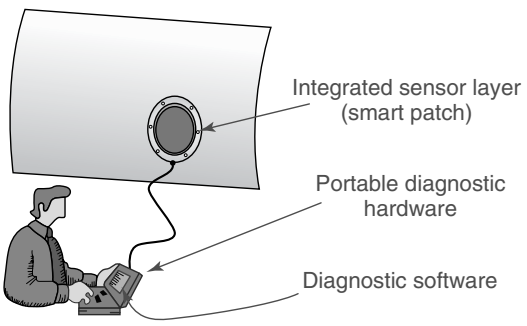


Figure 3. Overview of an off-board SHM system.

By using a passive onboard system, parameters such as impact time, location, and energy can be monitored. When the structure is struck by a sufficiently large force due to an external mass, the transducers (such as piezoelectric materials) that are bonded on the structure pick up the stress waves traveling through the structure. Upon triggering, all sensor measurements are recorded and stored in a data file for analysis. The data are immediately passed to a signal-processing and interpretation algorithm, which analyzes the measurements to estimate the location and also energy and/or peak force level of the impact.

Key issues for the passive onboard system include correct triggering, exclusion of electromagnetic interference, large hardware memory space; guaranteed continuous operation; and real-time monitoring. Simulation models need to be used to predict the possible damages on the basis of the impact location and energy. Similar issues may apply if the system is used for active onboard monitoring. The

diagnostic units integrated with the hosting structure will initiate, upon request, a damage scan of the area covered by the sensor network, collect and condition the data coming from the sensors, and send them through the Ethernet structure network to diagnostic software hosted on a computer for analysis. Key features, currently developed for an active onboard system, include miniaturized lightweight hardware, self-diagnostics and adaptive algorithms to automatically compensate for damaged sensors, reliable damage detection under different environmental conditions, and generation of probability of detection (POD) curves [11].

2 APPLICATIONS OF ONBOARD SHM SYSTEM

Some developments for potential applications of onboard SHM systems, including both active sensing and passive sensing, are summarized in this section.

2.1 Monitoring of impact damage in a composite pressure vessel

Fiber-reinforced composite materials are widely used in aerospace, automotive, shipbuilding, and other industries because of their high strength-to-weight and stiffness-to-weight ratios. However, composite materials, specifically those based on carbon fibers, lack plasticity when compared with metals. Whenever an accidental damage occurs, it is fairly difficult to determine the extent of damage without employing a fairly sophisticated NDT approach. However, it is important to have up-to-date information on the integrity of the composite structure. This is especially true for composite rocket motor cases because of the cost and liability associated with each launch failure. To ensure the safety and reliability of rocket components, they require frequent inspections to identify and quantify damage that might have occurred during manufacturing, transportation, or storage [12].

An integrated SHM system using a built-in sensor network for detecting impact damage in filament-wound composite structures such as storage tanks for space vehicles has been developed [13] and has been described from the sensing point of view in **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**. A prototype of a filament-wound composite bottle with embedded SMART Layers was fabricated successfully at NASA Marshall Space and Flight Center is shown in Figure 4.

Once a set of baseline data has been collected from the sensor network on the bottle, impact damage is introduced onto the bottle to look at its effect on the sensor signals. Another set of data for all paths is obtained after introduction of the impact damage. The scatter signals for all possible actuator–sensor diagnostic paths on the bottle can then be processed and displayed at the same time to produce a comprehensive image that illustrates very clearly the location and size of the damage, as shown in Figure 5.

2.2 Monitoring of impact damage in a large composite fuselage barrel

The issues discussed with the composite pressure vessel above also apply to composite aircraft fuselages, which are increasingly being designed to be built out of carbon fiber-reinforced polymers these days. As shown in Figure 6, a large composite fuselage barrel was fabricated as a proof-of-concept demonstrator for a new airplane development project at the Boeing Company, in 2001. This composite

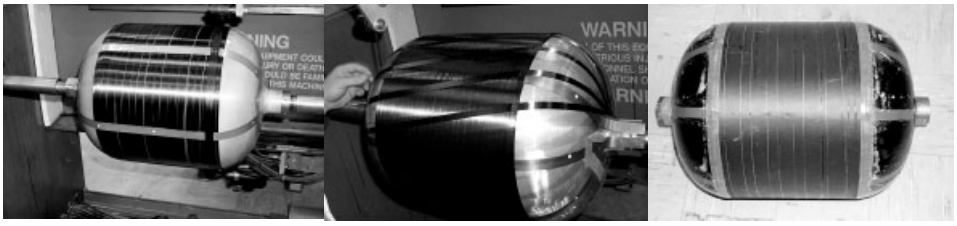


Figure 4. Filament-winding process incorporating the SMART Layers into the composite bottle.

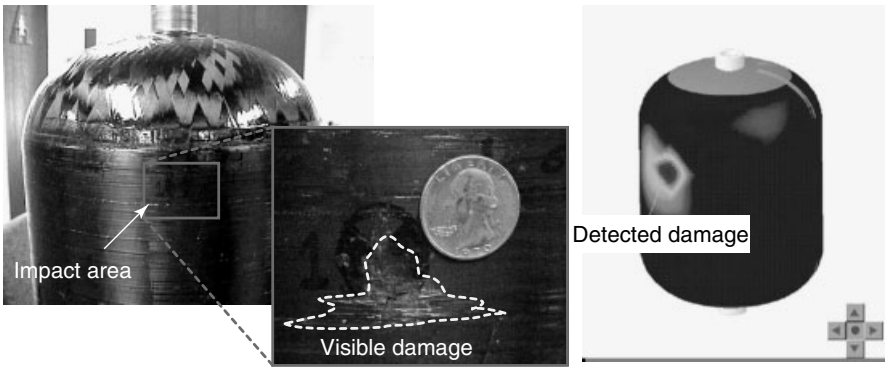


Figure 5. The damage identification result using signals collected by a network of built-in sensors.

structure provided a unique and realistic opportunity to investigate the use of SHM systems on a test specimen that represents a full-scale flight vehicle structure. Three big SMART Layers, of the type shown in Figure 7, were successfully installed and operated on this large-scale structure. The test demonstrated the feasibility of the sensor network-based active SHM system [14].

2.3 Detection of impact damage on a full-scale UCAV composite wing

The unmanned combat air vehicle (UCAV) is an affordable weapon system that increases tactical mission options for revolutionary new air power. Because of some significant advantages of composites, UCAVs, such as the X-45 UCAV, include a significant amount of composite materials. To ensure the safety and reliability of the UCAVs, it is very important to have up-to-date information on the integrity of composite structures; this is because, in wartime, structures are likely to be affected by battle damage, which is a fairly unpredictable damage condition that might have to be handled by a ground crew. Further issues in that regard are discussed to a wider extent in **Unmanned Aerial Vehicles**.

Figure 6. A large composite fuselage barrel.

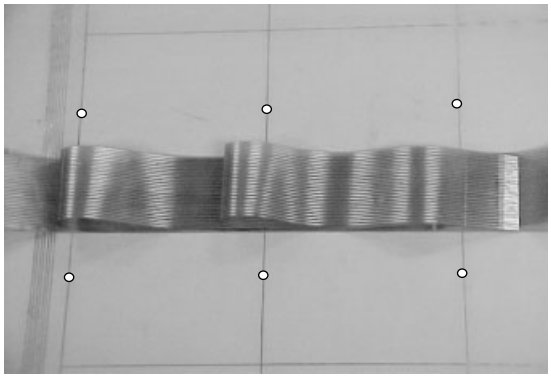


Figure 7. A 609 mm × 914 mm SMART Layer with 24 PZTs embedded.

The feasibility of using the SMART-Layer-technolog-based SHM system to detect impact and delamination damage in a full-scale UCAV composite wing has been evaluated at the Boeing Company in Seattle, Washington, in 2006. Twenty-one lead zirconate titanate (PZT) elements on SMART Layers were mounted on the upper skin of CAI-T4 wing to detect impacts on the skin, while another 40 PZT elements on eight SMART Layers were mounted on the skin around four paste-bonded joints to monitor the impact damages and any growth during fatigue testing. From these test results, it is clear that

the SMART-Layer-technology-based SHM system can successfully detect impacts and monitor impact damages in the UCAV composite wing under real environmental conditions. The basic concept of an SHM system for UCAVs is shown in Figure 8.

Figure 9 shows another example of the composite wing of a vehicle that has been instrumented with the SMART Layers for structural monitoring. The SMART Layers were designed and integrated into the wing during the manufacturing process itself, and used to detect changes in the structure when subjected to impact load. The wing was also shot at with a rifle [15].

2.4 Monitoring of cracks in liquid rocket engine pipes

Pipes of pressure vessels in liquid rocket engines can fail due to many causes including pitting, stress corrosion cracks, seam weld cracks, and dents resulting from internal or external impacts. The detection of defects and the monitoring of their growth is important, as it will address the need for safe as well as reliable advanced space exploration vehicle/propulsion systems. An active SMART Tape system for monitoring crack growth in a liquid rocket engine pipe has been developed and tested; this has been further described in **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**. Test results demonstrated that the system can detect a surface crack as small as 4 mm and a through-wall crack as small as 2 mm in the high-pressure engine pipe made of Alloy 718 in laboratory conditions [6]. It was also demonstrated that the developed system can withstand operational levels of vibration and shock energy on a representative rocket engine duct assembly, and is perfectly functional under the combined cryogenic temperature and vibration environment [16]. The setup for the combined cryogenic temperature and vibration test is shown in Figure 10.

Besides the applications mentioned above, the onboard active SHM system can also be used to

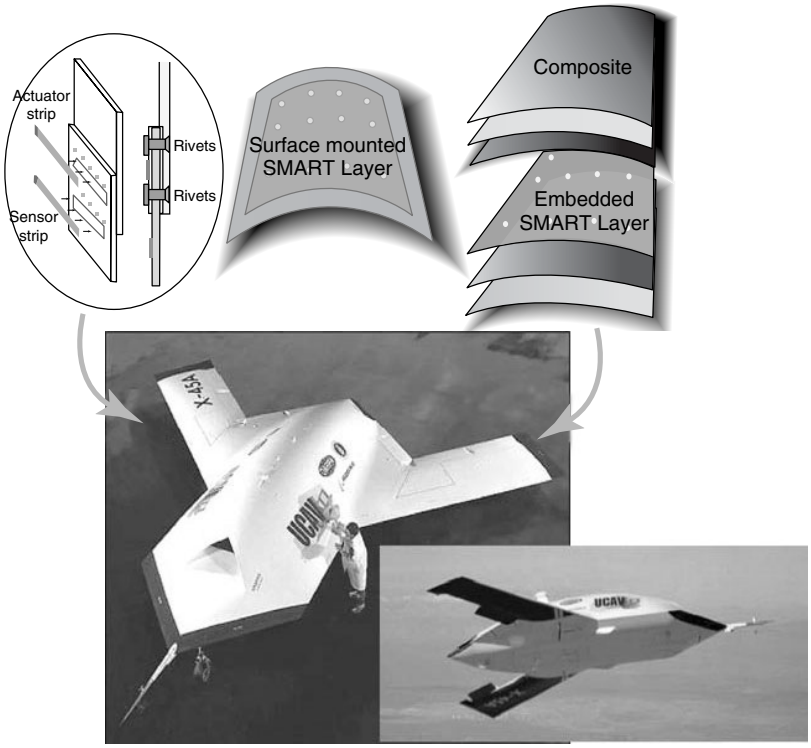


Figure 8. The concept of an SHM system for UCAV.

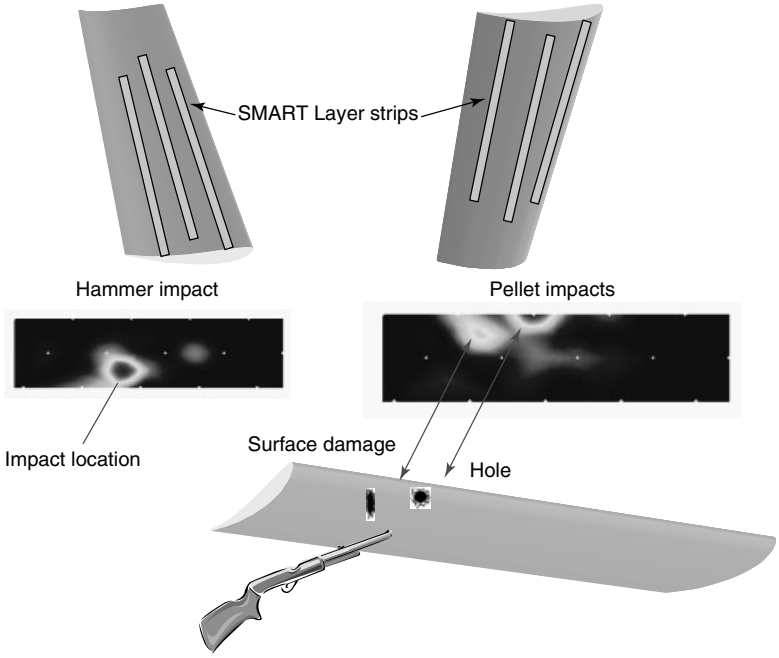


Figure 9. Monitoring the integrity of a composite wing.

monitor the integrity of structures in many other applications in both aerospace and civil engineering [5, 15, 17].

2.5 Monitoring of impacts on TPS panels

The TPS is an important component of space vehicles and provides protection from extreme high-temperature damage. Strong impacts on the TPS panels can cause severe damage to the panels and endanger the whole spacecraft as was made obvious by the Challenger accident. When a space vehicle starts reentry from outer space, the actual temperature is approximately 1650 °C (3000 °F) on the surface of the leading edge, and the temperature at the base of the TPS mounting bracket is approximately 175 °C (350 °F). Therefore, the sensor network has to be mounted at the base of the TPS panels to avoid the high temperature. A test performed for bolt-loosening detection by Stanford University on a TPS panel showed that the PZT sensors mounted at this location could function normally under a very challenging environment [18].

A demonstration of impact detection on the TPS panels is given below. As shown in Figure 11, a set of TPS panels designed and manufactured by Lockheed Martin Space Systems was used to build a test bed. There were a total of nine panels installed in the test structure surrounded by a metal frame. The TPS panels are mounted on a metal-stiffened back panel using brackets. The signal and ground traces for the sensors are implemented on a flexible dielectric substrate instead of conventional coaxial cables in order to reduce the weight for this application. The PZT sensors are individually mounted on the brackets. There are 16 sensors mounted in all. When an external impact is applied to a TPS panel, the mechanical stress waves propagate on the surface of the TPS panels, and then travel through the legs of the brackets to reach the PZT sensors.

The impacts were generated by using a hammer to hit the panel with different forces, in the range of 10–200 N. Both impact location and force can be determined by using the system identification method [19]. In order to monitor the impacts for a large area on a space vehicle, a multichannel-distributed impact detection system, as shown in Figure 12 [20], needs to be developed because the

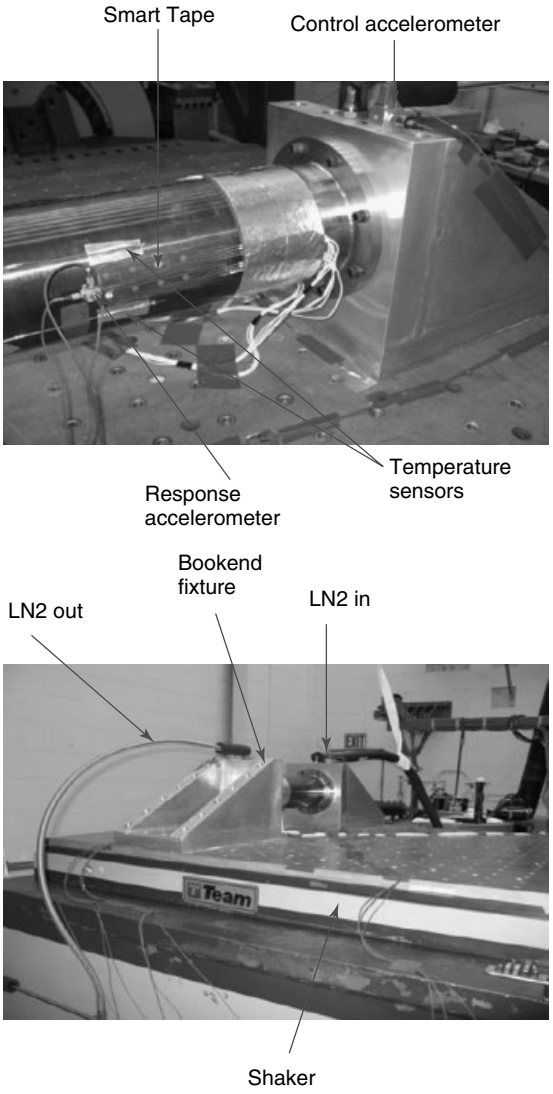


Figure 10. The setup for the combined cryogenic temperature and vibration test.

sensors required to cover the monitoring area might be up to hundreds.

2.6 Monitoring impacts on composite rocket motor cases

As another example, the onboard passive-sensing system for use in monitoring impacts on composite rocket motor cases was also developed and evaluated [21]. Figure 13 shows the implementation of the overall impact monitoring system. To demonstrate the impact detectability, including both location and force reconstruction, impact tests were conducted on a GEM60 rocket motor case instrumented with SMART Layers at Vandenberg Air Force Base (AFB).

Empirical approaches to structural characterizations and sensor network calibration, along with implementation techniques, were successfully evaluated. The simplistic empirical characterization approach, along with the robust/flexible sensor grids and the battery-operated portable logger, shows that the system can increase confidence in the composite integrity for new products progressing through manufacturing processes as well as existing assets that may be in storage or transportation.

3 APPLICATIONS OF OFF-BOARD SHM SYSTEM

Typical developments for potential applications of off-board SHM systems are summarized in this section.

3.1 Riveted lap joints in metallic structures

Off-board monitoring of metallic aircraft structures is possibly the most classical application of the SMART Layer system. A wide range of activity has been, and still is, related to monitoring riveted lap joints. A novel type of sensor layer that has become a standard in that regard emerged from a series of tests performed on multiriveted aircraft panels with multisite damage [22]. The novel sensor layer is the SMART Layer strip in which piezoelectric elements are placed at an equidistant pitch similar to the rivet pitch, as shown in Figure 14.

Since this requires a remarkable number of piezoelectric transducers to be placed, and hence an adequate amount of contacting on the SMART Layer, this led to the development of a stacked SMART Layer as shown in Figure 15. This configuration principle can even be applied along a fairly narrow spacing of rivet lines.

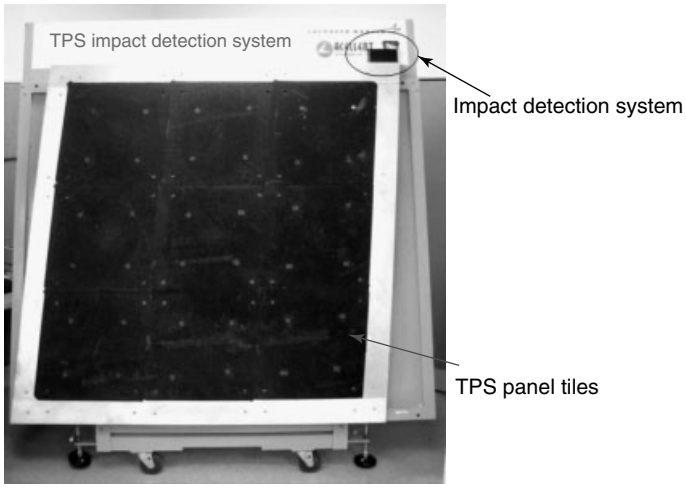


Figure 11. The TPS panel with the impact detection system.

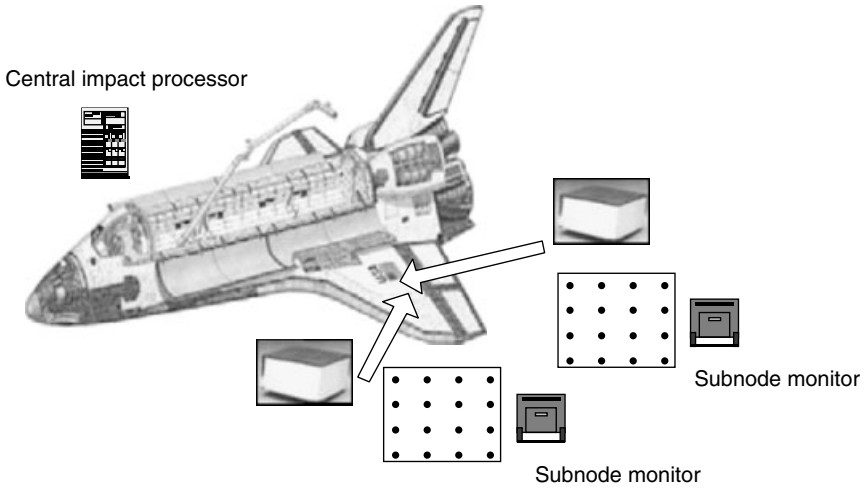


Figure 12. The distributed impact detection system.

Results obtained with this type of monitoring have shown that cracks of a length of 5 mm or more can be well detected, which is an adequate measure when compared to conventional NDT.

3.2 Monitoring of composite-bonded repair

Another application of the off-board SHM system in aerospace is the monitoring of composite-bonded repairs over fatigue-cracked areas. Metal structures such as fixed wing aircraft, helicopters, and other transportation systems are likely to develop fatigue cracks under cyclic loads and corrosive service environments. Repair methods such as bonded repair patches are used on hot-spot damaged regions. These repairs provide an efficient method for restoring the ultimate load capability of the structure. The bonded repairs offer several advantages over bolted repairs, including minimal changes to aerodynamic contours, weight saving, reduced cost, and formability to complex shapes [23–25]. A typical bonded patch repair is shown in Figure 16 [25]. Aircraft

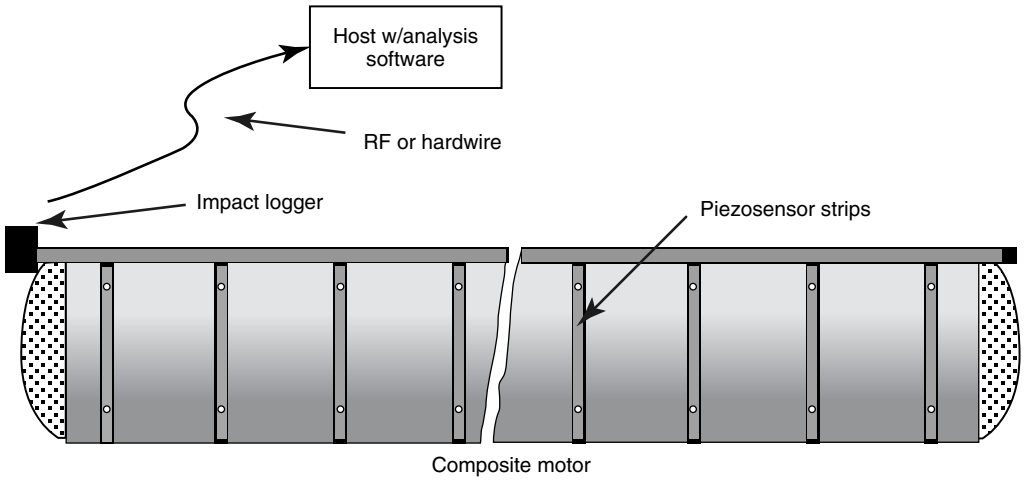


Figure 13. Overall system implementation for impact monitoring.

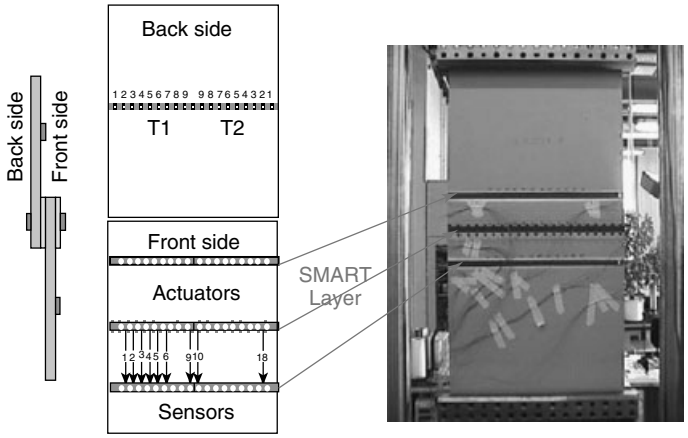


Figure 14. SMART Layer for monitoring multisite damage in riveted joints.

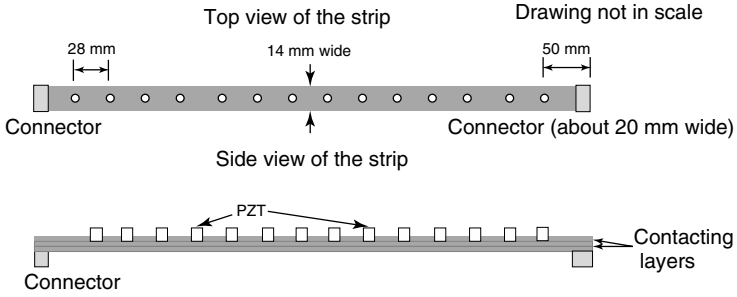


Figure 15. The stacked SMART Layer strip.

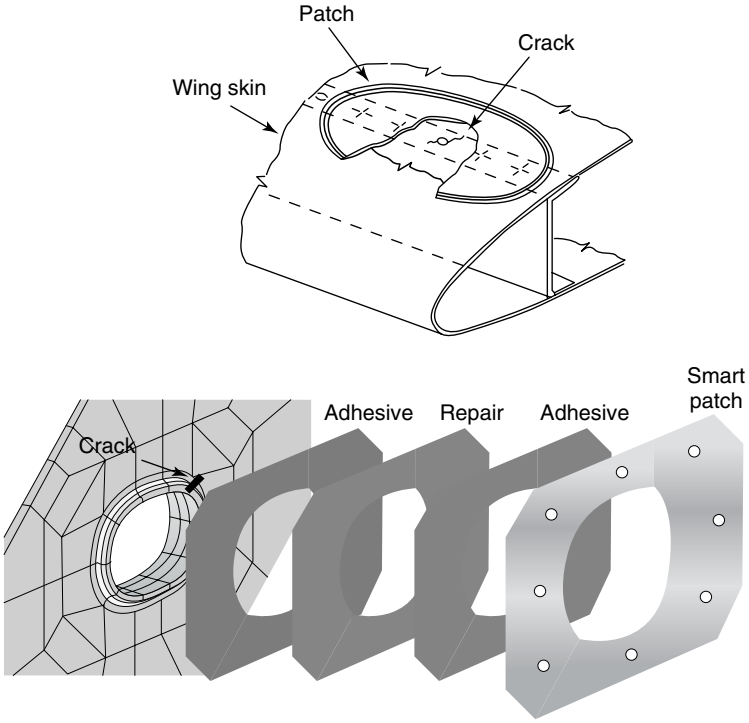


Figure 16. Bonded patch repair.

maintenance organizations generally require that the repaired structure undergoes inspection at the same time intervals as the unrepaired structure. This "fail-safe" approach essentially assumes that the repair is not effective in retarding crack growth of the original structure. In addition to potential bondline degradation, flaws and damage can have a detrimental effect on the integrity of the bond strength. Disbonds can arise from fatigue loading, impact events, or improper processing. Therefore, inspection of bonded repairs is an essential part of regular aircraft maintenance.

The SHM approach, such as also explained in further detail in **Design, Analysis, and SHM of Bonded Composite Repair and Substructure**, uses PZT sensors placed in, on, or around hot-spot regions to monitor damage initiation and damage growth. The sensor network built-in SMART Layer can be easily attached to existing aging structures without changing the local and global structural dynamics. The SMART Layer can also be embedded inside composite patches to closely monitor for internal flaws. These PZT elements in the layer can act as both actuators and sensors.

3.2.1 Monitoring of initial bond quality of bonded repair

Figures 17 and 18 show an example of SMART Layer technology applied to monitor the initial bond quality in a bonded repair coupon [26]. As shown in Figure 17, two specimens were prepared, one with fully bonded repair and the other with a 15 mm × 20 mm artificially induced disbond under one corner of the composite repair.

The signals from the specimen with a fully bonded composite patch were used as reference baselines. The damage index (DI), which was developed to extract features in sensor signals related to damage in the structure [27, 28], was calculated for each actuator–sensor path. Figure 18 shows an image generated by the DIs from all actuator–sensor paths. The white dots in the figure give the locations of PZT transducers. A heightened intensity of brightness indicates a bigger DI. From Figures 17 and 18, it is clear that the 15 mm × 20 mm disbond area can be well detected.

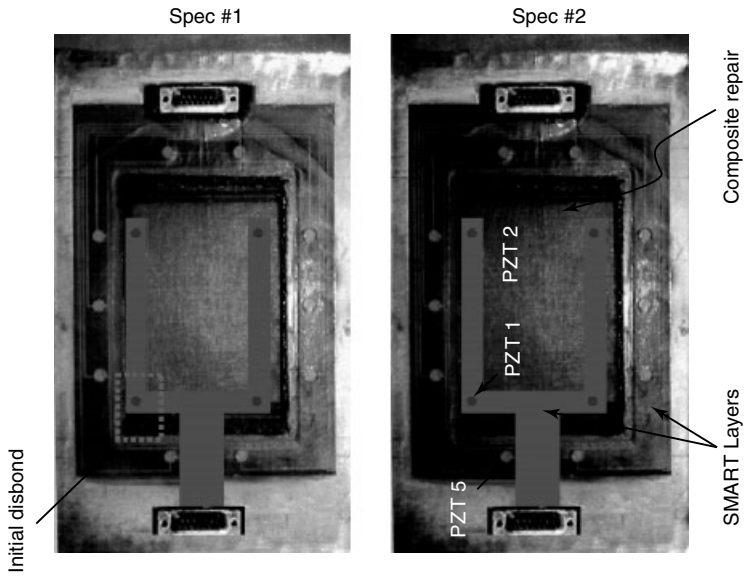


Figure 17. Test specimens with sensor network.

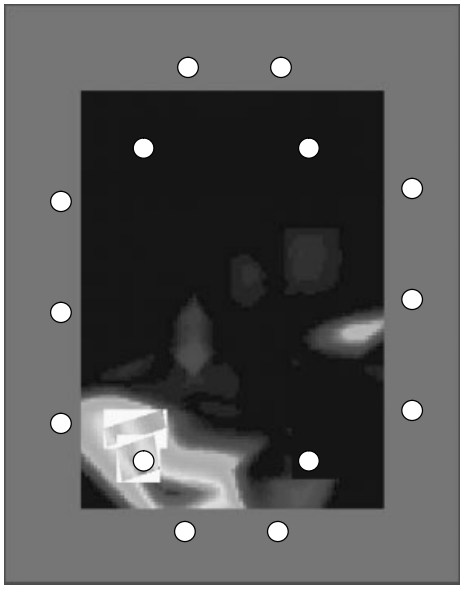


Figure 18. An image of detected disbond.

3.2.2 Monitoring of crack growth under bonded repair

Figure 19 shows another example of the active SHM system applied to detect crack growth under the bonded composite repair during fatigue cycling [26, 29]. The test was conducted at the US Wright-Patterson Air Force Base as part of a program to assess candidate SHM technologies for bonded composite repair patches. The results of fatigue testing can be viewed via diagnostic imaging shown in Figure 19, or quantified in a tabular form shown in Table 1.

To help quantify the damage size, a semiempirical method has been developed to give quantitative measurements of damage growth in real time, giving a few initial data points. Signal changes from each path are used to calculate DIs, from which DI curves are generated [26–28]. To calibrate the DIs to the damage sizes, several measurements of real damage size measured by conventional methods are used. The more the measurements used, the better are the estimates of the damage size. In this test, four measurements of the crack length by using an optical microscope were used to help quantify the crack size. The damage size estimates, along with the calculated uncertainty, are presented in Table 1.

As shown in Figure 20, the above-mentioned approach has also been used to monitor the repair in the center keel area of the F-16 fuselage station 341 bulkhead. More details about this application can be found in the literature [30] (also see **Validation of SHM Sensors in Airbus A380 Full-scale Fatigue Test**).

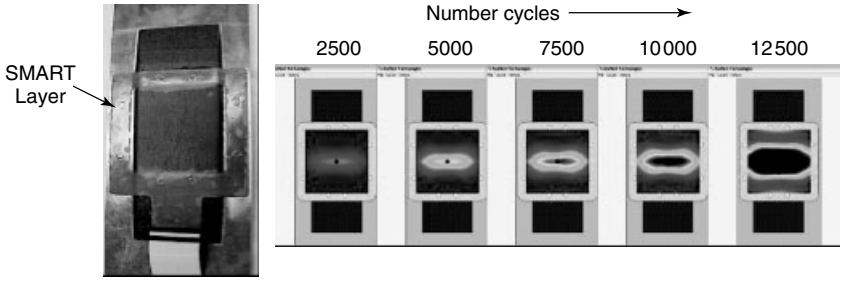


Figure 19. An image of damage growth under the bonded composite repair.

Table 1. Damage size estimates and uncertainity value

Test results

*** Bolted joint test results ***
Sensor system: SMART patch
Specimen configuration: A24
Specimen number: A24–8

Date	Time	Cycles	Measured crack length	Estinated crack length	Uncertainty
08/23/04	15:04:03	1000	0.158 in	------	------
08/23/04	15:56:31	2000	0.325	------	------
08/23/04	16:42:52	3000	0.762	------	------
08/24/04	09:11:36	4000	0.912	------	------
08/24/04	09:42:07	5000	0.993	------	------
08/24/04	10:30:23	6000	------	1.186 in	±0.099 in
08/24/04	11:01:55	7000	------	1.435	±0.131
08/24/04	11:41:04	8000	------	1.654	±0.188
08/24/04	13:16:22	9000	------	1.998	±0.243
08/24/04	14:04:37	10000	------	2.373	±0.329
08/24/04	14:26:01	11000	------	2.835	±0.417

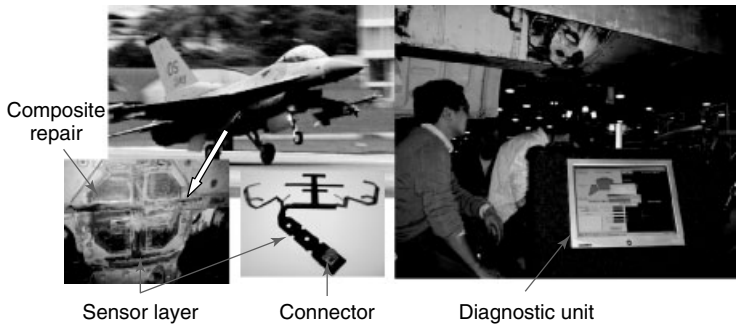


Figure 20. Health monitoring for bonded repair on F-16.

4 CONCLUSION

This article summarizes a series of aerospace applications of the SMART Layer technology, which have been differentiated as onboard and off-board systems. Onboard systems require all system elements of the SHM system to be included on board the air vehicle, since it requires monitoring during flight. An onboard system is unavoidable if the SMART Layer system is to be applied for load monitoring such as impact loads or in a passive mode such as with acoustic emission. The easier means of monitoring

and thus possibly the most attractive in the short term is the off-board system, since it only requires the SMART Layers to stay on board the air vehicle while the remaining signal generation and acquisition hardware with supporting software can be plugged in on demand. The application of the SMART Layer technology in aerospace is fairly wide and ranges from monitoring cracks in metallic, delaminations in composite structures, and repairs, in general, of conventional aircraft structures, to pipes, pressure vessels, or thermal protection shields for space vehicles. For all of those applications, the concept and functionality has been proven. The developed and finally qualified specific solutions demonstrated that the SMART Layer technology could work as established solutions in the future.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of US Army, US Air Force, Missile Defense Agency, DARPA, and NASA for sponsoring some of the developments listed in this article. The authors would like to thank Ms Irene Li, Dr Amrita Kumar, Dr David Zhang, and other colleagues at Acellent for their technical consulting and assistance for the development of the technology. The authors would also like to thank Mr Jim Gibson for his help in editing this article.

REFERENCES

[1] Staszewski WJ, Boller C, Tomlinson G. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons: New York, 2003.

[2] Boller C. Identification of life cycle cost reductions in structures with self-diagnostic devices. *Proceedings of the NATO RTO Symposium On Design Issues*. Ottawa, 1999; pp. 1–8.

[3] Chang FK. Ultra reliable and super safe structure for the new century. In *Proceedings of the 1st European Workshop on Structural Health Monitoring*, Balageas DL (ed). DEStech Publications: Lancaster, PA, 2002; pp. 3–12.

[4] Lin M, Qing X, Kumar A, Beard S. SMART layer and SMART suitcase for structural health monitoring applications. *Proceedings of SPIE on Smart Structures and Material Systems*. San Diego, CA, 2001; pp. 98–106.

[5] Qing X, Beard B, Kumar A, Ooi T, Chang FK. Built-in sensor network for structural health monitoring of composite structure. *Journal of Intelligent Material Systems and Structures* 2007 **18**:39–49.

[6] Qing X, Chan H, Beard S, Kumar A. An active diagnostic system for structural health monitoring of rocket engines. *Journal of Intelligent Material Systems and Structures* 2006 **17**:619–628.

[7] Zhang DC, Ouyang L, Qing P, Li I. A novel real-time health monitoring system for unmanned vehicles. In *Unmanned Systems Technology X*, *Proceedings of the SPIE, Volume 6962*, Gerhart GR, Gage DW, Shoemaker CM (eds). SPIE, April 16, 2008; p. 696217.

[8] Qing XP, Kumar A, Beard S, Yu P, Zhang D, Liu C, Hannum R. Advanced self-sufficient structural health monitoring system. In *Proceedings of the Third European Workshop on SHM: Structural Health Monitoring 2006*, July 2006, Granada, Spain, Güemes A (ed). DEStech Publications: Lancaster, PA, 2006; pp. 807–814.

[9] Zhang DC, Yu P, Beard SJ, Qing XP, Kumar A, Chang FK. A new SMART sensing system for aerospace structures. *Unmanned Systems Technology IX*, *Proceedings of SPIE, Volume 6561*, SPIE, April 2007.

[10] Zhang DC, Liu P, Beard S, Qing P, Kumar A, Ouyang L. SMART solutions for composite structures. In *Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*, *Proceedings of SPIE, Volume 6934*, Shull PJ, Wu HF, Diaz AA, Vogel DW (eds). SPIE, 2008; p. 69341C.

[11] Beard S, Liu B, Qing P, Zhang D. Challenges in Implementation of SHM. *Proceedings of the 7th International Workshop on SHM: SHM 2007 Quantification, Validation, and Implementation*. Stanford, CA, September 2007.

[12] Air Force Material Command, Space and Missile Systems Center, *Proceeding of the Graphite/Epoxy Rocket Motor Case Experience Sharing Meeting*. Launch Vehicles Program Office, May 5–7, 1998.

[13] Qing X, Beard S, Kumar A, Chan H, Ikegami R. Advances in the development of built-in diagnostic system for filament wound composite structures. *Composite Science and Technology* 2006 **66**:1694–1702.

[14] Dugnani R, Malkin M. Damage detection on a large composite structure. *Proceeding of the 4th International Workshop on Structural Health Monitoring*. Stanford University, September 2003.

[15] Kumar A, Wu HF, Lin M, Beard S, Qing X, Zhang D, Hamilton M, Ikegami R. Potential applications of SMART Layer technology for homeland security. In *Nondestructive Detection and Measurement for Homeland Security II*, *Proceedings of the SPIE, Volume 5395*, Steven R, Bar-Cohen, Y, Aktan AE, Wu HF (eds). SPIE, 2004; pp. 61–69.

[16] Qing XP, Beard SJ, Kumar A, Sullivan K, Aguilar R, Merchant M, Taniguchi, M. Performance of piezoelectric sensors based SHM system under combined cryogenic temperature and vibration environment. *Smart Materials and Structures* 2008 **17**(5):055010.

[17] Arritt B, Kumar A, Buckley S, Hannum R, Welsh J, Qing X, Wegner P. *Responsive Satellites and the need for Structural Health Monitoring*, *Proceedings of the SPIE, Volume 6531*, SPIE, 2007; p. 653109.

[18] Yang J, Chang FK. Verification of a built-in health monitoring system for bolted thermal protection panels, smart structures and materials. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, *Proceedings of the SPIE, Volume 5765*, Masayoshi T (ed). SPIE, 2005; pp. 769–780.

[19] Park J, Chang FK. System identification method for monitoring impact events. *Smart Sensor Technology and Measurement Systems Conference*, *Proceedings of SPIE, Volume 5758*, SPIE, San Diego, CA, 2005; pp. 189–200.

[20] Yu P. Real time impact detection system for thermal protection system. In *Proceedings of the 7th International Workshop on SHM: SHM 2007 Quantification, Validation, and Implementation*, September 2007, Stanford University, Chang F-K (ed). DEStech Publications: Lancaster, PA, 2007.

[21] Child JE, Kumar A, Beard S, Qing P, Paslay DG. Impact detection and analysis/health monitoring system for composites. In *Sensors for Propulsion Measurement Applications*, *Proceedings of SPIE, Volume 6222*, Korman V (ed). SPIE, 2006; p. 62220G.

[22] Boller C, Ihn J-B, Staszewski WJ, Speckmann H. Design principles and inspection techniques for long life endurance of aircraft structures. *Proceedings of the 3rd International Workshop on Structural Health Monitoring*. Stanford, CA, 2001; pp. 275–283.

[23] Belason EB. Status of bonded boron/epoxy doublers for military and commercial aircraft structures. *Composite Repair of Military Aircraft Structures: Proceedings of the 79th AGARD Structures and Material Panel Specialists' Meeting in Composite Repair of Military Aircraft Structures*. Seville, 1994.

[24] Chesmar EF. Metal bond and composite repairs: similarities and differences. *Composites'96 Manufacturing and Tooling Conference*. Society of Manufacturing Engineers, 1996; pp. 281–290.

[25] Stone RH. Field-level repair materials and processes. In *Composite Repairs*, SAMPE Monograph, No. 1, Brown H (ed). SAMPE-Soceity of Advancement of Material and Process Engineering, 1985; pp. 87–99.

[26] Qing X, Beard S, Kumar A, Hannum R. A real-time active smart patch system for monitoring the integrity of bonded repair on an aircraft structure. *Smart Materials and Structures* 2006 **15**:N66–N73.

[27] Ihn J, Chang FK. Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: I. Diagnostics. *Smart Materials and Structures* 2004 **13**(3):609–620.

[28] Beard S, Qing PX, Hamilton M, Zhang D. Multifunctional software suite for structural health monitoring using SMART technology. *2nd European Workshop on Structural Health Monitoring*. Munich, 2004; pp. 101–108.

[29] Beard S, Kumar A, Qing XP, Zhang DC, Patterson J. A smart patch system for monitoring of bonded Pairs. *Proceeding of the 5th International Workshop on Structural Health Monitoring*. Stanford University, September 2005.

[30] Malkin M, Qing X, Leonard M, Derriso M. Flight demonstration: health monitoring for bonded structural repairs. In *Proceedings of the Third European Workshop on SHM: Structural Health Monitoring 2006*, July 2006, Granada, Güemes A (ed). DEStech Publications: Lancaster, PA, 2006; pp. 167–175.

FURTHER READING

Kumar A, Roach D, Beard B, Qing X, Hannum R. In-situ monitoring of the integrity of bonded repair patches on aircraft and civil infrastructures. In *Advanced Sensor Technologies for Nondestructive Evaluation and Structural Health Monitoring II*, *Proceedings of the SPIE, Volume 6179*, Meyendorf N, Baaklini GY, Michel B (eds). SPIE, 2006; pp. 147–154.

Chapter 95
Commercial Fixed-wing Aircraft

Grant A. Gordon[1] and Christian Boller[2]

[1] *Research Technology Centre, Honeywell Inc., Phoenix, AZ, USA*
[2] *Saarland University & Fraunhofer Institute for Non-Destructive Testing, Saarbrücken, Germany (and formerly of The University of Sheffield, Sheffield, UK)*

1 INTRODUCTION

Aircraft are costly complex assets with significant maintenance support requirements. Structural health monitoring (SHM) [1–5] is part of a new proactive aircraft management approach to maintaining vehicle safety and significantly reducing costs. The ultimate vision of SHM is a system of permanently installed, continuously monitoring microsensors that provide timely information about the condition and integrity of safety-critical structures [6]. The application of SHM systems to aircraft is envisioned to not only provide real-time or near-real-time characterization of structural integrity for improved aircraft lifecycle management but also opportunities for adaptive flight control, enhanced vehicle mission flexibility, and changes in the approaches to aircraft design.

The significance of these SHM benefits varies according to the intent of the aircraft: transport, space exploration, or defense. In the case of commercial air transport, the challenge is to make the aircraft as profitable as possible by safely maximizing revenue-generating activities and minimizing support requirements. Any events that reduce aircraft operation, e.g., inspection, maintenance, holding loops, bad weather, accidents, or any other obstructions, detrimentally impact the aircraft's profitability and the operator's bottom line. A number of studies have examined the benefits of SHM [7–9] as well as the steps necessary to meet the regulatory requirements for introducing this technology [10, 11]. As a consequence of these findings, virtually every large aircraft manufacturer and aircraft subsystem manufacturer is exploring how the SHM technology vision can be implemented on real aircraft. Airbus [12–14] Boeing [15, 16], EADS Military Air Systems [17, 18], the US Air Force [19, 20], and Honeywell [10, 21] have all expressed their interest in this technology.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

This article presents some current views on the potential role of SHM in commercial fixed-wing aircraft. Although aircraft accidents due to structural failures have become increasingly rare, historically there were a number of incidents that have influenced the advancement of SHM approaches. The discussion starts with a review of how SHM was initiated in response to unforeseen structural failures and has evolved into a technology that, along with other health-management technologies, is leading the way toward a change in how aircraft are managed and maintained.

2 THE HISTORICAL MOTIVATION FOR SHM

2.1 Safe life and fail-safe design

Despite a generally excellent track record, there have been a number of milestone aviation accidents that have motivated and impacted the development of SHM system for fixed-wing aircraft. Fatigue failure has been a concern to aviation since as early as 1908 when the Wright brothers had to postpone their first powered flight due to fatigue failure of a propeller shaft. However, it was in January and April of 1954 that two De Havilland Comets drew worldwide attention to the problem of fatigue in the aircraft fuselage [22] as well as the aircraft structure in general. The Comet was the first commercial high-altitude jet transport plane offering passengers superior performance, higher cruising altitudes, and comfort. It was designed to support a significantly higher cabin pressure differential than its contemporaries—the propeller-driven transports. Yet two Comets crashed after they had undertaken only 1286 and 903 flights, following an explosive decompression of the fuselage. The fleet was grounded for extensive investigation, including full-scale pressurization tests, which revealed that unstable crack growth and an underestimation of notch sensitivity had caused the catastrophic failures. Fatigue failures of military aircraft were also occurring during this time.

Shortly thereafter, regulations were introduced to ensure that the *safe life* or service life of a vehicle could be demonstrated and validated using laboratory tests. To obtain certification, the platforms were tested to failure using simulated service loads conducted in a laboratory. Safe life was defined as the test life to failure divided by a safety factor of three or four to account for uncertainties. Once the safe life was reached, the vehicle was to be taken out of service. In addition to certification of safe life, *fail-safe* design principles were introduced in the late 1950s. By definition, fail-safe systems are designed to become safe when they fail or cease to operate. For aircraft, this doctrine mandates that the structure must be capable of withstanding significant damage before safety is compromised. Commonly, the requirements were met by employing multiple load paths and well-established residual strength requirements to accommodate failure or significant partial failure of a structural element. In addition, the structure was to be designed so that damage could be easily detected before safety was compromised.

2.2 Operational loads monitoring

A form of SHM known as *operational loads monitoring* first appeared in the mid-1950s in response to the crashes of the two De Havilland Comets. These systems are designed to help provide more effective fatigue life management. The loads monitoring approach is particularly relevant for military aircraft where the vehicles are often flown to their limits, experience operational damage, and unpredictable environmental conditions. Unfortunately, monitoring aircraft operational loads is far from an easy task, since they cannot be measured directly. The most popular way of monitoring them is by measuring strain. Strains, however, can be quite localized and therefore require the use of a large number of sensors to obtain an understanding of the entire vehicle. This leads to a prohibitively large increase in the aircraft's complexity and raises questions of sensor reliability and consequential overall aircraft reliability. As a result, many military aircraft manufacturers have adopted a strategy of inferring operational loads from global data collected by a small number of sensors and exploiting, to the maximum extent possible, any data from sensors that already exist on the aircraft for other reasons.

Early loads monitoring systems were therefore based on a small collection of accelerometers distributed over the aircraft. Probably the first to implement an accelerometer-based counting device was the UK Royal Air Force, who, in response to the Comet accidents, gradually equipped two-third of their active fleet with fatigue meters, Figure 1 [23]. This technology insertion also promoted and preceded a variety of developments in North America on the F-18 [24] and C-130 [25], and in Europe on the F-16 [26], AMX [27], and Tornado [28] (*see* **Fatigue Monitoring in Military Fixed-wing Aircraft**) aircraft. The more sophisticated systems used a series of g-meters, which counted the number of load exceedances for different g-levels. The system implemented in the Eurofighter Typhoon represents a modern state-of-the-art system and is based on one of two approaches: (i) a system comprised of 16 conventional strain gauges (UK and Spanish version) or (ii) a flight parameters-based system (German and Italian version) [29]. In either system, the time-domain data is fed into the digital loads model, which virtually calculates the load-time sequence at each specified location of the aircraft. This information can then be used to calculate the fatigue life consumed and hence the time when fracture is expected to occur.

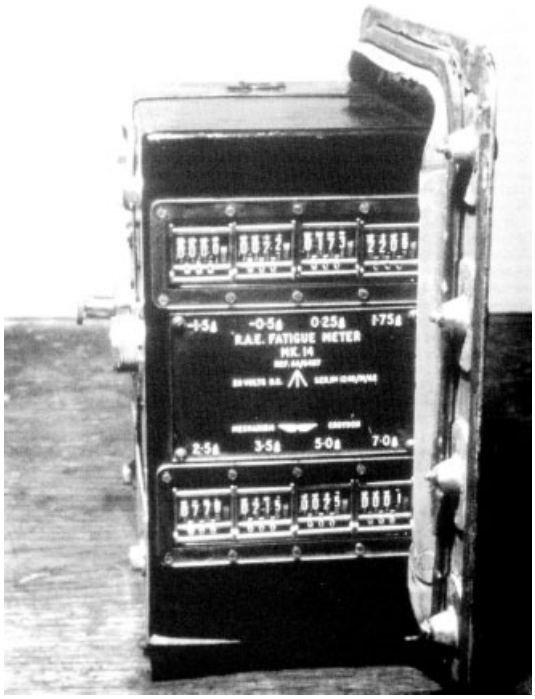


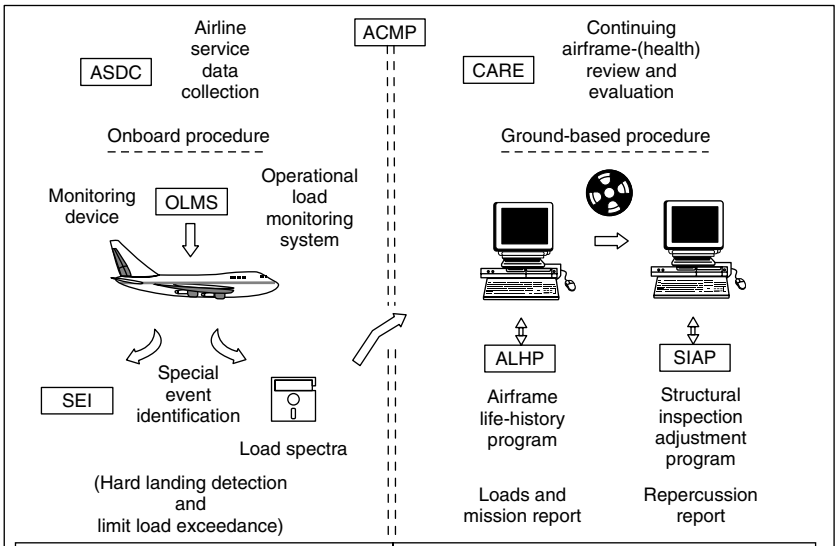
Figure 1. Fatigue meter used in UK RAF fleet since 1954. [Reproduced with permission from Ref. 23. © NATO, 1998].

Specifically, it has been the work around Tornado [28] that triggered Airbus, in the 1980s, to also explore an operational loads monitoring system (OLMS) [30] the logic of which is shown in Figure 2. This system, which considered monitoring hard landings and limit load exceedances was abandoned in the early 1990s. However, recent developments show that monitoring of hard landings and limit load exceedances is very relevant for commercial aircraft; more details can be found in other articles in this encyclopedia (*see* **Loads Monitoring in Aerospace Structures**; **Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft**; **Video Landing Parameter Surveys**; **Landing Gear**).

2.3 Damage-tolerance design

The safe-life principles dominated both military and commercial aircraft design until 1969 when the crash of a General Dynamic Corp F-111, a swing-wing fighter–bomber, caused engineers to rethink the safe-life approach. The F-111 had a safe life of 4000 flight hours and as such had been subjected to 16 000 h of laboratory fatigue testing, but the F-111 #94 failed after only 107 service hours while pulling 3.5 g, only half its design limit load. Investigation into the crash uncovered a troubling revelation, the failure fracture surface clearly showed that a flaw of significant size had been present on the first day that the airplane had entered service. Only a small extension of the flaw was necessary before catastrophic failure occurred. This, along with other evidence, caused the United States Air Force (USAF) to rethink the safe-life philosophy. A new approach emerged based on the belief that aircraft structures needed to be tolerant to an initial damage state.

The new *damage-tolerance* philosophy assumed that flaws existed in all critical locations. Structural integrity programs were implemented to find and repair damage before failure occurred, resulting in an approach that could extend the structural life beyond it design service life, Figure 3. The inspection intervals used in the maintenance programs were



SOURCE		PARAMETER
ADIRU	Air Data Inertia Reference Unit	ALT, MN, TAS, AOA, PTCR ROLR, LATG, VRTG
FMGC	Flight Manag. Guld. Computer	GW, CG
PRESS.	CONTR. Pressure Control System	PDC (cabin different pressure)
DMC	Data Management Computer	RALT
FQIS	Fuel Quantity Indication System	FUEL QUANTITIES WING (Inner and outer cell)
FCDC	Flight Control Data Computer	STAB, AIL ELEV, RH SPL 1-6
SDAC	System Data Analog Converter	RUDD
SFCC	Slat Flap Control Computer	SLAT/FLAP
FWC	Flight Warming Computer	FLIGHT PHASE
FDIU	Flight Data Interface Unit	TIME, DATE, AC-TAIL NUMBER

Figure 2. Concept of the operational loads monitoring system (OLMS) explored for Airbus A320 [30]. [Reproduced from Ref. 29. © John Wiley & Sons, 1991.]

based on models for crack growth driven under a presumed loading spectrum using initial flaw sizes defined in Military Spec 83444. These flaw sizes vary with location and geometry but are often referred to as 1.27 mm according to the specification of a surface-breaking crack. There is also a provision for designing structures that will not be inspected during their service life. This is an extremely useful design variable for accommodating areas where inspection is prohibitively difficult. These areas can still be classified as damage tolerant, provided they can be qualified for slow flaw growth, which requires that the flaw be incapable of growing to size where failure can occur during the design service life. Owing to the more stringent requirements, these sections are heavier than their inspectable counterparts. Since 1970, the USAF has required damage-tolerance programs for all its aircraft.

3 MODERN COMMERCIAL AIRCRAFT

3.1 SHM as an aid to the inspection process

Modern commercial aircraft also have to meet damage-tolerance regulations according to FAR 25.571—Amendment 96 with safe-life designs being used in limited cases for landing gear and certain helicopter applications. As discussed by a number of researchers, this provides an opportunity for SHM systems to modify the attendant inspection procedures. SHM can improve present inspection processes with onboard automated inspection technologies based on simple sensing devices permanently attached to the component of interest. As discussed in detail in other sections of

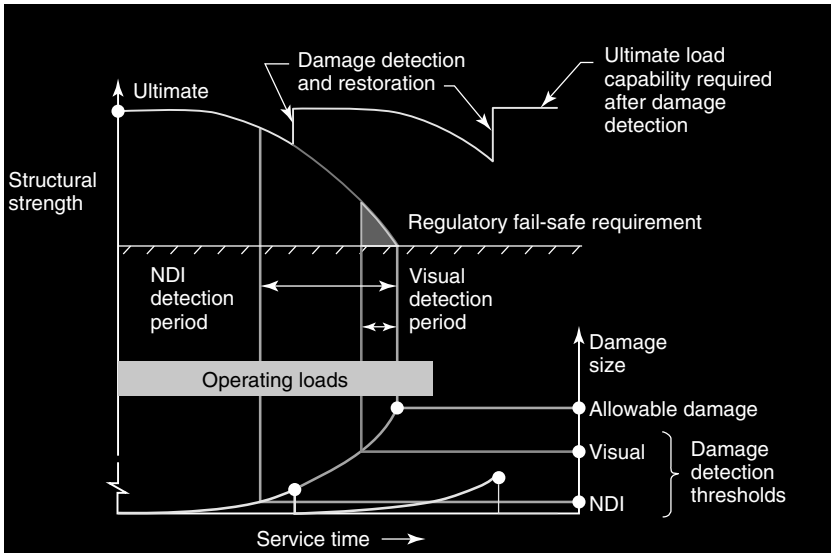


Figure 3. Damage-tolerance design philosophy and inspection approach.

this encyclopedia, such SHM sensing devices often operate on familiar principles explored and exploited by the nondestructive testing (NDT) community: ultrasonic [31–33], eddy current [34, 35], and acoustic emission [36, 37]. These and other novel approaches, such as comparative vacuum monitoring [35], fiber-optic sensors [38, 39], and electrical impedance methods [40], are just a few examples of approaches that hold the promise of improving safety and reducing the inspection burden. However, to succeed, it is commonly agreed that they must achieve certified maintenance credits that can be used to decrease the costs associated with maintaining an aircraft.

For modern air transport markets, vehicle design and maintenance requirements must satisfy both safety and economic considerations. Although inspections are mandatory when using a damage-tolerant design approach, they are also time consuming and subject to human error. Easy inspections and long inspection intervals are desired. Today, as in the past, the scheduled structural inspections of commercial jets rely primarily on the use of visual techniques. Thus, the inspectability and accessibility of the aircraft structure should allow visual methods to confidently make a damage assessment over a majority of the structure. As a result, aircraft designers try to limit the number of areas requiring directed inspections with sophisticated equipment and inspections in areas that are difficult to access. When inspections in these areas cannot be avoided, the components are designed to require infrequent inspection. Nonetheless, there are areas that are difficult to access where additional effort is required to dismantle subsystems and gain access. Examples include the frames and stringers of an aircraft fuselage located behind galleys, lavatories, or the air-conditioning system. These areas, and areas requiring more frequent inspection than originally anticipated, are good candidates to be considered for automated SHM.

3.2 Reconsidering structural design

Now consider the discussion above from a slightly different perspective, the design perspective. Damage tolerance is the approach used to ensure that the structure can withstand fatigue, corrosion, manufacturing defects, and accidental damage (AD) throughout the life of the aircraft and thus is critical in establishing the dimensioning criterion applied to the structural members. Allowable crack lengths are defined on the basis of how well they can be reliably detected. Regions that are difficult to inspect must be designed to support cracks that may grow for a substantial period of time before they can be reliably detected.

Examples include complex lapped joints where a crack can travel a significant distance before it emerges at an external surface where it can be reliably detected using the visual inspection protocol. In the case described above, the effort to dismantle galleys, lavatories, or air-conditioning systems to facilitate inspection of the underlying frames and stringers is high, and thus the designer chooses not to inspect these areas frequently and to account for this omission by designing to a more severe damage-tolerance scenario. Without regular inspection, the frames and stringers have to be considered as fully broken for the damage-tolerance analysis and any detectable initial crack found at the fuselage panel surface is assumed to experience higher stress loads and, therefore, propagate at high speeds.

Knowledge about the status of the internal members can be improved through the application of SHM and then be used in concert with an external inspection regiment. In this case, the damage-tolerance analysis and subsequent design could consider less severe crack scenarios e.g., a skin crack over an intact internal member [41]. This modified inspection approach could be used to increase the time between inspection intervals while maintaining the allowed stress levels, or increase the allowed stress levels while maintaining the inspection interval, Figure 4. Both circumstances are advantageous but the design benefit of being allowed to increase the stress levels would lead to a significant weight saving estimated to be between 13 and 20% [41]. Figure 5 shows a schematic comparison of allowable stress versus resulting inspection intervals for a fuselage panel with inspectable and noninspectable stiffening elements.

3.3 Aging aircraft

Another area where SHM shows significant relevance is the area of geriatric or aging aircraft. The problem of aging aircraft first gained widespread awareness in the aviation community after the May 14, 1977 Dan Air Boeing 707 crash. Investigation revealed that fatigue failure of a rear spar lead to the separation of the entire right-hand horizontal stabilizer just prior to landing. At the time of the crash, the aircraft had accumulated 47 621 flight hours of the intended design life goal of 60 000 flight hours. Although the stabilizer had been designed according to fail-safe principles, in practice, the area did not afford a significant and easily detectable damage condition before safety was compromised. The crash caused regulatory authorities to reconsider the problem of fatigue in older aircraft. As a result, they concluded that the existing inspection methods and schedules were inadequate for aging aircraft, and that supplementary inspections were needed.

The problem of aging aircraft was again brought to light by the Boeing 737 Aloha Airlines Flight 243 accident in 1988. The aircraft was old, having experienced more than 89 681 flights almost 15 000 more than the vehicles economic design life of 75 000 flights—although it should be noted that the maximum pressure differential were not reached for many of the flights. The Aloha accident was the result of many interrelated factors contributing to a heretofore unknown failure mechanism. Corrosion-induced disbonding at a cold, bonded skin splice allowed the cabin pressure loads to be transferred to joint rivets, which then induced multiple fatigue cracks at the rivet–skin interface. The numerous small fatigue cracks went undetected until the collinear

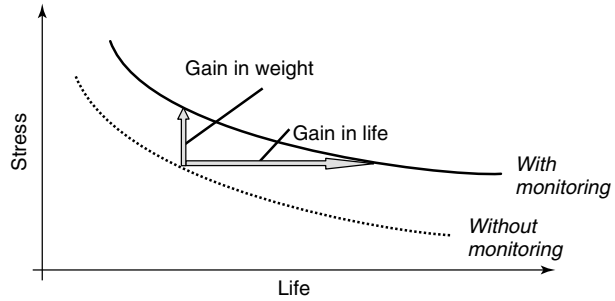


Figure 4. Comparison of stress-life behavior of fuselage skins with monitored and nonmonitored stiffening elements. [Reproduced with permission from Ref. 41. © DEStech Pub, 2007.]

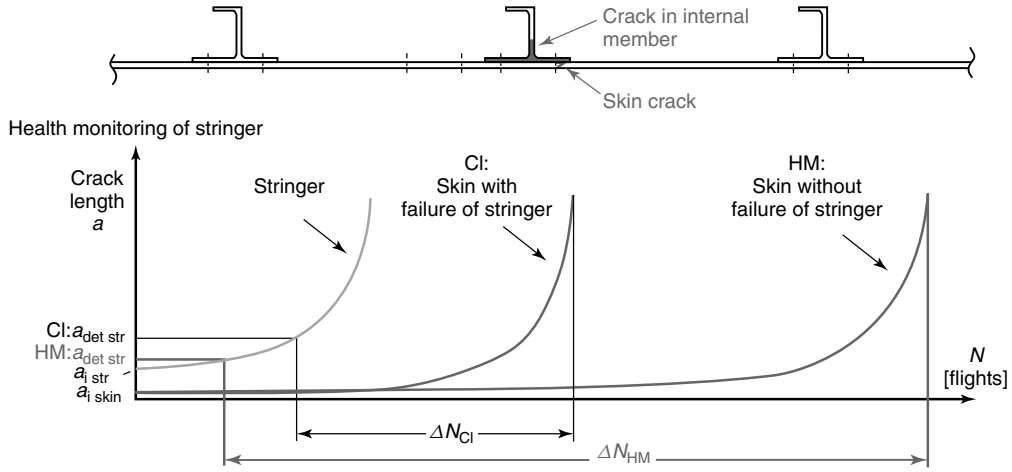


Figure 5. Comparison of fuselage skin crack propagation life for a nonmonitored and monitored frame, respectively. [Reproduced with permission from Ref. 41. © DEStech Pub, 2007.]

cracks rapidly linked up and caused explosive decompression of a major section of the fuselage. This damage mechanism came to be known as *multiple site fatigue damage* (MSD), which is a subset of *widespread fatigue damage* (WFD), itself defined as the simultaneous presence of multiple cracks of sufficient size and density to render the structure incapable of meeting damage-tolerance requirements. The newfound awareness to WFD lead to significant worldwide activities by manufacturers, operators, regulatory authorities, and researchers to develop new regulations and better understanding that would ensure the safety and integrity of aging aircraft. As a result, it is now a standard practice that many fracture-prone components be inspected at much shorter intervals if the aircraft is operated beyond its original design life. This significantly increases the amount of inspections needed in these circumstances.

The mean age of commercial and military aircraft fleets has been steadily increasing over the past few years. The relevance of this situation becomes clear in the commercial aviation sector where the demand for air transport capacity is high, but older aircraft must still be capable of operating in a cost-effective manner despite the increased maintenance load. Figure 6 shows the fleet age of the most common commercial aircraft types. This figure illustrates that operators often use airplanes well beyond their original design lives and that the majority of the planes are either in a mature or aged phase of their life cycle.

Aging aircraft is also of significant relevance to the defense sector. The average age of the B-52 bomber is 46.6 years and the KC-135 between 46 and 48 years Table 1. The USAF fleet now stands at an average age of 24 years and is expected to achieve an average of 26.5 years in 2012. The USAF would need to boost aircraft purchases by about 170 a year to reverse this trend. However, there are no such plans, so the military fleet will continue to age and there will be an ongoing need to economically maintain this aging fleet. Also, note that this situation is not exclusively an USAF problem but is repeated around the globe.

It is clear from the evidence and the preceding discussion that as aircraft age their maintenance requirements increase. In the Advisory Circular AC 91-60A, *The continued airworthiness of older airplane*, the US Federal Aviation Administration (FAA) notes that it is essential to have regular assessments of the airframe structural integrity as the aircraft ages. Furthermore, additional maintenance of the structural components will be required at structural points—a list of 28 such specific areas is given in the advisory circular. This increased burden includes inspection, repair, as well as engineering design support and results in

- shorter inspection intervals and hence more inspection effort, especially when the aircraft are designed fail safe;
- more frequent replacement of spare parts when components reach their end of life;

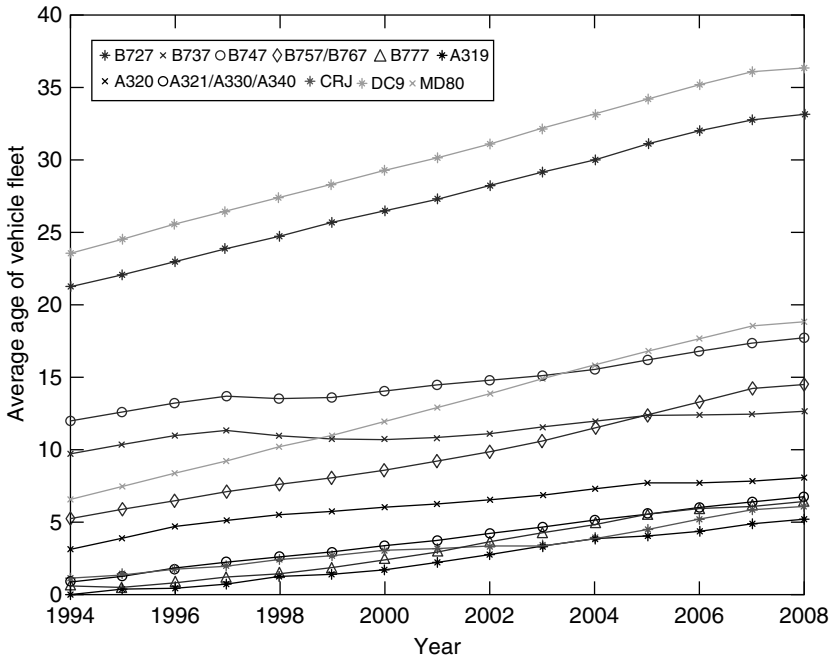


Figure 6. Fleet ages for common passenger jet aircraft.

- higher engineering effort to implement special repair due to poor spare parts availability.

The Air Transport Association (ATA) of America concluded that as an aircraft ages, the additional mandated or recommended maintenance requirements would increase the workload during heavy maintenance visits by up to 50% initially and 10–15% during subsequent visits [42]. However, they also noted that advanced technology methods could offset

Table 1. Aging military aircraft overview

Aircraft type	First service flight	Total in service	Average age
Boeing F/A-18 A/B/C/D Hornet/Super Hornet	1978/1983/1987	1203	13.6
Boeing B-52 A/B/H Stratofortress	1952/1955/1962	94	46.6
Boeing KC-135 E/R/T Stratotanker	1956	568	49/47 (E/R)
Dassault Mirage F1/C/CR/CT/D/E	1973–1992	309	
Dassault/Dornier Alpha Jet	1978	348	
Fairchild A-10 A/C Thunderbolt	1977/2007	364	
Lockheed-Martin C-130 A/B/E Hercules	1956/1959/1962	1338	44 (E)
Lockheed-Martin C-5 A/B Galaxy	1969/1980	111	37/20 (A/B)
Lockheed-Martin F-16 A/C/D Falcon	1979/1981/1989/1994	2982	16.7
Lockheed-Martin P-3 Orion	1959	368	28.5 (US Navy)
Lockheed-Martin U-2 Dragonlady	1956	33	
McDonnel Douglas F-15 Eagle	1974	968	25.5
McDonnel Douglas F-4 Phantom	1958	725	
MiG-21 (many variants)	1958—(Chinese version)	1528	
MiG-29 (many variants)	1985	1047	
Northrop T-38/A/C	1959/1961	636	
Panavia Tornado IDS	1974	505	

the need for more frequent scheduled inspections. SHM could be an important factor in improving the effectiveness of these inspections by enabling a focused, condition-based maintenance approach to be used.

The identification of critical structures is an evolutionary process and thus damage from critical fatigue or corrosion is often identified in unanticipated locations as the aircraft matures. To guarantee that this type of unexpected event does not lead to catastrophic damage, fleet leaders—the aircraft in a specific fleet that has accumulated the largest number of flights—are analyzed in detail and are used to trigger service bulletins that adjust the maintenance practices for all other aircraft in the fleet. For these scenarios, SHM is probably best applied once the aircraft has begun to manifest its operational-specific inspection needs. Thus, SHM has a role in improving safety and/or minimizing the inspection burden for both new and aging aircraft.

4 MAINTENANCE PRACTICES

4.1 Scheduled maintenance development

Up until this point, examples of metal fatigue failure as they relate to SHM have been discussed for historical and pedagogical reasons. But fatigue in metals is far from the only damage mechanism that influences aircraft structures. It is well recognized that there are three principal sources of aircraft structural damage: fatigue, environmental, and accidental damage (AD) Figure 7. The initial scheduled inspection requirements for these damage sources on a new aircraft are developed through an evaluation process defined in the (Maintenance Steering Group) MSG-3 guidelines document [43]. This document is published by the ATA, but is written collaboratively through the combined efforts of aircraft and engine manufacturers, regulatory agencies, airline, and government organizations. In the case of structural items, the assessment considers the significance of the items to continued airworthiness, susceptibility to various forms of damage, and the degree of difficulty in detecting damage to develop a scheduled structural inspection program.

Once the MSG-3 process is complete, the results are used to develop a maintenance review board report (MRBR), which outlines minimum inspection requirements for a particular airplane type certification. On the basis of the MRBR, the aircraft manufacturers create a maintenance planning data (MPD) document. Thus, one step for SHM to be accepted as part of the initial maintenance procedures is to be recognized within the MSG-3 document. New content in the MSG-3 that will outline the use of SHM is under discussion. Initially, the SHM system may collect data on a real-time basis but would probably

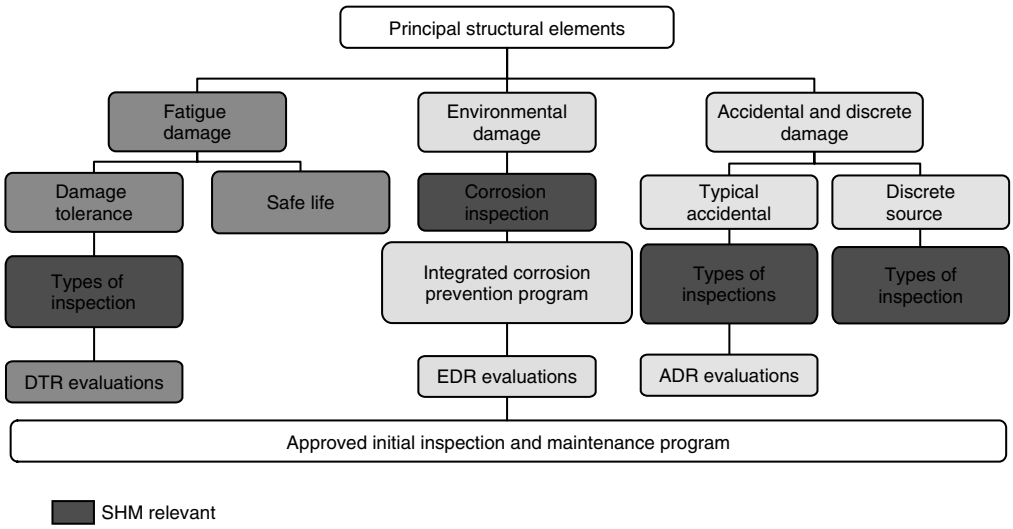


Figure 7. MSG-3 process for developing a maintenance program based on damage sources.

download the data at predetermined MRBR/MPD intervals similar to the traditional structural inspection periods. It is expected that until the SHM systems have been validated through comparative field tests, the traditional maintenance program, as outlined in the MPD document, will be used in tandem with an SHM system. Areas where SHM may become relevant in the MSG-3 process are highlighted in Figure 7.

4.2 Accidental damage

When assessing the inspection requirements for accidental damage, the MSG-3 recommends that all sources of damage such as impact from ground and cargo handling equipment, foreign objects, bird strikes, hail, runway debris, and the artefacts of human error be included in the assessment. This type of damage occurs for metallic as well as composite materials and can be of significant concern and frequency. While inspections due to fatigue damage may be unnecessary for nonmetallic materials, if the aircraft design has been based on a *no-damage growth* design philosophy, AD and environmental damage mechanisms are applicable to all aircraft when they first enter service.

It is important to note that some types of AD have a degree of predictability. AD due to ground vehicle and equipment collisions predominantly occurs in the same locations and therefore concentrates the need for inspections into an area of high rates of incidence. The same scenario of a high incident rate occurs for runway-induced damage from stone impacts in the vicinity of the landing gear and flaps damage due to hail. Impact damage on sensitive aerodynamic structures such as the radome and leading edges of the aerodynamic profiles bear special consideration, making these additional examples, along with the areas of high incidence rates, good candidates for potential SHM applications.

AD events often invoke the need for unscheduled maintenance. Unscheduled maintenance is a basic element of the continued airworthiness maintenance program developed by the commercial aircraft operators in addition to the scheduled maintenance program described above and approved by an appropriate regulatory agency such as the European Aviation Safety Administration (EASA) or FAA [44]. These programs address procedures for correcting discrepancies noted during scheduled maintenance tasks as well as performing unscheduled maintenance to correct issues that arise due to operational malfunction or abnormal operation. For example, the need to perform unscheduled maintenance can be invoked when the aircraft structure is exposed to unforeseen events such as hard landings, overweight landings, tail strikes, or lightning strikes.

Aircraft routinely depart from an airport within the acceptable maximum takeoff weight, but return before sufficient fuel is burned to reduce the aircraft weight below the maximum allowed landing weight. The pilot may, for example, discover the need to return to the airport immediately after takeoff without enough time to burn off sufficient fuel and avoid causing an overweight landing event. When an overweight landing occurs, an inspection of components such as the landing gear and landing gear mounting points is required. The occurrence of a hard landing, high drag, and side-load landings also invoke inspection procedures since the outcome from these events can vary in seriousness from simply causing mild passenger discomfort to situations resulting in serious vehicle damage. Currently, there are few planes that are equipped with even rudimentary aircraft condition monitoring systems (ACMSs) that can indicate the severity of a landing. Commonly, it is the pilot's decision alone, based on his/her impressions of the event, as to whether a structural inspection is necessary following a hard landing. Owing to the paucity of monitored data, guidance to determine if a high drag or side-load landing has occurred is given in terms of qualitative assessment criteria, such as "the airplane skidded on the runway sufficiently to make you think damage occurred." Load exceedance monitoring, aircraft condition monitoring, and other type of SHM systems could be used to help detect and classify the nature of a landing, as well as monitoring, assessing, and advising the maintenance crew on any damage sustained as a result of the event.

4.3 Environmental deterioration

As defined within MSG-3, environmental deterioration is the structural deterioration that occurs because of chemical interaction with the environment. Damage assessment programs need to cover

corrosion, stress corrosion cracking, and deterioration mechanisms that attack nonmetallic materials. After fatigue, corrosion is the second most prevalent reason for aircraft component failure, followed by overload and wear/abrasion. Yet in terms of expense, corrosion is the single most important measure affecting maintenance costs. As previously discussed, we again observe that the level of maintenance depends strongly on the age of an airplane. Breaking the service life of the aircraft into three phases: new, mature, and the aging or postdesign life, we find that for an aircraft experiencing 2500 flight hours per year, that it behaves like "new" for its first five to six years, and then enters a mature phase for five to six years, and beyond this, the airframe enters its aging phase. It is during the aging phase that an aircraft goes through the fastest rate of increase in the aging process, and incurs maximum maintenance expense [45].

There are various forms of corrosion: uniform, pitting, crevice, exfoliation, galvanic, and stress corrosion, and each poses a different problem to aircraft structures. The most pernicious of these is stress corrosion since it is mechanically assisted and leads to cracks that can undermine the component fatigue strength. As a result, *mass loss* is not a good indicator of the detrimental effect that stress corrosion cracking, nor, for that matter, pitting or intergranular corrosion can have on the integrity of a structure. Stress corrosion cracking was found to be particularly troublesome in some early high-strength aluminum alloys such as 7079 and 7075-T6.

Although modern design practices have made significant headway in the design of aircraft to control corrosion through the appropriate choice of materials, application of coatings, effective use of drainage, sealants, and corrosion-inhibiting compounds, there remain areas where inspection is difficult such as hidden corrosion and areas that require additional effort to gain access. Accidental water spillage around galleys and lavatories leads to enhanced corrosion, so they are target areas for SHM corrosion monitoring systems. The environment under which an aircraft operates is also a significant factor that drives corrosion. When the aircraft is operated in a severe environment, the FAA AC 43-4A [46] recommends that inspections are carried out every 15 days (calendar, not flight-hour based) in addition to the daily routine and preflight inspections. These inspections should include trouble areas and internal cavities, which often require removal of access panels and plates. SHM could be an important factor in improving the effectiveness of these inspections, but as we proceed, we will show that SHM is more than just another technique to perform inspection, it is a new approach to maintenance that has implications to a larger operational environment that includes logistic, supply chain, and fleet management.

5 INTEGRATED VEHICLE HEALTH MANAGEMENT

Integrated vehicle health management (IVHM) is defined variously as "the set of activities performed to identify, mitigate, and resolve faults with the vehicle" [47] as well as the act of "monitoring, assessing, and predicting the health of aircraft materials and structures using networks of sophisticated onboard sensors" [48]. Despite differences in the definition of IVHM, there is an agreement that SHM is part of an emerging set of health-management technologies that when combined can radically change how we operate and maintain aircraft. NASA believes that IVHM technologies have the potential to substantially improve aviation safety. Major airframe manufacturers expect that IVHM can reduce the effects of aging aircraft, reduce operational logistic footprints, reduce training requirements, more effectively manage operational objectives with maintenance requirements as well as enable a proactive maintenance response wherein the vehicle is maintained according to a condition-based maintenance (CBM) philosophy and not according to a schedule.

IVHM is more than just SHM and includes the critical element of prognosis. Consider the Joint Strike Fighter (JSF) Program, a example of state of the art aircraft IVHM. Existing and purpose-built sensors supply data that is preprocessed and then submitted to reasoners dedicated to each of the aircraft's major subsystems. There are area-level reasoners and system-level, model-based, reasoners hierarchically partitioned within the IVHM design. Prognosis data is used as input into ground-based management tools, which then optimize the maintenance and manage the

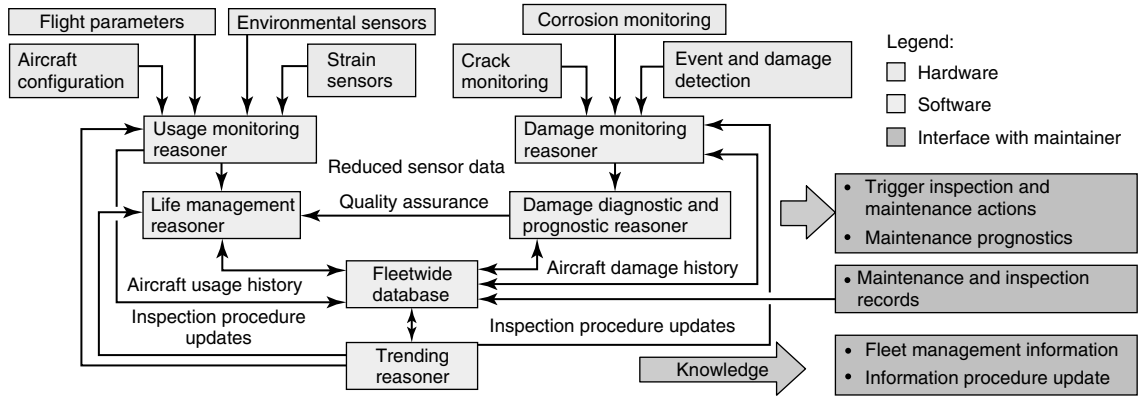


Figure 8. A Boeing commercial aircraft IVHM approach that incorporates SHM functionality. [Reproduced with permission from Ref. 49. © The Boeing Company, 2008.]

supply chain for individual vehicles as well as the fleet.

In the commercial sector, IVHM capabilities currently exist on Boeing 777 and 747-400 airplanes, but these IVHM systems do not presently include SHM functionality. Figure 8 shows a Boeing concept for including SHM prognostic capability as part of an IVHM system. A variety of sensor and flight parameters address the three fundamental aircraft structural damage mechanisms as shown in Figure 7. Load monitoring includes monitoring of strain, and a variety of environmental parameters such as temperature and various flight parameters. These data are then used to establish airframe exposure. Reasoners are used to perform both diagnosis and prognosis, which then drive ground-based decision algorithms for triggering inspections, maintenance actions, and planning at both the platform and fleet level. Figure 9 illustrates where the onboard components of such an SHM system would be included within the existing airplane health-management architecture.

The present Boeing 777 and 747-400 IVHM systems are third and fourth generations of vehicle health-management systems that began from humble beginnings, Figure 10. In the early 1980s, airplane systems and flight decks were transitioning from electromechanical instruments to software-intensive digital computers. As a consequence, significant technological evolution of the nature of the airplane system faults changed dramatically. A serious problem arose as mechanics were unable to use their traditional approach to troubleshooting. Equipment was being replaced under a *shotgun* mentality that lead to a large number of no *fault-found* when the replaced avionics item was returned for repair. In response, industry cooperated with the Aeronautical Radio Inc (ARINC) to develop a standard that provided guidance on how avionics equipment could

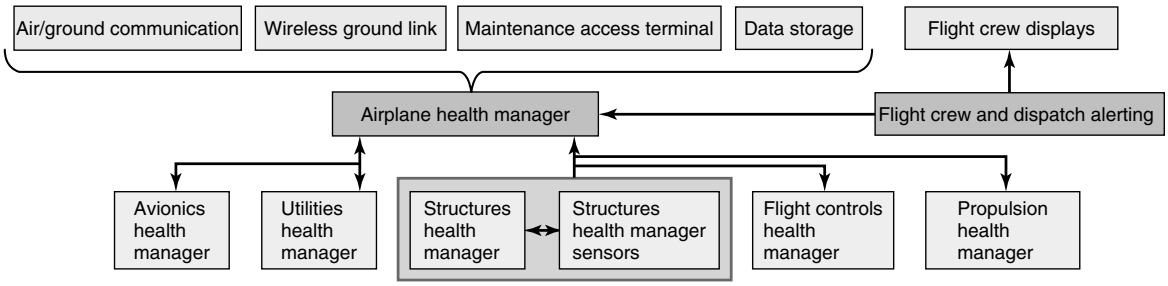


Figure 9. The schematic for the onboard portion of an integrated vehicle health-management approach. [Reproduced with permission from Ref. 49. © The Boeing Company, 2008.]

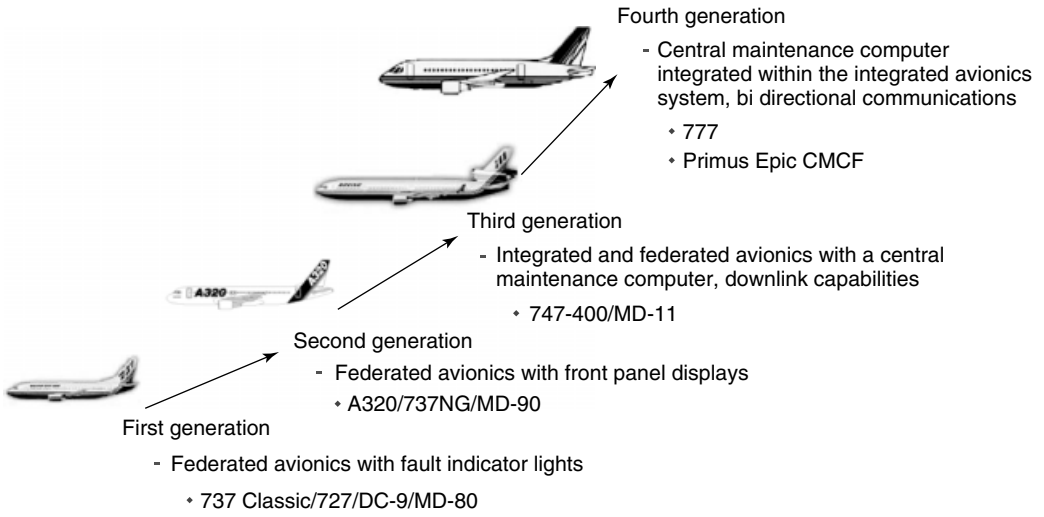


Figure 10. Vehicle health-management evolution in commercial aircraft.

make a self-check obviating the need for external test equipment: ARINC-604 *Guidance for design and use of built-in test equipment* [21].

The Boeing 747-400 pushed the concept forward by consolidating the symptoms from multiple line replaceable units (LRU) into a central maintenance computer (CMC), which could act as a centralized diagnostics reasoner. In the Boeing 777 onboard maintenance system, an Aircraft condition monitoring system (ACMS) provides additional capabilities in the form of simple prognosis. The ACMS is used to trend data and record specific data sets based on triggering conditions. This has proven to be useful for life extension of engines, monitoring fuel consumption, troubleshooting intermittent problems and trying to predicting failure events before they occur. The Boeing 787 will also use the Honeywell patented CMC and ACMS technologies within a networking infrastructure that allows airborne and onboard function to interact with ground-based tools. As indicated by Figure 9, the SHM functionality can be added into these IVHM architectures as a member system enhancing the value and providing a truly integrated and complete vehicle health-management system. It should be noted that alternative approaches and architectures for IVHM systems have been explored. Most notably, the Open System Architecture for Condition-based Maintenance (OSA/CBM) (*see* **Open Systems Architecture for Condition-based Maintenance**), based on the machinery information management open system alliance (MIMOSA) [49]. This approach has received significant research attention but it has yet to experience the same level of aerospace industry acceptance.

As the IVHM system grows in value, by providing a more complete awareness of the vehicle's health or a greater degree of future state prediction, so does the value of e-enabled operations. With propulsion systems leading the way, the powerful combination of IVHM, information technology, and modern network communications is poised to change the entire process of how the commercial air transport industry manage their assets. It will be an e-enabled maintenance revolution. Advanced knowledge of when a system or component is going to fail in combination with a mitigation strategy, such as positioning replacement parts where they can be installed with the least interruption to service, will result in dramatically fewer maintenance delays and flight cancellations. However, for e-enabled maintenance to succeed, it must be able to exploit a foundation of proactive and predictive maintenance, a requirement that SHM supports.

6 CONCLUSION

A number of examples have been given where SHM systems can be used to supplement or replace existing

inspection approaches to improve safety and/or minimize the inspection burden. Discussions on when in an aircraft's life-cycle SHM solutions should be introduced continue but a few conclusions seem clear. Load monitoring can be beneficially used when an aircraft enters service to monitor an aircraft's operational conditions and through numerical modeling estimate the accumulated loading experienced at any structural location. This monitoring would also ensure that loads exceeding the assumed design spectrum are appropriately noted and used to adjusted inspection procedures. AD that occurs randomly in time, but predictably in a certain area, is a scenario where SHM can be beneficially applied to both new and older planes. SHM could be an important factor in improving the effectiveness and reducing the burden of additional inspection that will inevitably result as the commercial aircraft fleets continue to age. As a member system within an IVHM implementation, SHM is a significant component of a truly integrated health-management approach where e-enabled maintenance could exploit the diagnostic and prognostic capabilities of the intelligent vehicle and dramatically change the aircraft support processes within a network centric model. In the future, SHM could play a significant role in how airplanes are designed, controlled, and tasked. SHM technologies will enable an intelligent vehicle scenario where the ability to automate, reconfigure, and adapt the vehicle has a major impact on operational scenarios.

REFERENCES

[1] Chang FK (ed). *Proceeding of the International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 1997, 1999, 2001, 2003, 2005, 2007.

[2] Balageas D, Boller C, Staszewski WJ, Gordon's G, Güemes A (eds). *Proceeding of the European Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2002, 2004, 2006.

[3] Staszewski WJ, Boller C, Tomlinson GR (eds). *Health Monitoring of Aircraft Structures*. John Wiley & Sons: West Sussex, 2003.

[4] *Structural Health Monitoring—An International Journal*. Sage Publications: London, 2002.

[5] Adams DE. *Health Monitoring of Structural Materials and Components: Methods with Applications*. John Wiley & Sons: West Sussex, 2007.

[6] Achenbach JD. On the road from schedule-based nondestructive inspection to structural health monitoring. *Proceeding of the 6th International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2007; pp. 16–28.

[7] Kent RM, Murphy D. *Health Monitoring System Technology Assessments: Cost Benefit Analysis*, NASA Technical Report NASA/CR-2000-209848, NASA, 2000.

[8] Bartelds G. Aircraft structural health monitoring, prospects for smart solutions from a European viewpoint. *Journal of Intelligent Material Systems and Structures* 1998 **11**:906–910.

[9] Hess R, Unger R, Reuter R. *On the Technical and Economic Feasibility of Load and Environmental Monitoring for Transport Aircraft Structural Health Monitoring*. Aerospace Vehicle Systems Institute: AFE Project 27, 2003.

[10] Foote P, McFeat J, Heimes F, Haugse E, Duke A, Hochmann D, Gordon GA. *Structural Health Monitoring Road Mapping Process*. Aerospace Vehicle Systems Institute, AFE Project 53, 2006.

[11] Munns TE, Beard RE, Culp AM, Murphy DA, Kent RM. *Analysis of Regulatory Guidance for Health Monitoring*, NASA Technical Report. NASA/CR-2000-210643, NASA, 2000.

[12] Beral B, Speckmann H. Proceedings of the 4th international workshop on structural health monitoring (SHM) for aircraft structures. *Workshop on SHM, 4th International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2003; pp. 12–29.

[13] Speckmann H, Henrich R. Structural health monitoring (SHM)—overview on technologies under development. *16th World Conference on NDT*. Montreal, QC, 2004; see also: http://www.ndt.net/search/docs.php3.

[14] Speckmann H, Roesner H. Structural health monitoring: a contribution to the intelligent aircraft structure. *Proceedings of the European Conference on NDT*. Paper Tu.1.1.1, 2006; pp. 1–7, see also: http://www.ndt.net/article/ecndt2006/papers~1.htm.

[15] Trego A. *Installation of the Autonomous Structural Integrity Monitoring System; 4th International Workshop on Structural Health Monitoring*, Chang FK, Güemes A (eds). DEStech Publications: Lancaster, PA, 2003; pp. 863–870.

[16] Trego A, Akdeniz A, Haugse E. *Structural Health Management Technology on Commercial Airplanes; Proceedings of the 2nd European Workshop on Structural Health Monitoring*, Boller C, Staszewski WJ

(eds). DEStech Publications: Lancaster, PA, 2004; pp. 317–322.

[17] Buderath M. Maintaining ageing military aircraft using the tornado fighter as an example. In *Structural Health Monitoring*, Balageas D (ed). DEStech Publications: Lancaster, PA, 2002; pp. 76–96.

[18] Buderath M. *Review the Process of Integrating SHM Systems into Condition Based Maintenance as Part of the Structural Integrity Programme; Proceedings of the 2nd European Workshop on Structural Health Monitoring*, Boller C, Staszewski WJ (eds). DEStech Publications: Lancaster, PA, 2004; pp. 307–316.

[19] Giurgiutiu V. *Damage Assessment of Structures—An Airforce Office of Scientific Research Structural Mechanics Perspective. III ECCOMAS Thematic Conference on Smart Structures and Materials*, Ostachowicz W, Holnicki-Szulc J, Mota Soares C (eds). Gdansk: Poland, July 2007; pp. 9–11.

[20] Derriso MM, Olson SE, Desimio MP, Pratt DM. Why are there few Fielded SHM systems for aerospace structures? *Proceedings of the 6th International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster PA, 2007; pp. 44–55.

[21] Bird G, Christensen M, Lutz D, Scandura P. Use of integrated vehicle health management in the field of commercial aviation. *NASA First International Forum on Integrated System Health Engineering and Management in Aerospace*. November 2005; see also: http://ti.arc.nasa.gov/projects/ishem/papers_pres.php.

[22] Schijve J. Fatigue of aircraft materials and structures. *International Journal of Fatigue* 1994 **16**:21–32.

[23] Armitage SR, Holford DM. *Future Fatigue Monitoring Systems*, NATO RTO-MP-7; Paper 2, 1998.

[24] Caron Y, Richard Y. *CF-188 Fatigue Life Management Program*, NATO RTO-MP-7; Paper 4, 1998.

[25] van der Hoeven AM. *CC130 Data Analysis System for OLM/IAT*, NATO RTO-MP-7; Paper 14, 1998.

[26] Spiekhout J. *F-16 Loads/Usage Monitoring*, NATO RTO-MP-7; Paper 13, 1998.

[27] Amabile P, Giacobbe T. *Proposal for the New Fatigue Management for the AMX*, NATO AGARD-CP-506; Paper 9, 1991.

[28] Krauβ A. Betriebslastenermittlung für Flugzeugentwurf und—entwicklung. *Proceedings of the 14th Meeting of DVM AK Betriebsfestigkeit, Rüsselsheim (in German)*, 1988.

[29] Hunt SG, Hebden IG. Validation of the eurofighter typhoon structural health and usage monitoring system. *Smart Materials Structures* 2001 **10**:497–503.

[30] Ladda V, Meyer H-J. *The Operational Loads Monitoring System OLMS*, NATO AGARD-CP-506; Paper 15, 1991.

[31] Chang FK. Smart layer: built-in diagnostics for composite structures. *Smart Materials and Structures—Proceedings of the 4th European and 2nd MIMR Conference*. Harrogate, 6–8 July 1998; pp. 777–781.

[32] Raghavan A, Cesnik CES. Review of guided-wave structural health monitoring. *The Shock and Vibration Digest* 2007 **39**:91–114.

[33] Gordon GA, Braunling R. Quantitative corrosion monitoring and detection using ultrasonic Lamb waves. *Proceedings of the SPIE Smart Structures and Materials 2005: Sensors and Smart Structures Technologies for Civil, Mechanical and Aerospace Systems*. 2005; Vol. 5765 pp. 504–515.

[34] Washabaugh AP, Zilberstein VA, Schlicker DE, Scheiretov Y, Grundy D, Goldfine NJ. Shaped-field eddy-current sensors and arrays. *Proceedings of SPIE Conference 4702 Smart NDE and Health Monitoring of Structural and Biological Systems*. San Diego, CA, 2002; pp. 63–75.

[35] Roach D, Rackow K, DeLong W, Yepez S, Reedy D, White S. *Use of Composite Materials, Health Monitoring and Self-Healing Concepts to Refurbish Our Civil and Military Infrastructure*. Report SAND2007-5547, Sandia National Labs, Albuquerque, NM, 2007.

[36] Kirikera GR, Shinde V, Schulz MJ, Ghoshal A, Sundaresan MJ, Allemang RJ, Lee JW. A structural neural system for real-time health monitoring of composite materials. *Structural Health Monitoring* 2008 **7**:65–83.

[37] Finlayson RD, Friesel M, Carlos M, Cole P, Lenain JC. *Health Monitoring of Aerospace Structures with Acoustic Emission and Acousto-Ultrasonics*. Insight, 43, 2001: see also http://www.ndt.net/article/wcndt00/index.htm.

[38] Zhou1 G, Sim LM. Damage detection and assessment in fibre-reinforced composite structures with embedded fibre optic sensors—review. *Smart Mater. Struct.* 2002 **11**:925–939.

[39] McAdam G, Newman PJ, McKenzie I, Davis C, Hinton BRW. Fiber optic sensors for detection of corrosion within aircraft. *Structural Health Monitoring* 2005 **4**:47–56.

[40] Park G, Inman DJ, Farrar CF. Recent studies in piezoelectric impedance-based structural health monitoring. *Proceedings of the 4th International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2003; pp. 1423–1430.

[41] Schmidt H-J, Schmidt-Brandecker B. Design benefits in aeronautics resulting from structural health monitoring. *Proceedings of the 6th International Workshop on Structural Health Monitoring*. DEStech Publications: Lancaster, PA, 2007; pp. 762–769.

[42] *Structural Maintenance Program Guidelines for Continuing Airworthiness*, ATA Report 51-93-01. Air Transport Association of America, Washington, DC, 1993.

[43] ATA MSG-3. *Operator/Manufacturer Scheduled Maintenance Development, Revision 2005. 1*. Air Transport Association of America, Washington, DC, 2005.

[44] AC 120-16C. *Continuous Airworthiness Maintenance Programs*. Federal Aviation Administration, Washington, DC, 1980.

[45] Koch GH, Brongers MPH, Thompson NG, Virmani YP, Payer JH. *Corrosion Costs and Prevention Strategies in the United States*, Report FHWA-RD-01-156. Federal Highway Admin, McLean, VA, September 2001, see also: http://www.corrosioncost.com/home.html.

[46] A8 AC 43-4. *Corrosion Control for Aircraft*. Federal Aviation Administration: Washington, DC, 1991.

[47] Aaseng GB. Blueprint for an integrated vehicle health management system. *IEEE 20th Digital Avionics Systems Conference*. 2001; pp. 3C1/1–1C1/11.

[48] National Research Council. *Decadal Survey of Civil Aeronautics, Foundations for the Future*. The National Academic Press: Washington, DC, 2006.

[49] MIMOSA Web Site: http://www.mimosa.org/, 2008.

Chapter 118
The Character of SHM in Civil Engineering

Helmut Wenzel
VCE Holding GmbH, Vienna, Austria

1 INTRODUCTION

Bridges are the flagships of civil engineering. They attract the highest attention within the engineering community. This is due to their small safety margins and their great exposure to the public. Early bridges were the backbone of powerful empires from China to Rome and the Incas in America. Currently, the transportation infrastructure is directly related to the economic success of a nation. Bridges are admired not only for their function but also primarily for their aesthetic impact. Imagine New York without bridges, Japan without the Honshu Shikoku project, or Europe without the Greatbelt Link. This article deals with the preservation and maintenance of these important elements of modern society.

Structural health monitoring (SHM) is the implementation of a damage identification strategy to the civil engineering infrastructure. Damage is defined as changes to the material and/or geometric properties of these systems, including changes to the boundary conditions and system connectivity. Damage affects the current or future performance of these systems.

The damage identification process is generally structured into levels:

1. damage detection, where the presence of damage is identified;
2. damage location, where the location of the damage is determined;
3. damage typification, where the type of damage is determined; and
4. damage extent, where the severity of damage is assessed.

Extensive literature has developed on SHM over the last 20 years. This field has matured to a point where several accepted general principles have emerged. Nevertheless, these principles are still being challenged and further developed by various groups of interest. The strategies in mechanical engineering or aerospace are taking different approaches. Nevertheless, the civil engineering community can considerably benefit from these efforts.

Separate approaches are necessary while considering each civil engineering structure as a prototype.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 HEALTH MONITORING FOR CIVIL ENGINEERING STRUCTURES

In civil engineering, the procedure and tools are best developed for bridges. Some kind of SHM always existed in this sector. Figure 1 shows how these procedures have developed from simple inspection routines to highly sophisticated monitoring campaigns.

The extent of monitoring mainly depends on the required results. Currently, five levels are used in order to determine the depths of investigation. These are as follows:

Level 1: rating
It represents the conventional assessment of the structure starting with a visual field inspection that provides a subjective impression of the condition of the structure. Some preliminary analytical investigation is performed to provide a rating as a basis for decisions. This would be the typical application of a bridge management system like PONTIS or DANBRO. Many bridge owners use databases to store the results.

Level 2: condition assessment
A rough visual field inspection has to be an element of any SHM campaign. After that a decision has to be made whether the conventional approach is

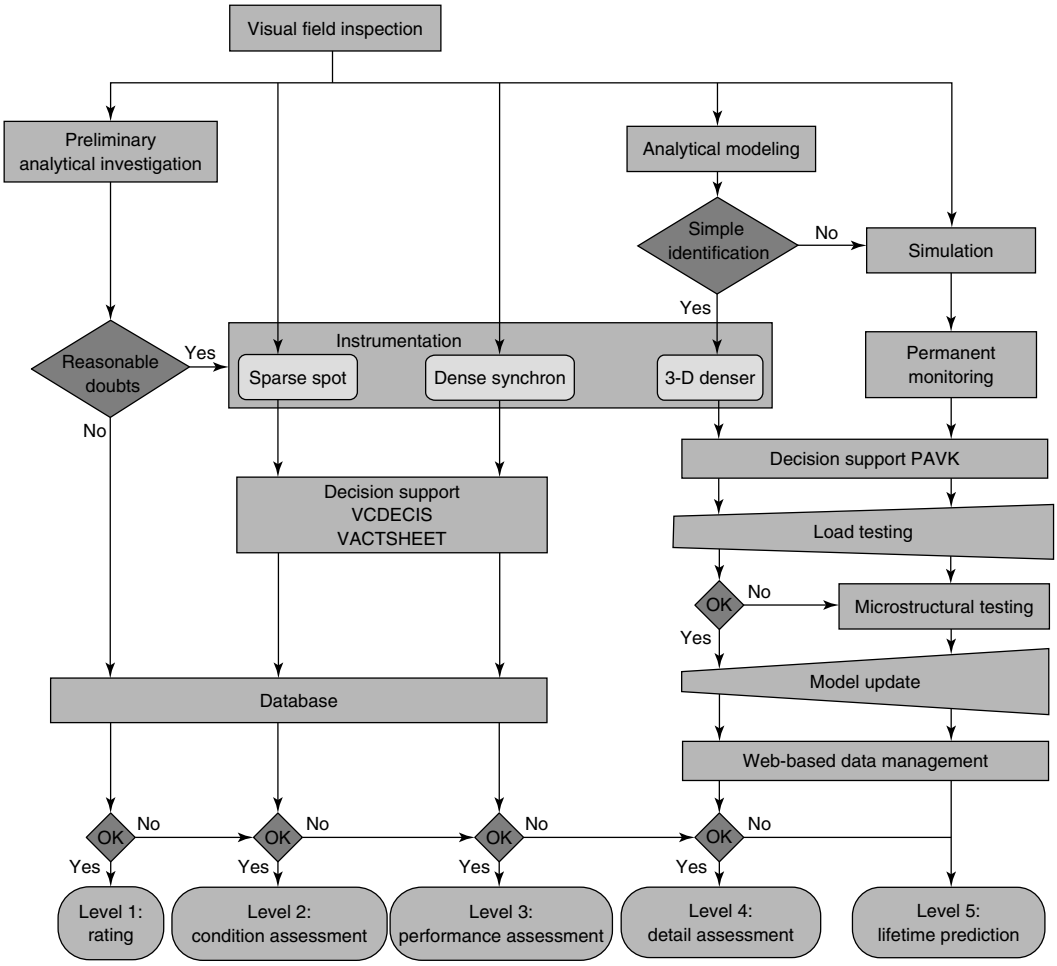


Figure 1. Typical hierarchical concept for SHM procedure for bridges.

satisfactory or an extended or even sophisticated additional approach is to be taken. This determines the type and quantity of instrumentation. For condition assessment, a simple instrumentation is sufficient and a simple decision support system will provide the necessary additional information. Storage and treatment of data should also be done in the existing database. A link to existing conventional tools is available. The monitoring can be performed at single spots only.

Level 3: performance assessment
This intermediate level uses the same procedure described under level 2. The level of assessment and performance elaboration in the decision support process is considerably higher as additional information like mode shapes is measured and elaborated. This provides additional indicators for the assessment and demonstrates the performance of the structure. It obviously requires a denser instrumentation and synchronous monitoring.

Level 4: detail assessment and rating
The next step is to establish an analytical model representing the structure. The model is compared with the monitoring results. In case that identification is simple, a step back toward level 3 might be taken. In case phenomena that cannot be explained from the records are detected, further steps have to be taken to deal with the situation. The most obvious thing is to introduce a permanent record over some period of time to capture the necessary phenomena valid for this specific case. Load testing also has been proven successful to establish performance parameters. With these results, a simple model update can be performed to assess the results and provide a rating. Extensive monitoring is required. The records shall cover at least 24 h, but shall rather be much longer to capture environment and traffic situations.

Level 5: lifetime prediction
For a serious lifetime prediction, the records available have to be long enough to cover at least three cycles relevant for the structure. This is normally in the order of three years. Simulation should be run from the analytical model to achieve a theoretical performance to be compared with. To handle the major quantity of data, special software for decision support is required. Load testing should be targeted and extensive. In addition, microstructural testing might be useful in order to look into the performance of individual elements of a structure. The update process will be extensive considering several conditions of the structure. This includes, in particular, the loaded and unloaded case and all the nonlinearities involved. In the case of reasonable doubts, this monitoring system shall be operated on-line and web based with a warning computed by the decision support. The final lifetime prediction can then be performed as described in **Condition Compensation in Frequency Analyses—a Basis for Damage Detection**.

The costs related to these procedures are given in Figure 2. These costs mainly depend on the extent

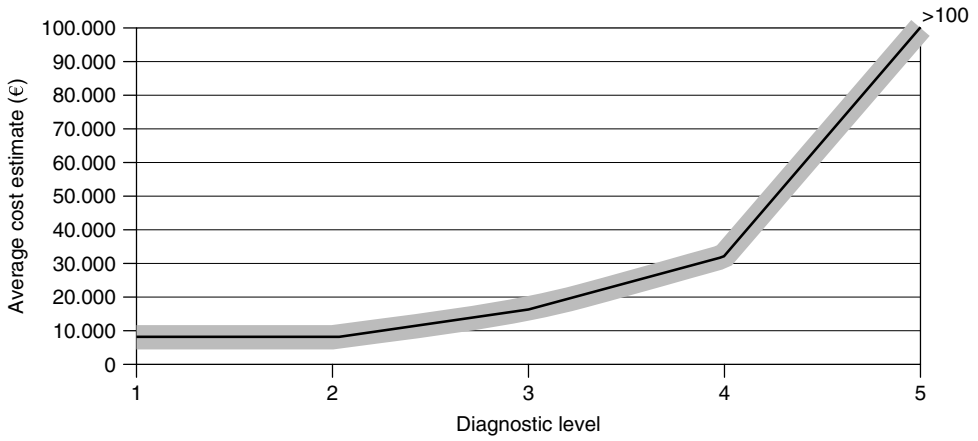


Figure 2. Cost of monitoring campaigns for a typical three-span bridge (€, base 2006).

of the monitoring campaign and the number of man hours to be invested in modeling, simulation, and update procedures. The figures provided in the graph are based on the prices in 2006 for a typical three-span bridge with an average length of 150 m. The prices can also be influenced by the number of spans, by the type of the structure, and also, in particular, by the condition for the monitoring campaign. It is expected that the prices will be rather reduced than increased. This can happen through the introduction of time-saving modeling procedures and sophisticated monitoring software. Nevertheless, these are still to be developed.

3 CLIENT REQUIREMENTS AND MOTIVATION

The construction sector is conservative. The implementation of new technologies needs a clear requirement and motivation to be accepted by owners and operators. It has been recognized that the current practice does not satisfy the needs of shrinking budgets and aging structures. Nevertheless, it satisfies valid codes and standards. Before a breakthrough in the implementation of new technologies can happen, the requirements and motivation have to be clearly understood and argued against potential clients.

There are three main drivers in the promotion of SHM. The motivation to apply and order services based on the new technologies are as follows:

- responsibility-driven motivation, which means that the new methods, to become standard applications, are supported by codes, standards, and guidelines;
- economically driven motivations, such as in situations where a ranking of structures to be rehabilitated is necessary because of insufficient budget available or the need to use a structure for a certain time period longer than designed for; and
- curiosity-driven motivations comprise those cases where clients would like to know more about their important and complicated structures. Results can also lead to better planning for future structures.

From the above-mentioned motivation, the following requirements can be derived; these are typical services requested from the technology providers.

- A certificate that a structure satisfies the requirements with respect to codes, standards, and guidelines comprises a main business opportunity. Many recommendations already consider the increase in maintenance periods in case these measurements are taken. The provision of such certificates by engineers is common practice in Europe. Other parts of the world do not apply this system. It has led to an impressive evolution of bridge technology in Europe, which has been exported worldwide. It creates the environment for quality construction.
- The transfer of liabilities and responsibilities for structures in terms of technical and operational matters takes place with the huge privatization drive that we can observe currently. Clients are systematically transferring the stock of structures into private hands. The new players involved are open to new applications that are able to support innovative and economic maintenance strategies.
- Special structures require special attention. The necessary top expertise may not always be available with every owner or operator. The top experts for each region require the newest technologies for their work.
- A shortage in capacities of personnel to carry out the routine maintenance and assessment works at the bridge stock also leads to new opportunities. As these services are normally tendered, new technologies might have an economic and technology edge.
- In case of emergency or accidents, the generation of a secure situation is desired by affected owners. Any assessment based on the results of measurements is more likely to be accepted than subjective assessment by the expert. The clients want to sleep well because somebody else is permanently watching and assessing their structures.
- *Ad hoc* assessment in case of doubts or emergency also comprises this application area. The subjective conventional assessment produces too many negative scores on structures and doubts are raised. A quantitative assessment is desired.
- The optimization of maintenance concepts requires input on which this process can be performed. The more data are available, the better the organization will be and better maintenance concepts will be available. The reduction of the remaining

risks helps to make decisions with lower safety margins.
- The determination of priorities, through a quantification based on measurements, helps to satisfy the growing demand in combination with shrinking budgets. This assessment can come up with better scores, minimizing the number of structures requiring immediate intervention. Decision support for investment planning can be offered on the basis of the above-mentioned services. Every new measurement improves the database and as such improves the quality of the results and supports the necessary decision making.
- Life-cycle cost determination helps to increase the periods where budgetary planning is necessary. The demand for retrofit and maintenance can be estimated over the whole life period of a structure or even a fleet of structures.
- The direct link of structural performance to the operation of a structure can be established. Very often information about an optimal speed or frequency in the traffic can be determined, which shall be used by the operation personnel of a transportation infrastructure and communicated to the drivers through telematic devices.
- Hot-spot identification technologies are very often requested in case the weakest point of the system or a significant accumulation of incidents is observed. Clients would like to know where to look first and what the background of certain phenomena could be.
- The prediction of structural performance for future loading scenarios is a further specific item requested, particularly when a nonlinear behavior can be expected, special expertise becomes necessary.
- Fleet observation when the number of structures is huge is desired to improve the quality of assessment. For this the conception has to be subdivided into stages depending on the depth of information required.

The selection of a suitable observation concept has to be based on mainly external factors. These are the number of structures to be observed in combination with the budget available. For this purpose, it is necessary to offer services on increasing quality levels. The levels can be subdivided into spot, periodic, permanent, and on-line assessment campaigns at structures. The respective features are as follows:

- A spot observation shall comprise a very quick measurement campaign with a few simple-to-handle sensors only. It shall bring information on the general condition of a structure in order to create a ranking.
- Periodic assessment means a measurement campaign on a structure, which is repeated after a specified period of time, to generate information on the performance over time. This spot information might comprise rather long periods.
- Permanent observation and assessment of structures becomes necessary when certain limits are passed. This observation allows a very detailed assessment based on permanent recordings and can help to implement quick decision making.
- On-line observation and assessment allows warning through electronic media, be it an SMS in the simple case or an on-line status through the Internet. Decisions might be taken by the computer based on the measurement data. These alert systems will only be applied at extremely critical structures.

In general, it has to be stated that clients need and desire a support of their work and not issues that makes it more complicated. Also, in this respect, the procedures have to be carefully watched and permanently improved. The information policy also plays a major role in the client–consultant relationship. The new methodologies are rather complex and require a deep understanding of structural dynamics, physics, and measurement techniques. Owing to the fact that this expertise is rarely available at the owners engineering department, the fear of being exposed to unknown black box applications has to be taken off their shoulders. On the other hand, they are spending considerable amount of money and would like to be informed frequently about progress and results. Therefore, we have to assure them that the technology part is in good and competent hands and that they will receive the necessary information they desire. The best results have been achieved with very simple reporting techniques. A periodic report received by e-mail comprising single page information is preferred. The example shown in Figure 3 provides such a typical weekly report. The main information is provided in a single window, where upper and lower normalized thresholds are given and the measurement results within this period are placed

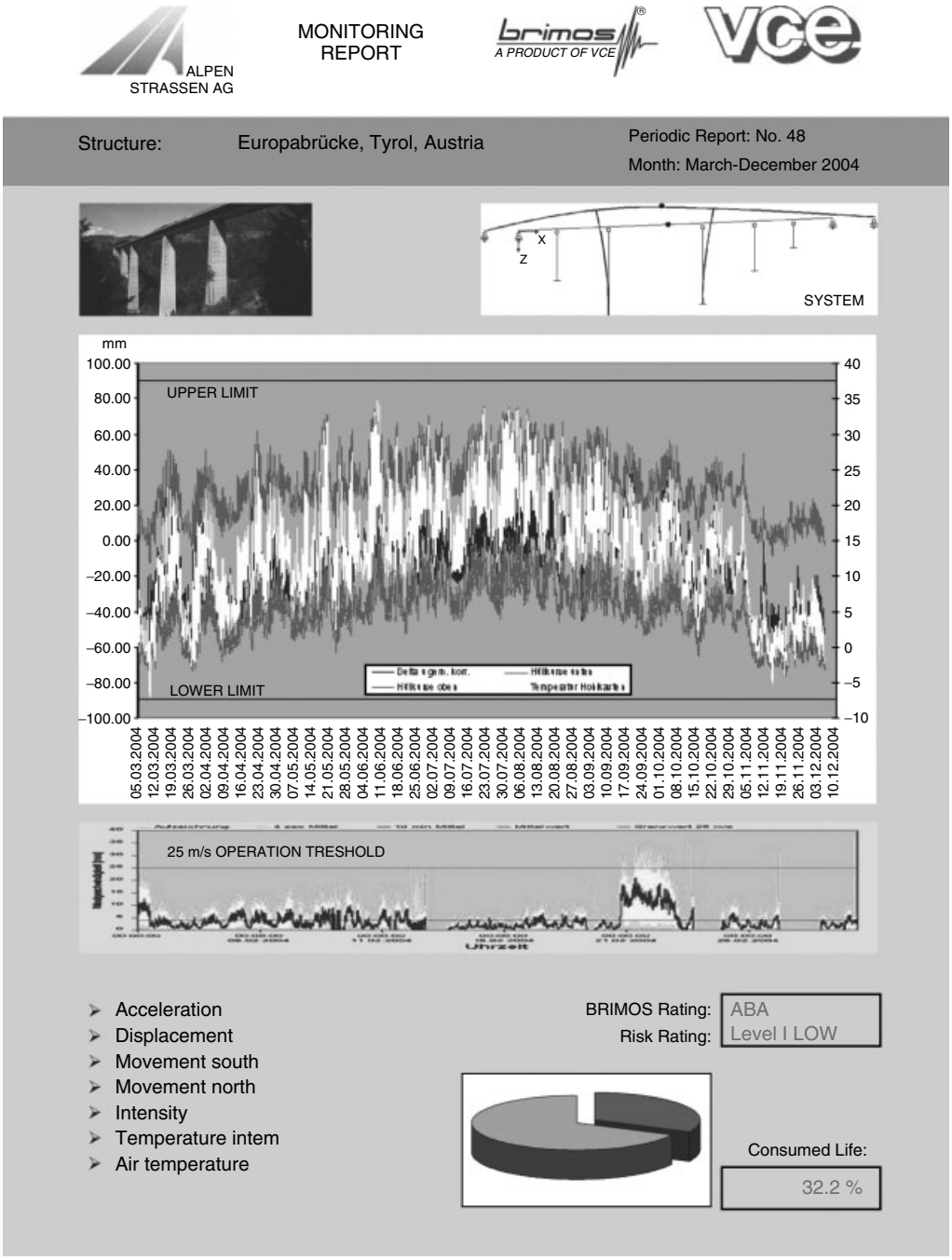


Figure 3. Periodic SHM report of a bridge.

within these thresholds. With one look at this graph, the personnel can immediately see whether any of the thresholds has been exceeded. The client is satisfied because all indicators are green and the ordered observation is permanently working.

The periodic report should provide on this single page the following information:

- A photo and a system plot of the structure under observation for easy and quick identification.
- A window with the periodic results placed within the relevant thresholds over the observation period.
- Eventually, a second window with special information required by the client, such as wind speed information or any other quantity desired.
- Finally, a rating shall be provided, which is based on the measurements taken in the reporting period. This rating shall enable the client to immediately see whether any changes have happened.
- Eventually, the specification of a remaining life capacity can be provided if the necessary data are recorded.

Besides this one-page record for the client, a scientific report shall be generated by the system for the expert. This shall enable a quick assessment of all the single measurements in order to obtain necessary expertise or learn from the performance. Every year, on an average, the system shall be calibrated with the information gained. This might also comprise a change in the rating and will update the remaining life capacity based on the existing knowledge.

FURTHER READING

Blevins RD. *Formulas for Natural Frequency and Mode Shape*. Van Regional Company: New York, 1979.

Cantieni R. *Dynamic Load Tests on Highway Bridges in Switzerland—60 Years Experience of EMPA*. Section Concrete Structures and Components, Report No. 211, Dübendorf, 1983.

De Roeck G, Peeters B, Maeck J. Dynamic monitoring of civil engineering structures. *Computational Methods for Shell and Spatial Structures IASS-IACM 2000*. Greece, 2000.

Forstner E, Wenzel H. *IMAC—Integrated Monitoring and Assessment of Cables*, Final Technical Report D33. IMAC Project, 2004.

Peeters B, De Roeck G. One year monitoring of the z 24-bridge: environmental influences versus damage events. *Proceedings of IMAC 18, The International Modal Analysis Conference*. San Antonio, TX, February 2000; pp. 1570–1576.

Veit R, Wenzel H, Fink J. Measurement data based lifetime-estimation of the Europabrücke due to traffic loading—a three level approach. *International Conference of the International Institute of Welding*. Prague, 2005.

Wenzel H, Pichler D. *Ambient Vibration Monitoring*. John Wiley & Sons: Chichester, 2005, ISBN 0470024305.

Chapter 120
The Influence of Environmental Factors

Helmut Wenzel
VCE Holding GmbH, Vienna, Austria

1 INTRODUCTION

The following analysis is based on investigations on the Europabrücke—a well-known Austrian steel bridge near Innsbruck, opened in 1963—which is part of one of the main alpine north–south routes for urban and freight traffic. A long-term preoccupation of VCE with BRIMOS (BRIdge MOnitoring System) on the Europabrücke (since 1997) led to the installation of a permanent monitoring system in 2003 [1]. Currently, the bridge is stressed by more than 30 000 motor vehicles per day (approximately 20% freight traffic). The superstructure is represented by a steel box girder (width = 10 m; variable height along the bridge length 4.70–7.70 m) and an orthotropic deck and bottom plate (Figure 1). This motorway bridge with six spans of differing lengths (the longest span of 198 m, being supported by piers with an elevation of 190 m) and a total length of 657 m comprises six lanes, three in each direction distributed over a width of almost 25 m.

To reach the already defined goals, a permanent monitoring system has been developed in a stepwise manner (Figure 2). It consists of 24 measuring channels (sampling rate 100 Hz) representing the accelerations of the main span, the pier, and the cantilever, and the dilatation, wind speed and direction, and temperatures of the abutment at several locations.

2 EXEMPLARY PROCEDURE OF ENVIRONMENTAL COMPENSATION

The bridge's reference sensor (3-D forced balance accelerometer) is installed within the main span, at a distance of 0.4 times of the span's length from pier II. At this base point, global stiffness and its dependence to several environmental influences are assessed (sampling rate = 100 Hz, file length = 330 s).

By evaluating the results (frequency spectra) of several measurements, telescoping them together and

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Figure 1. Europabrücke—overview.

viewing them from above (so-called trendcards), the following visuals are obtained (Figure 3), which exemplarily show the main span's relevant vertical stiffness patterns of a particular day with a distinctive progression of temperature.

For the sake of completeness, the corresponding frequency spectra themselves are shown in Figure 4, again over a period of this day. An individual procedure has been developed in [2], which contains some measurement preconditioning (offset elimination and band-pass filtering). To enable a more stabilized automatically performed peak picking in different ranges of frequency, the response spectra are smoothed in the course of frequency assessment.

The permanent monitoring system exhibits a remarkable loading impact, as the bridge is currently stressed by more than 30 000 motor vehicles per day (approximately 20% of them constitute freight traffic). By applying the previously described method to the reference sensor's measurement data for the whole day, a progression of stiffness, which consists of 281 single peaks is obtained (Figure 5a, b) and represents randomly occurring ambient and forced vibration conditions (scatter diagram).

The complementary relation between stiffness and the air temperature (registered at the bridge's base point directly above the pier II) is obvious and can be interpreted as a long sinusoidal wave of the main span in the vertical direction. In the course of the described procedure, it should be considered to omit band-pass filtering and replace it by a further optimized smoothing of the frequency spectra in order to stabilize the accuracy of peak picking.

3 STIFFNESS VERSUS TEMPERATURE

To describe the verified phenomenon mechanically, we focus on the temperature dependence of the roadbed's asphalt layer, since the change of steel characteristics under varying climate conditions is negligible. In the first step, a characteristic relationship of the dynamic Young's modulus in dependence of temperature is used [3].

Owing to this relation, temperature-sensitive asphalt layer is implemented into the cross sections of the global structural analysis model, leading to a distinctive progression of the flexural rigidity of the midspan (Figure 6a, b).

$$f_i = \frac{\lambda_i^2}{2\pi L^2}\left(\frac{EI}{m}\right)^{1/2} \tag{1}$$

According to the widely known equation (1), the frequency of vibration is proportional to the square root of the moment of inertia [4]. For that reason, a frequency curve needs to be generated (Figure 7b) for the next step, when the temperature-based stiffness path is eliminated from the overall trend (Figure 7a–c).

The obtained trend shows very clearly the remaining impact due to the freight traffic itself, which has maximum effect during the time between 5 a.m. and 10 p.m., when trucks pass over the bridge; this causes the two characteristic offsets during the course of the day as shown in Figure 8.

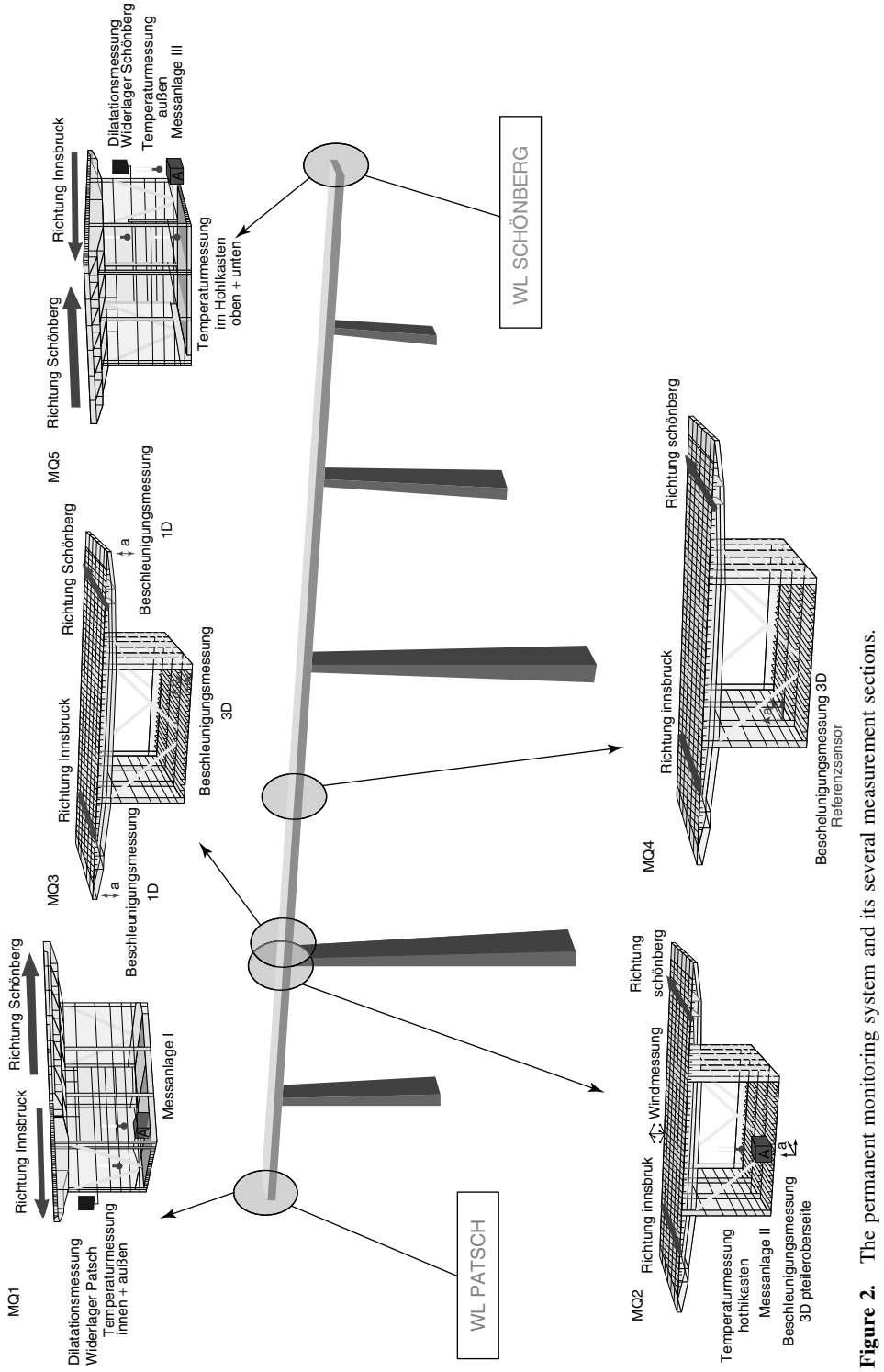


Figure 2. The permanent monitoring system and its several measurement sections.

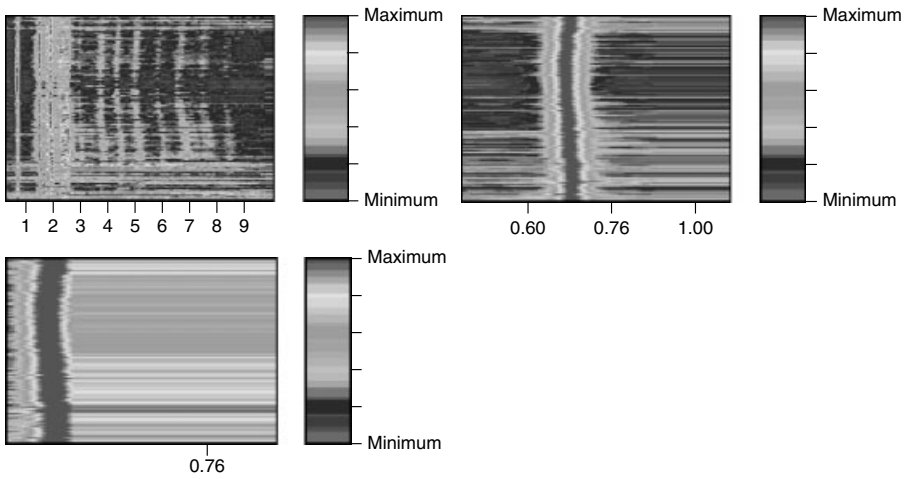


Figure 3. Trend of stiffness during one day: 0.30–10/0.30–1.10/0.60–0.80 Hz. The horizontal axis represents frequencies and the vertical axis represents time.

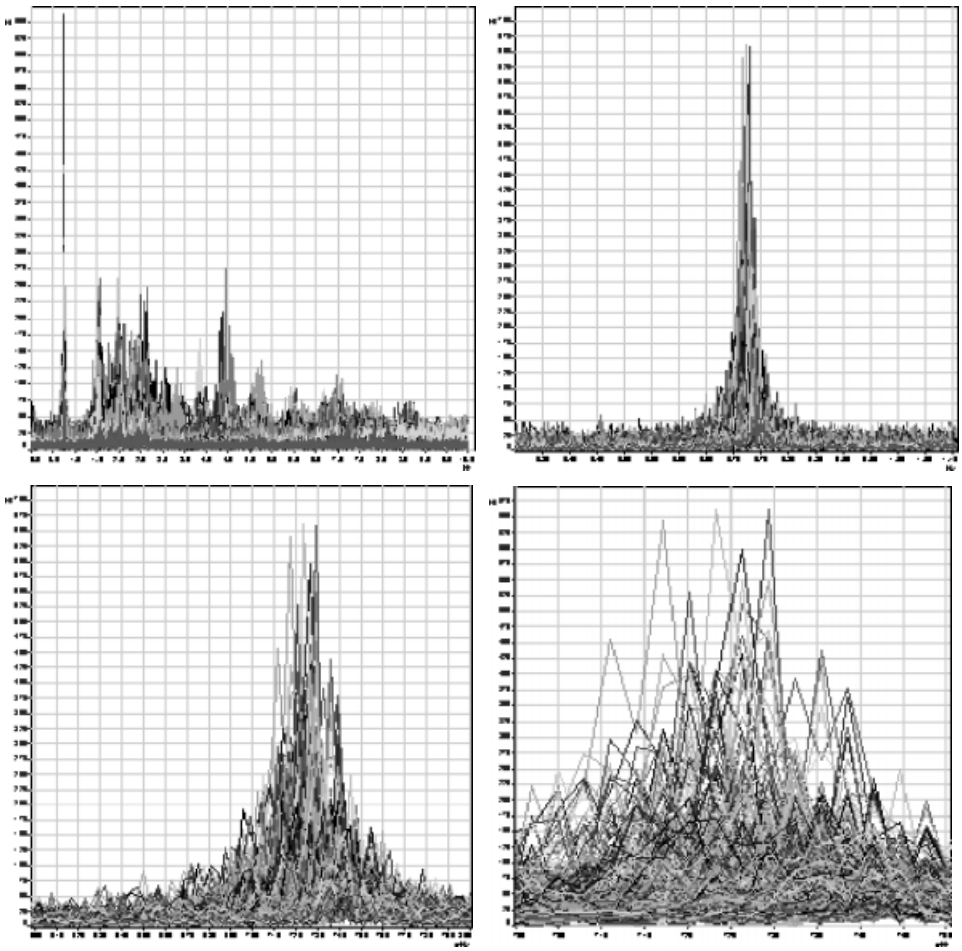


Figure 4. The front views of the trendcard for one day: 0.30–10/0.30–1.10/0.60–0.80/0.68–0.74 Hz.

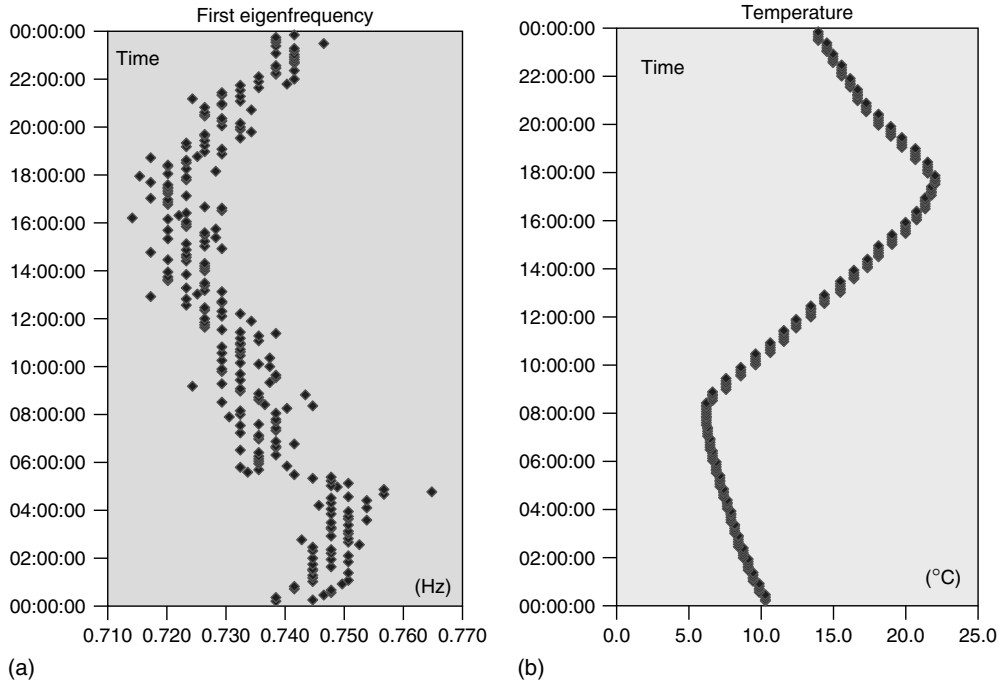


Figure 5. Pattern of the first eigenfrequency (a) and its obvious dependency on temperature (b); the shape of the figures correspond to each other: higher temperatures correspond to lower frequencies.

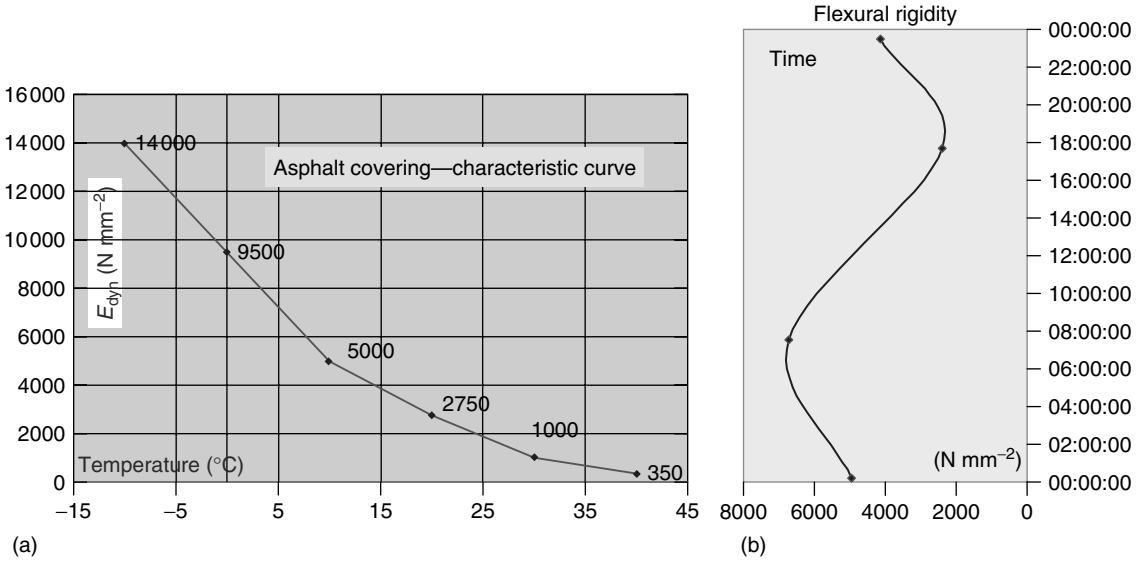


Figure 6. Progression of the asphalt layer's flexural rigidity in dependence of its temperature (a); development of the flexural rigidity over the period of one day (b).

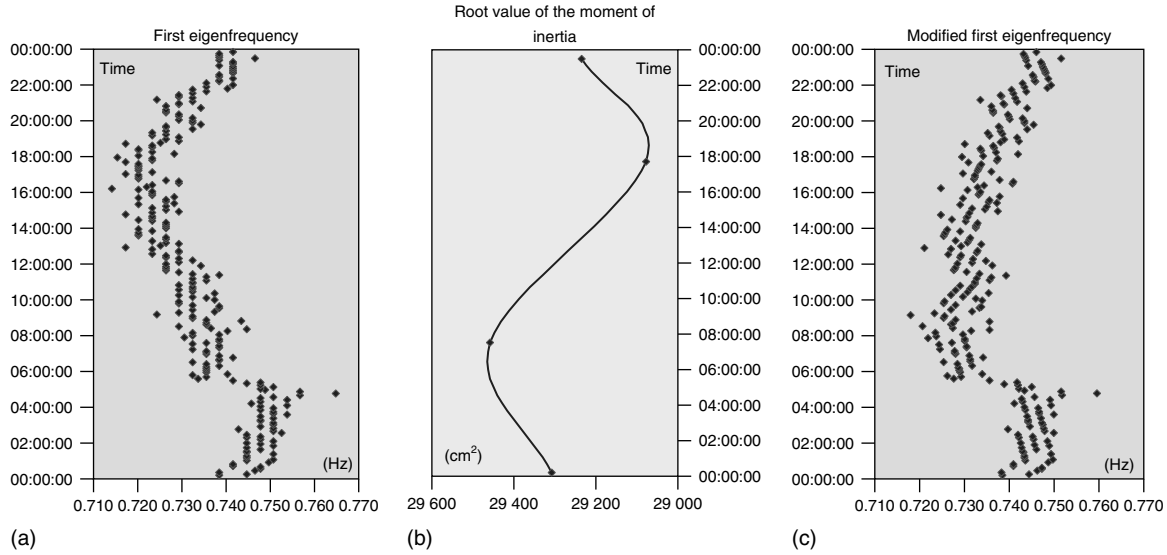


Figure 7. Pattern of the first eigenfrequency before (a) and after (c) compensation of temperature (b).

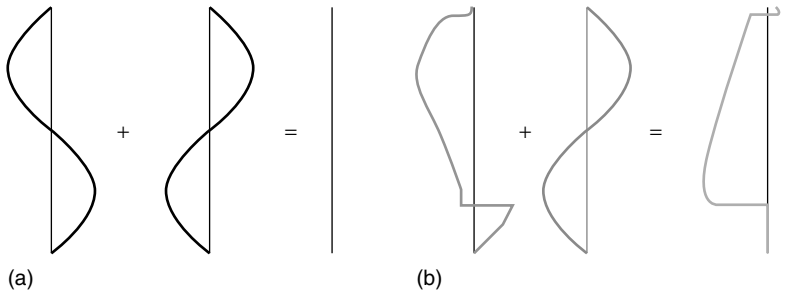


Figure 8. Comparison of expected (a) theoretical and (b) actual (the pattern considering traffic and other loads) consequences of temperature-compensated natural frequency patterns.

4 COMPENSATION OF ADDITIONAL (MOVING) MASSES

The modified trend of the stiffness of the main span includes a number of characteristics of the prevailing freight traffic progression. Unfortunately, traffic data from the competent authorities are available only per hour (Figure 9). For some introductive exploration on approximate additional mass compensation, further steps need to be taken.

The frequency of vibration, based on equation (1) again, is inversely proportional to the square root of the mass. This means that live loads cause an increase in effective mass, which leads to hourly calculated factors, to modify the fluctuating frequency. Owing to this relation, the scattered trend of frequency is straightened in dependence of modal contribution of trucks per hour (Figure 10).

In fact, the present configuration of the permanent monitoring system makes it possible to develop a more sophisticated and more reliable, method strictly based on the measurement data. Forced balance accelerometers located at a defined distance along the cantilever's outer edges—in both directions of the traffic flow—enable the verification of recurring truck passages and their related velocity and tonnage without any disturbance to the traffic. A dynamic freight traffic registration system was

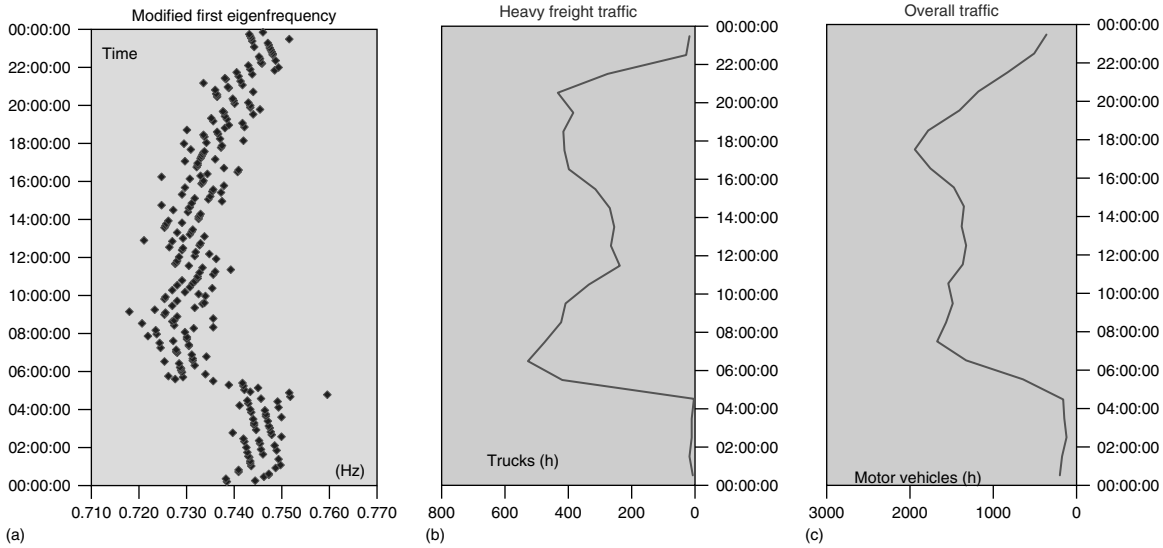


Figure 9. Modified pattern of stiffness (a), strongly affected by traffic loading (b, c) (moving additional masses as measured by the monitoring system).

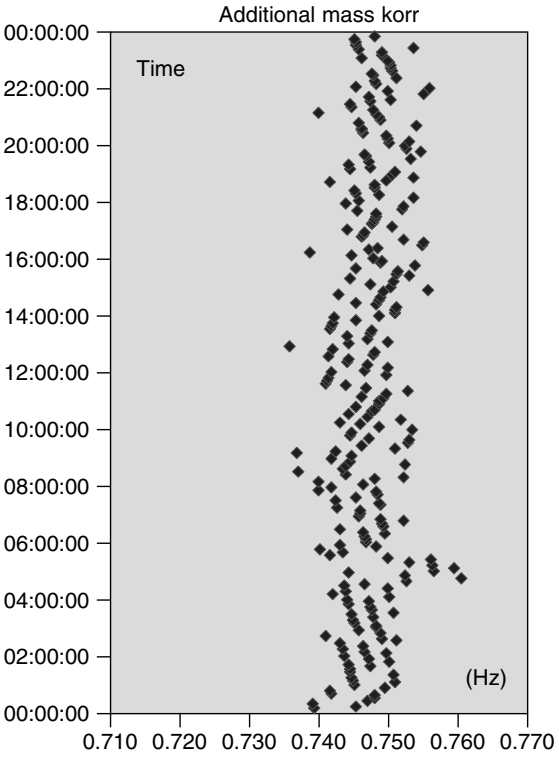


Figure 10. Stiffness pattern after approximate compensation of additional masses.

developed (a certain pattern-recognition procedure as introduced in [5]), which utilizes accelerometer-based methods to reproduce cantilever deformations. In this manner, the moving loads within each measurement file—passing the main span simultaneously—could be identified; this leads to a shifting of the single peak in the frequency response spectrum, which represents the registered time history of each measurement file (Figure 11).

5 CONCLUSIONS

As permanent monitoring systems produce huge amount of data, they have to be processed systematically in order to exploit the information fully. For easy handling, it is proposed that statistically based threshold levels are calculated (Figure 12). In this manner, continuous monitoring systems that provide information about changed modal parameters under "normal" operational conditions can be used to trigger warning and alarm levels with regard to damage.

The acquisition and elaboration of the measurements that are provided by the installed instrumentation allow the setting up of a structural behavior model that is considered to be the "regular model" (baseline model). The periodic analysis of

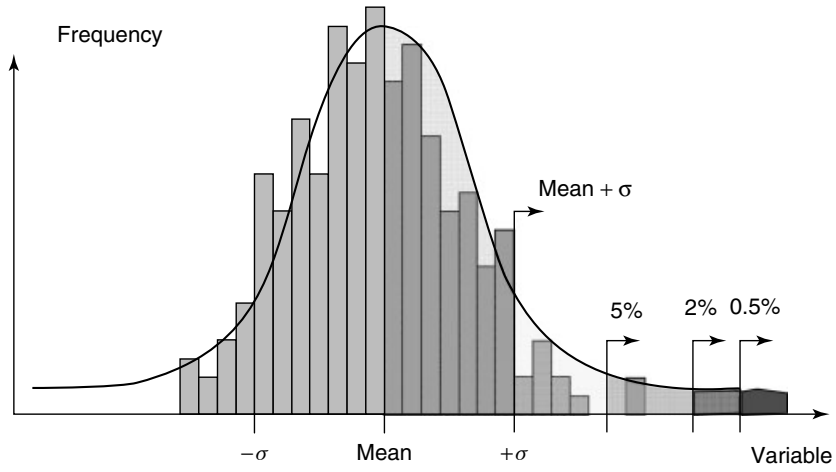


Figure 11. Histogram and best-fit distribution-based determination of threshold values.

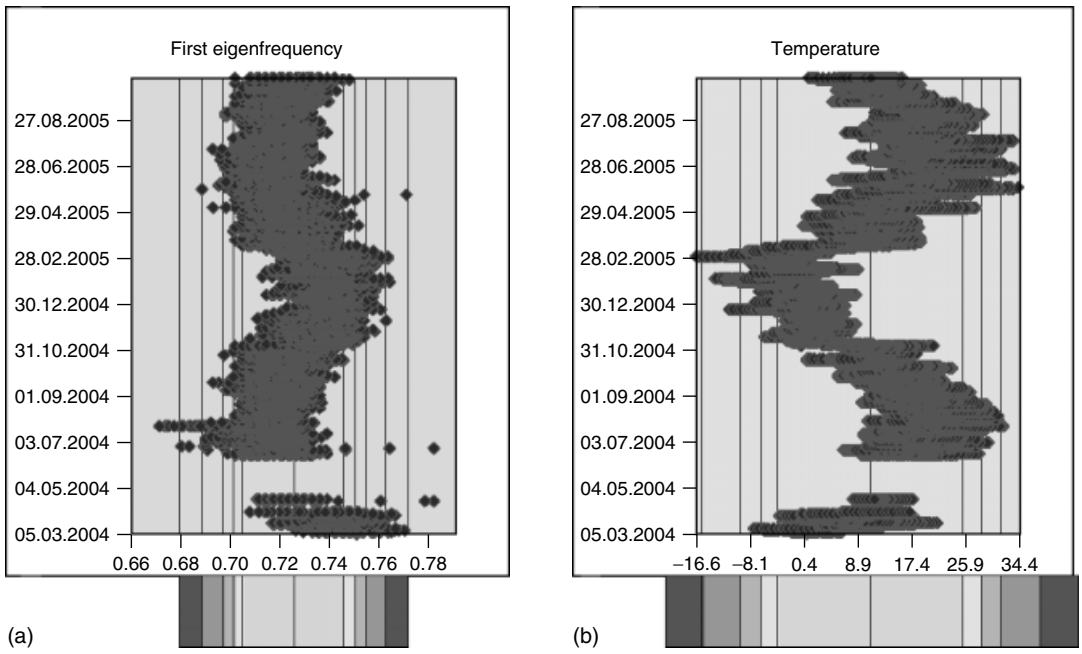


Figure 12. Response for 18 months using threshold levels (a) (statistical time history) with 5/2.75/2.5/1/0.135% probability of exceedance Related temperature (b).

the acquired measurements and their comparison with the baseline model help to point out indicators of potential structural damages. Periodic surveys of the cause quantities, moreover, allow setting up statistical models of the structural behavior, where the structural response is statistically correlated to the trend of the cause quantities. These models allow a control in time of the structural response by locating the "weight" of the cause quantities. The definition of alert threshold levels from an analysis of historical database (e.g., extreme-value analysis) may thus be illustrated.

6 OUTLOOK

The discussed approach—benefited by permanent monitoring—allows the expected goals to be reached, even if the results already obtained with the applied methods are somewhat approximate in nature. The results are quite promising, although the approach uses the effects of air temperature instead of that of structural elements. This approach represents an innovative method in the assessment of stiffness, which is appropriate for long-term application. The goal of generating frequency progressions over time without major environmental and operational impact has come within reach.

The procedure can be optimized progressively, once the already introduced cantilever-sensor-based approach is implemented.

REFERENCES

[1] Veit R, Wenzel H. Measurement based performance prediction of the Europabrücke against traffic loading. *Proceedings of the 16th European Conference of Fracture ECF16*. Alexandroupolis, 2006 ISBN 13-978-1-4020-4972-9.

[2] Haibach E. *Betriebsfestigkeit—Verfahren und Daten zur Bauteilberechnung*. 2. Auflage, VDI-Verlag: Düsseldorf, 2002.

[3] Hobbacher A. *Recommendations for Fatigue Design of Welded Joints and Components*. International Institute of Welding, doc. XIII-1965-03/XV1127-03, Paris, 2003.

[4] *ESDEP—European Steel Design Education Program: WG12 Fatigue, Lecture Notes*, Katholieke Universiteit Leuven.

[5] Wenzel H, Pichler D. *Ambient Vibration Monitoring*. John Wiley & Sons: Chichester, 2005. ISBN 0470024305.

Chapter 119
Ambient Vibration Monitoring

Helmut Wenzel
VCE Holding GmbH, Vienna, Austria

1 CONSERVATIVE DESIGN

The monitoring of over 400 bridges clearly shows different behavior in structures designed following different philosophies. Bridges designed conservatively are not affected by dynamic phenomena that generate concern or trigger damage. Bridges with designs that are constrained by economic considerations very often do not have any reserves to cover the extraordinary loads that they are subjected to in reality. The difference in dynamics becomes obvious in Figures 1 and 2. Conservative bridges show a high system damping with distinct characteristics. Economic designs very often result in resonance in certain areas. This resonance might have a very local and limited effect, but it leads to damage in structures over time.

From bridge management, we know that an additional investment in 10% higher quality makes a difference of over 200% in costs over the life cycle of a bridge of 100 years. Drastic examples include a bridge composed of single-span I-girders, designed to the limit, which has consumed 220% of the investment costs in retrofit over a period of 25 years. This may be compared with the resistance of a duly designed box girder bridge that survives a displacement of a single pier of 110 cm that could be retrofitted at reasonable costs.

2 EXTERNAL VERSUS INTERNAL PRESTRESSING

Damage found on grouted internal cables triggered a dramatic change in design philosophy. Some countries, like Germany, specified that new bridges have to be built using external cables only. This is for the purpose of inspectability and eventual replacement. The experience gained while testing 30 bridges built in the late 1950s and early 1960s showed that damage of the cables was found in only one of the structures.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

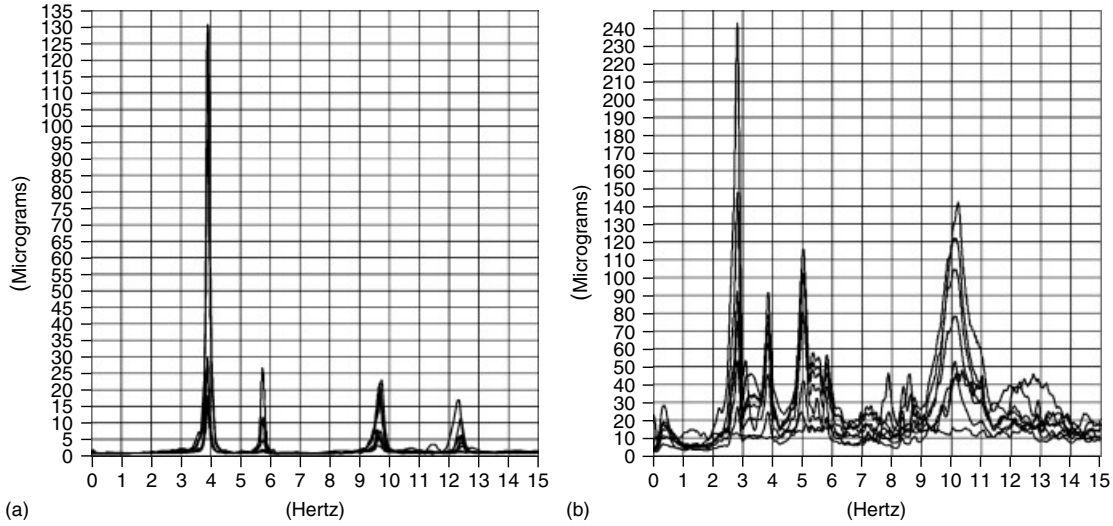


Figure 1. Spectrum of a sound bridge (a) and spectrum of a damaged bridge (b).

(a)

(b)

Figure 2. Resistant box girder (a) and costly I-girders (b).

In all other bridges where damage was suspected, no evidence of corrosion or wire breakage had been detected. The only damaged bridge also was far from malfunctioning. On the other hand, the bridges with external prestressing often show cracks in the anchoring parts and are often unequally stressed. The best results have been received on bridges where grouted tendons embedded in concrete in combination with external cables had been chosen.

3 INFLUENCE OF TEMPERATURE

The design codes for bridges provide clear instruction as to how to consider the temperature effect in bridge design. These instructions normally give a high and low temperature to be considered, and eventually a temperature gradient between the bottom and top of a structure. No reference is made to the type of bridge or the material used. Monitoring provides the chance to exactly record the actual effects of temperature on structures. The lessons learned are actually easy to accept:

- slender structures react very close to the provisions of the codes;
- stiff structures very often deviate considerably from the expected stress distribution;

- the temperature gradients actually recorded on stiff structures by far exceed the values of the design;
- temperature load cases can be the decisive load cases; and
- temperature changes do not trigger a linear behavior. A clear stiffening effect is recorded below 5 °C.

The stiffness of concrete bridges depends on temperature. The relation is given in Figure 3, which shows an almost bilinear condition measured on a classical posttensioned concrete box girder bridge. This has to be considered in the interpretation of monitoring results.

Steel bridges show quick reactions on changing temperatures. The records of a 5-m-high steel box girder are shown in Figure 4. There is a difference between heating or cooling periods. The sensors represent the outside temperature and the inside temperature on the bottom slab and the deck slab. The patterns shown here are representative.

Temperature changes in the annual cycle are rather homogeneous. Figure 5 shows the behavior of a concrete mass supported on bridge bearings in a railway tunnel. There is no influence of sunshine. Temperature conditions of the surrounding soil are rather stable, but the trains transport air from outside through the tunnel. The graph shows the homogeneous behavior of the structure over the years. The maxima and minima values measured are actually higher than expected. A concrete bridge (Figure 6) where concrete is used in the shape of a box girder with cantilever arms also shows a rather homogeneous cycle. The maximum and minimum temperatures according to the codes are never reached. The structure reacts moderately to warming or cooling. No clear daily cycle can be isolated.

A typical steel structure shows a rather violent reaction on temperature changes (Figure 7). The reaction is quite quick and produces strain in the system. This might be because of bearing friction, which is released suddenly causing a displacement of the structure. This effect can not only be particularly harmful to the expansion joints but also to the bearing. Particular bridges, which are not straight in plan, might develop exceptional forces into weak axes of the outfitting.

The consequence of these experiences should be an individual application of temperature loads depending on the type of the structures and conditions they have to bear. The following implications might be considered:

- to increase the loads from differential temperatures in stiff structures;
- to increase the temperature range considered in steel structures; and
- to look at the effects of quick temperature changes on the global behavior.

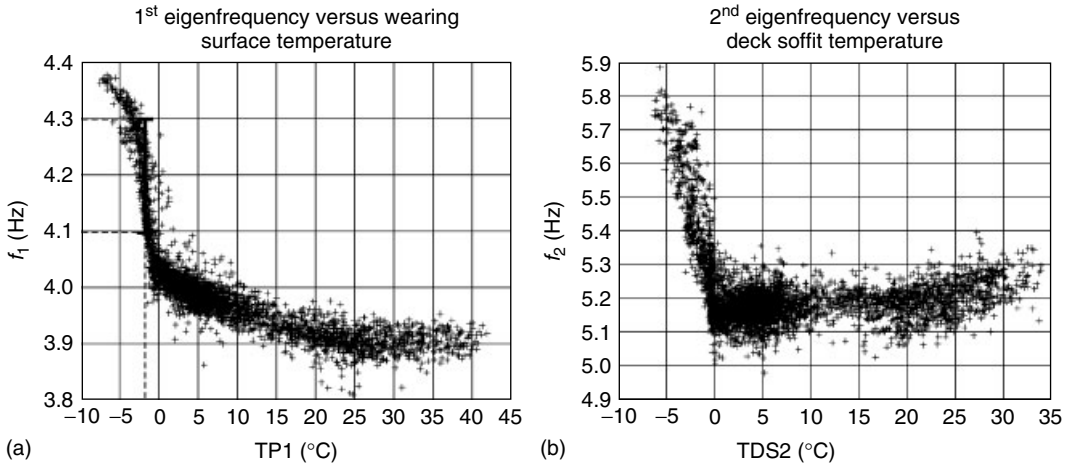


Figure 3. (a) First eigenfrequency versus wearing surface temperature [1] and (b) second eigenfrequency versus deck soffit temperature [1].

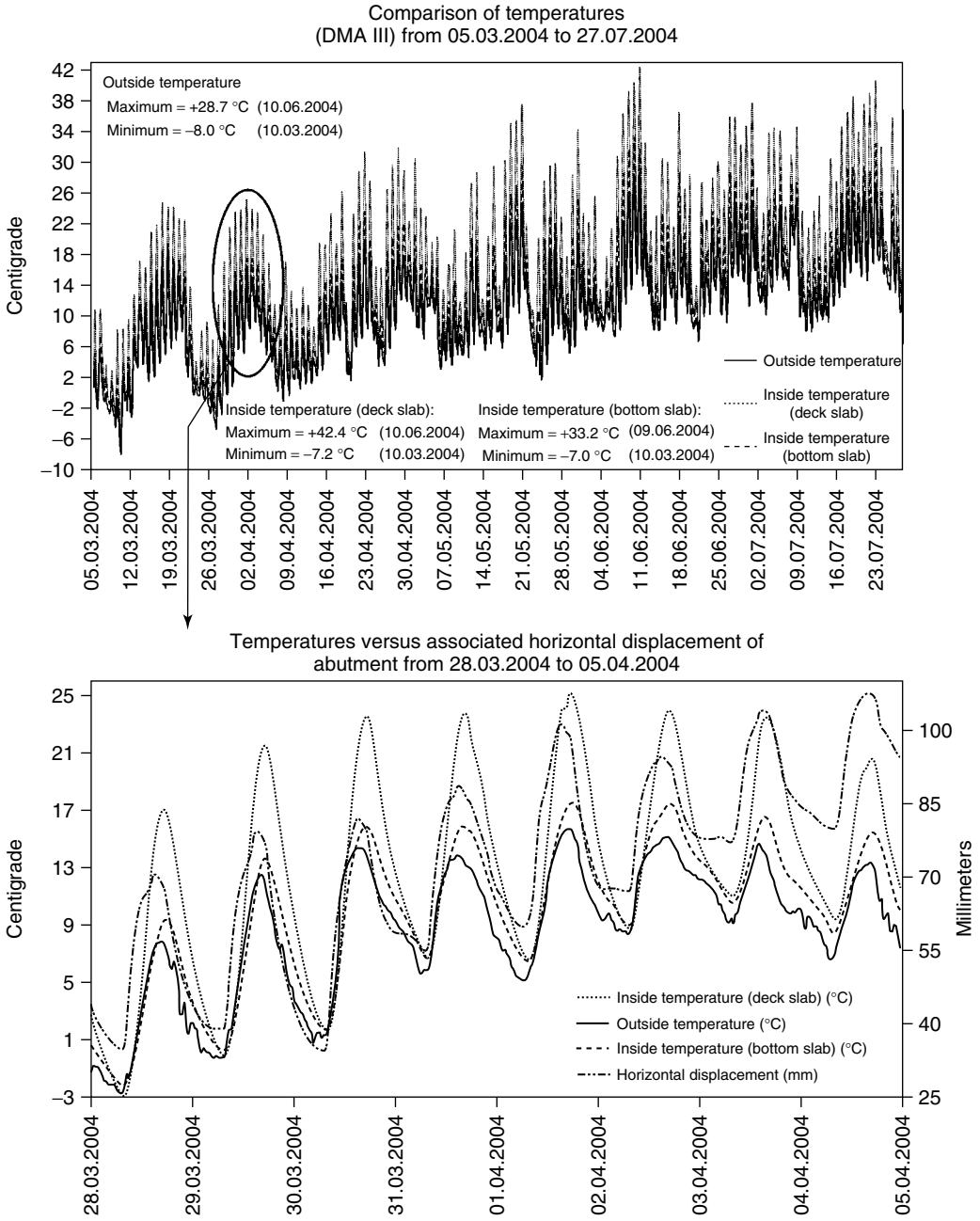


Figure 4. Representative temperature sensor records and longitudinal displacement of the steel bridges' abutment.

4 DISPLACEMENT

Bridges are flexible and displace under various loads. Displacements are calculated using structural models or finite element calculations (see Figures 13, 14, 15). Very often, these models do not reflect reality. A typical case is the displacement of a certain steel bridge due to temperature changes, which are shown

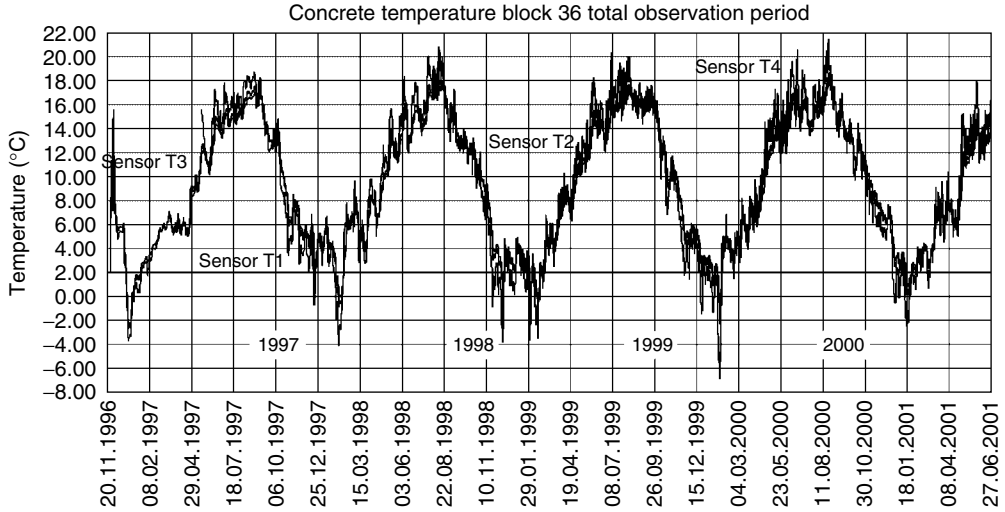


Figure 5. Long-run behavior of a concrete mass supported on bridge bearings in a railway tunnel.

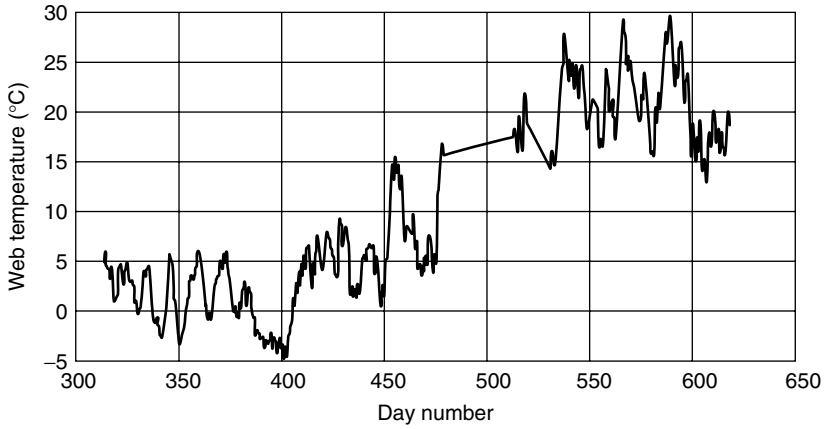


Figure 6. Variation of web temperature of the Z-24 bridge observed over a period of one year [2].

in Figure 8. Contrary to that, the monitoring results provided displacements according to Figures 9 and 10. Especially, the latter demonstrates the fact that the displacements of this steel bridge's abutment are approximately twice as much as those obtained from theoretical, linear elastic calculations. The difference is mainly related to the following facts:

- the stiffness of the columns depends very much on the degree of fixation of the pier in the foundation;
- bearings do not show a linear behavior at all times and rather tend to be stiff until a certain minimum force has been reached; and
- a certain stress limit has to be reached before restoring forces are activated, particularly when elastomeric bearings are provided.

In major bridges, sudden displacements of ± 50 mm have been recorded, which have to be attributed to sudden release of restoring forces of bearings. This displacement is normally within the regular limits of allowable displacement, but the sudden reaction might trigger secondary problems, such as restraints in the expansion joint. A number of failures of expansion joints have to be attributed to these phenomena. The frequency of such phenomena is

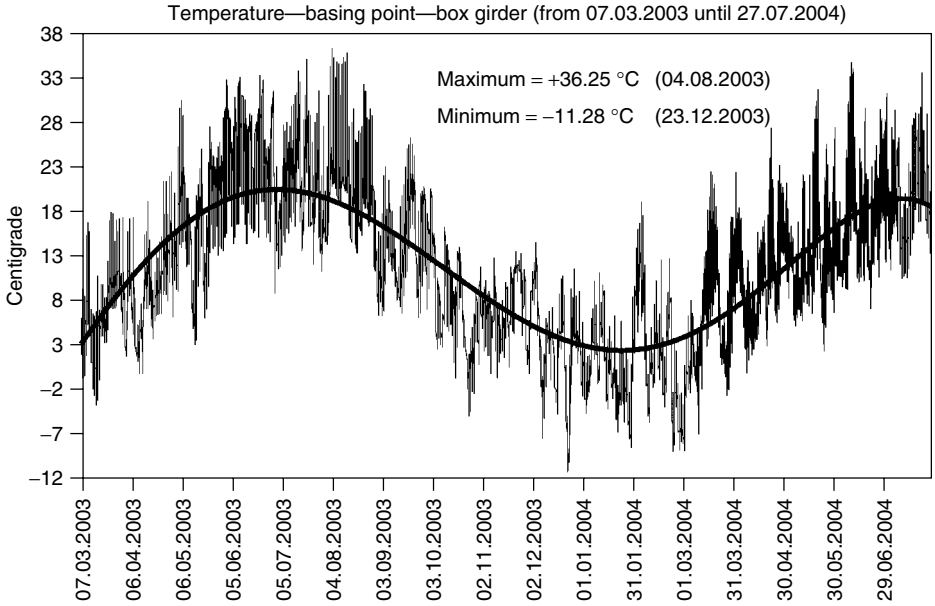


Figure 7. Temperature conditions inside the steel box girder at the *Europa Bridge* of the *Brenner Motorway*.

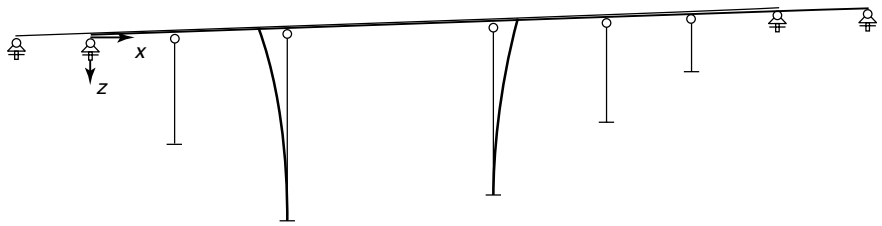


Figure 8. Displacement of a bridge due to temperature changes affecting only the superstructure.

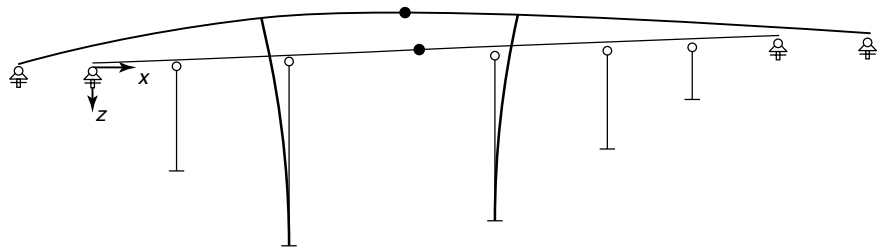


Figure 9. Displacement of a bridge due to temperature changes recorded by monitoring.

not sufficiently documented yet. In a six-month-long record of a major steel bridge, three such occasions have been detected (Figures 11 and 12).

The consequences from these records are that the realistic behavior of a structure can be found through monitoring, which might explain damage in the outfitting. The displacements calculated for bearings and expansion joints might be not enough to cover extraordinary events as described. The center of expansion of a structure can be dozens of meters away from the theoretical center and influence the design of bearings and expansion joints (Figure 9).

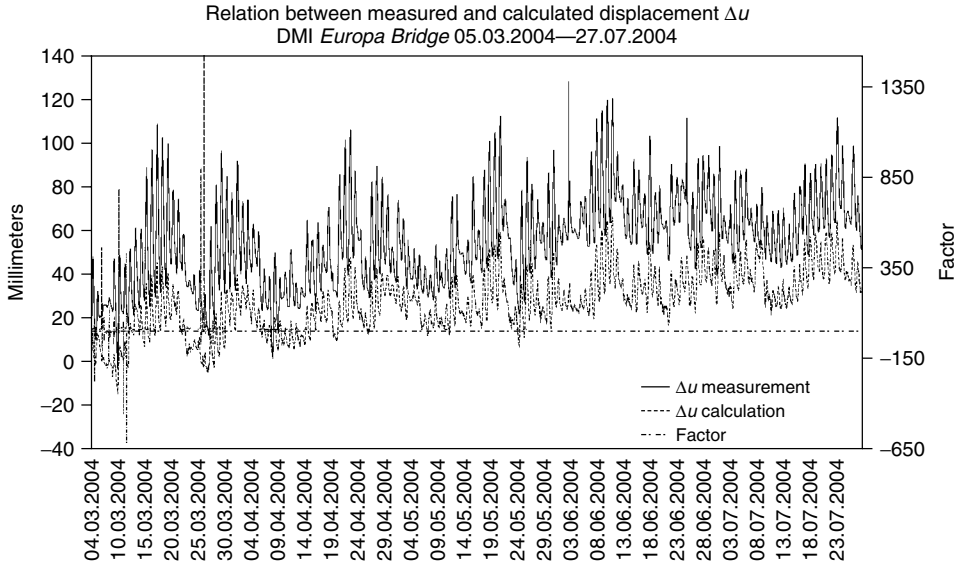


Figure 10. Comparison between measured and finite element (FE)-based displacements of a steel bridge's abutment.

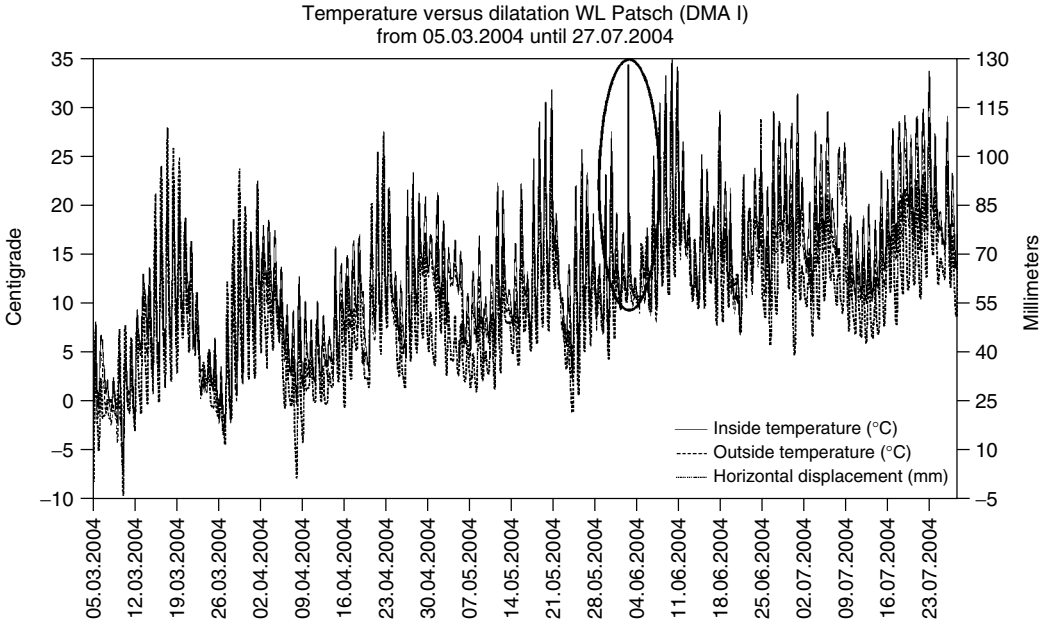


Figure 11. Uncommon, sudden reactions in the abutment's displacement recordings over a period of five months.

5 LARGE BRIDGES VERSUS SMALL BRIDGES

In the beginning, monitoring concentrated on large and important bridges. This has led to the impression that bridges normally perform very close to the theoretical behavior determined and based on the design assumptions. The subsequent assessment of small bridges showed that it is considerably more difficult to achieve good results, the smaller the

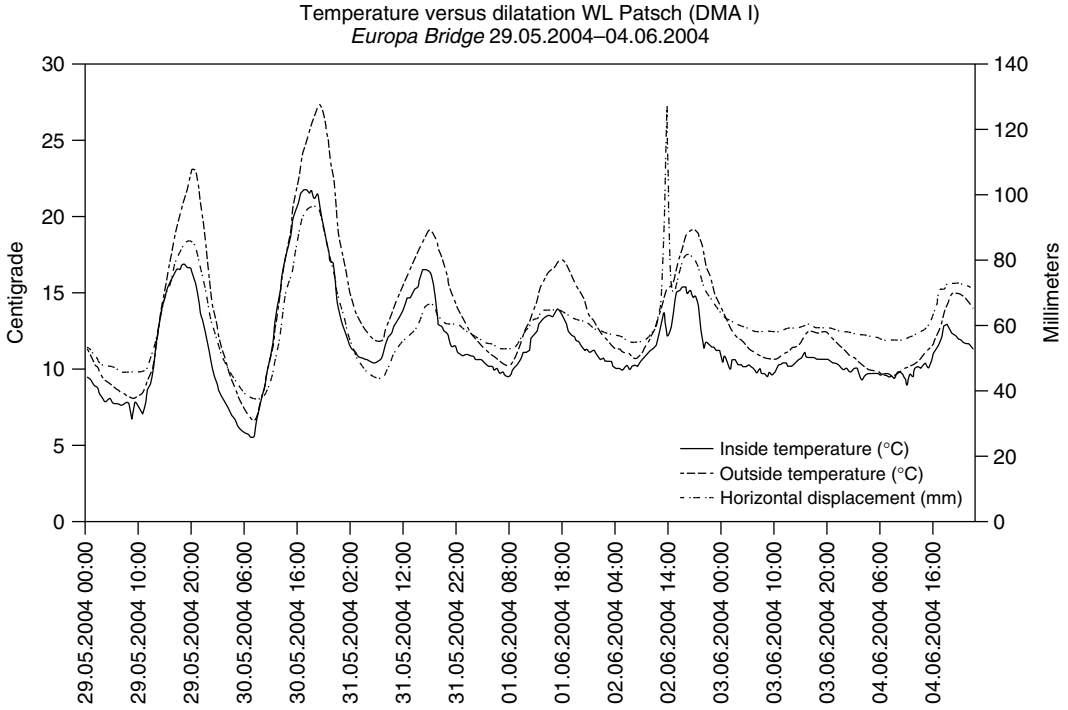


Figure 12. Focus on Figure 11 over a period of a certain week including one of the observed sudden reactions.

structure is, because of different approaches taken toward these, considered not so important, structures. Another fact is that boundary conditions are much clearer in large structures.

The lesson learned from monitoring is that even higher attention should be paid to smaller bridges and that a number of provisions of construction codes fit very well for large structures, but underestimate small ones. Here, in particular, the subject of temperature, as explained in another article, has to be highlighted. Furthermore, the correct modeling of the boundary conditions has to be taken care of.

6 VIBRATION INTENSITIES

The subject of resonance in pedestrian bridges is well known and taken care of. Frequencies close to resonance, particularly those of structural members, such as cantilever slabs, are not yet subject to consideration. Experience has shown that the evaluation of the vibration intensities measured for a structure can give considerable information on fatigue and related problems. The assessment of vibration intensities, therefore, can give indicators on the expected lifetime of a structure and on local problems to be expected on structural elements in the near future. It has been clearly demonstrated that bridges, where high vibration intensities have been recorded (Figure 16 and 17), most probably develop local problems in expansion joints, bearings, outfitting, and particularly in waterproofing.

7 DAMPING VALUES OF NEW COMPOSITE BRIDGES

Measurements taken at a number of new composite bridges show that the damping values determined at the newly built structures are considerably higher than the normal values of comparable concrete bridges or steel bridges (Figure 18). This might be attributed to the fact that the composite effect has to be established through a number of load cycles. After some time, the damping values of these bridges have been stabilized in normal ranges. Further conclusions on these phenomena have not been drawn yet, but

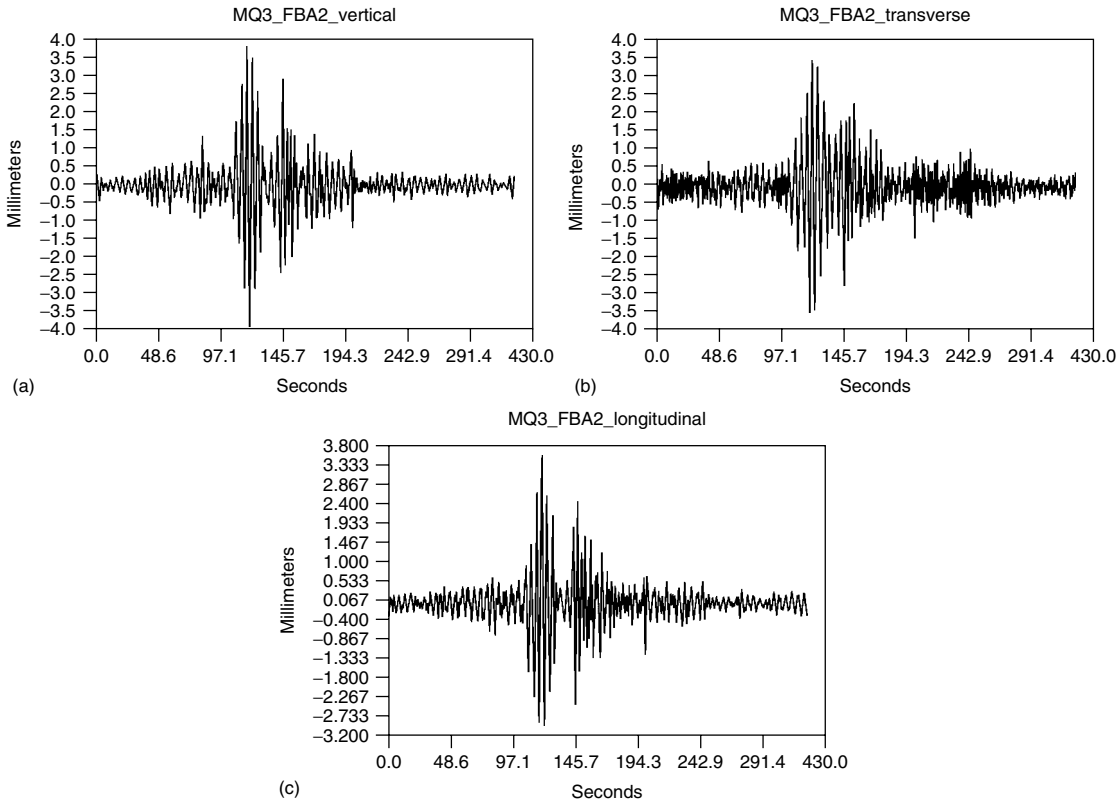


Figure 13. Relative displacement due to sudden occasions of restraint recorded with a 3-D acceleration transducer at the top of a 200-m-high pier subdivided into the vertical (a), transverse (b), and longitudinal (c) direction.

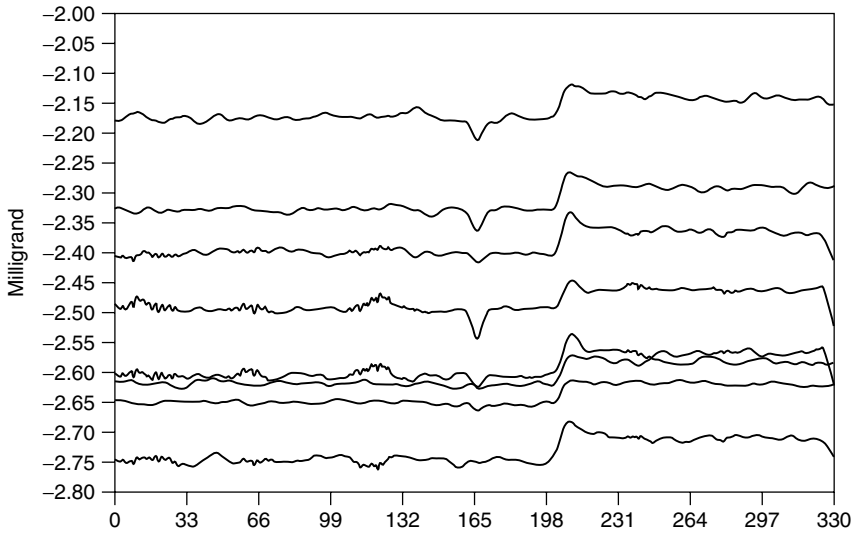


Figure 14. Displacement of the system's neutral axis due to bearing reset forces of the flyover *St. Marx* (basis: acceleration sensors).

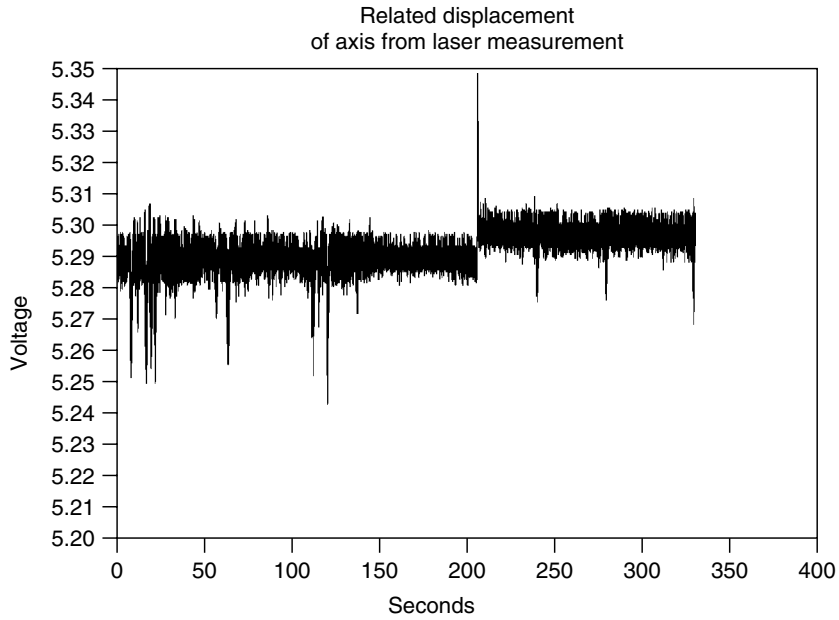


Figure 15. System displacement due to bearing reset forces of the flyover *St. Marx* (basis: longitudinal laser-displacement sensors).

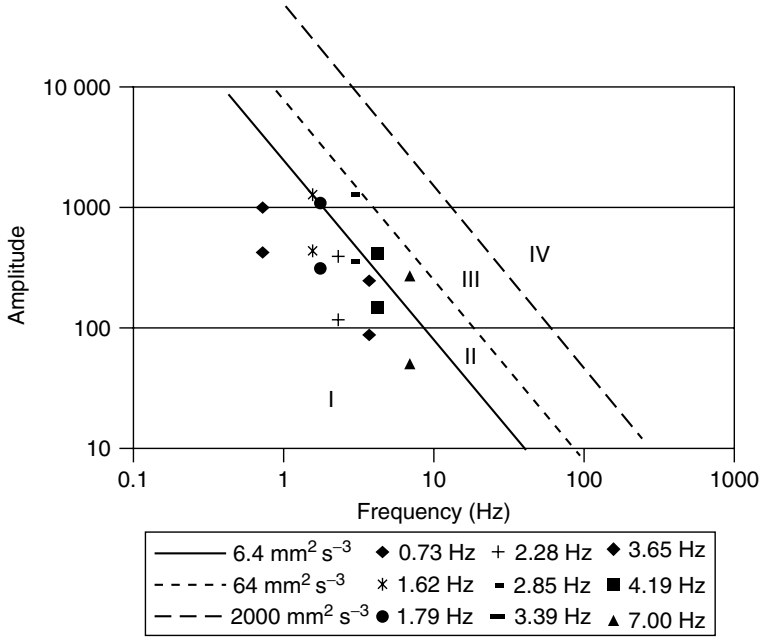


Figure 16. Intensity chart at the *Europa Bridge* of the *Brenner Motorway* (representing high vibration intensities).

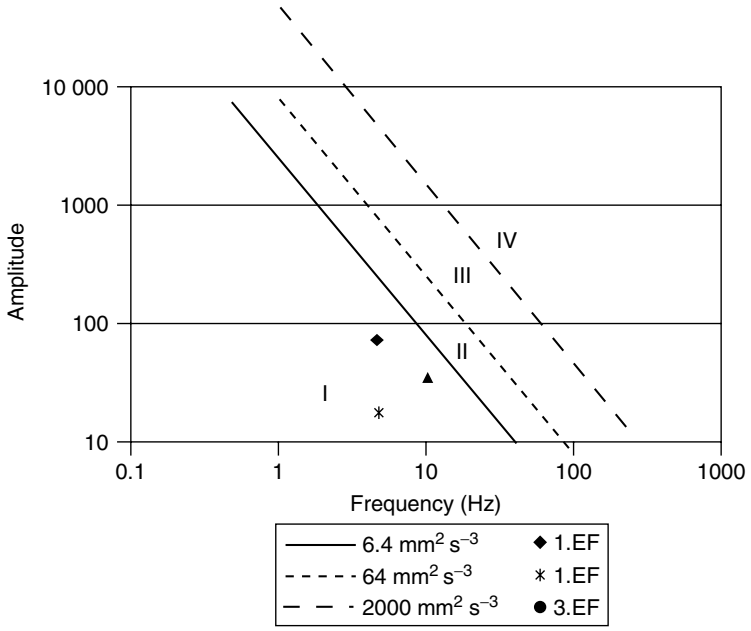


Figure 17. Intensity chart at the *S 36 Bridge* of the *A1 Motorway* (representing low vibration intensities).

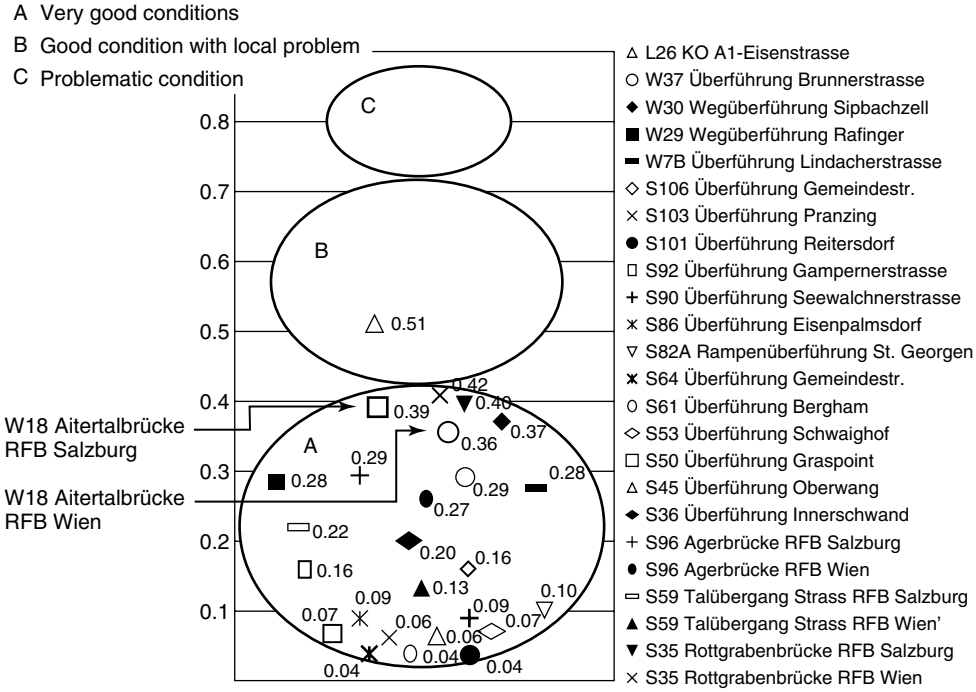


Figure 18. Classification of prestressed concrete and composite bridges according to their damping values.

it might be expected that a sharp drop in damping indicates eventual problems with bonding or the triggering of a hidden local damage.

8 VALUE OF PATTERNS

Certain elements of bridges exist in a repetitive form. It has to be expected that all members show the same dynamic performance under service. One of the valuable approaches of monitoring is to recognize patterns and to observe the performance of comparable components. Any deviation from the pattern indicates a malfunction or an extraordinary situation that can be identified.

As an example, the case of a concrete box girder bridge with a distinct cantilever is shown. The monitored performance of the cantilever minus the action of the global system provides information on the cantilever eigenmodes. Related symmetric modes can be determined and displayed. This should provide a distinct pattern, where every deviation indicates a problem. On the basis of colored frequency cards, so-called trend cards, the relevant cantilever eigenfrequencies have been determined, which are marked in Figures 19 and 20.

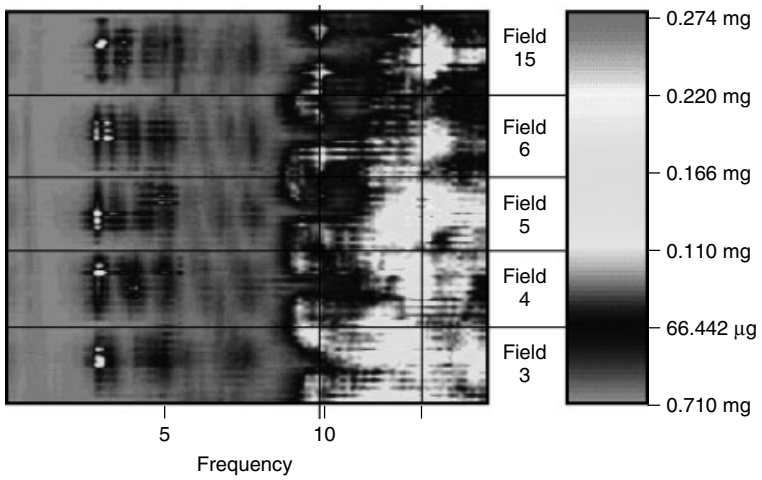


Figure 19. Course of frequencies at a certain concrete box girder bridge—structure south.

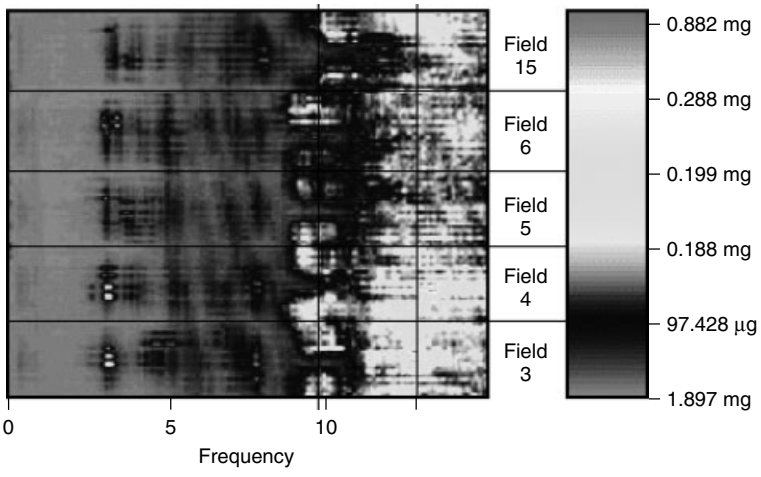


Figure 20. Course of frequencies at a certain concrete box girder bridge—structure north.

By comparing the response spectra of both box girders and their cantilevers, the share of cantilever vibration can be displayed directly as shown in Figures 21 and 22.

A detailed evaluation procedure analyzing the relation between the response spectrum and its energy content within the relevant frequency ranges leads to a certain behavior pattern of the cantilevers along the bridge. Deviations from this pattern are typically indications of irregularity.

Figures 23 and 24 show the pattern of an undamaged cantilever compared to a cantilever with minor corrosion damage of the transverse reinforcement.

This method is not good enough for detail localization of the problem, but it provides sufficient information on the quality of function of a structural element.

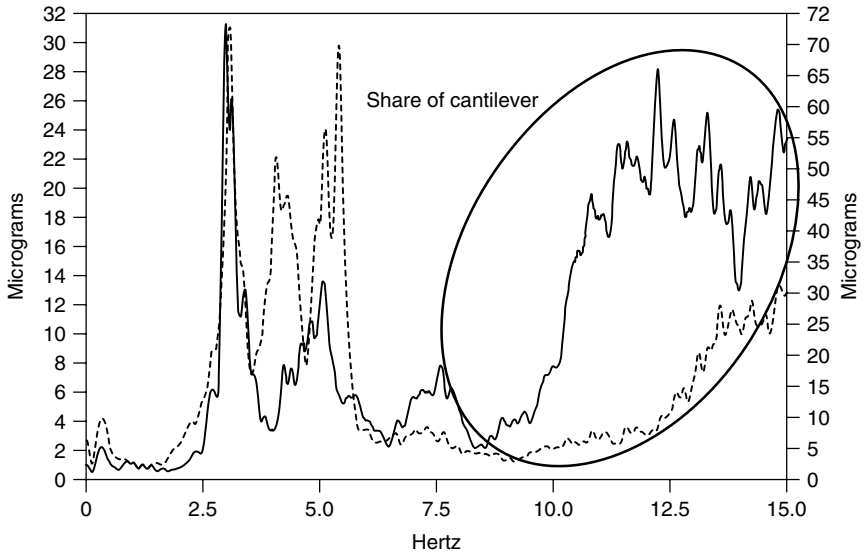


Figure 21. Spectrum of cantilever (continuous graph) and box girder (dashed graph).

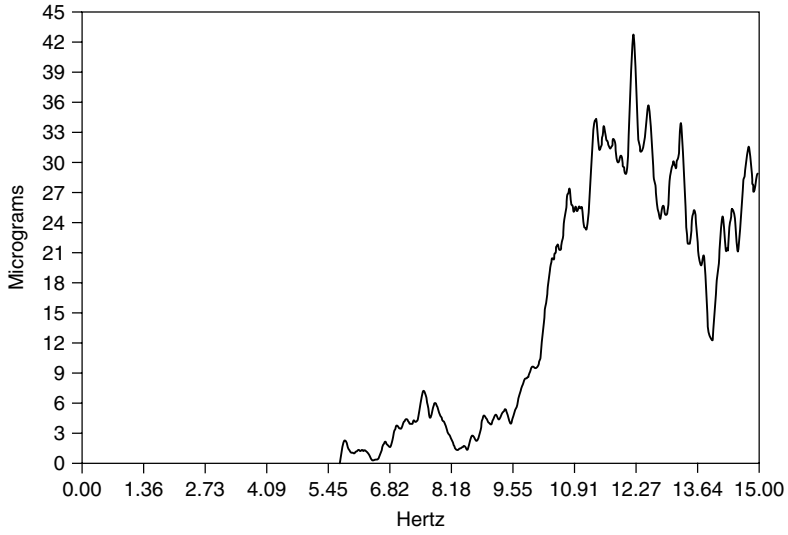


Figure 22. Response spectrum of cantilever vibration.

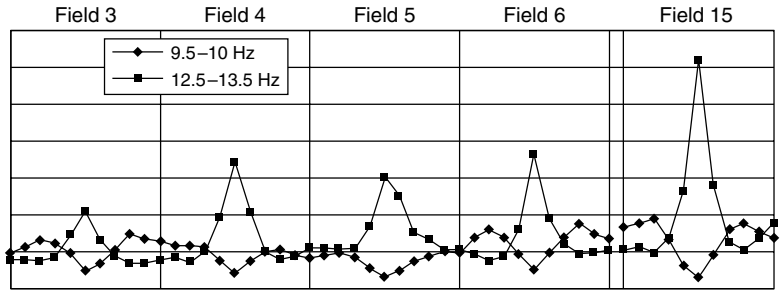


Figure 23. Acceptable behavior pattern of the cantilevers along the bridge.

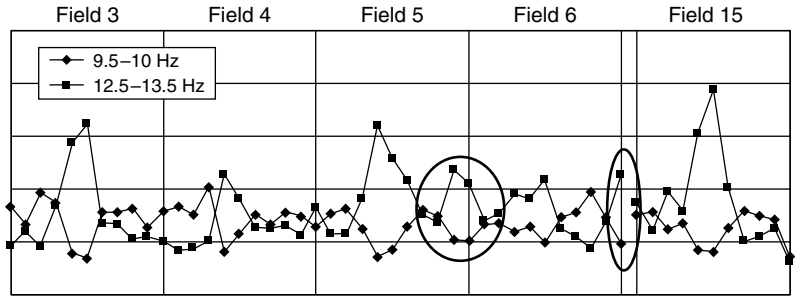


Figure 24. Behavior pattern of the cantilevers with indications of irregularity.

By a very quick and cheap test, it can be determined whether action is required.

8.1 Understanding of behavior

Complex bridge structures are often modeled in a rather simple way neglecting the behavior of the structure in the three-dimensional space. Monitoring is the recording of the actual behavior of a structure. This comprises eventual drift or strain from temperature, as well as eventual construction mistakes, such as wrong placement of bearings or nonrelease of restrainers. Figure 25 shows a case where a temporary fixture during construction has not been removed at the time of handover. The performance of the bridge has been considerably different than estimated. By monitoring this difference could be detected and immediate correction measures taken (Figure 25).

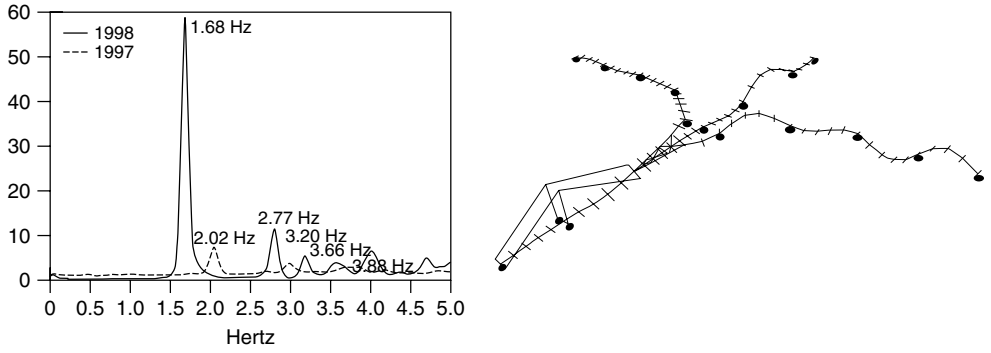


Figure 25. Frequency spectrum of *Inn Bridge Hall West* 1997–1998.

Figure 26. *Steyregg Bridge*.

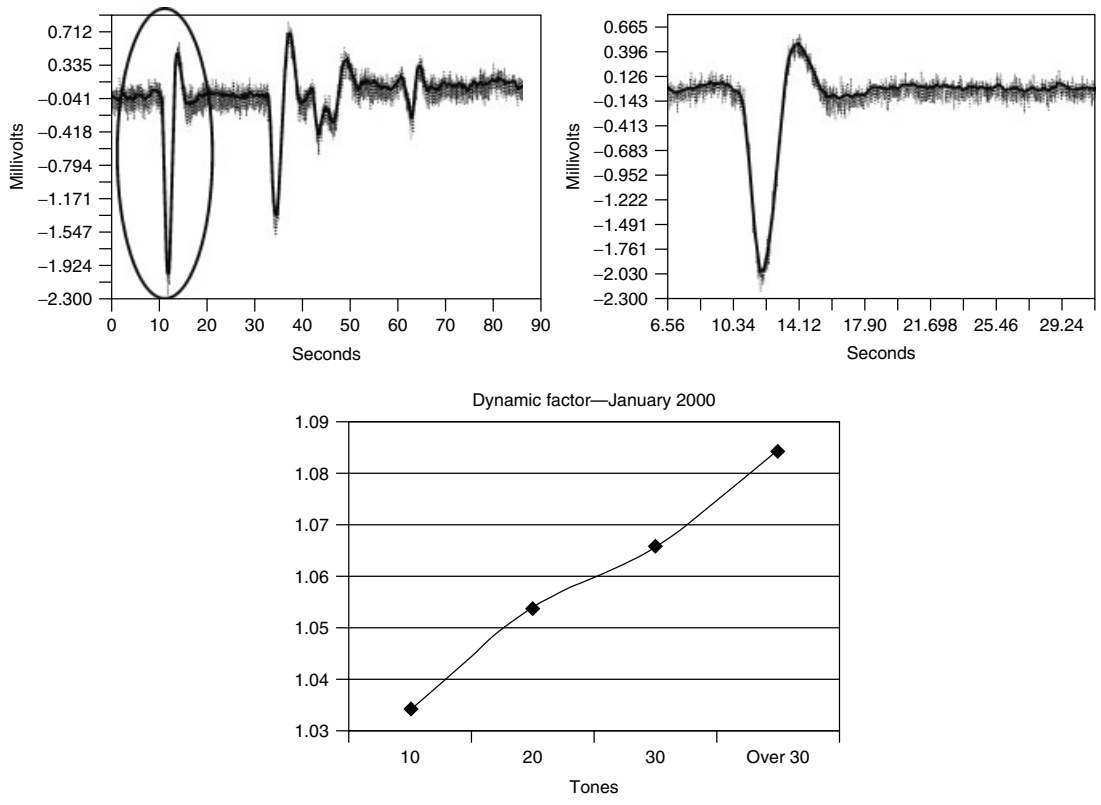


Figure 27. Dynamic factor of the flyover *St. Marx*.

Another important value is information on the actual displacement of a structure, particularly with regard to complicated cable-supported structures, where such displacements could generate problems in traffic clearance or related interfaces.

Another way of finding problems is to compare the expected behavior with the measured one. In case of stay cables, protected by steel tubes against vandalism, the contact of the cable to the tube has been found through monitoring. The effective

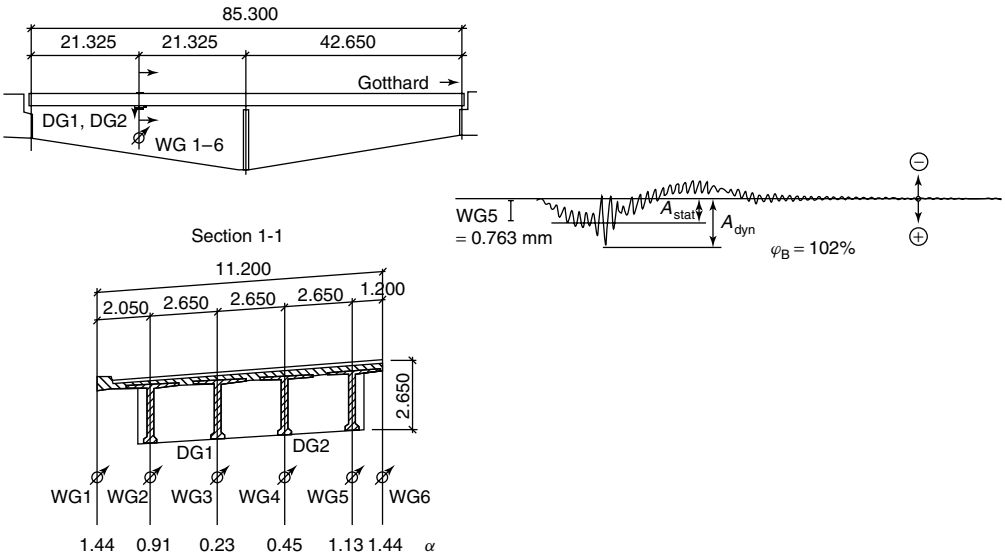


Figure 28. Dynamic factor of the *Boeschrüti Viaduct* due to induced impact loading. [Reproduced with permission from Ref. 3. © EMPA, 1983.]

vibration length of the cable has been shortened by this contact. Such a problem can lead to drastic damage at a cable providing a sharp edge, which introduces unintended bending. Monitoring is able to identify these problems (Figure 26).

9 DYNAMIC FACTORS

Current bridge design codes ask for dynamic factors mainly depending on the bridge span. The factor is considered to be 1.40 for components or directly effected members and varying between 1.00 and 1.40 depending on the span of a bridge. The lessons learned from monitoring are as follows:

- The dynamic factor provided by the code depending on the span length is actually conservative. All bridges so far showed smaller dynamic factors.
- The dynamic factor for components sometimes exceeds the values considerably. The record factor measured has been 2.20.
- The dynamic factor is also considerably dependent on the speed of the vehicles. This can eventually be controlled by speed limits.

The consequences are that overloaded vehicles that drive slowly will not produce harmful stresses. The low increase always has to be seen in conjunction with eventual speed effects. Consequently, dynamically sensitive elements should be avoided in design.

Another lesson learned is that the dynamic behavior also depends on the type of structure designed. Bridges with box girders (Figure 27) are considerably less vulnerable to dynamic effects than bridges of other types of design (Figure 28).

The dynamic vulnerability of a structure depends on the acting mass. This is also clearly shown in monitoring records. Concrete bridges with a mass of $1.5\,t/m^2$ or more are very little affected by dynamic amplification. Continuous girders react less violently to any impact. Elements with major differences in stiffness produce an inharmonic behavior unfavorable for the structure.

REFERENCES

[1] Peeters B, De Roeck, G. One year monitoring of the z 24-bridge: environmental influences versus damage events. *Proceedings of IMAC 18, The International Modal Analysis Conference*. San Antonio, TX, February 2000; pp. 1570–1576.

[2] De Roeck G, Peeters B, Maeck J. Dynamic monitoring of civil engineering structures. *Computational Methods for Shell and Spatial Structures IASS-IACM 2000*. Greece, 2000.

[3] Cantieni R. *Dynamic Load Tests on Highway Bridges in Switzerland—60 Years Experience of EMPA*. Section Concrete Structures and Components, Report No. 211, Dübendorf, 1983.

[4] Wenzel H, Pichler D. *Ambient Vibration Monitoring*. John Wiley & Sons: Chichester, 2005, ISBN 0470024305.

Chapter 121

Long-term Monitoring of Dynamic Loads on the Brandenburg Gate

Werner Rücker

Division VII.2 Buildings and Structures, BAM Federal Institute for Materials Research and Testing, Berlin, Germany

1 INTRODUCTION

The Brandenburg gate (Figure 1) is the best-known German national monument. It was constructed by the architect Carl Gotthard Langhans between 1789 and 1791. The architect orientated the design (Figure 2) of the gate on ancient buildings. The construction was erected with sandstone. It consists of six pillars with an architrave section on the top. Above the architrave, there is the so-called quadriga, a very well-known work of art. To the left and right of the gate there are two gatehouses, which were also constructed like Greek temples. The total dimensions of the gate are 22 m high, 65 m wide, and 11 m deep (without quadriga). During World War II, the gate was heavily damaged. After the war, the reconstruction was carried out until its completion in 1959 by the former German Democratic Republic (GDR).

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 THE PROBLEM

After the reunification of Germany in 1989, the senate of Berlin was interested to open the gate again for traffic use. Ahead of a political decision on that, an inspection of the structure was made. The inspection revealed the extent of damage on the building itself as well as the zones where the gate is connected to the neighboring two gate houses. There were unusual settlements and vertical and horizontal cracks (perpendicular and along the gate) visible at the foundation structure and the architrave. As can be seen from the examples given in Figure 3, the crack opening is about 2 mm and more. The shear deformation of the architrave is limited by a shear tension band. Figure 4 shows the construction details and the cracked clamping system. Based on the inspection results, the following questions were raised:

Figure 1. The Brandenburg gate.

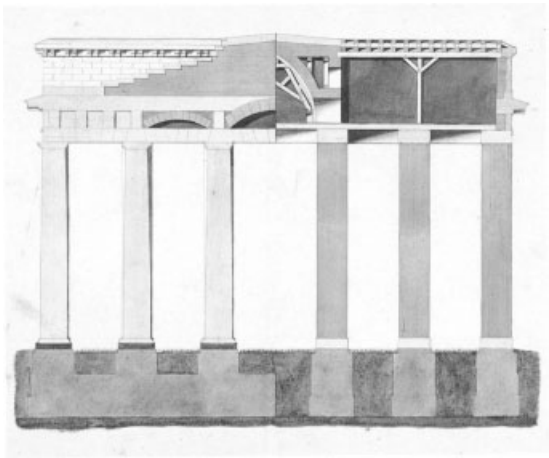


Figure 2. The architect's drawing.

- Is the railroad and road traffic around the gate the main cause of the damage?
- Is it possible that the damage can be exacerbated by traffic?
- Are there reasons other than traffic possibly responsible for the damage?

In order to answer these questions, the responsible city administration asks for an investigation concept. The results of the investigation should serve as a basis for the development of the best technical and economical rehabilitation measures and also for the decision to open the gate for traffic or not.

3 IMPACTS TO THE STRUCTURE

As mentioned above, the most important impacts to the structure seem to be road and railroad traffic.

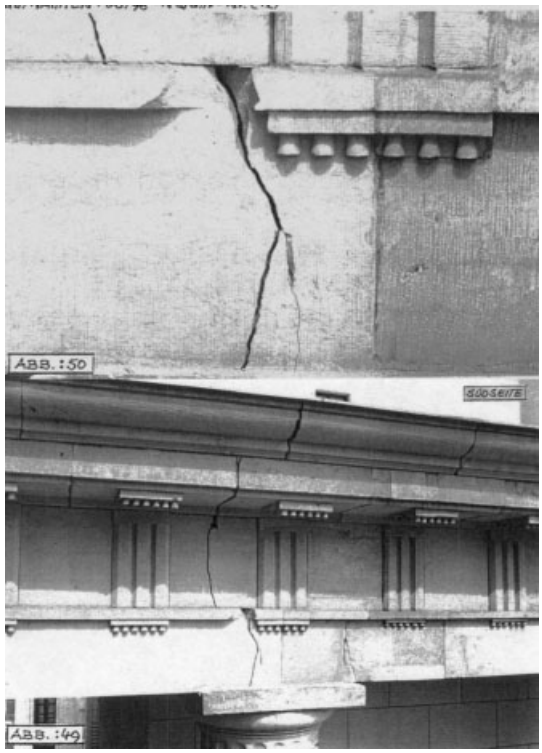
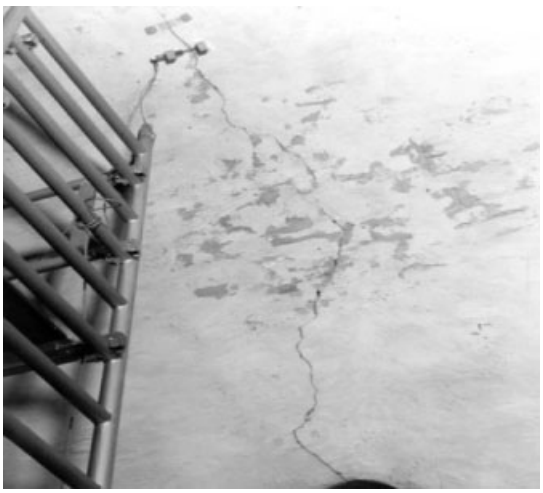


Figure 3. Observed cracks at the architrave.

The traffic generates dynamic forces, which are transmitted to the gate and the neighboring gate houses through the soil. Owing to the interaction effects between the building and the soil, the vibration transmitted from the soil to the foundation will be normally reduced; but in the ongoing structure, amplification depending on the modal behavior and damping of the structure will take place. If the vibrations reach threshold values, cracks and other damage can occur. However, the settlement of soil under and beneath the foundation can induce great stress in the structure, resulting in cracks. Besides traffic, construction work in the surrounding area of the gate can also generate vibrations, causing additional damage. Since the gate is located in the middle of two long straight-line streets, the influence of wind may also contribute to the damage. A special event that takes place every year in the surrounding area of the gate is the "love parade". During the event, a number of concerts happen and create high air pressure (Figure 5). This may also have an influence on the gate because the very low frequencies of the

(a)



(b)

Figure 4. Examples for observed cracks Tension band (a), Architrave wall (b).

music may excite Eigen frequencies and Eigen modes of the gate. As known from the measurements on bridges and towers, the temperatures also have a great influence on the development of cracks and other damage to the structure. Therefore, the influence of the following impacts (see Figure 5) to the structure has to be considered in total:

- railroad and road traffic
- construction work
- wind
- love parade and
- temperature.

4 THE INVESTIGATION PROGRAM

To characterize the static and dynamic behavior of the structure and to assess the different impacts and their damaging effects to the structure, a three-level investigation program was developed. The program consists of the following elements:

- modal analysis of the complete building;
- definition and derivation of threshold values; and
- permanent monitoring of the structure at certain significant spots for acquisition of all significant influences.

Through modal analysis (*see* **Modal–Vibration-based Damage Identification**), the global vibration behavior of the Brandenburg gate and the gate houses was examined. The results provided the basis for the following permanent vibration measurement and the permanent condition monitoring with which all the mentioned effects on the building were to be measured in quantity and character (*see* **Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge**). The interaction of dynamic excitation with known cracks should have been examined at the same time. From those results, the reasons for the cracking between the gate and the neighboring gate houses should also be investigated. In Figure 6, the measurement layout for the investigation by modal analysis is shown. Measurement points were located at the foundation, at the pillars, and at the architrave. For the measurement, velocity transducers were used. The operating frequencies of these transducers are approximately between 0.5 and 1000 Hz.

Furthermore, possible changes in the dynamic behavior of the Brandenburg gate (if there are any) should be discovered and quantified (*see* **Risk Monitoring of Civil Structures**). For this reason, the changes in the width of known cracks, strains in special structural elements like the diagonal anchor in the Attica area, and local temperature were supervised at selected locations. In addition to the influence of all kinds of traffic (road and rail), other excitation processes like tunnel drilling and pile driving beneath the foundation of the gate and special events like the love parade have also been measured by installing and operating a condition monitoring

Figure 5. Examples of impact to the Brandenburg gate (traffic (a) and love parade (b)).

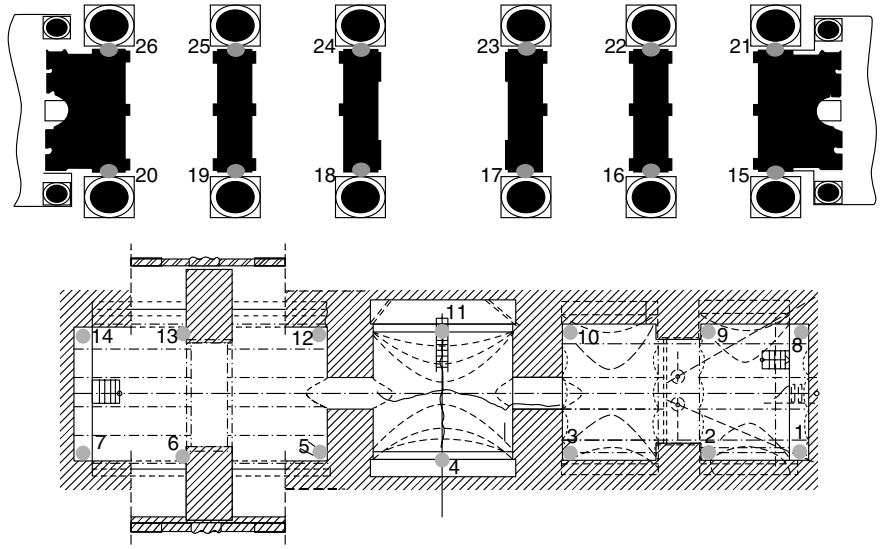


Figure 6. Measurement layout for modal analysis at the Brandenburg gate.

system (*see* **Ambient Vibration Monitoring**). In all, over 48 significant points were observed permanently (Figure 7). Most of the evaluation of the data was done in situ. The evaluation includes statistical quantities like maximum, minimum, RMS values, and their deviations. Frequency evaluation by FFT was also done at the site (*see* **Statistical Time Series Methods for SHM**).

5 THRESHOLD VALUES

In order to assess the measured dynamic quantities, threshold values have to be specified. For this purpose, national and international standards, own experiences on comparable buildings, as well as results of the modal analysis were used. Threshold values are defined as permissible values for the vibration velocity in vertical direction at the foundation and in both horizontal directions (i. e. direction "north-south" and direction "west-east") at the architrave. Besides the maximum value of vibration, the kind of vibration and the number of maximum values consecutively are also regarded. Following these criteria, the threshold values given in Table 1 were prescribed for permissible vibration velocities, depending on the kind of excitation process [1].

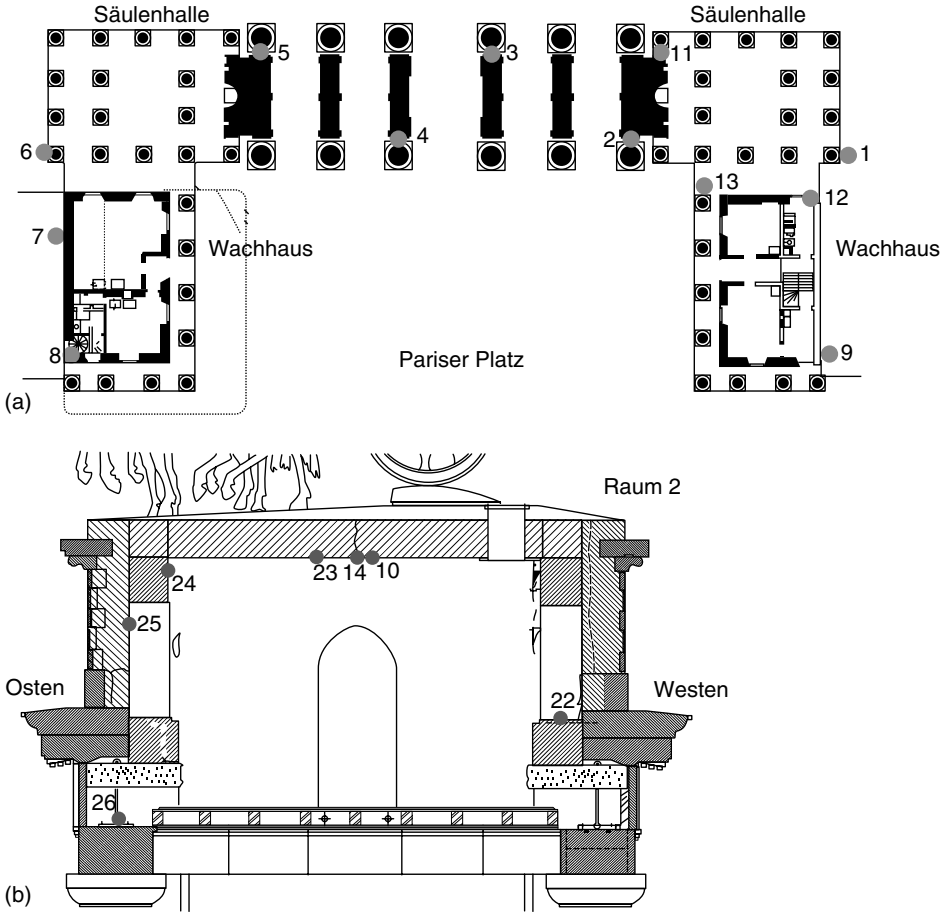


Figure 7. Measurement layout for monitoring at the foundation (a) and the architrave (b).

Table 1. Threshold values

Measuring level	Max V (mm s^{-1})	Manner of use	Permitted transgressions
Foundation	≤ 1	Durable	None
Architrave	≤ 2	Durable	None
Foundation	≤ 1	Temporal	Maximum 100%[(a)]
Architrave	≤ 2	Restricted	Maximum 50%[(a)]

[(a)] Threshold values are permitted to exceed for the duration of 1 min at the most by a frequency of three times in 24 h at the most.

6 MAIN RESULTS OF THE MODAL ANALYSIS

For the experimental modal analysis of the Brandenburg gate, ambient vibration excitation has been used. The extraction of modal data (Eigen frequencies, mode shapes, and damping) has been done by software with "output only" option. The essential results of these examinations are given below.

The main natural frequencies (Figures 8 and 9) of the Brandenburg gate with the accompanying main

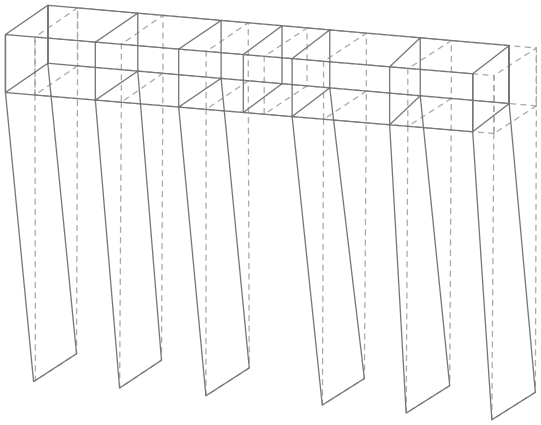


Figure 8. First Eigen mode with 1.77 Hz.

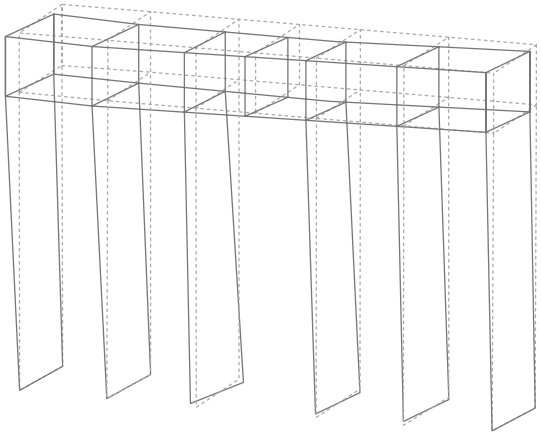


Figure 9. Second Eigen mode with 2.44 Hz.

movement directions are as follows:

$f_1 = 1.77$ Hz (rotation)
direction north "German Reichstag"

$f_2 = 2.44$ Hz (rotation)
direction east "Unter den Linden"

$f_3 = 3.17$ Hz
direction east "Unter den Linden"

$f_4 = 7.26$ Hz
direction east "Unter den Linden"

The Eigen modes to the first Eigen frequencies f_1 and f_2 have a small damping value (<1%). The damping of a structure is normally influenced by the so-called material damping and by radiation damping in the soil (influenced by movement of the foundations). The values of radiation damping are much higher than the values of material damping. Both effects result in the observation that measured damping is normally much higher than the well-known values of material damping. Therefore, in the present case, these modes can be excited easily by wind and also by heavy trucks and buses, which have first Eigen frequencies close to the Eigen frequencies to the gate.

The dominant natural frequencies of the two gate houses are not equal and differ also with those of the gate. This leads to different movements between the gate, the two gate houses, and the connection zone. In order to avoid additional cracks between the gate and the gate houses, the stiff connections that have been added to the gate at the reconstruction after the Second World War are not good measures and have to be removed.

Higher Eigen frequencies and Eigen modes with frequencies above 7–8 Hz belong to bending modes of the architrave, in general. Those modes can be excited by all vibration sources mentioned above.

7 MONITORING RESULTS

According to the results of the modal analysis, a program for the permanent condition monitoring of the gate was developed. A monitoring system having 48 measurement channels was used for the supervision task [2]. In general, vibration values, crack widths at certain locations at the gate, the two gate houses, and the transition zone between the buildings as well as temperatures were measured and recorded permanently. The measurement locations were chosen according to the results of the experimental and numerical modal analysis. From the permanent monitoring, the following results were achieved.

The relevant frequency domain of the measured vibration amplitudes is between 1.4 and 25 Hz (Figure 10). The greatest vertical vibrations at all measurement points at the gate arise from traffic excitation (mainly buses and trucks), while the horizontal movements can be explained mostly by free vibration. The influence of road traffic excitation is much more important than the influence of railroad traffic due to the S-Bahn. The reason for that is that the S-Bahn

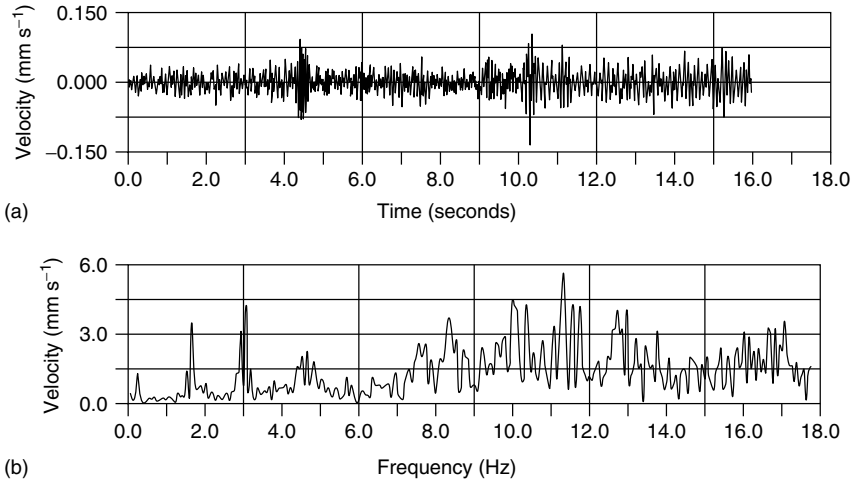


Figure 10. Time signal (a) of passage of a bus with associated frequency spectrum (b).

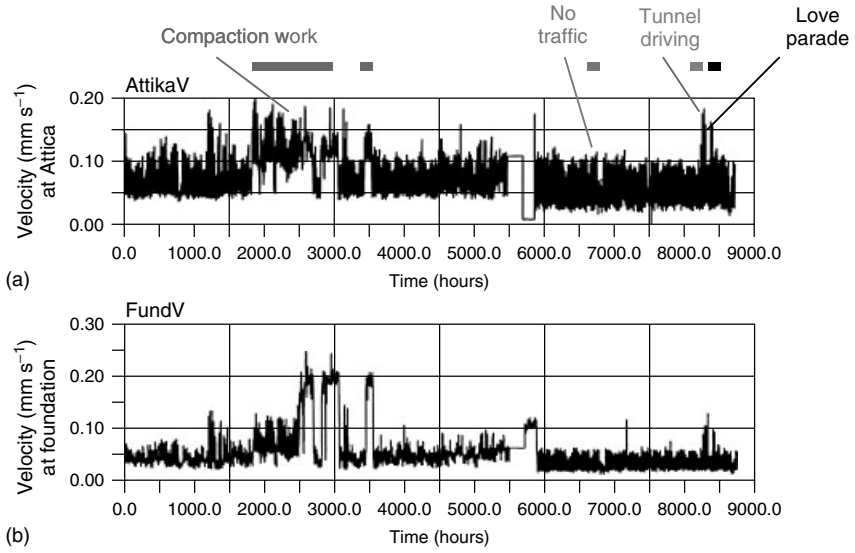


Figure 11. Comparison of vibration values of different impacts.

tunnel is only close to the southern part of one of the gate houses and therefore the vibrations transmitted through the soil are highly damped.

In Figure 11, a comparison between the different measured impacts to the gate is given. The measured impacts are railroad and road traffic, compaction work, and tunnel drilling near the gate and, last but not least, the influence of the love parade. As can be seen, the most important vibration occurs at the Attica of the gate. There is an amplification of the vibration from the foundation to the upper part of the structure. Regarding the maxima vibration values shown in Figure 11, one can see that the values initiated by the love parade and tunnel drilling are even more important than the values coming from the "normal" road and railroad traffic.

In general, the vertical vibrations are much greater than the horizontal ones. Regarding the horizontal

vibrations, those in the eastern direction (Unter den Linden) are normally much higher than those in the northern direction (German Reichstag). The maximum amplitudes in every measurement direction depend on the respective traffic events, which are given by the traffic conditions, vehicle type, and speed.

The supervision of the crack widths leads to the result that temperature is the most important cause (Figure 12). For all three monitored cracks, the long-term change of the crack width correlates strongly with the measured building temperature. This relation is about 0.1 mm/°C and behaves reciprocal to the temperature, which means that the crack opens

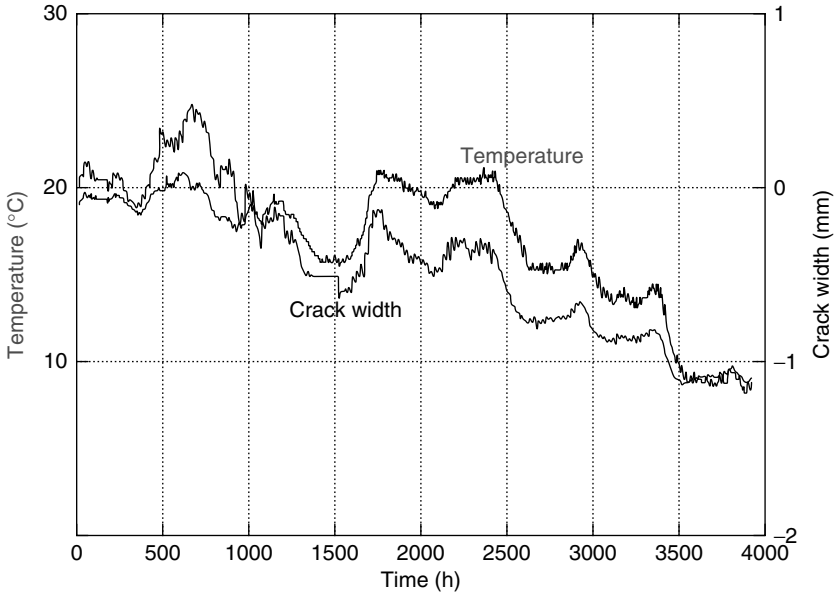


Figure 12. Crack monitoring.

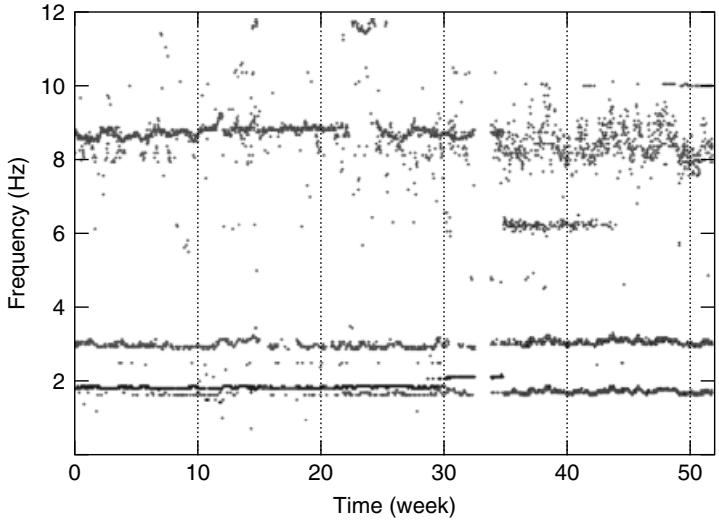


Figure 13. Frequency tracing diagram.

with decreasing temperature and vice versa. Another association between the measured crack width and the observed natural frequencies is that of a reduction in the natural frequencies when there is an increase in crack width.

For the global monitoring of the condition of the structure, a frequency observation was used. This frequency observation is done by evaluating "peak value frequency spectra" at constant time intervals and plotted versus time. This leads to a frequency tracing diagram as given in Figure 13. Natural frequencies are represented in this diagram as horizontal lines. Relevant damage of the structure is shown by deviation of the horizontal lines. Under normal conditions, the damage must be strong enough to have an indication in the frequency tracing diagram. However, in the present case, there is a significant shift in the first Eigen frequency approximately in the "33rd" week of monitoring.

If we look at the values for the first five Eigen modes (Table 2), we see that there is a decrease in all natural frequencies. Searching for the reason, we recognized that the changes occurred during the process of tunneling of a metro line beneath the foundation of the Brandenburg gate. The Eigen modes (Figure 14) belonging to the first natural frequency, was orientated in the direction of the "German Reichstag" (north) before the beginning of the work but after completion of the tunneling work, a dominant movement in the direction of "Unter den Linden" (east) can be seen.

Table 2. Eigen frequencies before and after the tunnelling work

Mode number	Eigen frequencies (Hz)	
	Before the tunnelling work	After the tunnelling work
1	1.77	1.59
2	2.44	2.29
3	3.17	2.90
4	7.26	5.80
5	8.91	8.3

Also, if we consider the mean average values of the foundation vibrations obtained by the monitoring system, it can be seen that these values are much higher than they were earlier. This observation supports the assumption that the dynamic behavior of the structure has changed because of the tunneling work (Figure 15).

In order to give an explanation for the observed behavior, simple numerical systems for the gate and the soil (including the interaction effect) were constructed and analyzed. This analysis reveals that the observed changes in the modal behavior are initiated by a decrease of the stiffness in the soil under the gate foundations.

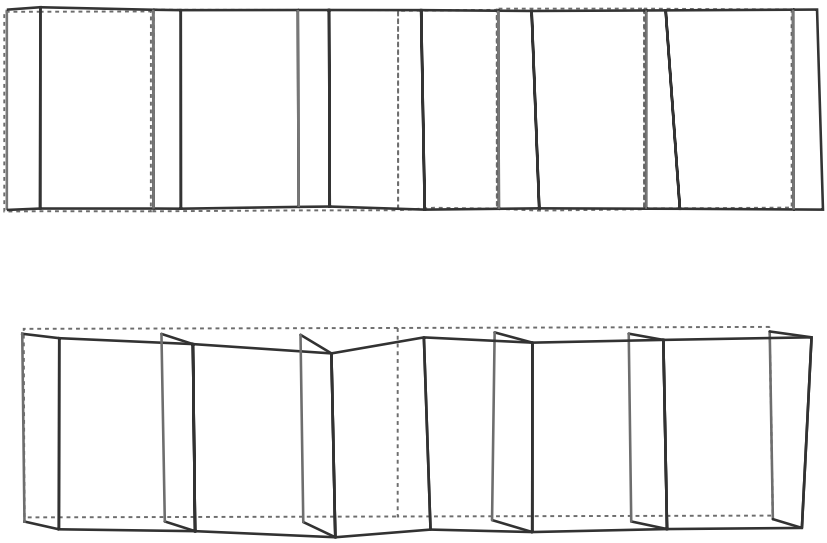


Figure 14. Mode shape for frequency 1.7 Hz before and after the completion of tunneling work.

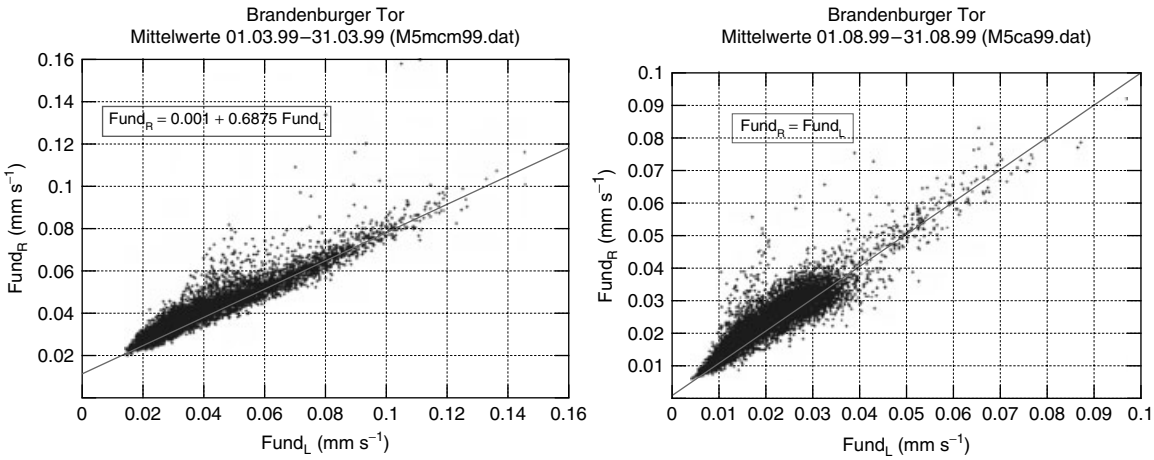


Figure 15. Regression of foundation vibrations between the left and right parts before and after the tunneling work.

8 CONCLUSIONS

After the reunification of Germany in 1989, the senate of Berlin was interested to open the Brandenburg gate again for traffic. In the context, a technical inspection of the gate was carried out, which revealed a lot of cracks and other damage at the gate and the neighboring gate houses. By an investigation program, the causes for the damage were found. To characterize the static and dynamic behaviors of the structure and to assess the different impacts and their damage effects to the structure, a three-level investigation program was developed (modal analysis, derivation of threshold values, permanent monitoring). First of all, a modal analysis of the complete building was done. Threshold values were defined by standards, own experience, and the results of the modal analysis. An important result of the modal analysis was that the first Eigen frequencies have a small damping value and can coincide with the first natural frequencies of heavy trucks and buses. Also, wind can excite the structure easily in this low-frequency domain. Another outcome of analyzing the mode shapes is that the existing stiff connection between the Brandenburg gate and the neighboring gate houses has to be removed to avoid further cracks between them. A comparison made between the different excitation sources shows that the vibration impacts induced by events like love parade are more damaging than the impacts of the normal road and rail traffic. Furthermore, it has been demonstrated that frequency tracing together with the monitoring of relevant mode shapes may be a significant damage indicator of the structure.

REFERENCES

[1] Rücker W, Rohrmann R, Said S. Monitoring and assessment of structures under changed loading conditions. *SAMCO Summer Academy 2005 on Structural Assessment, Monitoring and Control*. Austria, 5–9 September 2005, Zell am See 2005.

[2] Rücker W, Rohrmann R, Said S, Schmid W. *Dynamische Verfahren zur Sicherheitsüberwachung von Brückenbauwerken*, D-A-CH Tagung 2003, 18-19.09. 2003, Zürich, 2003; pp. 35–42, Hrsg.: SIA, ISBN 3-908483-74-3.

Chapter 122

Development of a Monitoring System for a Long-span Cantilever Truss Bridge

F. Necati Catbas[1] and A. Emin Aktan[2]

[1] *Civil and Environmental Engineering Department, University of Central Florida, Orlando, FL, USA*
[2] *Drexel Intelligent Infrastructure Institute, Drexel University, Philadelphia, PA, USA*

1 INTRODUCTION: MONITORING OF LONG-SPAN BRIDGES

Structural health monitoring (SHM) is a paradigm that enables an integrated systems approach for a reliable measurement-based understanding of the loading environment of major bridges, and how their structural systems carry their loads as they operate and fulfill their functions [1]. The SHM paradigm enables a comprehensive and integrated evaluation of the entire spectrum of performance expected from a major bridge. Management of bridge operations, response to accidents and emergencies, routine inspections, preventive maintenance as well as any structural repair or retrofit may be integrated and optimized in a rational manner based on objective, quantitative criteria customized to a specific bridge.

Currently, most of the major decisions by managers of long-span bridges are made on the basis of the visual inspections. There is evidence that visual biannual inspections of major bridges cost significantly while restricting operations for many months. Yet these inspections may miss many of the early initial signs of deterioration and damage even when these may be visible, as there are natural limitations to the ability of even experienced human eyes to scan hundreds of members and connections that may have dimensions in the order of a hundred feet.

Many bridge engineers concur that there are too many limitations and shortcomings in the current approaches to inspection, evaluation, maintenance, rehab and retrofit design, and construction of existing major bridges. Applications of SHM to existing major bridges that exhibit premature aging, distresses, and performance problems and/or to bridges that have aged beyond their anticipated design life cycles

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

offer exceptional payoff. Whether the conventional approaches to the maintenance management of such bridges are effective, especially after their aging, is a valid question. Another important question is that whether the effective implementation of SHM techniques and technologies are in such a way that the root causes of problems are detected, solutions are formulated, and finally repair and retrofit are developed to mitigate existing or future problems [2].

2 STRUCTURAL HEALTH MONITORING OF A LONG-SPAN TRUSS BRIDGE

In this study, an overview of a comprehensive monitoring project of the longest cantilever truss bridge in the United States is presented. For long-span bridges, some of the most pressing challenges are solving performance problems that are related to objectionable movements and geometry changes, displacements, vibrations, and visible signs of aging. To respond to deterioration, distress and damage to materials, elements, and connections, it is essential to first clearly identify the root causes before it is possible to design the most effective maintenance or renewal measure that can mitigate the root cause. In most cases, monitoring over an extended time may be necessary for definitively identifying the root cause(s) and mechanisms leading to symptoms of deterioration or damage. A lack of durability and serviceability limit state performance of many long-span bridges often translate to major increases in life-cycle cost, and properly designed monitoring may offer an excellent payoff by considerably reducing life-cycle cost.

2.1 Description of a long-span truss bridge

The Commodore John Barry Bridge (CBB) (Figure 1) spans the Delaware River between Chester, Pennsylvania and Bridgeport, New Jersey. The bridge has five traffic lanes and currently serves more than six

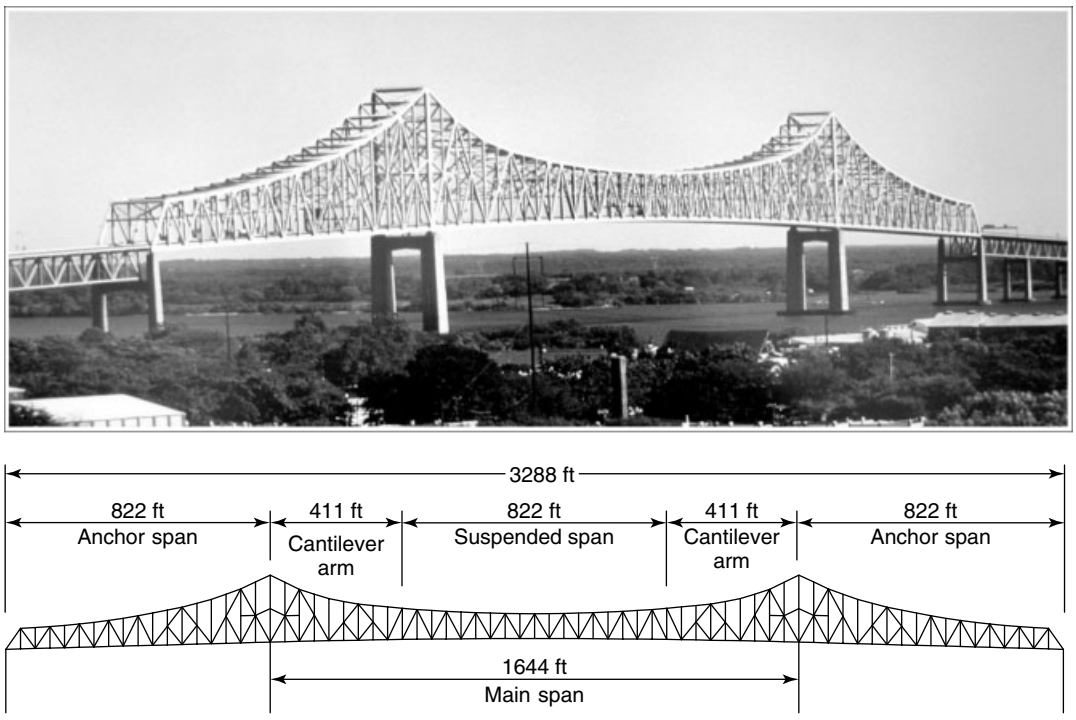


Figure 1. Commodore Barry Bridge through-truss structure (Overall Length is 1002 m with 501 m main span length).

million vehicles annually, a significant percentage of which is heavy truck traffic. It was opened to traffic in 1974 as the longest cantilever steel truss bridge in the world with a main span length of 501 m (1644 ft) and a total bridge length of 4240 m (13 912 ft). The focus of the study and subsequent discussions are directed to the principal long-span through-truss component shown in Figure 1.

The substructures of the through-truss comprised of four reinforced concrete piers that are shown in the photo in Figure 1. The piers were constructed on pile foundations. The two principal trusses of the through-truss are spaced 22 m (72.5 ft) apart. Each truss has 73 panel points spaced at 14 m (45.7 ft) intervals. The top and bottom chords of the trusses are constructed from welded box sections. A combination of welded box and I-sections are used for the vertical and diagonal truss members.

Lateral "wind" bracing is provided by K-bracing at the top and bottom chord levels, and by portal and sway frames located at various panel points throughout the structure. The suspended span of the bridge is connected to the cantilever arms via vertical hangers, which are pinned at their upper and lower extremities. Truss members with axial and rotational releases transition the top and bottom chords between the suspended span and the adjacent cantilever arms. The floor system of the bridge is an 20 cm (8 in.) thick lightweight reinforced concrete deck that is composite with nine steel beams laterally spaced at 2.1 m (6.9 ft). The beams are continuous over the floor beams in either four-span or five-span increments. Figures 2–4 further illustrate various aspects of the structure.

2.2 Purpose and expected outcomes from SHM

Given the general characteristics of the long span, the specific objectives of the study were established. A research work plan was designed to accomplish the integration of experimental, analytical, and information technologies within a coherent health monitoring approach for the following objectives:

1. Conceptualizing the structural systems of the bridge, including the full recognition of the

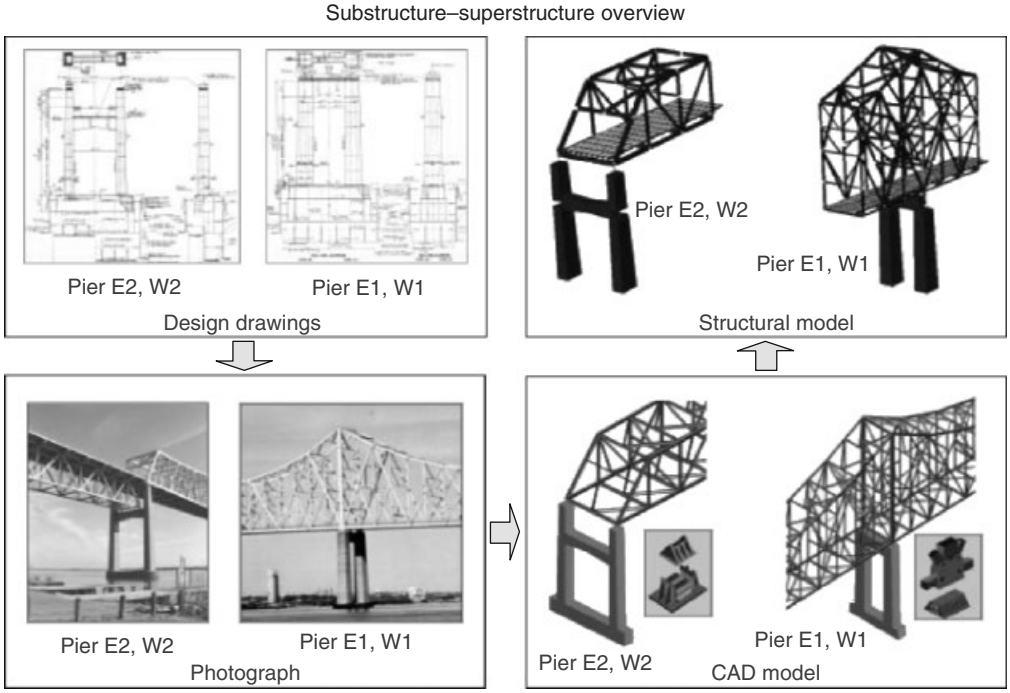


Figure 2. Design drawings, photos of bridge components, 3D CAD and FEM of the bridge.

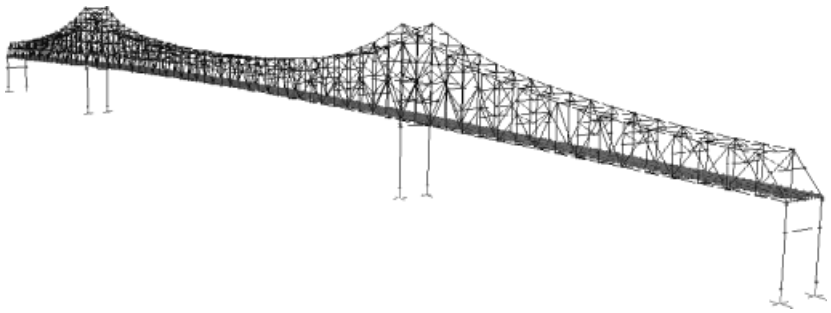


Figure 3. Finite element model of the Commodore Barry Bridge.

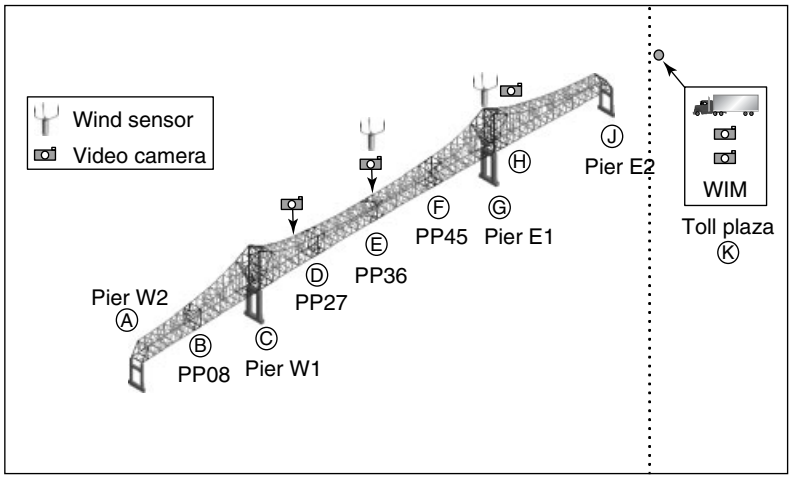


Figure 4. Sensor distribution and critical measurement locations.

natural–environmental and the sociotechnical systems that impact the bridge performance.

2. Analytical model development and calibration for detailed simulations and evaluations as well as for designing the monitoring system.
3. Development and demonstration of a complete SHM framework with Internet-based real-time monitoring modules, various sensors, data-acquisition systems, data management and communication protocols, and data analysis methods.
4. Development of various intermittent and long-term monitoring studies on the bridge to address issues such as operating vibrations, damper condition, electroslag weld evaluations, long-term behavior of critical structural members, and phenomena that is related to the performance of the concrete deck, which showed undesirable levels of cracking and deterioration.

3 ANALYTICAL ISSUES

3.1 Bridge characterization

Analytical model development for bridge characterization required a thorough review of the design and shop drawings for the bridge, inspection reports, and any additional relevant reports or documentation for the bridge to identify the current state of the structure including its performance and maintenance history [3]. Several site visits were necessary to visually examine and verify the condition and locations of complex member and connection details, retrofits to the structure, boundary conditions, and to establish access requirements for instrumentation. Photographic documentation of the critical structural details such as movement systems is generated. Using the existing design drawings and photographs, a 3D

visualization of the entire bridge as well as the critical structural details is generated by computer aided design (CAD) software (Figure 2).

A 3D finite element model (FEM) of the bridge, as shown in Figure 3, was constructed to assist in identifying the critical regions and behavior mechanisms of the bridge's structural systems and to estimate the limits of the forces, strains, tilts, displacements, and accelerations that may be necessary to measure. The FEM is calibrated through system identification procedures to permit reliable simulations based on the data from a health monitoring implementation. The data needed for system identification of the bridge and for subsequent calibration of the FEM were obtained from controlled experiments conducted on the bridge. These experiments included ambient vibration monitoring of the through-truss spans and a controlled load test using heavy cranes.

The calibrated FEM, which better reflects the true measured behavior of the bridge, may serve a number of purposes including accurate load rating analyses, vibration mitigation studies, vulnerability evaluations, maintenance/retrofit designs, as a benchmark for evaluating future changes in condition, as a foundation for evaluating system reliability and condition indicators for health monitoring, and as a starting point for any nonlinear analysis for failure limit state evaluations [3, 4].

4 LONG-TERM MONITORING ISSUES

4.1 Measured phenomena

The phenomena that should be monitored can be broadly considered in two categories. The phenomena that are understood or known, but cannot be reliably measured, analytically modeled, and accounted for in the design of the bridge, but that may affect the life-cycle performance of the bridge are in the first category. For example, the impact of settlements can be modeled and the intrinsic stress distribution can be identified by means of instrumentation or analytical modeling. The phenomena that are not clearly understood or modeled and accounted for (if at all) in design, but which may serve as causative effects for deterioration and damage are in the second category. For example, vibrations and displacements of the floor systems under various live load and wind combinations, and possible impacts of temperature shocks on movement and support systems are the phenomena that can be discovered only by long-term monitoring.

For the long-term health monitoring implementation of the CBB, the possible impacts of humidity, wind, temperature, radiation, long-term movements, tilts, slips and settlements on the intrinsic strains, and forces are measured [5]. Nonlinearity and nonstationary of boundary and continuity conditions and energy-dissipation mechanisms are recognized in the design of the health monitoring system.

The CBB health monitoring system was designed to integrate "vision" and "mechanical responses", the former by capturing streaming digital video images that monitor the traffic moving over critical areas of the bridge and the latter in the form of temperature, displacement, tilt, strain, and acceleration measurements distributed as shown in Figure 4.

4.2 Sensor characteristics, types, and quantities

In order to monitor the phenomena that have been identified in the earlier stages of the structural health monitoring (SHM) design, a number of different sensors are selected after extensive calibration and laboratory evaluation studies. The measurement phenomena are grouped into three distinct categories: (i) traffic, (ii) weather; and (iii) bridge response as given in Table 1.

4.3 Measurement and communication infrastructure

The CBB health monitoring system permits a combination of continuous, event-based, and time-based programmable as well as manually controlled online data-acquisition modes for interrogating specific sensor clusters. Since data in a long-term health monitoring implementation is generally obtained over large spatial (kilometers) and temporal spans (years to decades), and along a wide frequency band (various sensors may be interrogated once every quarter hour for ambient temperatures all the way to gigahertz frequency for acoustic emission

Table 1. Phenomena to be monitored with sensors and locations

Phenomena	To be monitored	Sensor description	Location
Traffic	Image streams	Real-time video image streams	D, E, J
	(weigh-in-motion) WIM	Bending plate scale	K
Weather	Air temperature	Self-contained weather station	G
	Humidity		
	Solar radiation		
	Wind speed	Ultrasonic sensor	E, G
Bridge response	Live load strain	350 Ω weldable	C, D, E, F
	Live load strain	350 Ω weldable	
	Environmental strain	Vibrating wire sensor	A, C, D, E, F, J
	Environmental strain	Vibrating wire sensor	
	Temperature	Thermistor	A, C, D, E, F, G, J
	Tilt	Vibrating wire sensor	A, C, D, F, G, J
	Displacement	Vibrating wire sensor	A, D, F, J
	Acceleration	Capacitive sensor	B, D, F, G
	Acceleration	Piezoelectric sensor	
	Fiber-optic strain	Bragg grating	

sensors), space-time stamping and synchronization of the output from many different data-acquisition systems distributed through a bridge is a challenge. Data of different modalities (e.g., image streams vs a single strain reading) acquired by different systems add to the challenges of accurately time-stamping and synchronizing data. The sensing system for the CBB is interrogated over a local area network, the architecture of which is illustrated in Figure 5.

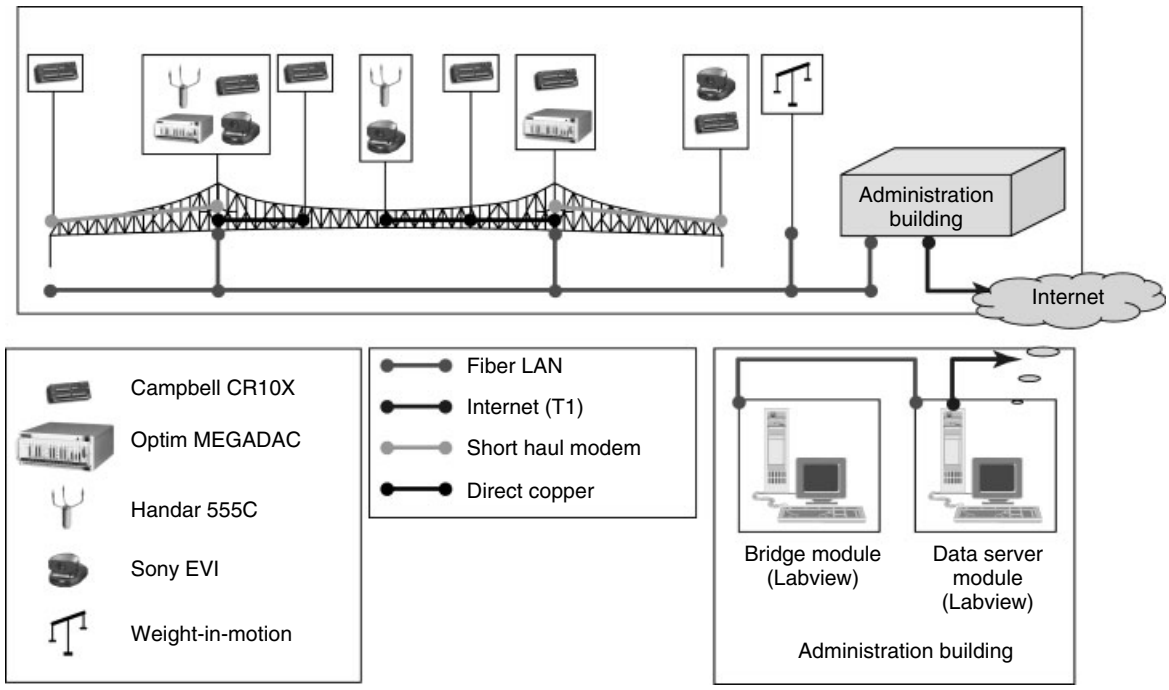


Figure 5. Measurement and communications infrastructure at the Commodore Barry Bridge.

The figure further depicts how the various data-acquisition system components are distributed and networked along the bridge. The data-acquisition systems are controlled, synchronized, and integrated by the software developed in Labview. Wind, temperature, radiation, and humidity interrogation is done with a data-acquisition system from Handar–Vaisala, vibrating-wire-based displacement, tilt, and strain sensors capture variations in the intrinsic responses over the long term and are interrogated by a dedicated data-acquisition system from Campbell Scientific, Inc., and the cameras are controlled by yet another dedicated system. The high-bandwidth strain, displacement, and acceleration sensors for high-speed responses captured over short-time increments are interrogated by a system from Optim Electronics, Inc.

4.4 Monitoring protocol, automated data acquisition and archival, and user interface

An authorized operator may take control of the CBB health monitoring system at anytime; however, the system is designed to operate in a programmed mode in which the inputs due to weather and traffic, and the entire set of vibrating-wire sensors are continuously interrogated at low frequency. The high-frequency sensors operate on time- or event-based triggered modes. For example, the system may be triggered to acquire and archive data from a subset of the complete sensor suite on the bridge, during the morning and evening rush hours when traffic levels on the bridge are highest, at midnight when traffic levels are very low, when the wind speed reaches a certain threshold value, or when a heavily loaded truck is detected by the weight-in-motion system. The frequency and duration of data and image collection, their processing, evaluation and dismissal, archival, presentation to a manager and/or alarm protocols will be eventually transformed to intelligent agents after researchers can more reliably establish the bounds of normalcy and possible indications or precursors of anomalies in operation or structural behavior. Clearly, data quality assurances, processing and archival practices represent the major information technology–related challenge in regard to health monitoring of a major bridge. Many tests are essential for data quality assurance even after the best possible sensor and data-acquisition design, operation, processing and archival practices are to be followed. Redundancy requirements in the application of sensors, integration of different types of sensors and measurement systems, calibration of the health monitoring system in the field by controlled testing, and, most importantly, justifying the output of any sensor based on the physics of the measured phenomena are techniques for data quality assurance.

The value of health monitoring applications, especially for operational and emergency management in conjunction with engineering purposes, is realized only by the visual display of critical images and data online in real-time (or near real-time). A challenge is in the integration and graphical display design of critical data streams so that users and owners may conceptualize the phenomena reflected in the measurements, in order to make timely decisions. Health monitoring design should involve the owners and engineers in charge of the operations, maintenance, and management of the bridge for maximum benefit. User communication, information and alert protocols, and training and maintenance support needs are major challenges related to monitor-user-organizational interface design. Figure 6 illustrates the interface designed for viewing real-time images from the bridge and information from the weight-in-motion system as well as the weather station.

4.5 Brief overview of findings

Approximately, 16 accelerometers were used to simultaneously capture the traffic-induced vibrations and their attenuations or amplification through the deck, the railing, the floor system, and the trusses. Frequency-domain transformations and cross correlations revealed the dominant input vibrations occurred at a frequency band of 5–15 Hz and with an amplitude of about 0.25 g at the deck, attenuating to an amplitude of 0.15 g at the truss lower chord. Impact-modal analysis of several railing elements revealed that the railing fundamental frequency in the lateral and vertical directions was about the same, 10 Hz, and this coincided with the frequency of the input excitation. The coupled lateral–vertical resonance of the railing elements caused extensive damage to the railing.

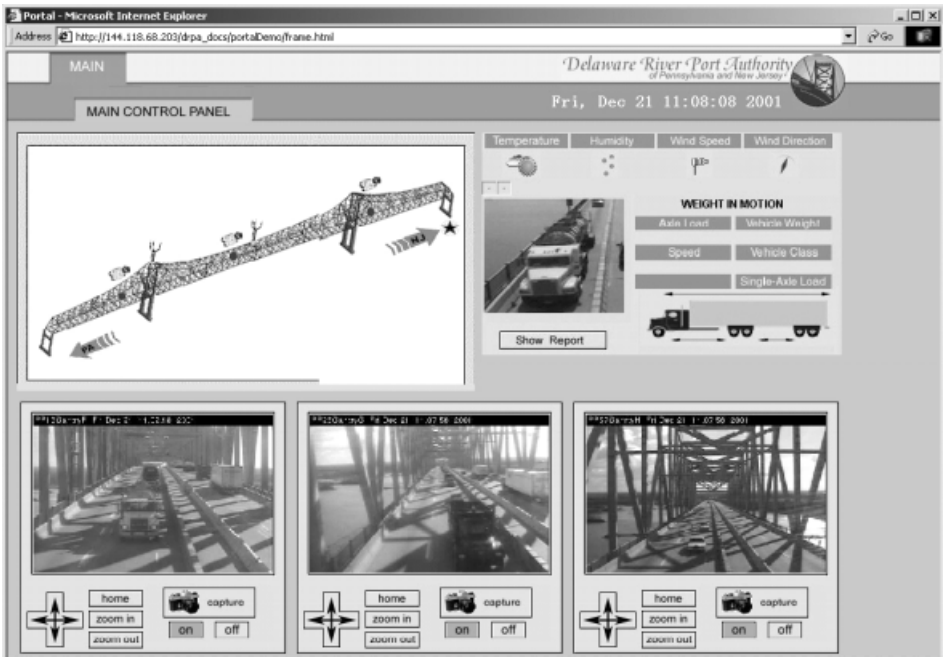


Figure 6. Internet-based user interface of the CBB health monitor system.

The time-histograms of member and damper acceleration responses monitored intermittently indicate that acceleration amplitudes that were about 0.15 g without the dampers are being reduced to about 0.06 g with the dampers. The importance of the dampers in controlling operating vibrations was therefore clear from monitoring these results. Dampers were found to have been well tuned to the member frequencies for optimum energy dissipation. The chemical and physical properties of the neoprene material were investigated further by cutting samples for chemical analysis. The results did not reveal any deterioration. Although the damper and member vibrations are yet to be monitored under sufficiently high wind that would cause wind-induced excitation of a member, it is clear that the dampers are highly effective and critical for controlling truss vibrations.

Temperature changes and solar radiation were observed to be the most significant load effect on the trusses. For example, Figure 7 shows the annual change in the intrinsic strains of one of the hangers, together with the temperatures recorded during the year. Strains are observed to vary significantly in the order of 41 kPa (6 Ksi) or more during days when large temperature changes occur within a short time. An annual seasonal variation of up to 69 kPa (10 Ksi) is observed in the intrinsic strains correlating perfectly with the temperature. This is a considerable stress when calculated dead load stresses varied between 138 kPa and 207 kPa (20 and 30 Ksi) at most truss members.

A closer scrutiny of the measured strain and temperature histograms indicated that the hanger intrinsic strains were affected by the complex movement and force-release systems at and in the vicinity of these members. A distinctly unsymmetric behavior of the long-term strains of the two instrumented hangers was attributed to a difference in the behavior of the movement systems at their respective boundaries on the north and the south trusses. In addition, an out-of-plane behavior was noted in the hangers that were expected to be concentrically loaded due to radiation and temperature changes.

In addition to continuous long-term monitoring, intermittent monitoring under operating traffic and known truck loads were conducted. Figure 8 shows a crane that was positioned with another crane in static configurations as well as crawled along the bridge for this test that required closure of the bridge.

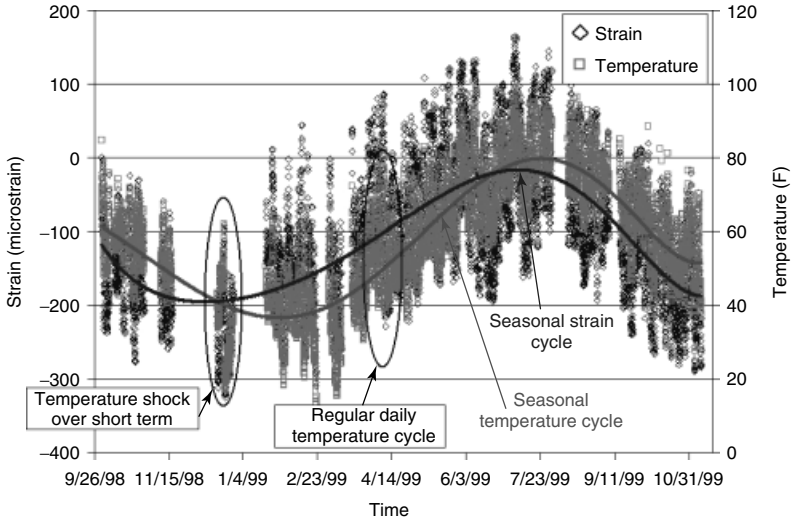


Figure 7. Long-term changes in intrinsic hanger strains with temperature.

Two 480 kN (108 Kip) cranes were used for loading and 52 high-speed strain gauges were used for recording critical member strains during the tests. Both the floor system and the truss responses were captured as the cranes were statically positioned at critical locations. Following this, a crane crawled on each of the five lanes throughout the bridge. By conducting the tests between midnight and 4 a.m., it was possible to maintain reasonably constant ambient conditions. Figure 4 shows the influence coefficients obtained for one of the hangers and a lower chord member at midspan by allowing the crane to crawl along each lane. These influence coefficients may be further normalized by decomposing them into influence coefficients corresponding to single-axle loads, and these may serve as an excellent index capturing the as-is structural behavior.

In addition to truck load test, ambient vibration tests were conducted to determine the global vibration characteristics and to provide data for finite element calibration. To maximize the spatial resolution of the mode shapes, one-half of the bridge was instrumented. It is possible to develop multiple instrumentation grids, which are then numerically spliced. However, roving the sensors reduces the reliability in the data and the roving option was eliminated. In order to verify that the test instrumentation grid will be adequate and for optimizing the sensor locations, a finite element analysis of the bridge was conducted. Analysis results helped to estimate the fundamental frequencies of the bridge, the frequency band of interest as well as the frequency spacing. 32 PCB Model 393C ICP and 13 PCB capacitive accelerometers were placed on the main truss using magnets. Two ICP signal conditioners and a capacity signal conditioner, in conjunction with an Agilent Technologies VXI system with three 16-channel E1432A boards, were used for data acquisition. The data sets were collected for 6-min intervals. For each sensor channel, time and frequency plots are defined using HP DAC Express software. This allowed real-time monitoring and quality control of the data both in time and frequency domains as it was acquired.

When the dynamic characteristics, i.e., mode shapes and frequencies, are compared, it is seen that the nominal FEM underestimated the global stiffness of the structure by as much as 40% (assuming that the inertia is properly modeled, a 20% discrepancy in frequency would correspond to over 40% discrepancy in stiffness). Obviously, unless a FEM is properly tested, calibrated and verified, it is not advisable to make critical decisions based on simulations. In addition, there were discrepancies between the measured stresses at critical elements and the FEM results. After understanding the impact of various assumptions on boundary and continuity conditions on the dynamic properties, the pier stiffnesses were increased substantially, and all

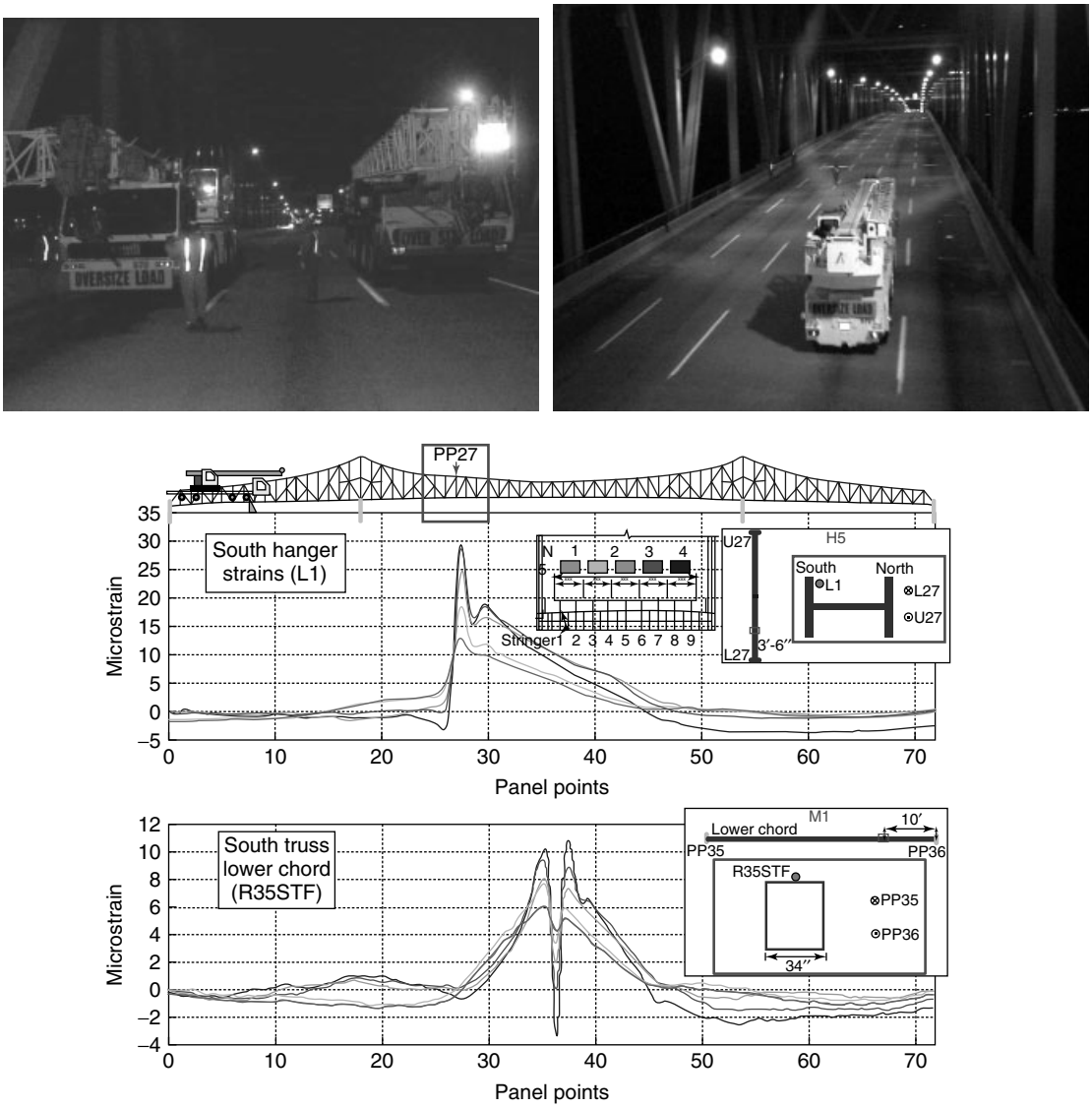


Figure 8. Loading cranes and sample influence lines from controlled testing.

the movement mechanisms were fixed before the model dynamic properties approached the measured properties. After a sufficiently close correlation was obtained between the measured and simulated global dynamic properties, the strains at the critical elements obtained during the controlled load testing were used for local calibration of the floor system element and connection stiffnesses. When a fully composite behavior of the deck and the stringers and full continuity between the stringers and the deck were simulated, the rms of the error, between measured and simulated strains at 20 measurement locations, was reduced from 55 to 24%. The assumptions on global and local stiffness and continuity brought the rms of the errors in the first six frequencies from 28 to 1.5%. Hence, the model was considered to be sufficiently calibrated. This FEM would be suitable for many purposes as discussed earlier.

5 CONCLUDING REMARKS

The health monitoring paradigm has been defined and described, potential benefits from its various application modes are suggested, and an ongoing application to a major long-span bridge has been summarized. The health monitoring applications to the CBB have provided a rare wealth of data and information about the loading mechanisms and behavior of the bridge, such as documenting the dynamic frequency band of the input excitation and how this creates near-resonance with various elements of the floor system. An in-depth intuitive understanding and quantitative documentation of the loading effects of wind, temperature, and traffic as well as the corresponding strains and forces were possible. The challenge is in developing monitoring systems that will help solve problems such as deck performance, which might be owing to many different factors such as material properties and excessive vibrations.

It is important that although the monitor system offered great promise as a technology that would potentially assist the bridge owner to develop and implement an integrated operational and maintenance management plan to minimize its life-cycle cost, this would have required a major investment into technology maintenance and technology management for the owner agency. At present, such a high level of technology maintenance and management expertise may be too costly for many of the bridge agencies in the United States. Both societal and organizational reforms are needed before technology-intensive paradigms such as health monitoring for operational and maintenance management may become a common application in the United States.

This brief review of the project articulated the need for an integrated systems approach to technology development and for the leveraging of technology in order to make a difference in how we manage major bridges. Integrated systems approach requires that we correctly identify, measure, understand, and then incorporate, in any management decision, the interactions between natural, sociotechnical, and constructed systems that govern the performance of all infrastructure components and systems. While technology advances are relatively easier to accomplish, these cannot be of great use unless they are accompanied by organizational and societal advances. The importance of technology management in infrastructure agencies is a most important issue following the experiences of the writers.

ACKNOWLEDGMENTS

The authors would like to acknowledge the invaluable contributions to the research made by their many multidisciplinary colleagues, researchers, and students at the Drexel Intelligent Infrastructure and Transportation Safety Institute. The research described in this article was sponsored by the Delaware River Port Authority and the Federal Highway Administration (FHWA). The support of these agencies is gratefully appreciated. The authors would especially like to acknowledge Messrs. Box, Faust, McCulloch and Bistline from DRPA for their collaboration. The authors also thank Dr Chase and Dr Ghasemi of FHWA for their continued support and interest over the years. Finally, the authors acknowledge National Science Foundation (NSF) for various grants that supported the authors' research on health monitoring of bridges.

REFERENCES

[1] Aktan AE, Catbas FN, Grimmelsman KA, Pervizpour M. *A Model Health Monitoring Guide for Major Bridges*, Report to FWHA, DOT/FHWA Solicitation: DTFH61-01-Q-00072, September 2002.

[2] Catbas FN, Pervizpour M, Grimmelsman KA, Aktan AE. The health monitoring paradigm for infrastructure systems. *Proceedings of the Workshop on Structural Health Monitoring and Diagnostics of Bridge Infrastructure*. University of California, La Jolla, San Diego, CA, 7–8 March 2003.

[3] Aktan AE, Catbas FN, Turer A, Zhang ZF. Structural identification: analytical aspects. *Journal of Structural Engineering, ASCE* 1998 **124**(7):817–829.

[4] Catbas FN, Ciloglu SK, Hasancebi O, Grimmelsman KA, Aktan AE. Limitations in structural identification of large constructed structures. *Journal of Structural Engineering, ASCE* 2007 **133**:1051–1066.

[5] Barrish RA, Grimmelsman KA, Aktan AE. Instrumented monitoring of the Commodore Barry bridge. *Proceedings of the Fifth International Symposium on Nondestructive Evaluation and Health Monitoring of Aging Infrastructure*. The International Society for Optical Engineering: Newport Beach, CA, March 2000; Vol. 3995, pp. 98–111.

Chapter 123

Modular Architecture of SHM System for Cable-supported Bridges

Kai-Yuen Wong[1] and Yi-Qing Ni[2]

[1] *Highways Department, Government of Hong Kong, China*
[2] *Department of Civil and Structural Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong, China*

1 INTRODUCTION

Bridge health monitoring is the tracing of the structural health conditions of the bridge in terms of the physical parameters categorized as environmental loads and status, traffic loads, bridge features, and bridge responses by reliably measured data and evaluation techniques, in conjunction with inductive reasoning and experience so that the current and expected future performance of the bridge, for at least the most critical limit events, can be predicted or evaluated. Bridge health monitoring system has been adopted in the past decade to monitor and evaluate the structural health conditions of cable-supported bridges in Hong Kong [1–5]. The bridge health monitoring system in Hong Kong is referred to as *wind and structural health monitoring system* (*WASHMS*). Bridge health monitoring system is currently considered as an integral part of bridge operation, bridge inspection, and bridge maintenance and has been included as a standard mechatronic system in the design and construction of most large-scale and multidisciplinary bridge projects such as Stonecutters Bridge (SCB) in Hong Kong [6], and Sutong Bridge [7] and Donghai Bridge in mainland China [8]. The experience gained in the design, installation, operation, maintenance, and development of the WASHMS for Tsing Ma Bridge (TMB), Kap Shui Mun Bridge (KSMB), Ting Kau Bridge (TKB), and the cable-stayed bridge in Hong Kong side of the Hong Kong-Shenzhen Western Corridor (HSWC) has a significant influence on the design of new structural

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

health monitoring systems (SHMSs) in Hong Kong and in mainland China. The WASHMS in Hong Kong, which is based on modular design concept, has been improved particularly in the aspect of data interpretation, health evaluation, and data management and such improvements have been incorporated in the design and installation of the WASHMS for SCB (SCB-WASHMS) [6].

2 SYSTEM ARCHITECTURE OF WASHMS

The modular concept of WASHMS [9–11], which is devised to monitor structural condition and evaluate structural degradation as it occurs rather than to detect structural failures, is composed of six integrated modules, namely, module 1—sensory system (SS); module 2—data acquisition and transmission system (DATS); module 3—data processing and control system (DPCS); module 4—structural health evaluation system (SHES); module 5—structural health data management system (SHDMS); and module 6—inspection and maintenance system (IMS). Figure 1 illustrates the modular architecture and input/output block diagram of WASHMS. Of these six modules, modules 1–3, which are the key modular systems for the execution of real-time structural health monitoring, are composed of different types of SSs, data/video logging systems, cabling network systems, servers and software facilities for data acquisition, transmission, and processing. The schematic layout of typical connections among modules 1–3 is shown in Figure 2. Modules 4 and 5, which are key modular systems for the execution of offtime structural health evaluation, are composed of servers, workstations, and software facilities for data interpretation, health evaluation, data management, and generation of reports. Module 6 is a set of portable computers (that store the system design information and operation and maintenance manual of WASHMS) and tools for carrying out inspection and minor maintenance of the WASHMS itself.

2.1 Module 1—sensory system (SS)

The SS that refers to the sensors and their corresponding interfacing units for input signals gathered

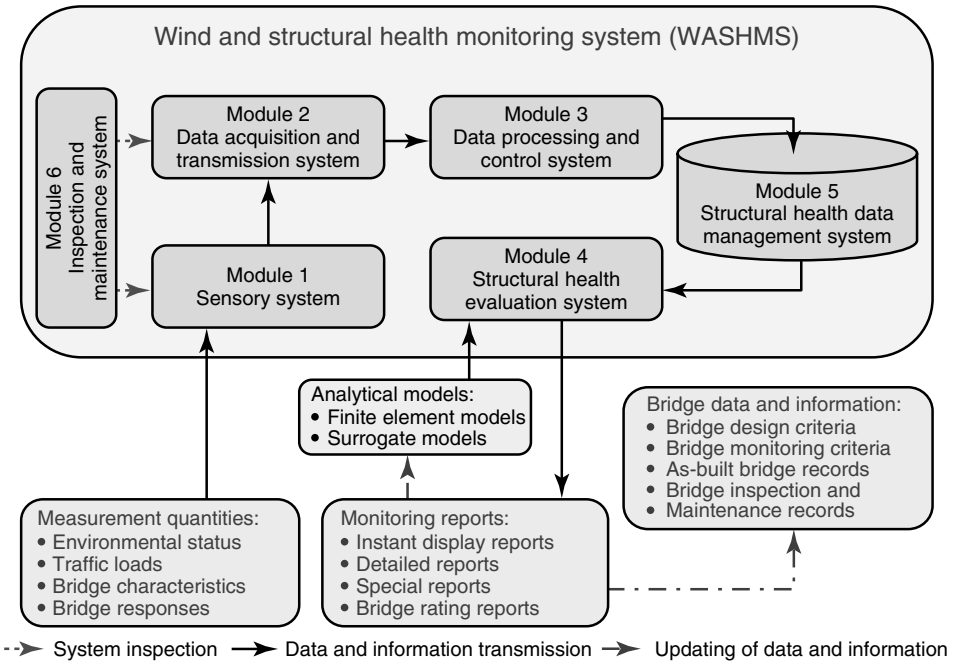


Figure 1. Modular architecture and input/output block diagrams of WASHMS.

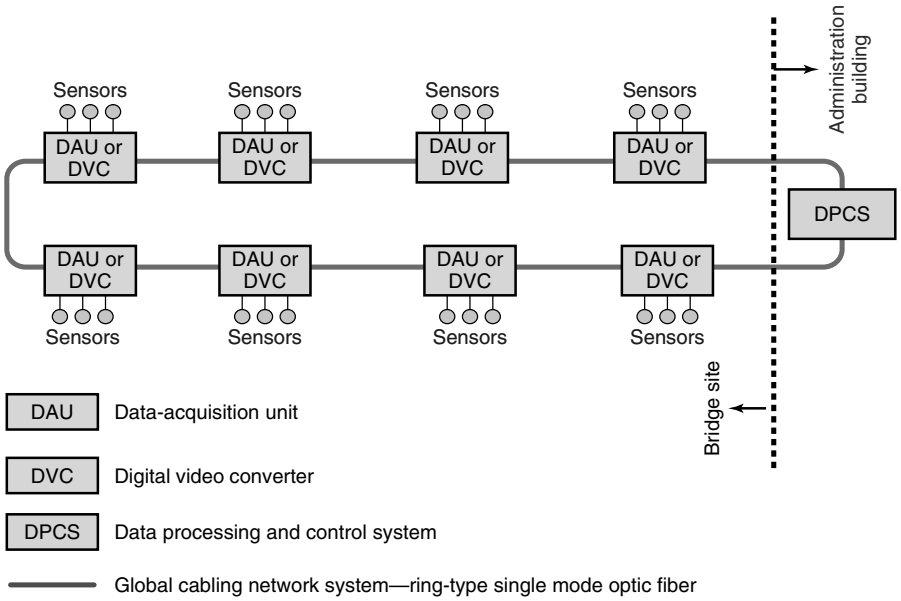


Figure 2. Schematic layout of module 2—data acquisition and transmission system.

from various monitoring equipment and sensors is categorized into four groups: (i) sensors for monitoring of environmental loads and/or status, which include anemometers (three-dimensional ultrasonic type and two-dimensional propeller type), temperature sensors (for the measurement of temperatures in respective air, asphalt pavement sections, structural concrete sections, structural steel sections, suspension cables, and stay cables), corrosion cells, hygrometers, barometers, and rainfall gauges; (ii) sensors for monitoring of traffic loads, which include dynamic weigh-in-motion stations, digital video cameras, and dynamic (weldable foil type) strain gauges; (iii) sensors for monitoring of bridge characteristics, which include fixed and removable/portable servo-type accelerometers, global positioning systems, level sensing stations, and dynamic strain gauges; and (iv) sensors for monitoring of bridge responses, which include dynamic strain gauges, static (vibrating wire type) strain gauges, displacement transducers, global positioning systems, tiltmeters, fixed servo-type accelerometers, buffer sensors, bearing sensors, and elastomagnetic sensors.

In the design/selection of the types and locations of SSs, the following six basic criteria should be fulfilled: (i) the SS should have the ability to capture the local- and system-level responses, which could be correlated or compared with the design values; (ii) the SS is required to integrate the predictive modeling and data interrogation processes with the sensing system design process; (iii) the SS should have the function to acquire data in a consistent and retrievable manner for long-term statistical data processing and analysis; (iv) all sensors should be chosen from the contemporary commercially available sensors that best match the defined sensing performance requirements; (v) the SS should include additional measurements by removable/portable sensors to quantify changing operational and environmental conditions; and (vi) at key locations, different types of sensors should be deployed so that cross calibration of sensors could be carried out.

The major parameters monitored by each type of sensors are listed in Table 1. Figures 3–7 illustrate the layouts of the SSs in TMB, KSMB, TKB, HSWC, and SCB on their respective full three-dimensional finite element models, which are built by MSC-PATRAN. The layouts of the SSs as shown in Figures 3–7 are deployed or arranged in such a manner that the measured raw data can be used to derive the information or parameters as listed in the fourth column of Table 1. These derived information and parameters are then used to compare/correlate with (i) the corresponding bridge

Table 1. List of required sensory systems and physical parameters for processing and derivation

Monitoring category	Physical quantity	Required types of sensory systems	Physical parameters for processing and derivation
Environments	Wind load monitoring	• Ultrasonic-type anemometers • Propeller-type anemometers • Barometers[a,b] • Rainfall gauges[a,b] • Hygrometers[a,b]	• Wind speed and wind direction plots (time-series data) • Wind speeds (mean and gust) and directions (histograms) • Terrain factors and wind speed profile plots • Wind rose diagrams • Wind incidences at deck level • Wind turbulence intensities and intensity profile plots • Wind turbulent time- and length-scale plots • Wind turbulent spectrum and cospectrum plots • Wind turbulent horizontal and vertical coherence plots • Wind response and wind load transfer function • Wind-induced accumulated fatigue damage • Histograms of air pressure, rainfall, and humidity
	Temperature load monitoring	• Platinum resistance temperature detector (RTD) type for temperature measurements in structural steel, concrete, asphalt pavement, and air • Thermocouplers for cables	• Effective temperatures in towers, deck, and cables • Differential temperatures in deck and tower • Air temperatures and asphalt pavement temperatures • Temperature response • Temperature load transfer function
	Seismic load monitoring	• Fixed servo-type accelerometers	• Acceleration spectra near tower and anchorage • Deck and tower response spectra • Seismic response and seismic load transfer function
	Corrosion status monitoring[a,b]	• Corrosion sensors[a,b] • Hygrometers[a,b]	• Potential risk of rebar corrosion in concrete towers, concrete piers in side spans, and concrete deck in side spans
Traffic loads	Highway traffic load monitoring	• Dynamic weigh-in-motion stations (bending-plate type) • Dynamic strain gauges • closed circuit television (CCTV) cameras[c–e] • Digital video cameras[a,b]	• GVW spectrum in each traffic lane • AW spectrum in each traffic lane • Equivalent number of SFV spectrum in each traffic lane • Equivalent number of SFA spectrum in each traffic lane • Highway-induced accumulated fatigue damage—SFV • Highway-induced accumulated fatigue damage—SFA • Overload vehicles detection • Traffic composition in each traffic lane • Traffic load response and traffic load transfer function

Table 1. (*continued*)

Monitoring category	Physical quantity	Required types of sensory systems	Physical parameters for processing and derivation
	Railway traffic load monitoring[c,d]	• Dynamic strain gauges[c,d] • CCTV video cameras[c,d]	• Bogie loads in each line of train • Train loading spectrum • Equivalent standard load (train) spectrum • Train-induced accumulated fatigue damage • Train load response and train load transfer function
Bridge features	Static influence coefficient monitoring	• Level sensing stations[c,d] • global positioning system (GPS)[b–e] • Dynamic strain gauges	• Lane stress history in each traffic lane—each vehicular type • Stress range of each type of vehicle in each traffic lane • Influence surfaces for combined deck plates and troughs • Line stress history of each type of train • Stress range of each type of train • Influence coefficients at tower tops and deck midspan
	Global dynamic characteristics monitoring	• Fixed and portable servo-type accelerometers	• Global bridge modal frequencies • Global bridge vibration modes • Global bridge modal damping ratios (derived) • Global bridge modal mass participation factors (derived)
Bridge responses	Cable forces monitoring	• Portable servo-type accelerometers	• Cable frequencies and hence cable forces • Cable damping ratios
	Geometry monitoring	• GPS[b–e] • Level sensing stations[c,d] • Displacement transducers • Servo-type accelerometers • Static strain gauges[c]	• Thermal movements of cables, deck, and towers • Wind movements in cables, deck, and towers • Seismic movements in deck and towers • Highway load movement in deck and cables • Railway load movement in deck and cables[c,d] • Creep and shrinkage effects in concrete towers[a,b]
	Stress monitoring	• Dynamic strain gauges • Static strain gauges[a,b] • Elastomagnetic sensors[b]	• Stress historical plots of instrumented components • Stress demand ratio plots of instrumented components • Principal stress plots of instrumented components • Force demand ratios plots of concrete–steel interfaces
	Fatigue life monitoring	• Dynamic strain gauges	• Total accumulated fatigue damage and hence, remaining fatigue life due to combined load effects • Accumulated fatigue damage and hence remaining fatigue life estimation due to individual load effects

(*continued overleaf*)

Table 1. (*continued*)

Monitoring category	Physical quantity	Required types of sensory systems	Physical parameters for processing and derivation
	Articulation monitoring	• Dynamic strain gauges • Displacement transducers • Bearing sensors[(b)] • Buffer sensors[(b)]	• Stress histories in bearings[(c,e)] • Stress demand ratios in bearing • Motion status in movement joints • Stress and motion status in buffers[(b)]

GVW, gross vehicular weight; AW, axle weight; SFV, standard fatigue vehicle; SFA, standard fatigue axle.
[(a)] HSWC.
[(b)] SCB.
[(c)] TMB.
[(d)] KSMB.
[(e)] TKB.

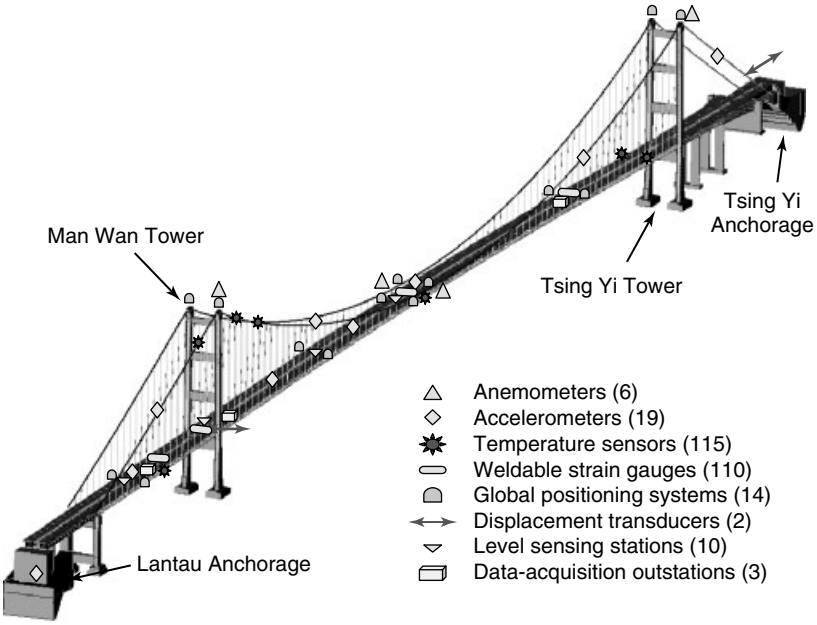


Figure 3. Layout of sensory system and data-acquisition system in Tsing Ma Bridge.

design information and parameters for detection of any significant deviation from design values; (ii) previous similar measured/derived values for detection of any abnormal or adverse structural performance; and (iii) analytical results from numerical or physical models for estimating the extent of damage, if any. Portable or removable servo-type accelerometers are also deployed with the main purposes of (i) calibration of the full three-dimensional finite element model of the bridge and (ii) extraction of high-order frequencies and mode shapes from time history acceleration data for facilitating future damage detection works. The deployment of the measurement system should also be able to quantify the changing operational and environmental conditions and to provide information for developing future loading prediction models.

2.2 Module 2—data acquisition and transmission system (DATS)

The DATS is composed of four subsystems, namely, data-acquisition system (DAS), local cabling network

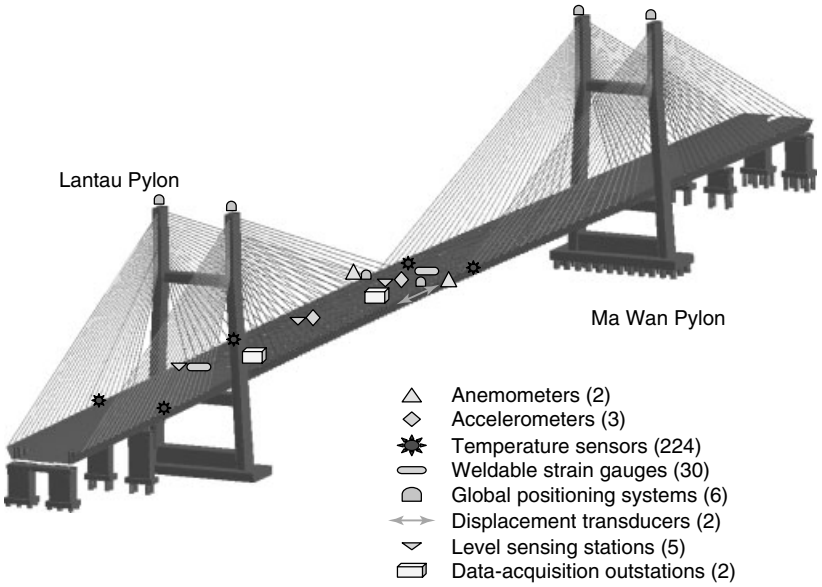


Figure 4. Layout of sensory system and data-acquisition system in Kap Shui Mun Bridge.

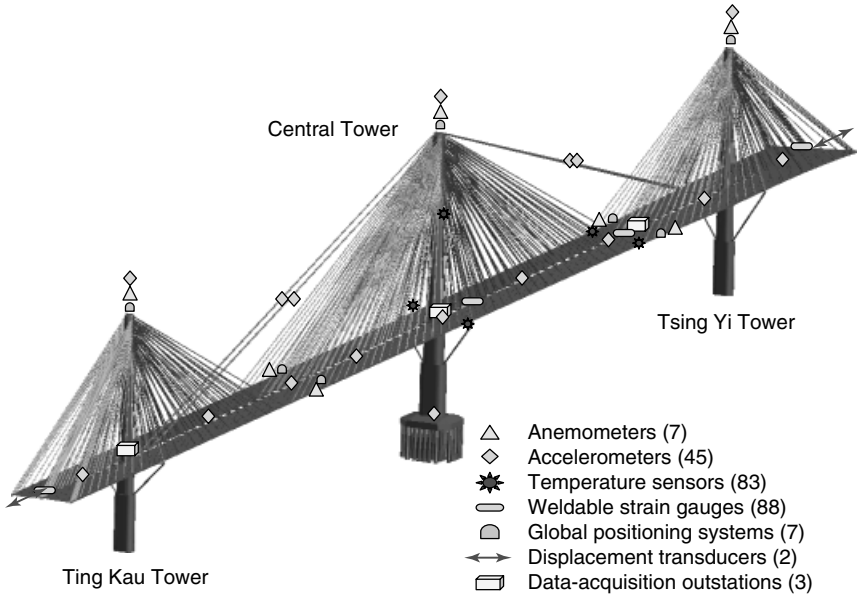


Figure 5. Layout of sensory system and data-acquisition system in Ting Kau Bridge.

system (LCNS), global cabling network system (GCNS), and commercial cabling network system (CCNS). The DAS is composed of fixed data-acquisition units (DAUs), portable DAUs, and digital video converters (DVCs) for collection of respective random and digital video signals. All DAUs and DVCs are PC-based equipment. The fixed DAUs and DVCs are permanently installed in the bridge deck and bridge towers for collection and processing of the signals received from SS (excluding corrosion cells).

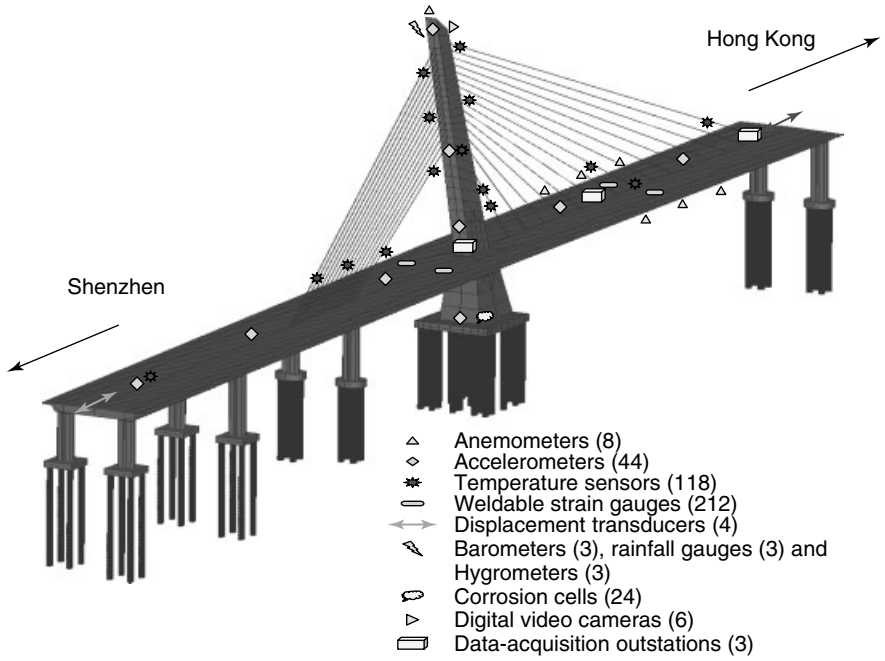


Figure 6. Layout of sensory system and data-acquisition system in Hong Kong–Shenzhen Western Corridor (Hong Kong side cable-stayed bridge).

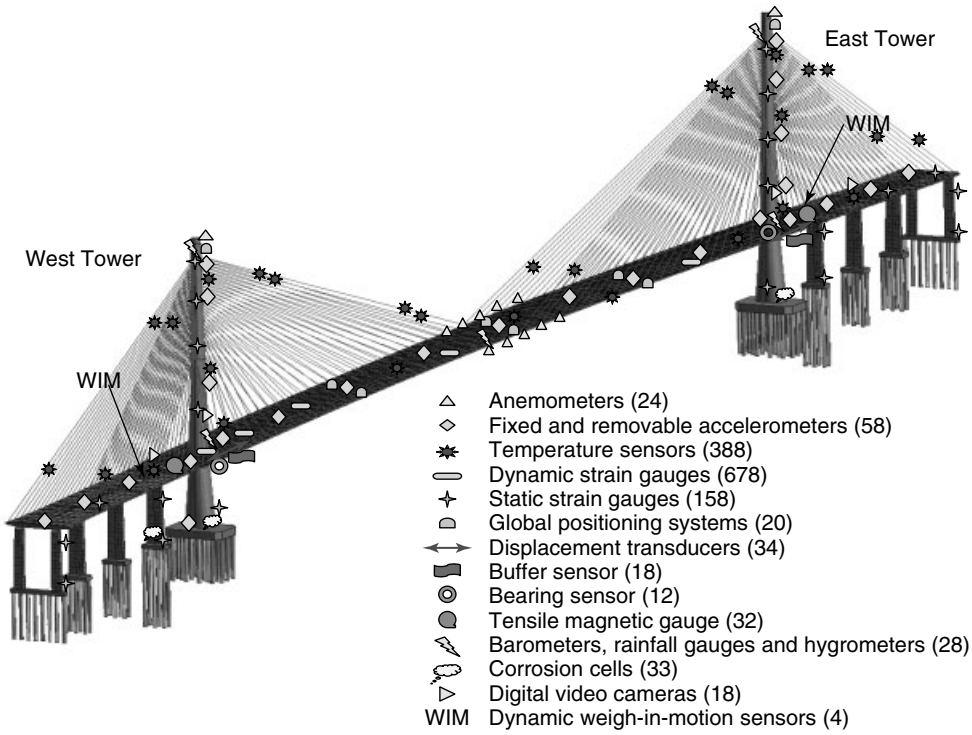


Figure 7. Layout of sensory system and data-acquisition system in Stonecutters Bridge.

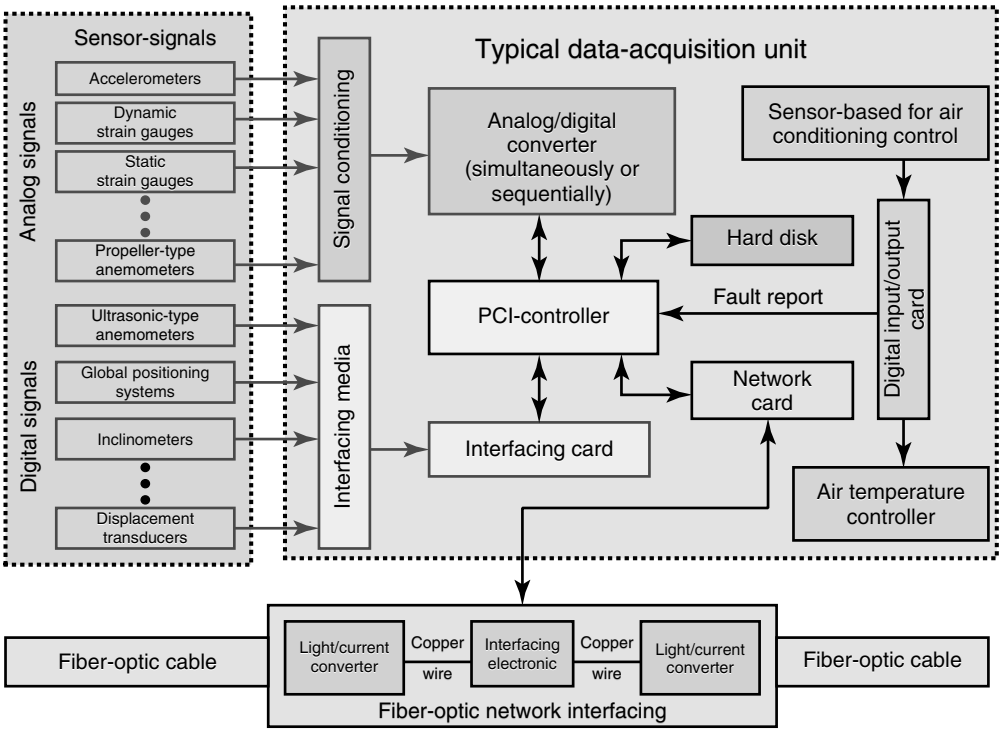


Figure 8. Schematic layout of typical connection among DAU, SS, and GCNS.

The portable DAUs are used to collect signals from portable servo-type accelerometers and corrosion cells during ambient vibration measurements and/or specified field measurement works.

Figure 8 shows the schematic layout of a typical connection among SS, DAU, and GCNS (fiber-optic cable). The major components in the DAU are the peripheral components interconnect (PCI)-controller, the signal conditioning device, and the analog-to-digital converter, and the proper selection/design of these components are the key steps to obtain measurement data with high quality. The LCNS is composed of two local cabling networks, namely, the copper cabling network for transmission of the signals from SS (excluding global positioning systems and digital video cameras, which are transmitted by fiber-optic cables) to DAUs for random signals and DVCs for digital video signals, as shown in Figure 2.

The GCNS is composed of two backbone cabling networks, namely, the random signal transmission cabling network for transmission of digitized signals (excluding digital video cameras) from individual DAUs to DPCS-1 and the digital video signal transmission cabling network for transmission of digital video signals from individual DVCs to DPCS-2. Both backbone cabling networks are ring-shaped single-mode fiber-optic cabling networks with a data transmission capacity of 1 Gbps.

The CCNS is the leased high-speed line with a data transmission rate of not less than 40 Mbps for data communication (i) between the data-acquisition outstations in bridge site of HSWC and the bridge monitoring room in West Control Building and (ii) between the bridge monitoring room in the Tsing Yi Administration Building at North Tsing Yi and the bridge monitoring room in West Control Building in South Tsing Yi (or Tsing Ma Control Area).

2.3 Module 3—data processing and control system (DPCS)

Figure 2 shows that the measured data collected from module 1 are preprocessed and transmitted by module 2 to module 3 or DPCS, which is composed

of two high-performance servers, namely, DPCS-1 and DPCS-2 for data processing and control of random (digital) signals and video (digital) signals, respectively. The DPCS is devised to carry out four operational functions, namely, system control, system operation display, bridge operation display, and post processing and analysis of data. Figures 9 and 10 illustrate the respective functions of DPCS-1 and DPCS-2 in block diagrams.

2.4 Module 4—structural health evaluation system (SHES)

Figure 2 shows the comparison part on evaluation criteria or the offtime structural health evaluation that is devised to be taken up by the SHES or module 4, which is composed of two high-performance servers (one mainly for MSC and MATALB software tools and the other mainly for ANSYS and MATALB software tools) equipped with appropriate software tools to carry out the following operational functions:

- **Finite element software interfacing capability**

That is, finite element models built by MSC/PATRAN can be transferred to and executed in ANSYS software without any manual modifications, and vice versa.

- **Integration of finite element models and measured data**

The automatic input of measured data into relevant finite element models and under relevant (predefined) solvers for execution is therefore required.

- **Analytical and experimental modal analyses**

The automatic extractions and plots of the global dynamic features of global bridge structural system and local bridge components from the measured time history acceleration data are required. Figure 11

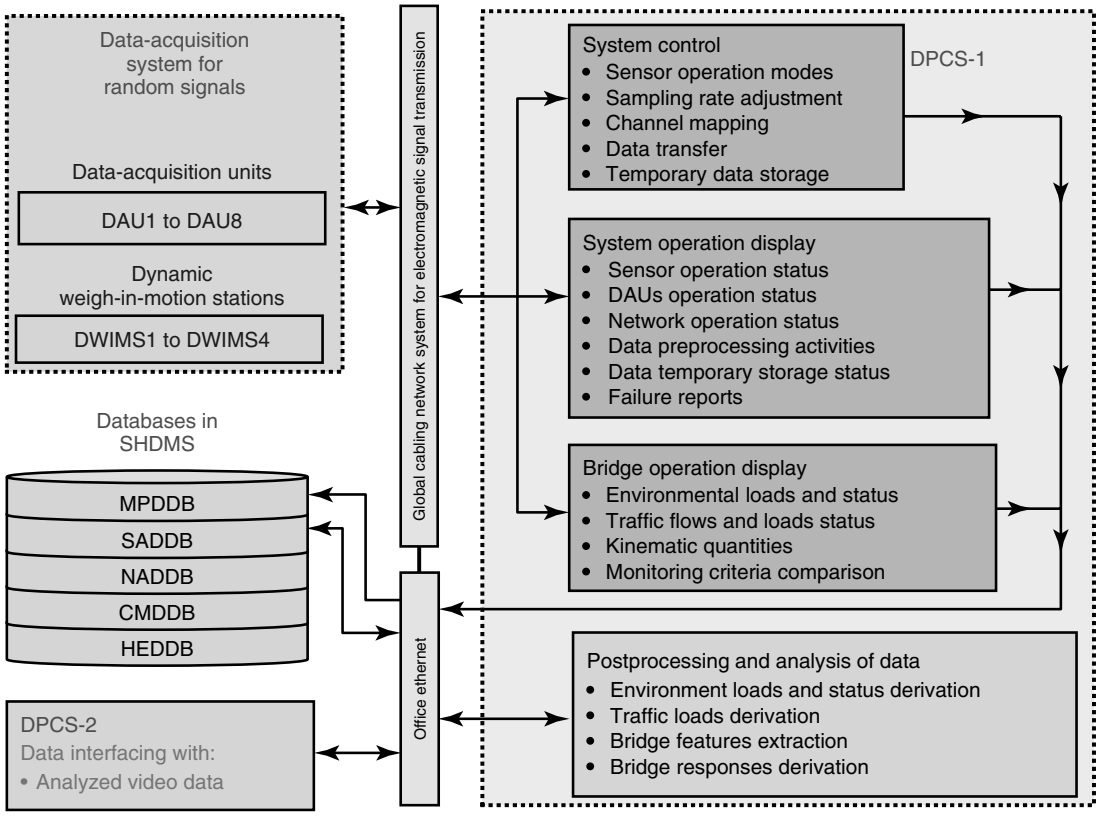


Figure 9. Functional block diagram of DPCS-1—module 3 for random signals.

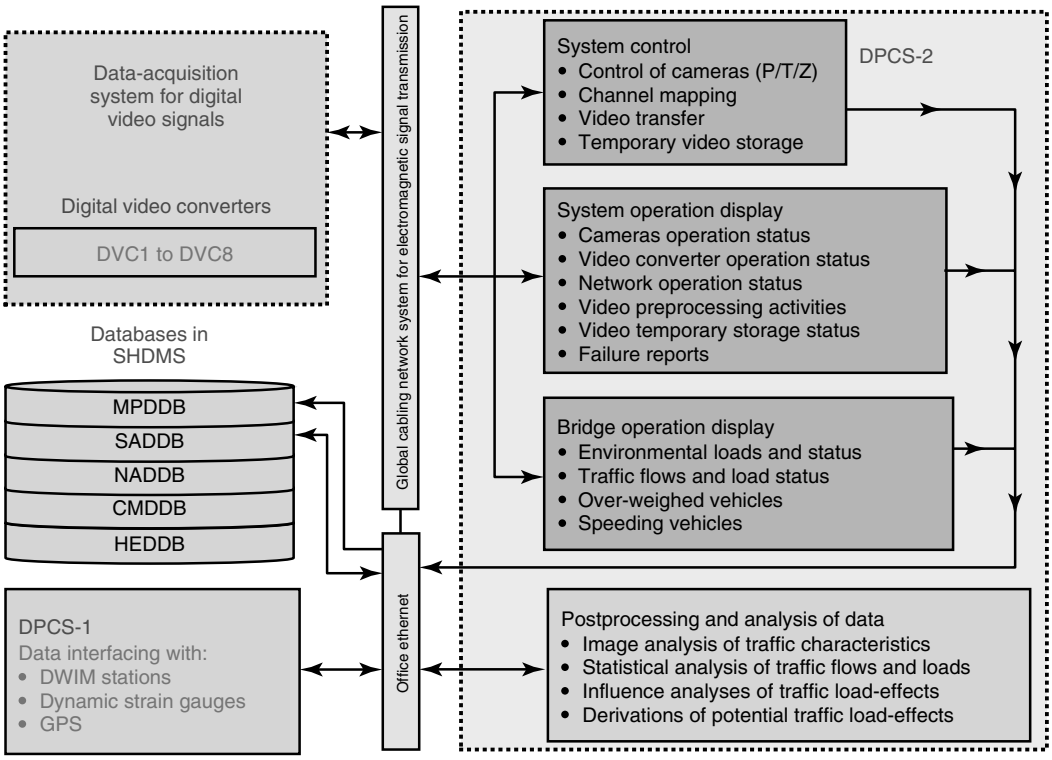


Figure 10. Functional block diagram of DPCS-2—module 3 for video signals.

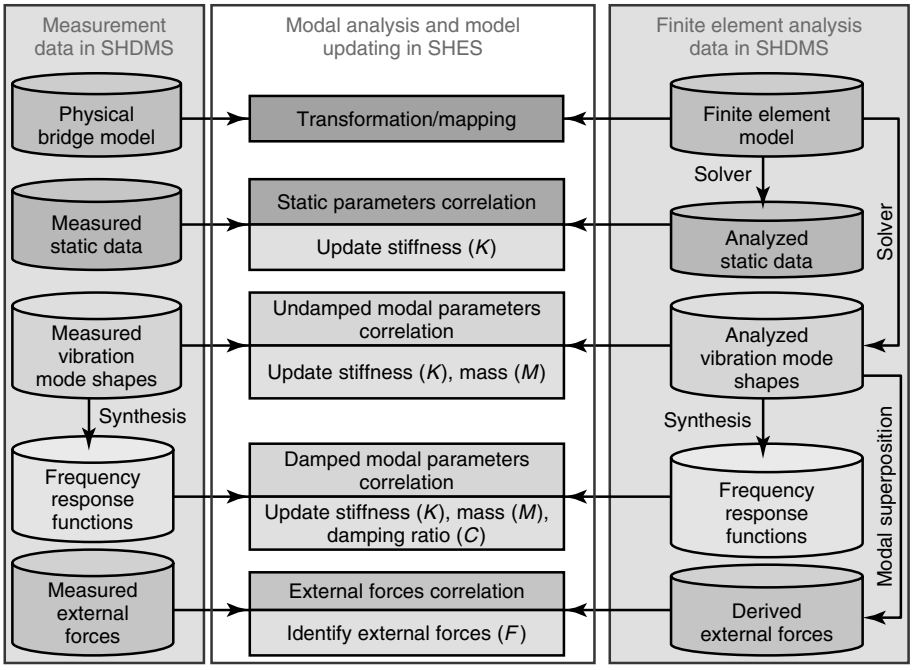


Figure 11. Modal analysis and model updating in SHES by interfacing with SHDMS.

shows the typical modal analysis and model updating works in SHES by interfacing with SHDMS.

- **Structural diagnosis modeling analyses**

For any key structural component or location with a measured and/or derived stress and/or displacement that exceeds 60–75% of its predefined monitoring criteria, structural diagnosis or assessment of the current state of the bridge structure and its history will be carried out in order to assess any adverse effects on the global structural system based on the current and historical environmental, and operational loading conditions. For the sake of facilitating such assessments, it is required to perform automatic execution of finite element analyses on those prebuilt models and under predefined typical types of structural analysis such as geometric nonlinear analysis of traffic and temperature loads, random response analysis of dynamic wind seismic loads, etc.

- **Structural prognosis modeling analyses**

The aim of structural prognosis is to assess the ability of the bridge structural system to carry out future loading conditions (derived on the basis of past and current measured loads) or extreme events. The applications of structural prognosis in WASHMS are (i) to predict/assess the remaining fatigue life of structural steel components based on a combination of finite element analyzed results, BS5400: Part 10: fatigue assessment rules, and measured strain results from WASHMS [12]; (ii) to investigate the different types of potential failure modes and the associated predictable and unpredictable loading conditions that cause damage and subsequent failure; (iii) to determine the potential consequences of each failure event or multiple events acting simultaneously; and (iv) to facilitate the planning of scheduled inspection and maintenance activities.

- **Visualization of analyzed results**

In order to increase the efficiency and accuracy in identification and quantification of abnormal features or defects, all the analyzed, measured, and derived results are presented in comparative plots (with animation, where necessary) and tabulated in the matrix form.

The execution of the above operational functions requires the development of a customized finite element interfacing software system (FEISS) to manipulate the operation of different software tools (including both finite element analysis tools and random data processing and analysis tools) under different hardware and software operating platforms. The FEISS is composed of five modules, i.e., modules A, B, C, D, and E, as shown in Figure 12. Module A includes the prebuilt finite element models of (i) full 3D global bridge model, (ii) full 3D local foundation bridge model, (iii) 3D global spine-beam or gird-beam bridge model, and (iv) full 3D local segmental bridge models. Module B includes the preconfigured finite element analysis types such as normal mode analysis, linear static analysis (influence coefficients' determination), nonlinear static analysis, random response analysis (buffeting and seismic responses determination), fatigue analysis, and impacting analysis. Module C includes the finite element and statistics solvers such as MSC-NASTRAN, ANSYS-Vertical Physics, MATLAB data analysis suite, etc. Module D includes post-posting and display software for generation of analyzed results and reports. Module E includes the interfacing and control software for the execution of (i) automatic/manual retrieval of measured/analyzed data from the relevant database in module 5 to module 4 for processing/analysis; (ii) automatic/manual inputting of the retrieved data into relevant prebuilt finite element or statistical model/models for predefined types of finite element or statistics analysis; and (iii) automatic/manual display and storage of the analyzed results.

2.5 Module 5—structural health data management system (SHDMS)

The SHDMS is composed of a high-performance server equipped with data management software, and is the interfacing platform for the interoperability of data and information so that the efficiency of fusion of data and information for decision making can be significantly enhanced.

The following five major databases are devised to be executed in the SHDMS for interfacing works:

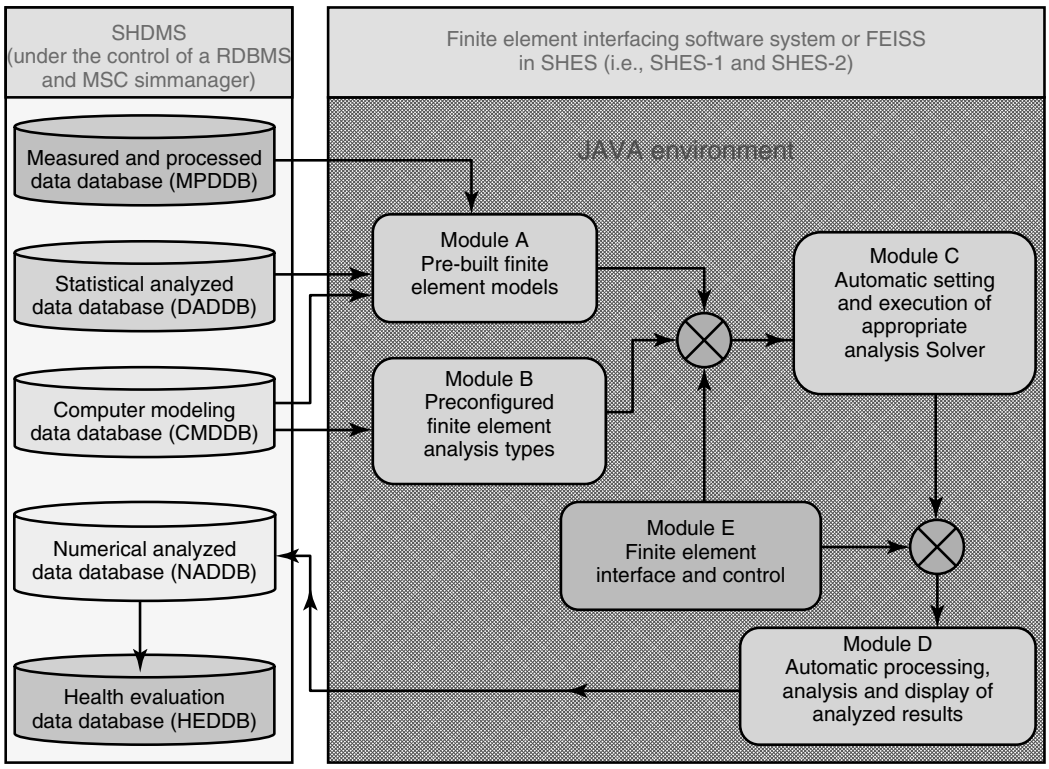


Figure 12. Layout of FEISS in SHES and its interfacing with the databases in SHDMS.

- **Measured and processed data database (MPDDB)**
All time-series data obtained from measuring sensors.

- **Statistical analyzed data database (SADDB)**
All data generated from signal/data processing and analysis software tools such as MATLAB—data analysis suite, NI—data processing and reporting, SD tools, etc.

- **Numerical analyzed data database (NADDB)**
All finite element analyzed/output data generated from finite element analysis software tools such as MSC-NASTRAN, ANSYS-Vertical Physics, MIDAS, LUSAS, etc.

- **Computational modeling data database (CMDDB)**
All finite element modeling/input data generated from finite element analysis tools such as MSC-PATRAN, ANSYS-Preprocessor, ANSYS-Workbench, etc.

- **Health evaluation data database (HEDDB)**
All updated structural health monitoring and evaluation criteria and concise monitoring and evaluation results of environmental loads and status, traffic loads, bridge features, and bridge responses.

These five databases are manipulated and managed by a data warehouse system, which is customized basing on "IBM DB2 UDB Warehouse Enterprise", and is equipped with data management and data analysis tools for integrating enterprise-wide corporate data into a single repository from which users or engineers can easily run queries, perform analysis, and produce reports. The data warehousing system in SHDMS is devised to carry out the following functions:

1. Systematic cleansing, reconciliation, derivation, matching, standardization, transformation, and conformity of data and information from all data source systems such as DPCS servers and SHES servers as shown in Figure 13.

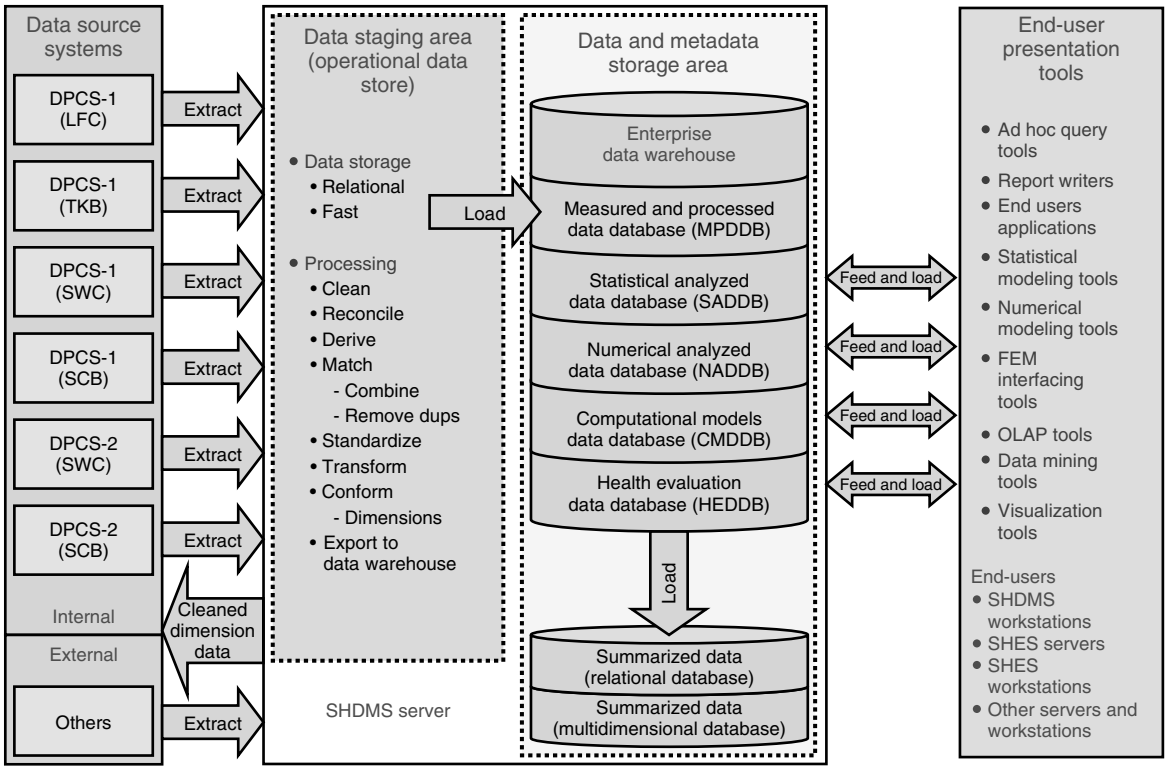


Figure 13. Architectural layout of SHDMS and its interfaces.

2. Manipulation of all types of correlation analyses and features extraction plots, by online analytical processing tools and appropriate data mining tools, based on all data and information generated from the software tools in DPCS-servers and End-users' servers and workstations as shown in Figure 13.
3. Creation of data marts (summarized data in Figure 13), based on the results of aforementioned correlation analyses and plots, for the execution of the following monitoring and evaluation works:
   (a) reporting the current and future loading conditions (such as wind, temperature, seismic, and traffic) acting on the bridge;
   (b) reporting the current and future corrosion status on specified bridge components;
   (c) reporting the current and future structural health conditions of the bridge in terms of the physical parameters as listed in the fourth column of Table 1;
   (d) planning of scheduled bridge inspection and maintenance activities with bridge maintenance team; and
   (e) updating/calibrating the bridge rating system and computational (numerical and statistical) models for processing and analysis of data and information, where necessary.
4. Forming the center of data interrogation and metamodeling for bridge health diagnosis and prognosis through the integration of data and information from both measurement and computational systems. (Metamodels refer to the functional forms of statistical-based models, finite element models, neural networks, etc.)

2.6 Module 6—inspection and maintenance system (IMS)

The IMS is composed of two notebook computers (IMS-1 and IMS-2) and a tool-box (IMS-3). Its

function is devised to carry out inspection and maintenance works on SSs, DAUs, display facilities, LCNSs, and GCNSs. All information (drawings and records) regarding system design, system installation, system operation, and system maintenance is stored and operated in IMS-1 and IMS-2. The IMS-3 is a tool-box for carrying out inspection and minor remedial works.

3 OPERATION OF WASHMS

The system operation block diagram of WASHMS is shown in Figure 14, where the monitoring of kinematic quantities refers to the monitoring of bridge features and bridge responses. The figure shows two levels of monitoring, namely, the sensor-based comparison of the measured results and the monitoring criteria, and the model-based comparison of the derived results and the evaluation criteria. The former refers to the comparison of the measured results with the predetermined monitoring criteria (i.e., at about 60−75% of the design values at serviceability limit state), whereas the latter refers to the comparison of the derived results with the evaluated criteria (i.e., at 100% of the design values at serviceability limit state). The criteria for monitoring and evaluation are defined and calibrated in accordance with the updated requirements of the damage types as defined in the criticality and vulnerability ratings, i.e., structural damage (due to structural actions), environmental damage (corrosion), accidental damage, and wearing damage [13]. In Figure 14, it is shown that if the measured results exceed the monitoring criteria, structural diagnosis and prognosis works will be carried out. Both the structural diagnosis models and structural prognosis models are finite-element-based and/or empirical/statistical-based models, of which the former is used to assess the current and/or historical state of the global bridge structural system

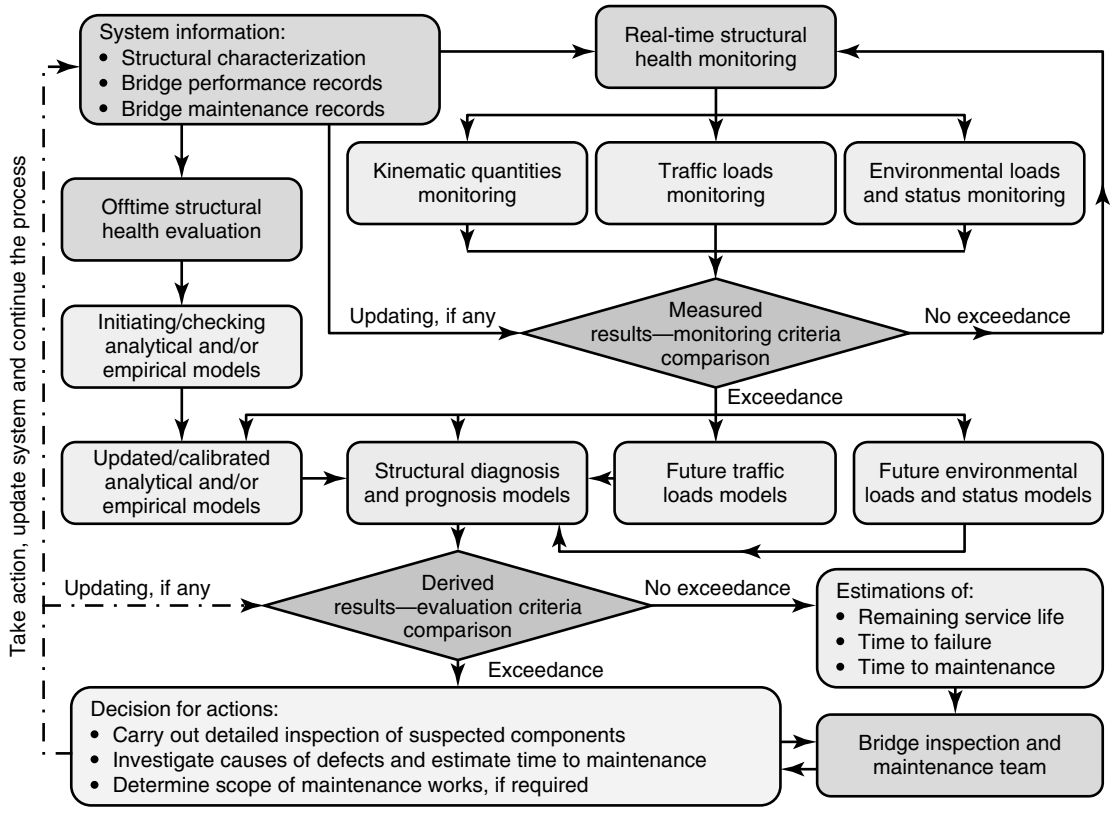


Figure 14. System operation block diagram of WASHMS.

and/or local structural components whereas the latter is used to predict the consequences under the assumed future operational and environmental loads and status.

4 CONCLUSIONS

This article has described a modular architecture that has been used in the design of SHMS for cable-supported bridges in Hong Kong. The architecture of the SHMS described consists of six integrated modules, namely, SS, DATS, DPCS, SHES, SHDMS, and IMS. Each module is well defined and encapsulated. In order to ensure the reliability of measured results, different types of sensors should be deployed at the same key locations and/or components where large displacements and stresses are expected to occur so that the measured results can be validated through correlation among themselves.

Since the hardware configuration of the DAS has a significant influence on data quality, the performance requirements on the linearity, temperature drift, accuracy, direct current resolution, bandwidth, etc., of the signal conditioning and the data-acquisition device should be identified and quantified. Appropriate customized software systems should also be developed and configured to process or derive the measured data in the data formats applicable to bridge health monitoring and evaluation.

Structural health monitoring and evaluation works should be executed through the correlation analyses and features extractions of (i) measured and analyzed results, (ii) current and previous measured results, (iii) previous and updated analyzed results, (iv) and derived results (from analysis and/or measurements) and assigned bridge performance criteria. As the correlation analyses and features extractions involve the synchronized processing of two or more data files/sets, the use of data warehouse system equipped with online analytical processing tools and appropriate data mining tools will facilitate the automatic execution of such synchronized data processing and analysis works.

It is concluded that the WASHMS for cable-supported bridges should at least be able to monitor the loading and structural parameters set by the bridge designer so that the bridge performance under current and future loading conditions can be evaluated, and such evaluated results should be able to facilitate the planning of bridge inspection activities, and be able to determine not only the cause of the damage but also the extent of remedial work, once the damage is identified.

REFERENCES

[1] Highways Department, *Wind and Structural Health Monitoring System, Particular Specification for the Electrical and Mechanical Services in Lantau Fixed Crossing*, Highway Contract No. HY/93/09, 1993.

[2] Highways Department, *Wind and Structural Health Monitoring System, Particular Specification for the Construction of Ting Kau Bridge and Approach Viaduct*, Highway Contract No. HY/93/38, 1993.

[3] Highways Department, *Wind and Structural Health Monitoring System for Lantau Fixed Crossing and Ting Kau Bridge*, Consultancy Agreement No. CE 75/94, 1997.

[4] Wong KY, The wind and structural health monitoring system (WASHMS) for cable-supported bridges in Tsing Ma Control Area. an invited paper, *Proceedings of the IFAC Conference on New Technology for Computer Control*. Hong Kong, 2001.

[5] Highways Department, *Wind and Structural Health Monitoring System, Appendix W of the Particular Specification for the Construction of Hong Kong—Shenzhen Western Corridor*, Highway Contract No. HY/2002/21, 2002.

[6] Highways Department, *Wind and Structural Health Monitoring System, Section 33 of the Particular Specification for the Construction of Stonecutters Bridge*, Highway Contract No. HY/2002/26, 2002.

[7] Dong X, Zhang Y, Xu H, Ni YQ. Research and design of structural health monitoring system for the Sutong Bridge. In *Structural Health Monitoring 2005: Advancements and Challenges for Implementation*, Chang F-K (ed). DEStech Publications: Lancaster, PA, 2005, pp. 1736–1742.

[8] Sun L, Dan D, Sun Z, Health monitoring system for Donghai Bridge in Shanghai, *Proceedings of the Asia-Pacific Workshop on Structural Health Monitoring*. Yokohama, Japan, 2006.

[9] Wong KY. Instrumentation and health monitoring of cable-supported bridges. *Journal of Structural Control and Health Monitoring* 2004 **11**:91–124.

[10] Wong KY, Recent development of structural health monitoring system. a keynote paper, *Proceedings of the International Workshop on Integrated Life-Cycle Management of Infrastructure*, The Hong Kong

University of Science and Technology: Hong Kong, September 2004.

[11] Wong KY. Design of a structural health monitoring system for long-span bridges. *Structure and Infrastructure Engineering* 2007 **3**:169–185.

[12] Wong KY, Stress and traffic loads monitoring of Tsing Ma Bridge. *China Bridge Congress 2007*, 28–30 March. Chongqing, 2007, organized by Merisis.

[13] Wong KY, Criticality and vulnerability analysis of Tsing Ma Bridge, *Proceedings of the International Bridge Conference on Bridge Engineering*. The Hong Kong Institution of Engineers: Hong Kong, November 2006.

Chapter 124
Monitoring of Bridges in Korea

Hyun-Moo Koh[1], Hae-Sung Lee[1], Sungkon Kim[2] and Jinkyo F. Choo[1]

[1] *Department of Civil and Environmental Engineering, Seoul National University, Seoul, Korea*
[2] *Department of Structural Engineering, Seoul National University of Technology, Seoul, Korea*

1 INTRODUCTION

1.1 Bridge construction activities in Korea

Civil engineering projects constituted the backbone of the development and economic growth of Korea. In the domain of transportation infrastructures, particularly bridge structures, construction activities have been restlessly undertaken in the peninsula. The volume of road bridges in Korea increased rapidly during the three decades following the industrialization period of 1970s, when massive investments were done to strengthen the transportation network. The 9332 bridges covering 268 km in 1970 augmented to 22 937 bridges in 2006 with a total length of 1987 km. Figure 1 illustrates the evolution of the bridge stock and total length in Korea from 1970 to 2006 [1].

During this relatively short period, bridge engineering in Korea has achieved outstanding advances that resulted in the construction of numerous cable-stayed and suspension bridges since 2000. Youngjong Bridge, the first three-dimensional self-anchored suspension bridge in the world, is representative of the remarkable progresses made in Korea.

Since Korea is a peninsula surrounded on three sides by the sea and includes 70% of mountainous area, bridges are the key components for the development of its transportation network. Recently, bridge construction activities have been revitalized by the ambitious plan of the government to link some of the 3000 islands of the peninsula to the mainland (Figure 2) in order to promote balanced development of the national territory and renew urban environment

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

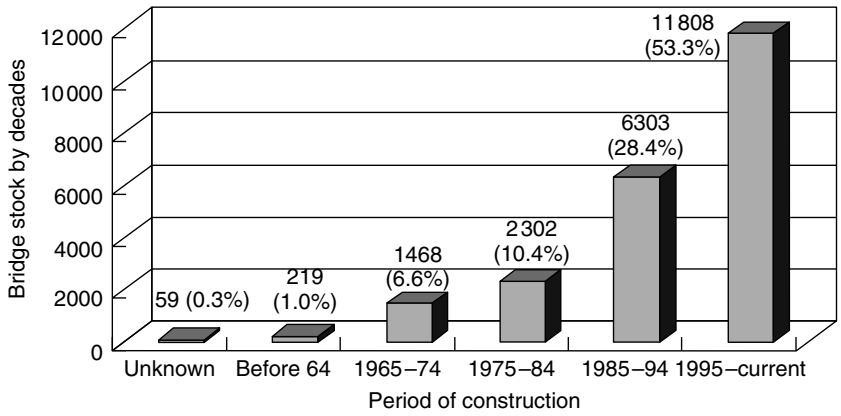


Figure 1. Bridge stock by decades in Korea.

[2, 3]. For example, the province of Jeollanamdo is undertaking construction projects of 102 sea-crossing bridges of which 32 have already been completed, 21 are under construction, and 48 are under planning.

Accordingly, the construction of bridges in the coastal areas is involved in the national amendment promulgated in August 2001. This amendment assigns the extension and construction of national roads no. 2, 24, and 77 with more than 65 km only in the maritime sections including about 50 major bridges (Figure 2) [2, 3].

Among these bridges, Incheon Bridge illustrated in Figure 3 is an offshore circular expressway with a total length of 12 343 km that will link Incheon International Airport to the new city of Songdo in the southern region of Incheon South Port. Its construction (Figure 4) began in October 2004 and the bridge will be completed in October 2009. The cable-stayed bridge $(80 + 260 + 800 + 260 + 80\,\text{m})$ has been designed to secure a navigation clearance of 715 m and overhead clearance of 74 m for the passage of shipping, and will rank at the fifth position among the longest cable-stayed bridges in the world.

Figure 5 presents a rendering (view) of the future 1545 Bridge giving access to Yeosu National Industrial Complex from Gwangyang Port through Myo Island in the Province of Jeollanamdo. The suspension bridge has been designed to allow crossing of 18 000 TEU container ships. The bridge with its main span of 1545 m will be the third longest bridge in the world after Akashi and Great Belt East bridges.

1.2 Initiation of bridge monitoring in Korea

Issues related to the lifetime and durability aspects of bridge structures are of critical importance. A lifetime of at least 100 years is now targeted while designing a bridge. Accordingly, active development and applications of structural health monitoring (SHM) systems for major bridge structures are continuously implemented as a tool to sustain such lifetime perspective in terms of safety, durability, and performance. Modern and integrated monitoring systems are actually introduced in newly built bridge structures since the design stage. Automatic measurement of instrumented civil engineering structures is now widely applied for behavior monitoring during construction in field as well as long-term monitoring for lifetime assessment of bridge structures. Efforts tending toward the increase and upgrade of the monitoring efficiency and performance through sensor-based bridge monitoring systems (SBBMSs) are also continuously undertaken to provide advanced innovative functions.

Korea implemented active development and application of SHM for bridge structures since the early 1990s. This activation was initiated primarily by the tremendously increasing number of deteriorated infrastructure systems, mostly built during the industrialization boom of 1970s, which often resulted in tremendous lack of quality. In addition, the successive collapses of Haengju Bridge during its construction and Sungsu Bridge in 1994, only 15 years after its completion, opened the eyes

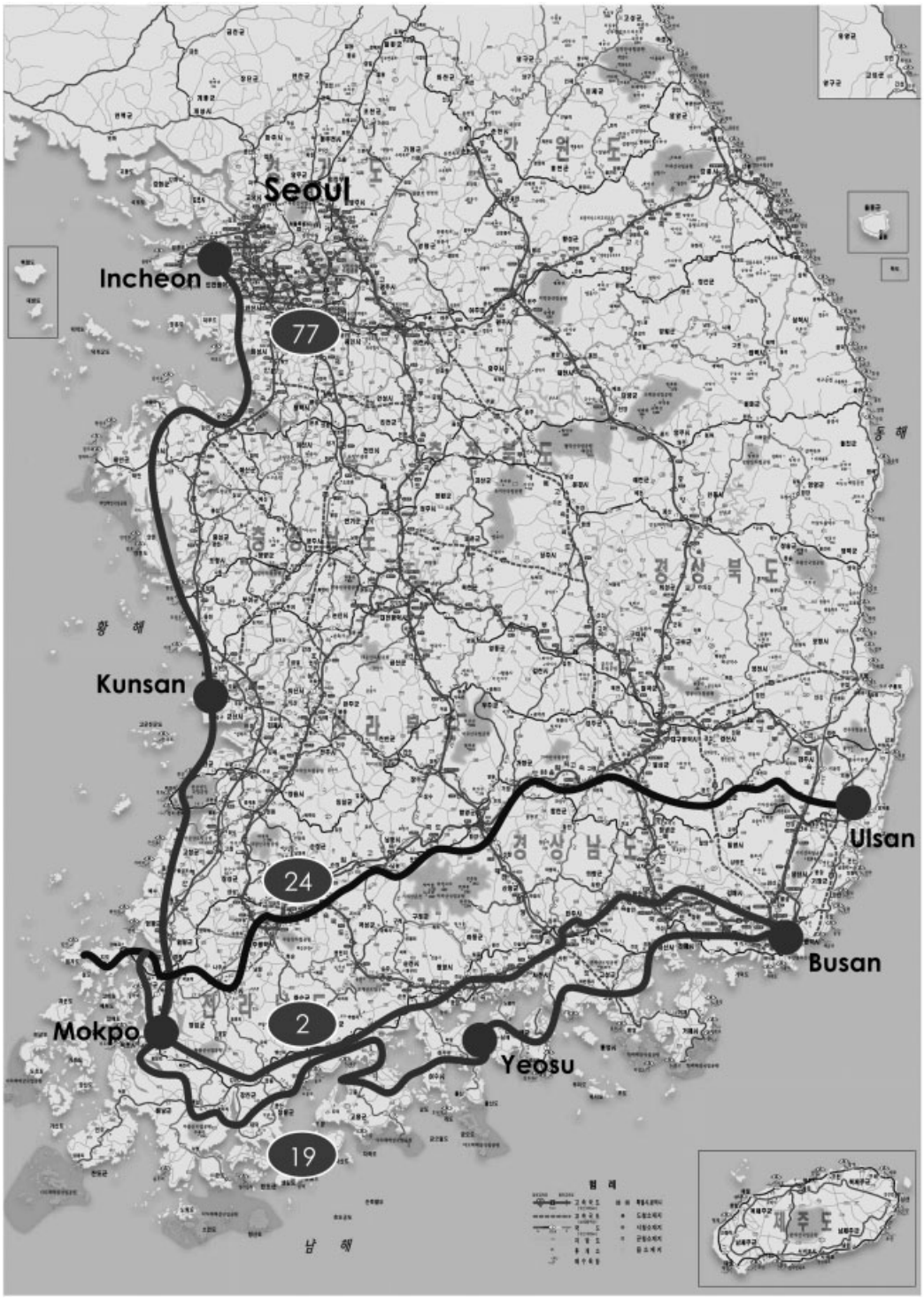


Figure 2. Extension of the national roads in the southwestern coast of Korea.

Figure 3. Rendering of the Incheon Bridge.

Figure 4. View of the Incheon Bridge under construction.

of the public to the importance of the management of bridges together with the increasing recognition of the potential devastating disruption of infrastructure

Figure 5. Rendering of 1545 Bridge.

systems due to manmade and natural hazards, i.e., typhoons and earthquakes. Consequently, attention has been first focused on bridge structures as the most expensive and vulnerable asset of the transportation network directly affecting the economy and public safety.

These accidents, combined with the fact that a large number of bridges before and during 1970s still remains in operation today, led the governmental authorities to issue more stringent requirements on bridge management and operational programs, including systematic visual inspection, instrumentation, load capacity tests and field measurements for design and construction verification, and long-term performance monitoring and assessment.

Figure 6 shows the evolution of maintenance investment together with the health level of the

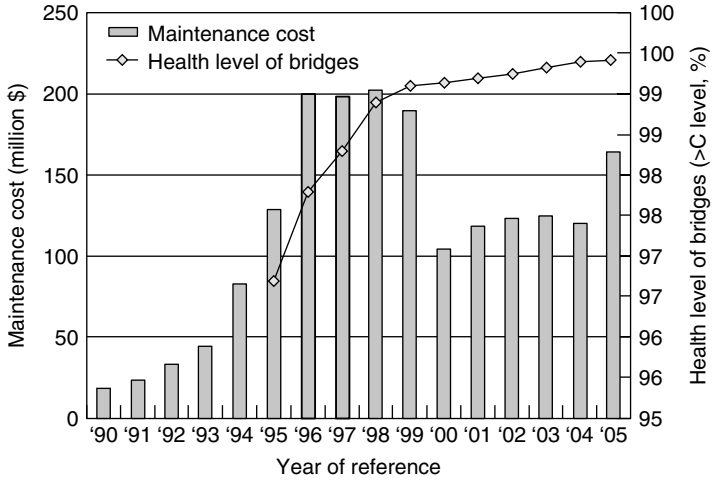


Figure 6. Records of maintenance and health level of the Korean bridge stock. [Reproduced from Ref. 5. © Korea Institute of Construction Technology, 2002.]

Korean bridge stock since 1995. "C level" stands for operational health level under normal conditions [4, 5]. Until the mid-1990s, attention was paid on the construction of bridges rather than on their maintenance. After the collapses of the Haengju and Sungsu Bridges, stress was given on the importance of bridge maintenance. Accordingly, systematic maintenance system was adopted, which made it possible to rationally invest maintenance costs within the budget limits and prevent large catastrophes. As a result, maintenance costs monotonically decreased since 1995 together with remarkable improvement of the health of the bridge stock even in a short period of time.

1.3 Evolution of bridge monitoring in Korea

In order to deploy a successful monitoring system, the requirements are proper instrumentation, reliable signal processing, and knowledgeable information processing. Along with the rapid and massive progresses made in the domain of IT for infrastructures, complete and integrated monitoring systems were systematically installed in all major bridges in Korea (Figure 7). Accordingly, the evolution of bridge monitoring system in Korea can be classified into three generations in terms of developing stage and functionality, which places Korea as one of the leading countries in bridge monitoring today.

The system corresponding to the first generation is characterized by a stand-alone field system consisting of sensors, field hardware, and online transmission to a computer on field. In the second generation, this stand-alone system evolved into an overall bridge management integrated system involving two kinds of integration: operational integration where multiple stand-alone systems operate together and functional integration where bridge monitoring system operates with different systems such as bridge management system (BMS) or vehicle monitoring system [6]. The third generation, also called *future system*, will provide advanced innovative functions like sensor fusion, reliable massive signal transmission, automated surveillance, adaptive signal processing, etc. Many research efforts are now led to develop the third generation system and enhance the performance of the current system by introducing new sensing techniques, power generation from bridge vibration, web-based operating system, or wireless signal transmission (*see* **Wireless Sensor Network Platforms**; **Web-based SHM**; **Microelectromechanical Systems (MEMS)**).

2 KOREAN BRIDGE MANAGEMENT SYSTEM (KOBMS)

Total BMS attempts to include inspection, evaluation, estimation, and rehabilitation of bridges in a systematized organization, which integrates SHM systems installed in bridges. BMS is an information-oriented system, which aims at the global supervision of all the information gathered in every bridge so as to help the supervisor in deciding current and future requirements for optimal management and rehabilitation of bridges (*see* **Maintenance Principles for Civil Structures**). Following this, to perform scientific and rational management and rehabilitation of bridges, the Korean Ministry of Construction and Transportation (MOCT) together with the Korean Institute of Construction Technology (KICT) developed the Korean bridge management system (KOBMS) in 1995, which has been in operation for ordinary road bridges since then.

The hardware of KOBMS is constituted by regional networks and high-performance and high-capacity computers to efficiently manage the huge volume of data gathered in bridges. Such network operates interactively by linking the KICT, road facilities of the MOCT, five regional offices, one R&D office, and 18 road management offices since 1996.

The software of the KOBMS (Figure 8) is composed by a database (DB) recording archives related to bridges, a program computing the investment priorities, a rehabilitation and retrofit techniques DB, a tunnel DB, a program outputting the current state of bridges, and a decision-making system performing the essential functions of the BMS. Especially, the BMS DB stores and compiles about 230 items per bridge including its characteristics, structure, inspection records, load-carrying capacity, etc. Investment

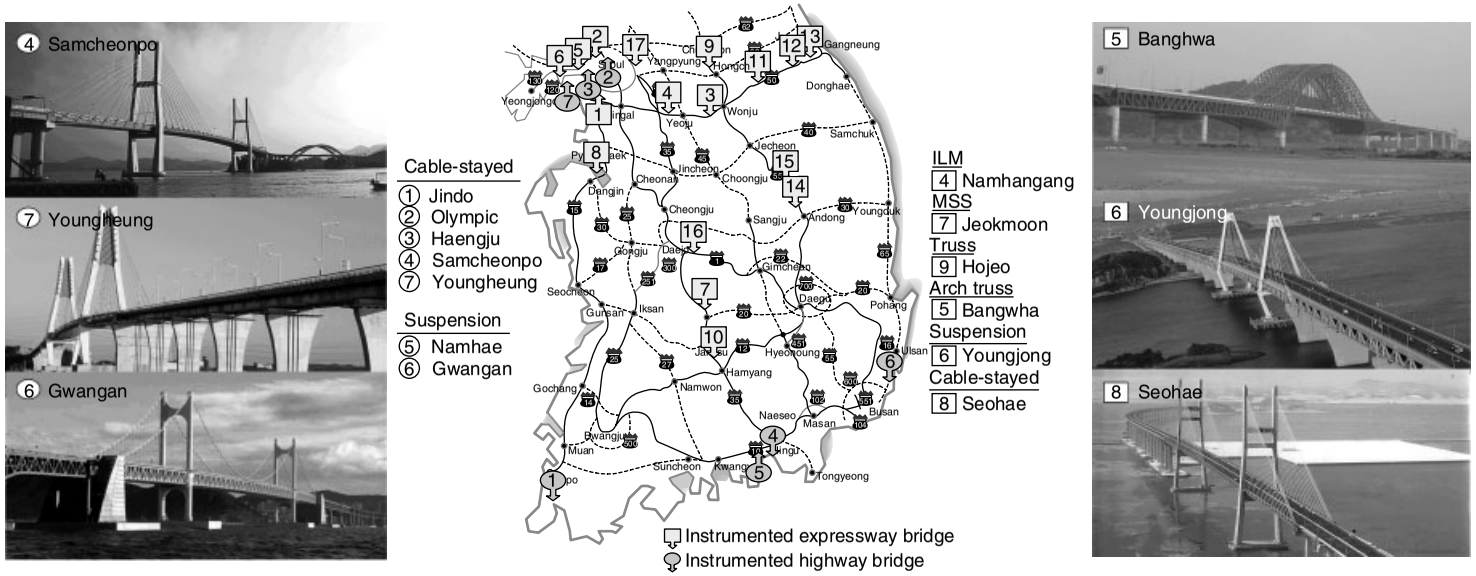


Figure 7. Major instrumented bridges of the national expressway and national highway networks of Korea.

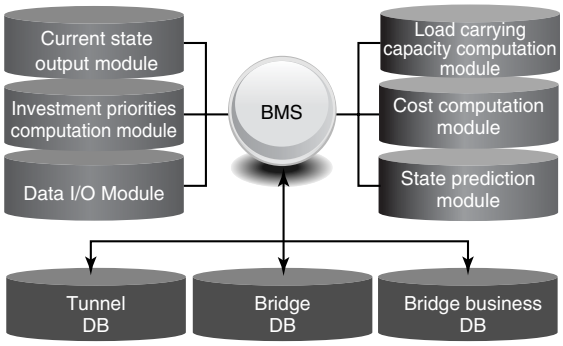


Figure 8. Software of the KOBMS.

priorities, as a basic function of the BMS, stand for the process by which one decides the sequence of bridges to be considered for management and rehabilitation in the budget planning. The KOBMS DB arranges 64 basic rehabilitation and retrofit techniques for bridges and defines 120 types of damage that may occur in bridges, so as to automatically select an appropriate rehabilitation and retrofit basic technique for each type of damage. Using the KOBMS DB, a report on the current state of the bridges is published annually, and it provides statistics of the basic parameters of the bridges and their current state according to their classification [7].

The KOBMS made it possible to efficiently manage the huge volume of bridge data in a systematic and effective organization. Its exploitation currently helps in systematizing state evaluation and records management of rehabilitation and retrofit for each bridge as well as for the whole bridge stock. The state of bridges has seen significant improvement while large reduction of the budget invested for their management, being rationally shared, has been obtained [8].

On the other hand, even if the KOBMS selects optimal maintenance measures for the whole bridge stock, it does not necessarily mean optimal measures for individual bridges. Therefore, developing a project level BMS that reflects the characteristics and conditions of each individual bridge networked with the DB collected from the SHM system of the bridge is currently under discussion for implementation. The introduction of an expert system considering the severity and range of eventual damage or anomaly occurring in the bridges is also under plan.

3 FIRST GENERATION: SHM SYSTEMS FOR EXISTING BRIDGES

The first applications of health monitoring systems in Korea began for existing bridges in order to collect field data by full load scale capacity tests for design verification and, subsequently, evaluate the health of the structures. Immediately after the collapse of the Sungsu Bridge in 1994, this first generation of monitoring systems was applied in the existing bridges of which two representative examples are described hereafter.

3.1 Namhae Bridge

Namhae Bridge, which was opened to traffic in 1973, is the first suspension bridge erected in Korea. It is a three-span suspension bridge ($128 + 404 + 128$ m) with main cables made of parallel wires and a stiffening girder consisting of welded steel box with an orthotropic steel deck. The joints of the streamlined box were welded on site. The bridge connects Namhae Island with the mainland in the southern coast of the peninsula.

At the time of its design, the load-carrying capacity prescribed in the codes was 30% smaller than the current one. Moreover, since its completion, the bridge experienced excessive vehicle weights due to the proximity of a large industrial zone. Accordingly, a safety evaluation conducted after 20 years of service in 1993 [9] revealed major defects, such as fatigue cracks at welding and corrosion of steel members. Accordingly, short-term strengthening and rehabilitation were carried out, and a long-term maintenance and health monitoring project were scheduled to extend the service life of the bridge and prevent further loss of load-carrying capacity. The objective was to monitor the structural responses of the bridge focusing on the identification of its deteriorating rate over a long period of time (*see* **Damage Measures**; **Principles of Structural Degradation Monitoring**; **Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control**).

According to the project, the initial status of the bridge was evaluated by geometric survey and static loading test [10] prior to logging its long-term

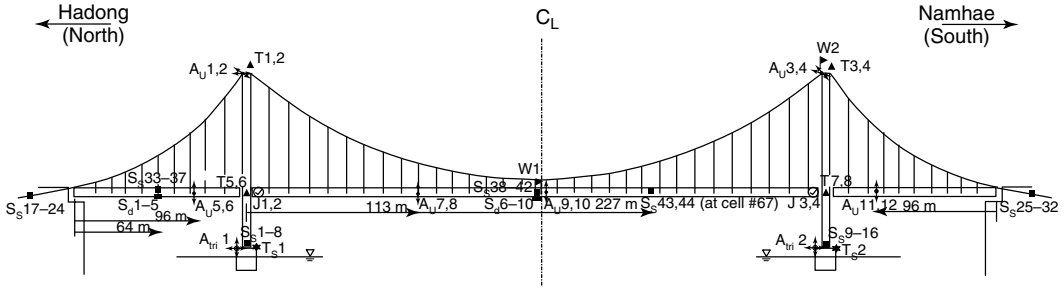


Figure 9. Overall layout of the instrumentation of the Namhae Bridge.

Table 1. Characteristics of sensors installed in the Namhae Bridge

Sensor		Symbol	Quantity	Characteristics
Static	Tiltmeter	▲(T)	10	Electrolytic biaxial (16 channels)
	Tiltmeter	✸(T_s)	2	Submersible biaxial (4 channels)
	Strain gauge	■(S_s)	44	Vibrating wire (44 channels)
Dynamic	Data logger		2	Modem
	Accelerometer	◂▸	12	Uniaxial, force balanced (12 channels)
	Strain gauge	▬(S_d)	10	Electric resistance (10 channels)
	Accelerometer	✥	2	Triaxial, force balanced (6 channels)
	Jointmeter	⊘(J)	4	LVDT (4 channels)
	Anemometer	⚑(W)	2	3D-propeller (4 channel)
	DAQ system		2	GPS system

behavior by health monitoring. The deck geometry obtained from the survey was compared with the intended geometry of the original design. Thereafter, the bridge was instrumented with the health monitoring system illustrated in Figure 9. The monitoring system consists of 110 channels of both static and dynamic sensors such as strain gauges, accelerometers, tiltmeters, jointmeters, and anemometers (Table 1). In 1999, the dynamic measurement system installed in the bridge was used to identify its natural frequencies through vehicle loading test and ambient vibration test (AVT) *see* **Ambient Vibration Monitoring**. The whole acquisition of measured vibration data on the bridge took almost 2 weeks. A few natural frequencies were obtained through simple impact tests by identifying peaks in the power spectral density functions, and the monitoring system was also used to evaluate the effect of thermal variation on the change of the fundamental frequencies during 9 months. Impact test appeared to be adaptable, if only lower bending modes were of interest and traffic-induced vibration was found to be utilizable for this type of bridge. In addition, it was shown that when dynamic properties obtained from measured data were used to assess damage, variation of factors, such as temperature, should be considered [11].

3.2 Jindo Bridge

Jindo Bridge (Figure 10) is a three-span (70 + 344 + 70 m) semi-harp cable-stayed bridge. The bridge, completed in May 1984, has been the only way linking Jindo Island and the mainland at Haenam in the southwestern coast of the peninsula. The bridge has today a twin that is 20 years its junior with

Figure 10. View of the first and second Jindo Bridges.

the completion of a matching cable-stayed structure alongside. At the time of the design of the first bridge, design live load was DB-18 (total weight 34 tonf) corresponding to the second-grade bridge according to the categories of the current design codes. Close investigation in 1996 including load and vibration tests resulted in the closing of the bridge for trucks exceeding the design load.

An AVT was carried out in 1998. Mode shapes as well as natural frequencies were found to give a more precise description of the current bridge state (*see* **Ambient Vibration Monitoring**). During the AVT, the acquisition was carried out without traffic restriction, and the dynamic properties were computed from these ambient vibrations under normal traffic. With six accelerometers, a couple of weeks were needed to acquire all the vibration data [12].

Similarly to Namhae Bridge, though the traffic-induced vibration was used, very smooth and reasonable spectra were obtained. Finally, 26 natural frequencies and corresponding 16 mode shapes were found by AVT. No significant noise was found on the mode shapes, which demonstrated that traffic-induced vibration could be used for this type of bridge, since the mass of traffic is relatively small compared with that of the structure and the traffic could excite the structure with enough energy in the frequency band of interest. Such results are believed to provide a precious reference for future monitoring. Figure 11 draws the conceptual diagram of the SHM system installed in the Jindo Bridge.

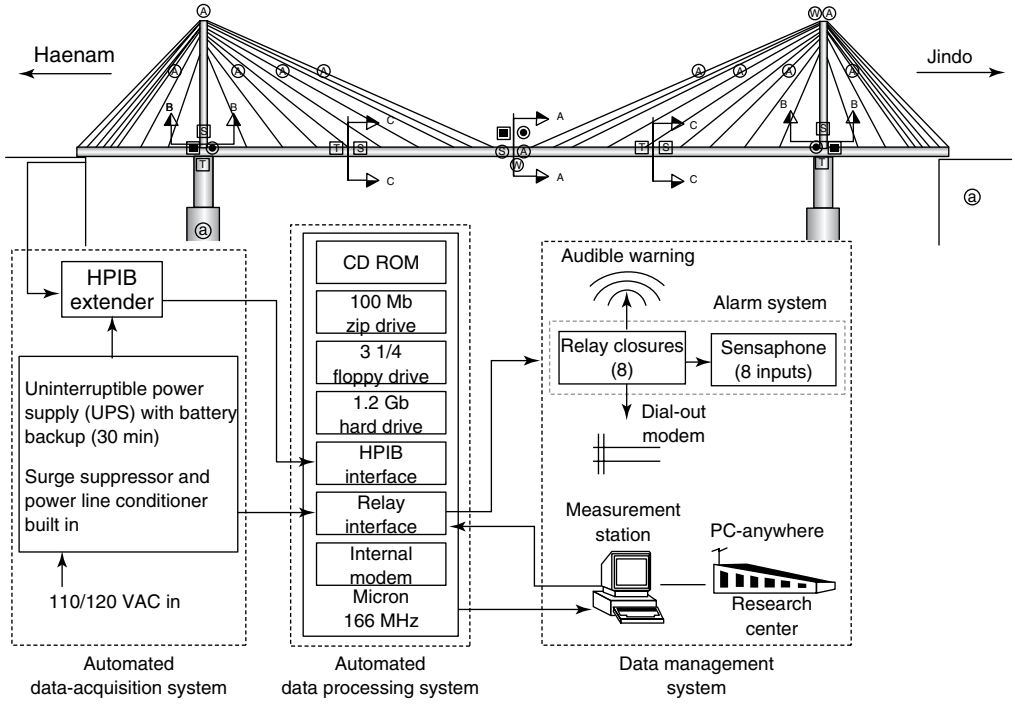


Figure 11. Conceptual diagram of the health monitoring system in the first Jindo Bridge.

4 SECOND GENERATION: INTEGRATED SHM SYSTEMS FOR NEW BRIDGES

Today, every newly built long-span bridge in Korea is equipped with modern monitoring systems. Unlike earlier applications of health monitoring systems installed in existing bridges, where conventional sensors, loggers, and transmission methods were used and individual systems served each bridge independently, this second generation of health monitoring systems exploits modern technologies from sensing to processing, i.e., many sensors and data-acquisition systems that measure the behavior of the bridge during its construction become part of

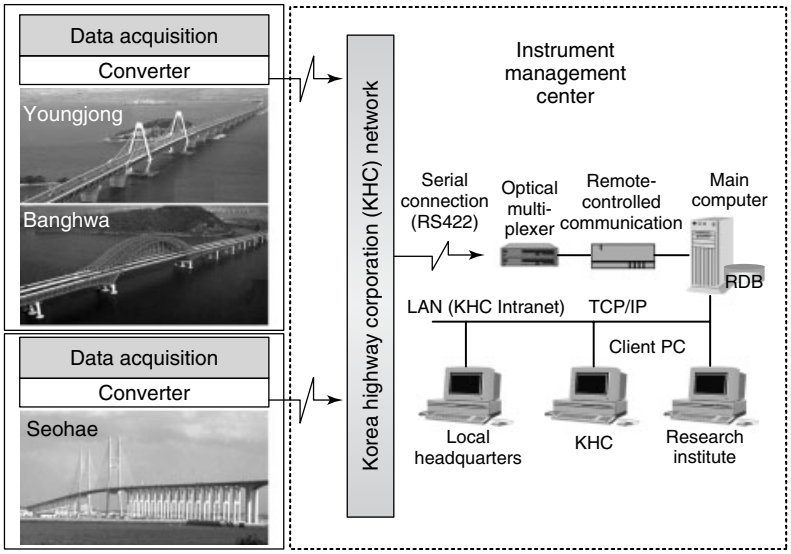


Figure 12. Integrated operating system of the Seohae, Youngjong, and Banghwa Bridges.

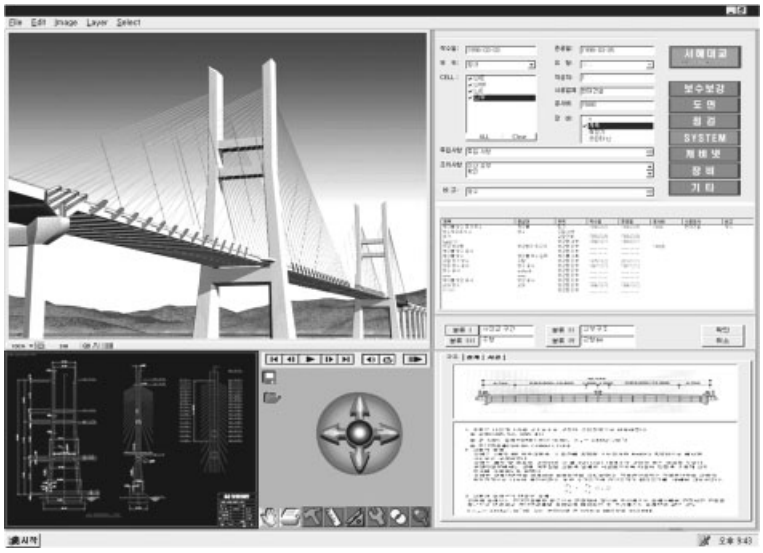


Figure 13. Functional integration of SHM system with BMS (Seohae Bridge).

the long-term health monitoring system, in order to monitor long-term performance and durability of the bridges in the scope of systematic inspection and maintenance programs.

This second generation of health monitoring system has been applied to newly built bridges since the design stage, by integrating the systems of several bridges together to reduce costs and significantly increase the management efficiency. This integrated system includes BMS for systematic decision making and budgeting of inspection, estimation, rehabilitation, and repair (*see* **Usage Management of Civil Structures**). The integrated system for Seohae, Youngjong, and Banghwa Bridges may be cited as the best example of the current monitoring system. The data collected at each bridge are processed exclusively at each field station for real-time monitoring and alarming sudden abnormal behavior. Data that are useful for long-term evaluation of bridge condition, as well as periodical inspection data, can be transmitted through high-speed internet line to the management center located far away from the site (Figure 12). Once data are collected at the center, integrated BMS (Figure 13) handles them to classify, store, and retrieve. This integrated BMS is able to itemize bridge maintenance details and manage status assessment, rating, repair, and strengthening histories.

Figure 14. Aerial view of the Youngjong Bridge.

4.1 Youngjong Bridge

Youngjong Bridge (125 + 300 + 125 m), completed in November 2000, is part of the Incheon International Airport Highway, which connects Seoul and the Incheon International Airport. Being the first bridge that foreign visitors see when arriving in Korea, particular attention has been paid on its design with unique features such as three-dimensionally profiled suspension cables, self-anchoring, and double decks for both automobile and train traffic (Figure 14).

A total of 393 sensors, including static and dynamic strain gauges, and 23 data loggers are distributed over the bridge (Figure 15, Table 2).

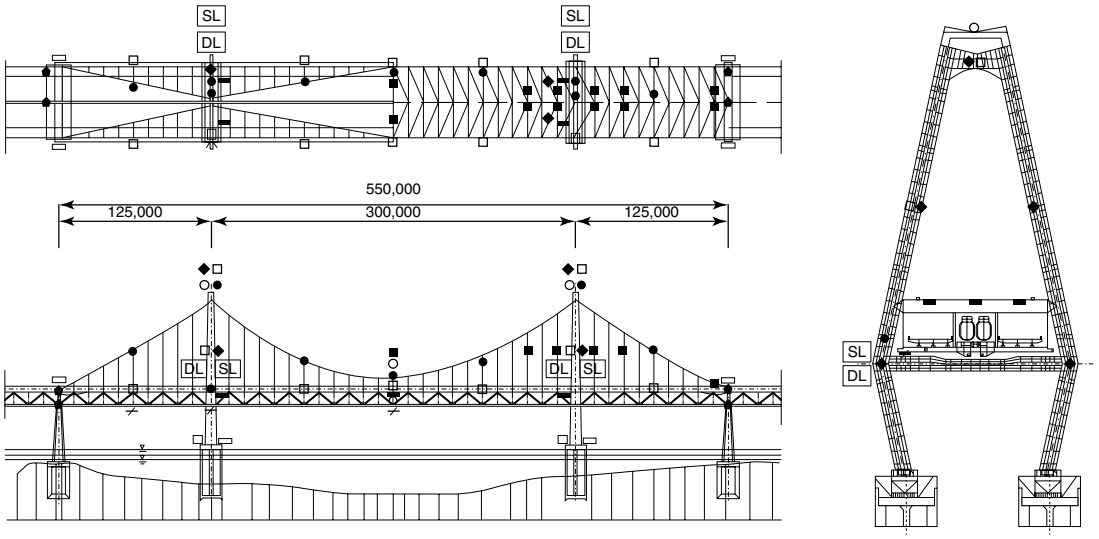


Figure 15. Overall layout of the instrumentation of the Youngjong Bridge.

Table 2. Characteristics of sensors installed in the Youngjong Bridge

Symbol	Sensor	Quantity	Location
●	Thermometer	33	Cable, member, tower
=	Static strain gauge	122	Anchor, deck cross section, link shoe
━	Dynamic strain gauge	175	Deck cross section, etc.
◆	2D tiltmeter	10	Tower inclination
■	1D accelerometer	12	Cable
□	2D accelerometer	14	Tower top, deck
□	3D accelerometer	3	Tower foundation
○	Anemometer	4	
×	Laser displacement sensor	3	
⬟	Potentiometer	4	Expansion joint
SL	Static data logger	2	
DL	Dynamic data logger	2	

Table 3. Dynamic characteristics of the Youngjong Bridge identified by forced vibration test

Dynamic characteristics	Natural frequency		Damping ratio	
	Design	Measured	Design	Measured
Flexure	0.422	0.487	0.03	0.06
Antisymmetric flexure	0.716	0.810	0.03	0.03
Torsion	0.781	1.060	0.02	0.023
Antisymmetric torsion	1.195	1.700	0.02	0.05

The hardware system was designed to collect data remotely, and the software system was developed to process data and display results in a custom-designed format. The monitoring system has been completed in 2001 and a huge volume of signals have been collected up to date. These signals were carefully analyzed for verifying the system performance and for assessing the bridge health [13].

The SHM system was exploited at first to identify the dynamic properties of the bridge before opening, by means of field loading test using two vibrators to generate flexural and torsional vibrations of the bridge. Comparison of the measured data such as natural frequencies, vibration modes, and damping ratios with design values in Table 3 showed good correspondence attesting for the reliability of the bridge [6].

During the system stabilization period, signals showed regular pattern of fluctuation along with the daily and seasonal temperature changes (*see* **Model-based Statistical Signal Processing for Change and Damage Detection**). Some typical signal patterns are described hereafter [13].

Accelerometers were mounted on 12 hangers to evaluate tension forces. Frequencies computed from acceleration responses, measured under ambient vibration during 15 months, were used to evaluate

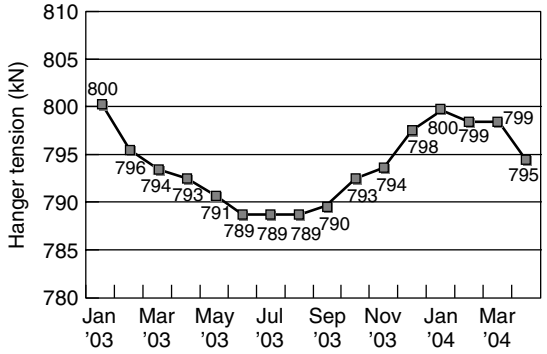


Figure 16. Seasonal variation of the average hanger tension observed during 15 months in the Youngjong Bridge.

hanger tension forces. The resulting average tension force for the whole set of hangers displayed a pattern corresponding to seasonal variation (Figure 16). The same seasonal pattern, according to the variation of temperature, could also be observed for the vertical and lateral displacements at midspan of the stiffening girder measured by means of laser displacement sensor (Figure 17). Joint displacements at both ends were also seen to present quasi-linear relationship with temperature, with a displacement averaging 46 mm for a thermal variation of 10 °C in ambient temperature. Acceleration data measured under ambient vibration were exploited to analyze the dynamic properties of the bridge, and the frequencies of the first and second modes were measured to be 0.494 and 0.831 Hz, which correspond fairly to the results of field vibration tests (0.487 and 0.810 Hz) listed in Table 3.

Long-term responses of the Youngjong Bridge were thus verified to be governed by daily and seasonal variations of the temperature, which revealed that the bridge behavior is as expected.

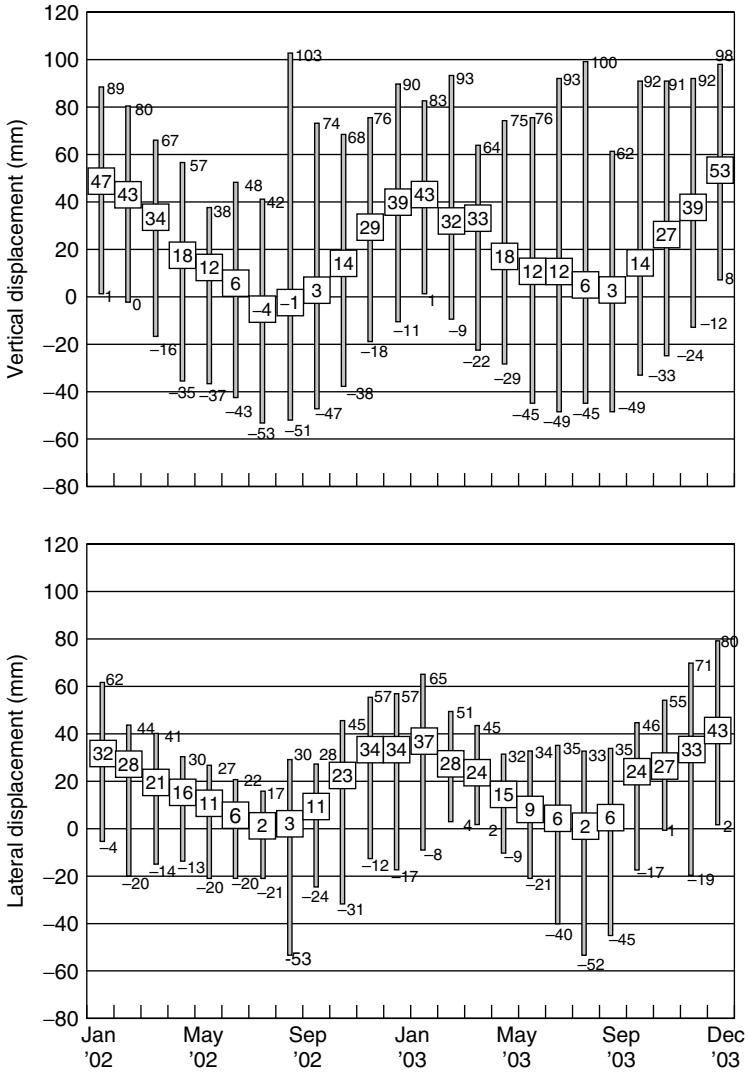


Figure 17. Vertical and lateral displacements at midspan observed during 2 years in the Youngjong Bridge.

4.2 Seohae Bridge

Until the completion of the Incheon Bridge ($80 + 260 + 800 + 260 + 80$ m) in 2009, the Seohae Bridge ($60 + 200 + 470 + 200 + 60$ m, Figure 18) will remain the longest cable-stayed bridge in Korea. Its five spans are constituted by stiffened steel girders with precast slab. Figure 19 schematizes the history of the development of the monitoring system of the bridge beginning from its planning to its operation. More than 180 sensors of 10 types are actually installed in the major parts of the cable-stayed (Figure 20, Table 4) PSM and FCM approach bridges [6, 8].

The structural behavior of the cable-stayed bridge was observed and analyzed during the first 2 years following its completion. Results showed that the annual variation of the vertical deflection in the stiffening girder satisfies the allowable design limit with a range of -320 to 30 mm and that the deflection due to live load presents a range of 189.7 mm, which represents only 25% of 808.8 mm, the design limit. The stress range in the stiffening girder due to live loads showed good correlation with the volume of traffic monitored during 2 years. Stress margin appears to remain considerable since measured stresses represent only 5–12% of the design stress. Accordingly, it seems that the actual highway bridge design specifications in Korea are producing excessively conservative structures.

The thermal deformation of expansion joints at the extremities of the bridge was verified to exhibit correlation of about 96% with theoretical predictions (Figure 21). The tensioning force in the cables ranged from 95 to 104% of the initial value. Seohae Bridge

Figure 18. View of the Seohae Bridge.

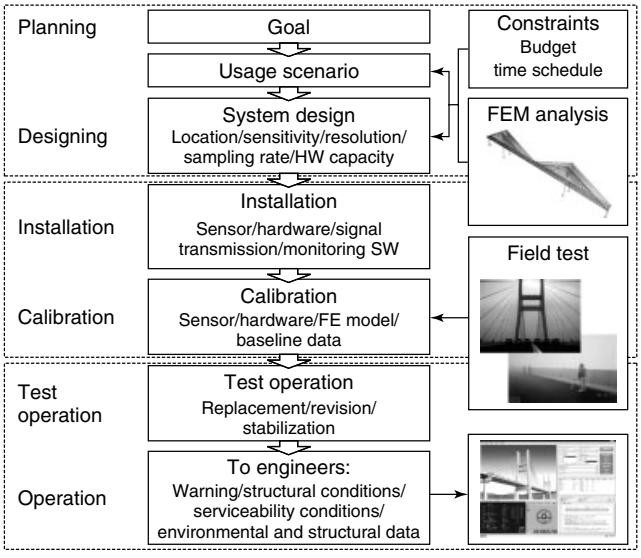


Figure 19. History of the development of the monitoring system of the Seohae Bridge.

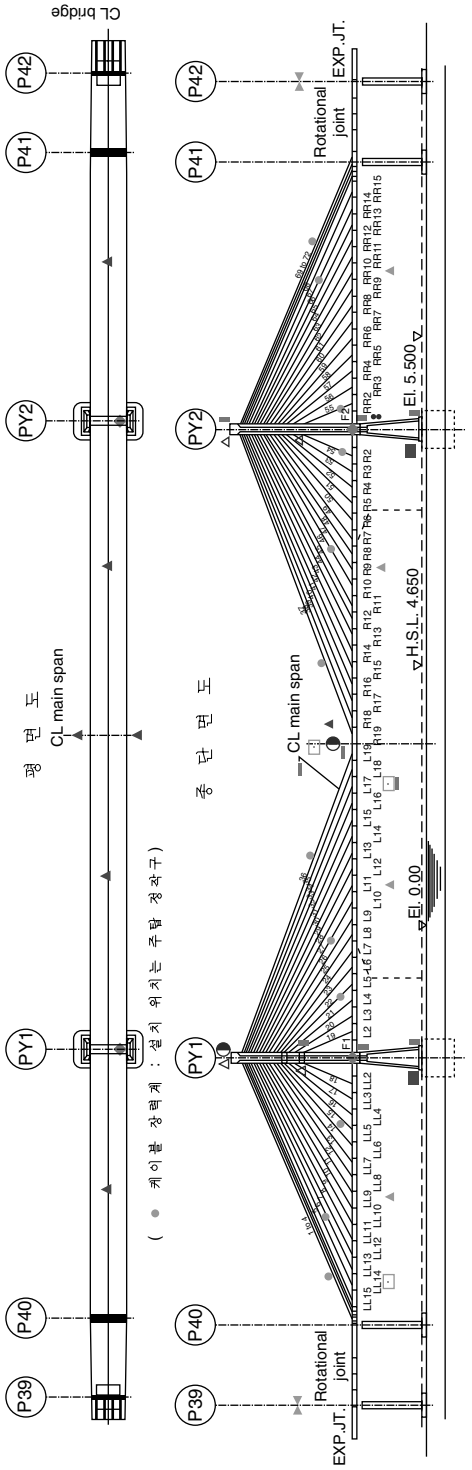


Figure 20. Overall layout of the instrumentation in the Seohae Bridge.

Table 4. Characteristics of sensors installed in the Seohae Bridge

Symbol	Sensor	Quantity
◑	Anemometer	2
●	Cable tension force (acceleration)	12@2
▲	Accelerometer (deck)	6
△	Accelerometer (tower)	4
▮	Tiltmeter	6
▬	Thermometer	14
▣	Static strain gauge	12
▣	Dynamic strain gauge	82
⁘	Laser displacement sensor	4
⋈	Jointmeter	10
◆	Field control box	4
■	Accelerometer for earthquake	2

thus appears to be healthy in view of its long-term behavior [14].

5 THIRD GENERATION: SENSOR-BASED BRIDGE MONITORING SYSTEMS

More recently, efforts tending to increase and upgrade the monitoring efficiency and performance of the second generation, that is SBBMSs, have also been effectively introduced in newly built bridges.

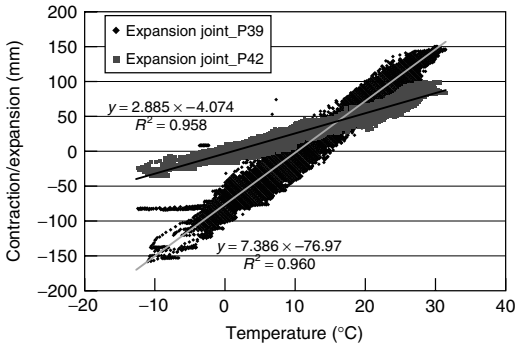


Figure 21. Thermal expansion/contraction at the end expansion joints of the Seohae Bridge.

The purposes of SBBMS are to provide information (i) to assess the behavior of the bridge, (ii) to ensure serviceability and safety during its service life, and (iii) to help design, construction, and maintenance. Application of SBBMS can be found in Gwangan and Samcheonpo Bridges.

The hardware system performs measurement and data acquisition of the bridge behavior by remote sensing, using sensors and data loggers, and the software system achieves data processing, storage, analysis, and display in customized form (Figure 22).

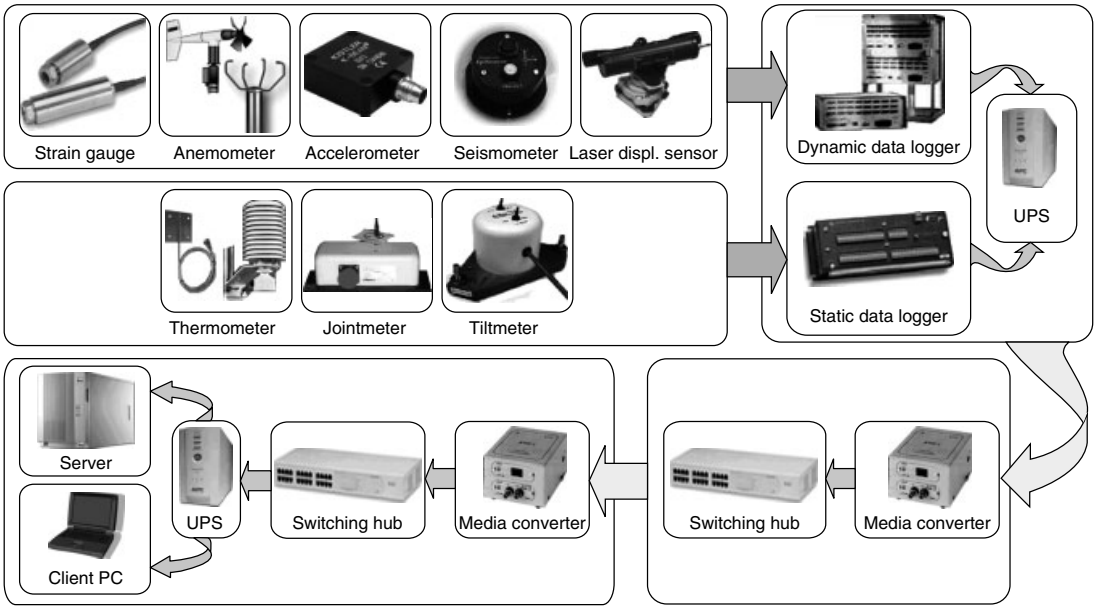


Figure 22. Organization chart of the monitoring system installed in Gwangan Bridge.

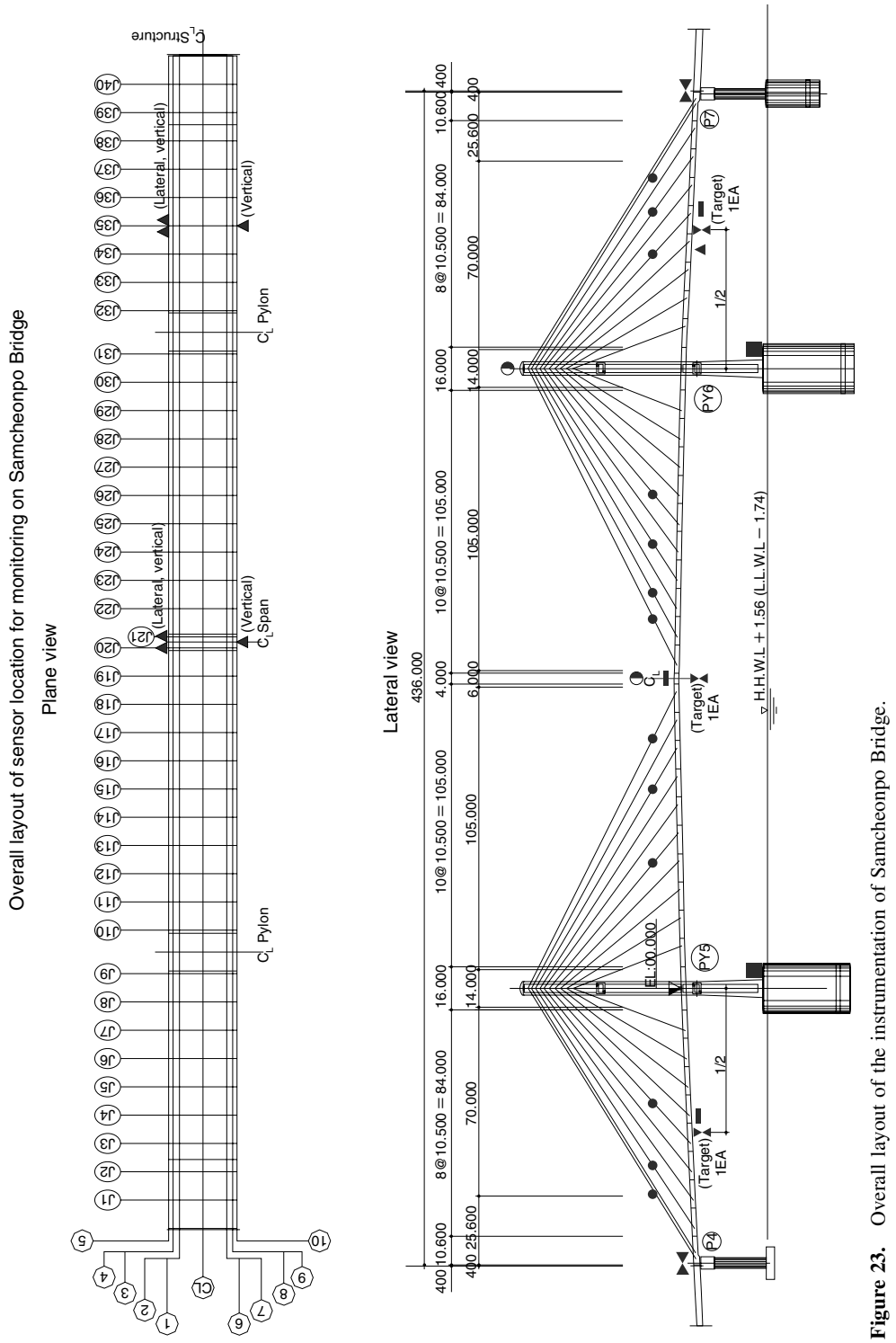


Figure 23. Overall layout of the instrumentation of Samcheonpo Bridge.

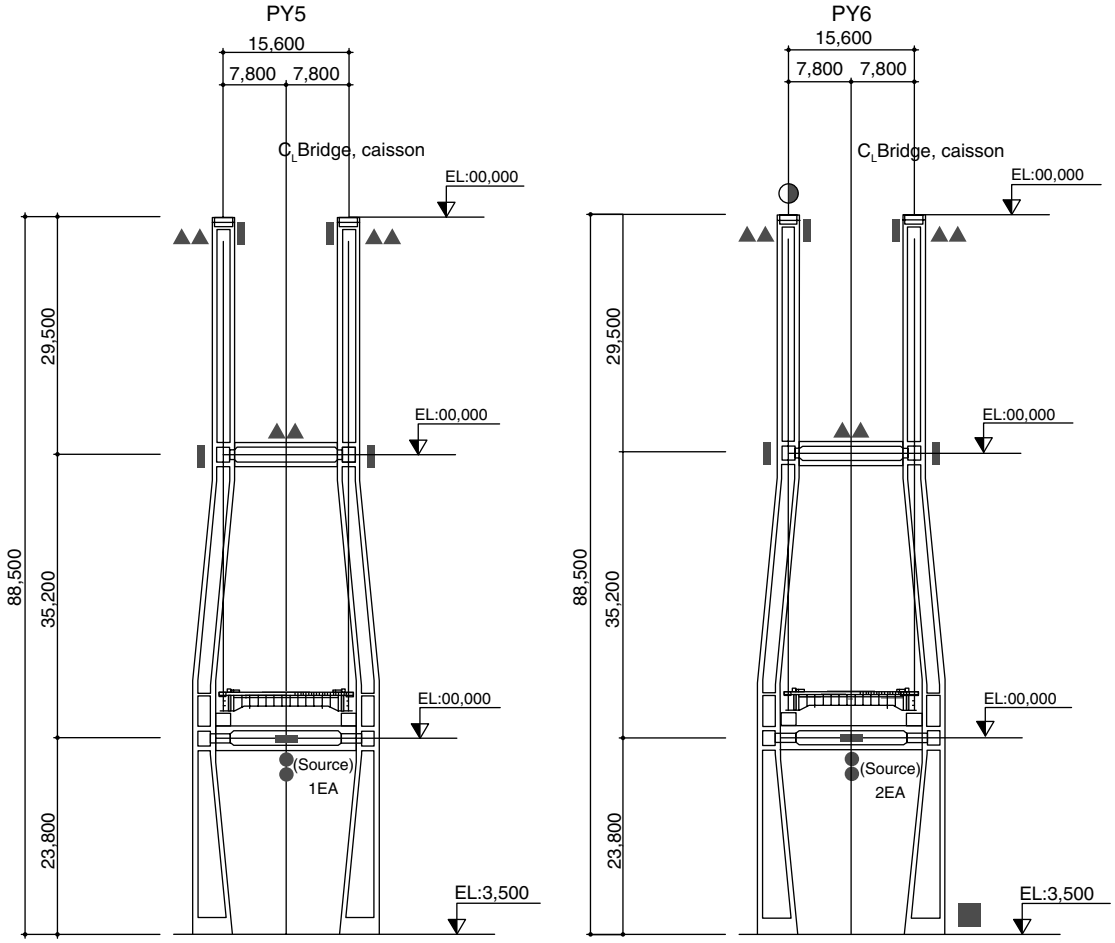


Figure 24. Overall layout of the instrumentation in the pylons of Samcheonpo Bridge.

Table 5. Characteristics of sensors installed in Samcheonpo Bridge

Symbol	Sensor	Quantity	Location
◑	Anemometer	2[(a)]	PY6, midspan
●	Accelerometer	26[(a)]	Cables
▲	Uniaxial accelerometer	18[(a)]	Deck, pylon strut
▮	Biaxial tiltmeter	8[(b)]	PY5, PY6
▬	Resistance temperature detector (RTD)	5	
■	Seismometer	2[(a)]	PY5, PY6
⁝	Laser displacement sensor (transceiver)	3[(a)]	PY5, PY6
⧗	Laser displacement sensor (target)	3[(a)]	Midspan, midside spans
⋈	Jointmeter (wire type)	2[(b)]	Expansion joint, shoe

PY, pylon.
[(a)] Dynamic.
[(b)] Static.

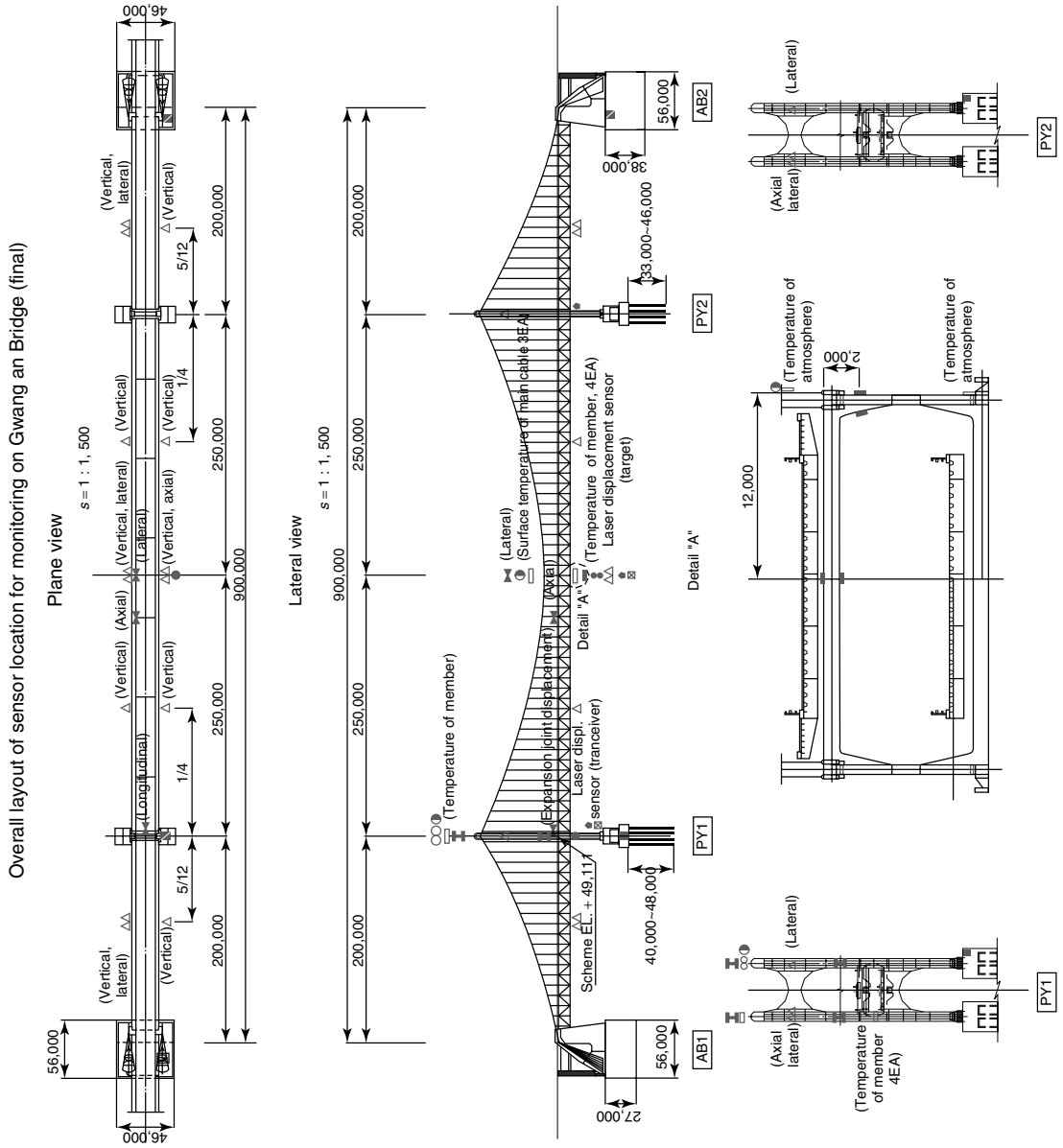


Figure 25. Overall layout of the instrumentation of Gwangan Bridge (Ab, Abutment).

5.1 Samcheonpo Bridge

Samcheonpo Bridge is a three-span cable-stayed bridge with composite girder, located in the Hallyeo maritime national park and connecting Sacheon city to Changsung Island [15]. The health monitoring hardware system for the management of Samcheonpo Bridge (Figures 23 and 24 and Table 5) is similar to that of Gwangan Bridge presented later.

5.2 Gwangan Bridge

Gwangan Bridge, the central part of Gwangan principal road, is located in front of Gwangan town beach and is the longest suspension bridge in Korea with its 900-m overall length. It is an earth-anchored suspension bridge with a double-deck warren truss girder carrying roadways. The pylons are steel towers where the main cable is sustained with its stiffening girder at 105-m height.

The health monitoring hardware system for the management of Gwangan Bridge was designed to perform real-time monitoring of its structural behavior. Composed by dynamic monitoring instruments like laser displacement sensor, anemometer, accelerometer, and by static sensors such as tiltmeter, thermometer, and jointmeter, the monitoring system processes signals, analyzes data, and stores the data acquired from the sensors in the monitoring center. Figure 25 illustrates the location of the instrumentation and Table 6 lists its specifications.

The SHM system was used to produce alarm/warning during the crossing of Maemi typhoon in September 2003. Measurement of the wind speed at the pylon and midspan of the bridge (Figure 26) helped to make decision of blocking and reopening of the bridge to traffic so as to ensure public safety during the typhoon. Traffic was blocked with respect to an average wind speed at midspan of $20\,\mathrm{ms}^{-1}$ and was reopened when this speed was less than $10\,\mathrm{ms}^{-1}$.

Health monitoring after the crossing of the typhoon was also performed using the measured inclination of the pylon (Figure 27) and displacement at the expansion joints. The corresponding natural frequencies (Figure 28) were computed and the results showed that natural frequencies remained within safety limits, which made it possible to conclude that the bridge was not affected by the typhoon.

Table 6. Characteristics of sensors installed in Gwangan Bridge

Symbol	Sensor	Quantity	Location
◑	Anemometer (propeller type)	3[a]	PY1, midspan
○○	Anemometer (ultrasonic type)	1[a]	PY1
▭	RTD (3ea. for atmosphere, 11ea. for members)	4	
□	Thermometer	14[b]	Truss, PY1, cable, air
■	Seismometer	2[a]	PY1, AB2
●	Laser displacement sensor (transceiver)	1[a]	Midspan
● ●	Laser displacement sensor (target)	1[a]	Midspan
⊢	Tiltmeter	4[b]	PY1
⋈	Jointmeter (wire type)	3[b]	Expansion joint, shoe
—	Strain gauge	4[a]	Truss
△	Accelerometer (uniaxial)	20[a]	PY1, PY2, truss
	Uniaxial accelerometer (portable)	—	For tensile force measurement
⬟	Dynamic data logger	3[a]	
⊠	Static data logger	2[b]	

ea., each.
[a] Dynamic.
[b] Static.

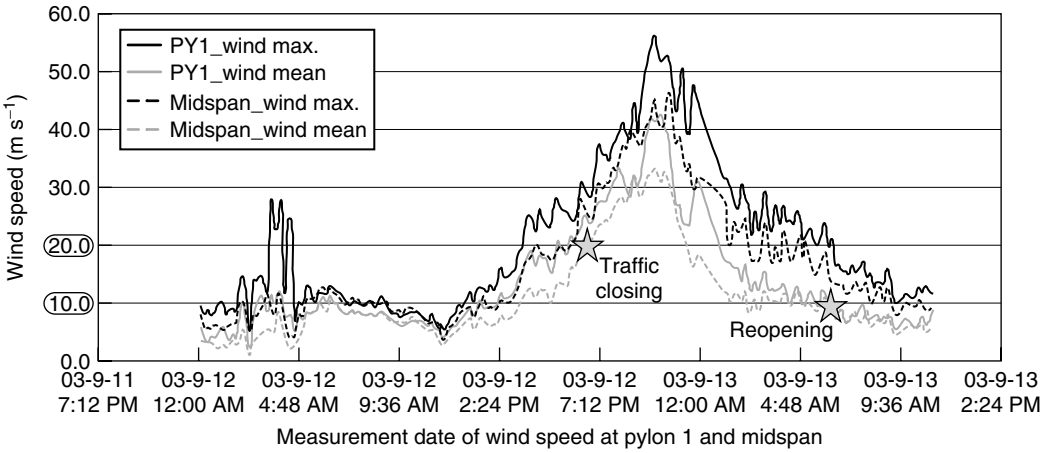


Figure 26. Wind speeds measured in the pylon and midspan of Gwangan Bridge during Maemi typhoon in 2003.

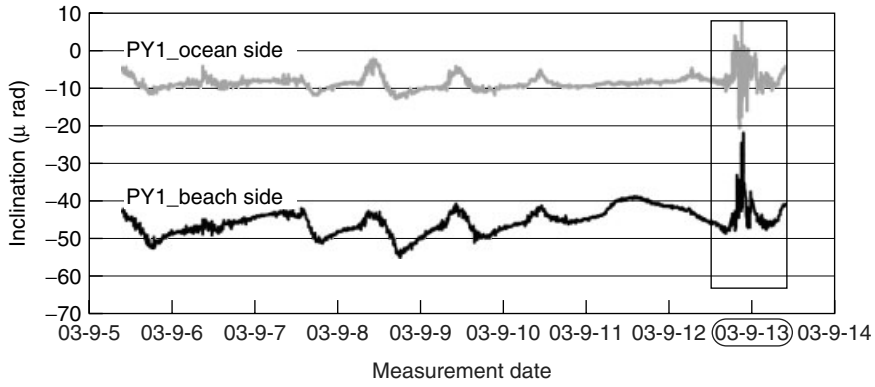


Figure 27. Pylon inclination of Gwangan Bridge measured during Maemi typhoon.

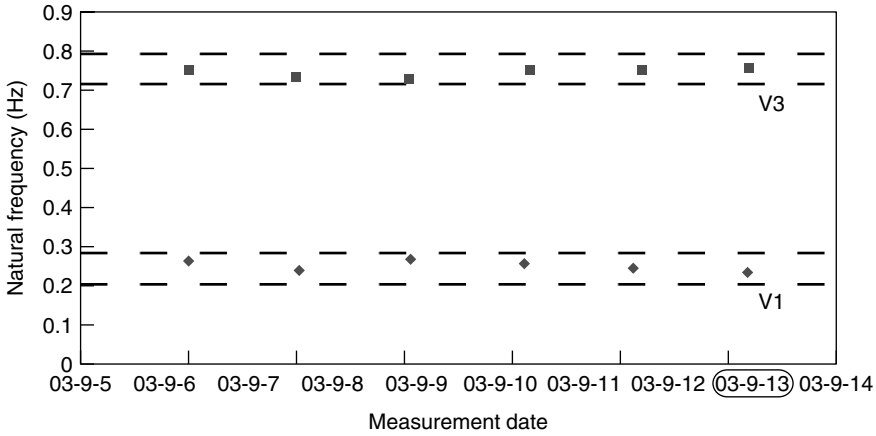


Figure 28. Natural frequencies obtained from pylon inclination of Gwangan Bridge measured during Maemi typhoon.

6 CONCLUSIONS

Remarkable developments have been made in health monitoring within a very short period of time, which has enabled to obtain more reliable information on the actual state of the structures. This is a challenging task where domestic and international researchers are exerting efforts to develop and enhance the performance of the current system by introducing new sensing techniques, power generation from bridge vibration, web-based operating system or wireless signal transmission in order to achieve advanced innovative functions like sensor fusion, reliable massive signal transmission, automated surveillance, adaptive signal processing, etc.

Recent development and applications of SHM for newly built bridges have been addressed through a review of the evolution of bridge instrumentation in Korea, which make Korea as one of the leading countries in bridge monitoring today. The observations obtained through integrated SHM system as well as SBBMS were seen to be exploited effectively for verifying design, calibrating analytic model, assessing long-term behavior, and giving alarm when abnormal behavior is detected.

Although health monitoring system itself plays an important role in studying real behavior of structures in real environment and in reducing uncertainties in further design process, it has been combined with other technologies such as SI and damage detection theory, BMS, and artificial intelligence (AI) to give overall estimation of bridge condition and to make proper maintenance decision and, finally, to lengthen the service life of structures. Current researches are focusing on implementing decision algorithm for repair and strengthening method, priority, and budgeting as well as on improving hardware performance of health monitoring system. Accordingly, current researches focus on developing systematic decision algorithm for repair and strengthening, and improving SHM hardware performance so as to lengthen service life of bridges and ensure serviceability and safety.

REFERENCES

[1] Korea Ministry of Construction and Transportation, *Records of the Current Status the Korean Bridge Stock*, 2006 (in Korean).

[2] Koh HM. Recent research and development on bridge technology in Korea. *Proceedings International Symposium on Sea-Crossing Long-Span Bridges*, Korean Group of IABSE: Mokpo, February, 2006; pp. 81–93.

[3] Koh HM, Choo JF. Advanced bridge research and monitoring activities in Korea (Invited lecture). *Proceedings SAMCO Summer Academy*, Zell am See, September, 2005.

[4] Korea Ministry of Communication and Transportation, *Final report on Bridge Management System and Management in 2002* , 2003, (in Korean).

[5] Korea Institute of Construction Technology, *Report on Bridge Management*, 2002, (in Korean).

[6] Koh HM, Choo JF, Kim S, Kim CY. Recent application and development of structural health monitoring systems and intelligent structures in Korea (State-of-the-art report). *Proceedings 1st International Conference on Structural Health Monitoring and Intelligent Infrastructure*, Tokyo, Japan. Balkema Publishers: Lisse, The Netherlands, November 2003; pp. 99–111.

[7] Korea Institute of Construction Technology, Current Status of Bridge Management System, *Workshop on Maintenance System of Public Facilities*. Korea, June 1996.

[8] Koh HM, Kim S, Choo JF. Recent development of bridge health monitoring system in Korea. *Proceedings North American Euro Pacific Workshop for Sensing Issues in Civil Structural Health Monitoring*. Oahu, HI, November 2004. Springer: Dordrecht, The Netherlands, 2005; pp. 33–42.

[9] Korea Ministry of Construction and Transportation, and Korea Institute of Construction Technology, *Namhae Grand Bridge Safety Evaluation Report*, 1993.

[10] Hyundai Institute of Construction Technology, *Geometry Evaluation and Monitoring System of the Namhae Suspension Bridge*, 1996.

[11] Kim CY, Kim NS, Yoon JG, Jung DS. Monitoring system and ambient vibration test of Namhae suspension bridge. *Proceedings of SPIE 5th Annual International Symposium on Nondestructive Evaluation and Health Monitoring of Aging Infrastructure*. SPIE: Newport Beach, CA, March 2000; Vol. 3995, pp. 324–332.

[12] Kim CY, Kim NS, Cho EK, Yoon JG. Ambient vibration tests on two cable-supported bridges in Korea. *Proceedings IABSE Conference on Cable-Supported Bridges*. IABSE Report: Seoul, June 2001; Vol. 84, pp. 268–269.

[13] Kim S, Kim CY, Lee J. Monitoring results of a self-anchored suspension bridge. *Proceedings North American Euro Pacific Workshop for Sensing Issues in Civil Structural Health Monitoring*. Oahu, HI, November 2004. Springer: Dordrecht, The Netherlands, 2005, pp. 475–484.

[14] Park CM, Park JC. Evaluation of structural behavior using full scale measurements on the Seohae cable-stayed bridge. *Proceedings Annual Conference of Korean Society of Civil Engineers (KSCE)*. Korea, 2003; pp. 571–576.

[15] Yoon JG, Lee J, Kim JI. FVT signal processing for structural identification of cable-stayed bridge. *Proceedings 2003 Fall Workshop of Korean Society for Noise and Vibration Engineering (KNSVE)*, Korea, 2003.

Chapter 125
Bridge Monitoring in Japan

Masato Abe[1] and Yozo Fujino[2]

[1] *BMC Corporation, Mihama-Ku, Chiba, Japan*
[2] *Department of Civil Engineering, University of Tokyo, Tokyo, Japan*

1 INTRODUCTION

Development of bridge monitoring in Japan has been highly influenced by the geographical and socioeconomical conditions of Japan. Natural disasters such as earthquakes and typhoons are some of the major concerns for civil engineering construction. Therefore, sensing technology for natural disasters is well developed and various bridges have been instrumented for decades to evaluate these uncertain loading associated with natural disasters. Although based on similar sensing technology, these instrumentations are mainly focused on environmental and load effect on structures, but not necessarily aimed to detect damage or other structural health monitoring purposes. Historically, Japanese monitoring has laid more emphasis on environmental measurement for verification of design assumptions.

Owing to aging of infrastructure built during the post-WWII rapid economic growth between 1955 and 1975, evaluation of structural performance and/or damage of existing structures has become more and more important for rational and efficient stock management. Structural health monitoring is currently attracting wide interest as one of the key solutions to this problem. At the same time, social concern for safety is higher than ever for public infrastructure following recent failures of major bridges in Japan, as well as in the United States. Therefore, monitoring technology not only to keep average stock condition for stock management but also to prevent extreme failure event is strongly needed. Although these two requirements appear to be similar, practical implementations for these two needs are different. Stock management would focus more on the average condition of the entire stock, whereas safety and risk management would focus on extreme values, which are very unlikely to occur. These two features of health monitoring are treated separately in the current article.

In Section 2, background of development for Japanese monitoring technology is summarized to

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

include the following: (i) environment and natural disasters, (ii) stock management, and (iii) risk and safety. In the following sections, examples of implementation of monitoring for bridges are explained under these three categories.

2 BACKGROUND OF DEVELOPMENT

2.1 Environment and natural disasters

Japan is located in the seismically active and typhoon-prone area of the Pacific Rim. In the 1995 Kobe earthquake, more than 6000 people were killed and major infrastructure was heavily damaged as shown in Figure 1 [1].

Figure 2 shows the number of deaths, and the damage costs for public facilities in Japan since 1950 [2]. The decreasing tendency of the number of deaths and slight increase in loss in public facilities can be observed in Figure 2. Extreme values for both human and property losses are observed at 1995, due to the Kobe earthquake. Although the damage cost for public facilities is only a fraction of the entire damage loss including the private sector whose statistics are not available on consistent bases, it is considered to give the representative indicator for the entire damage.

Figure 3 shows the government expenditure related to disaster prevention. Steady increase in the disaster reduction budgets is observed, especially after the 1995 Kobe earthquake. Because annual public-facility damage is about 1–2 trillion yen and the disaster reduction budget is about 3–4 trillion, every year, approximately 4–6 trillion yen is spent for natural disasters by the public sector of in Japan.

Figure 1. Damage after the 1995 Kobe earthquake.

Sensing technology forms the basis for not only evaluation of meteorological risk, but it has also been playing a major role in seismic risk evaluation. Seismometers are densely distributed throughout Japan, and seismic intensities are reported right after the earthquake [3]. Tokyo Company Gas has its

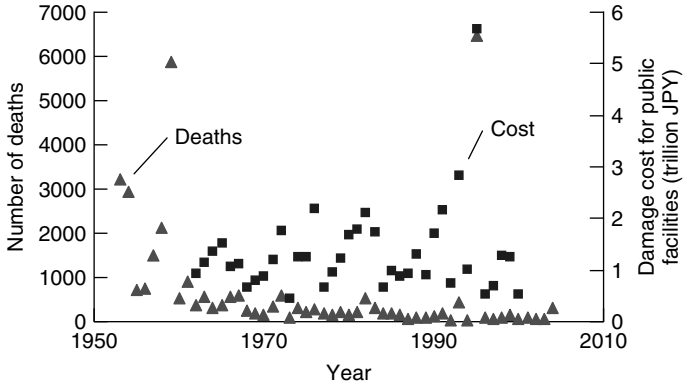


Figure 2. Damage loss due to natural disasters in Japan.

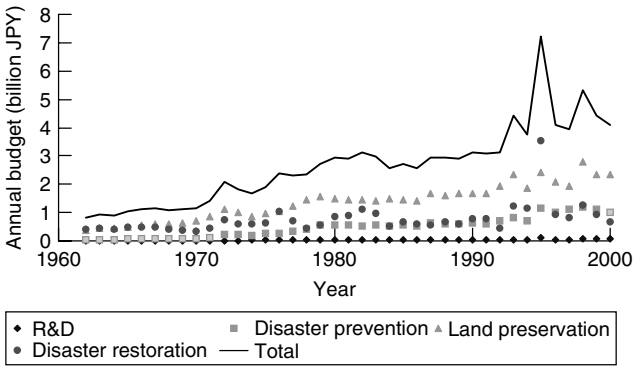


Figure 3. Government expenditure for disaster reduction.

own network of seismic sensors and utilizes the information to shut down the network to prevent fire and explosion after major earthquakes [4].

From 2007, earthquake early warning is issued by the Japan Meteorological Agency [5]. The earthquake early warning system automatically calculates the focus and magnitude of the earthquake and estimates the seismic intensity for each location by detecting the ground motion (i.e., the P-wave, or the preliminary tremor) near its focus. An earthquake early warning is then given within a few seconds to a few tens of seconds before the arrival of the S-wave, or principal motion. This technology was first developed for seismic protection of Shinkansen trains several decades ago, and is now applied nationwide [6].

Another geographical condition, which influences the Japanese infrastructure, is that the country consists of mountainous islands. Therefore, many of the bridges are built near seashores or to cross channels such as the world's longest Akashi–Kaikyo Bridge (Figure 4). These bridges are exposed to severe environmental deterioration conditions with high chloride-ion density and humidity. In the Akashi–Kaikyo Bridge, a dry-air injection system (Figure 5) has been installed to protect the main cable by humidity control [7]. Figure 6 shows other sensors in the monitoring of the Akashi–Kaikyo Bridge [8].

Because of the geography, land use is also concentrated, and severe traffic loadings are often observed, especially in metropolitan areas as shown in Figure 7.

These environmental conditions impose a burden on infrastructure, forcing additional spending to keep the integrity of the infrastructure system. Reduction in this expenditure and enhancement of safety would naturally be the motivation for structural health monitoring.

Figure 4. Akashi–Kaikyo Bridge.

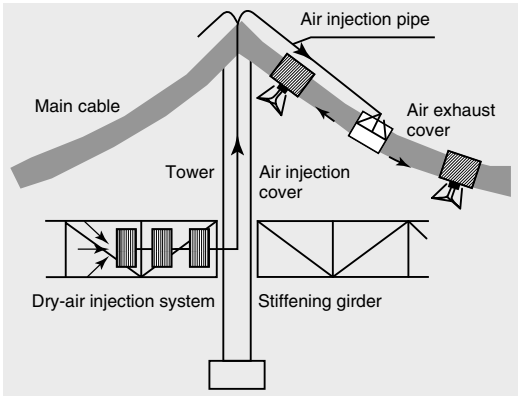


Figure 5. Dry-air injection system of the Akashi–Kaikyo Bridge.

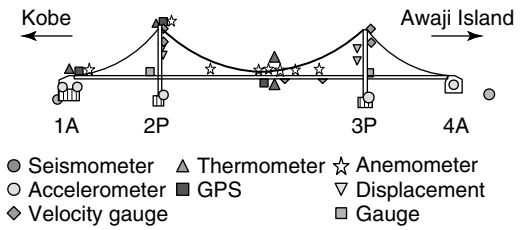


Figure 6. Monitoring system of Akashi–Kaikyo Bridge.

Figure 7. Metropolitan Expressway.

2.2 Stock management

Japanese economy developed rapidly between 1955 and 1975, and a large amount of infrastructure was built in this period. Figure 8 shows a comparison of bridge construction between Japan and the United States [9]. On an average, the US bridges are 10-years older, but the construction of the Japanese bridges took place over a shorter, concentrated period of time. This concentration may induce simultaneous deterioration of a large portion of the stock and would lead to high social cost.

Japanese National Railway (JNR), now divided and privatized, had systematic inspection procedures with data accumulation. Figure 9 shows an example of such data for the period 1963–1971. Annual repair spending and deficient stock are normalized with respect to the total infrastructure stock of JNR. It can be seen that deficient stock is kept at 10–15% of total stock by spending about 1% on repair [10].

Figure 10 shows the total infrastructure stock and GDP from 1955 to 2000 [11, 12]. Because GDP is the flow in economy, while infrastructure is the stock, i.e., integration of flow, a steady increase of stock is observed. As explained by Figure 9, a certain percentage of stock, e.g., 1%, is required each year to maintain the condition of the stock, a larger portion of GDP would be required to in this effort. However, as shown in Figure 11 [13], rehabilitation expenditure with respect to the entire stock is reducing each year, which implies the requirement for more efficient technology to maintain the condition of the stock.

2.3 Risk and safety

Recent finding of severe damage at the Kisogawa Bridge (Figures 12 and 13) on the major trunk highway caused high public awareness on the stock condition and safety of bridges. Because truss structure is theoretically a statically determinant structure, this failure can be considered to near collapse. Detection of this kind of fatal failure would be required to assure safety.

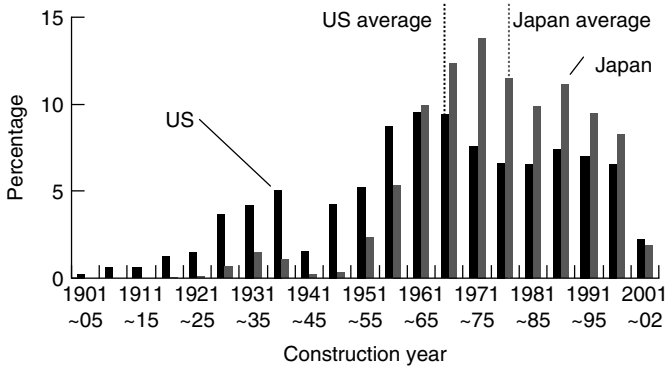


Figure 8. Bridge construction distribution.

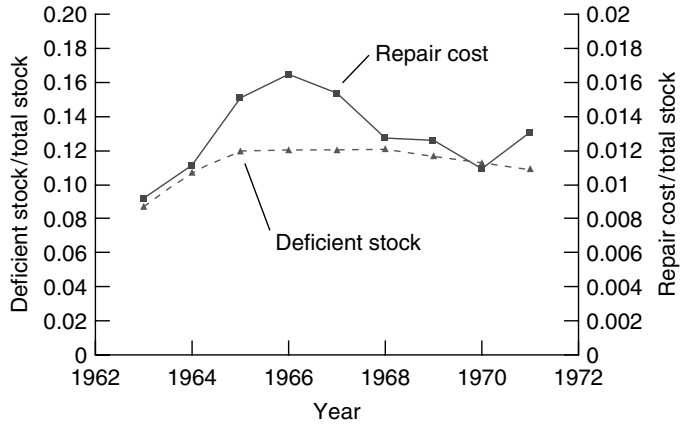


Figure 9. Repair cost and deficient stock at JNR.

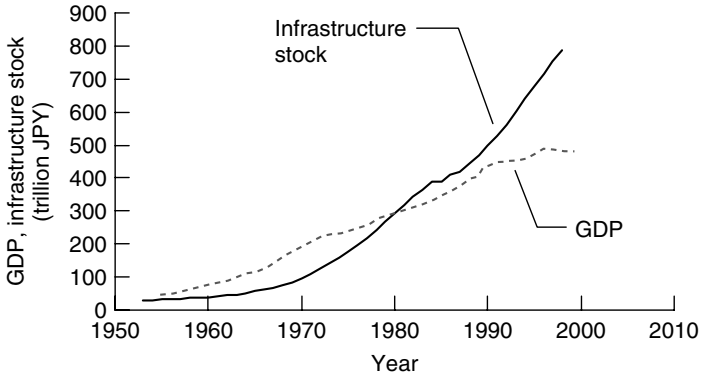


Figure 10. Infrastructure stock and GDP.

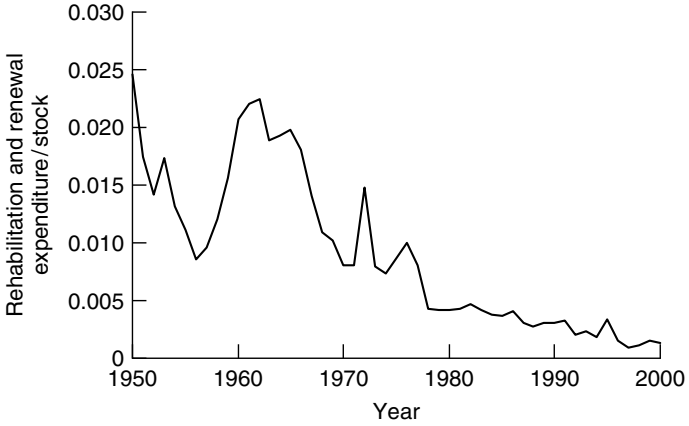


Figure 11. Fraction of rehabilitation and renewal expenditure with respect to the stock.

Figure 12. Kisogawa Bridge.

Figure 13. Kisogawa Bridge defect.

3 MONITORING FOR ENVIRONMENT AND DISASTERS

3.1 Earthquake

Because of the high intensity of seismic activities in Japan, monitoring for seismic response has been widely employed, especially for bridges with special features such as long-span bridges [8], and bridges with new technology such as base-isolated bridges [14].

One such example is the Yokohama Bay Bridge (Figure 14) [15]. A dense seismic measurement system has been installed, as shown in Figure 15. The system identification method based on the system realization using information matrix (SRIM)

Figure 14. Yokohama Bay Bridge.

that utilizes correlations between base motions and bridge accelerations to identify coefficient matrices of a state-space model is applied to several seismic records.

In addition to global behavior such as amplitude dependence of damping ratios, variations of local components were also observed. The identified results of the Yokohama Bay Bridge revealed two types of the first longitudinal mode, where the main difference was the relative modal displacement between the end piers and girder (Figure 16). In design, the end piers and girder are connected by link-bearing connections (LBCs) whose essential function is to prevent the large inertial force of a superstructure from being imparted to substructures during large excitation. For this purpose, the LBC is expected to function as a longitudinal hinge connection to indicate that the girder and pier caps work as separate units in design.

The first type of longitudinal mode exhibited a large relative modal displacement as evidenced by $\varphi = 0.80$ and $\varphi = 0.82$ for the left and right end piers, respectively. The φ factor is defined to express the relative motion between the pier cap and the girder. The shape characteristics and frequency of this mode were very close to that of the analytically obtained the first longitudinal mode. The large relative modal displacement suggested that there might be a hinge mechanism.

The other mode was also identified. The smaller relative modal displacements between the end piers and the girder of this mode (i.e., $\varphi = 0.28$ and $\varphi = 0.25$ for the left and right end piers, respectively) suggested that the LBC had yet to function as full-hinged connections, causing stiffer connection

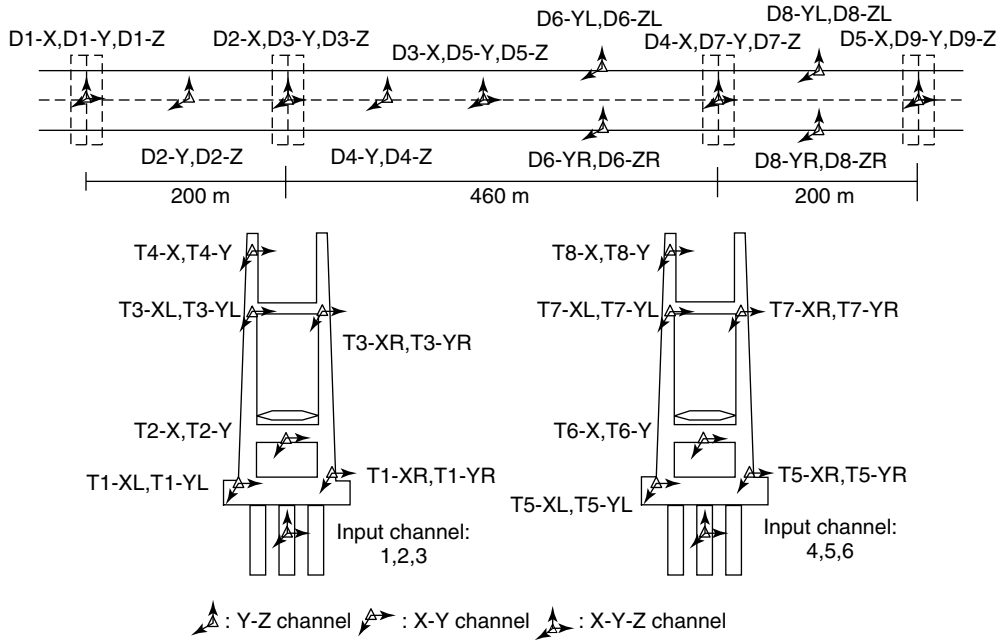


Figure 15. Seismic monitoring at the Yokohama Bay Bridge.

with higher natural frequency. These results indicate that performance of LBC depends on the amplitude of earthquake excitation and does not always follow the analytical prediction.

Although the study is focused on comparison based on design and analysis, the result implies that damage or malfunctioning of the bearing could be detected by the monitoring system. In addition, this result is reflected to on-going seismic retrofit of the bridge, connecting the girder end to the footing by cables (*see* **Design Principles for Civil Structures**; **Soil–Structure Interaction and Seismic Effects**).

3.2 Wind

Wind-induced ambient vibration is measured by dense array at the Hakucho Bridge for a few weeks under various levels of wind speed (Figures 17 and 18) [16]. Ambient vibration measurement is an important tool to evaluate the integrity of in-service structures (*see* **Ambient Vibration Monitoring**; **The Influence of Environmental Factors**). The measured data clearly show the quadratic relationship between wind velocity and response as shown in Figure 19.

The applied structural identification method consists of two steps: identification of vibration modes and inverse analysis of structural properties from the identified modes. For modal identification, the method treats the structure as a multi-input–multioutput system, distinguishing noise from true modes and employing ambient vibration measurement. For the identification of structural properties, assumptions on proportionality of damping, previous estimation of structural damping/stiffness, and numerical iteration are not required. The results verify that the method can precisely determine the characteristics of not only the lower modes, but also the higher modes, and can effectively detect changes in the structural properties. Figure 20 shows the identified aerodynamic stiffness and damping, which are usually measured in wind tunnels, but difficult to obtain *in situ*.

4 MONITORING FOR STOCK MANAGEMENT

Monitoring is expected to improve conventional inspection procedure and rationalize stock management (*see* **Maintenance Principles for Civil Structures**).

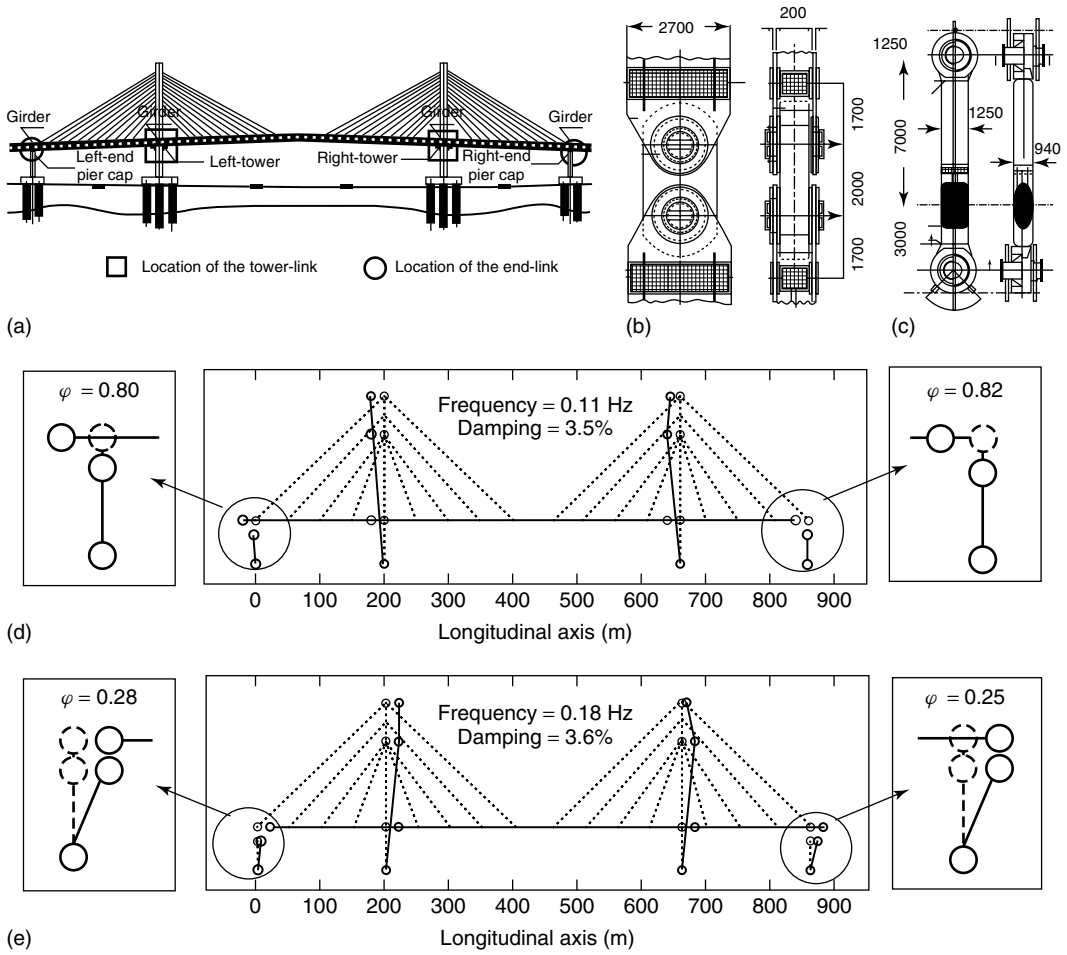


Figure 16. (a) Location of link-bearing connection of the Yokohama Bay Bridge; (b) typical LBC at the tower; (c) typical LBC at end piers; two typical first modes of the Yokohama Bay Bridge identified from the main shock at 17 min 57 sec; (d) hinged–hinged mode: $\varphi_{\text{left}} = 0.80$, $\varphi_{\text{right}} = 0.82$; (e) fixed–fixed mode: $\varphi_{\text{left}} = 0.28$, $\varphi_{\text{right}} = 0.25$.

Figure 17. Hakucho Bridge.

In this context, several examples of monitoring are briefly introduced: (i) workhorse bridges; (ii) preventive maintenance; and (iii) advanced routine inspection.

4.1 Workhorse bridges

For efficient stock management, it is essential to monitor short to medium span bridges, which are the major portion of the stock. Not only the cost associated with sensor hardware, processing and interpreting the large number of data is also a great

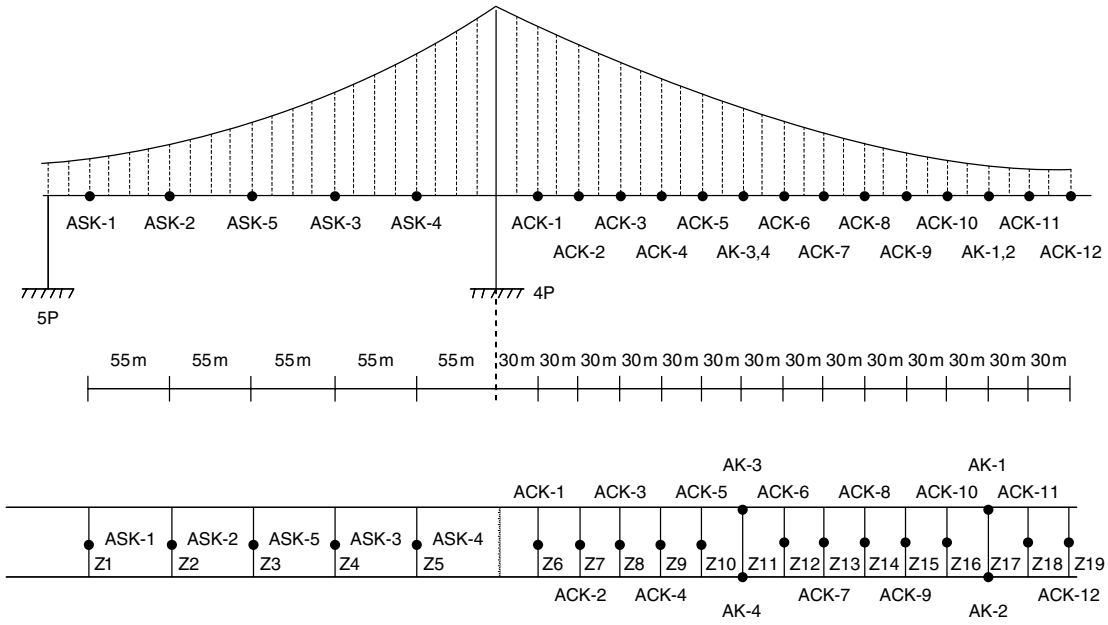


Figure 18. Accelerometer installation.

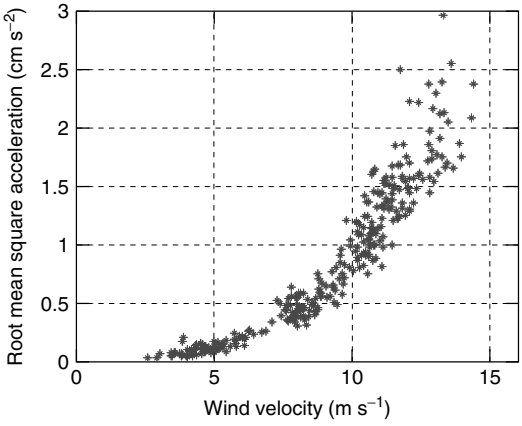


Figure 19. Wind velocity and response.

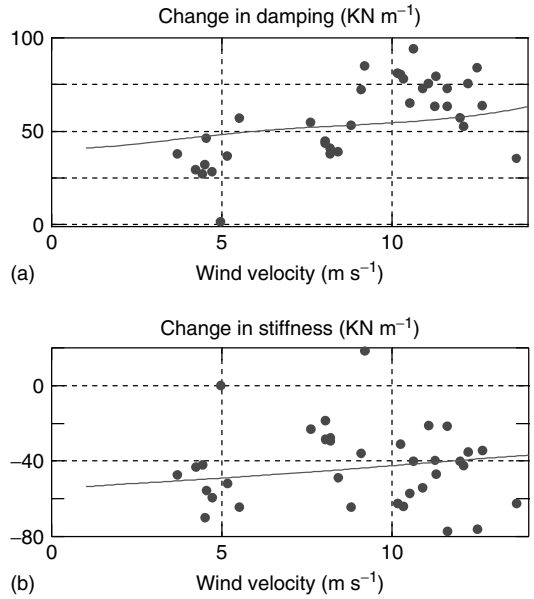


Figure 20. Identified changes in aerodynamic force. (a) Aerodynamic damping, (b) aerodynamic stiffness. ●: Identified value; —: wind-tunnel experiment.

challenge for engineers. The study in [17] demonstrates how to apply some available algorithms to the data measured by a practical structural health monitoring system and how to evaluate the results for the sample ordi- nary medium span bridge of Figure 21. The sensor location at the bridge is shown in Figure 22. Identification of dynamic characteristics, tempe- rature effect, and fatigue damage evaluation (*see* **Fatigue Life Assessment of Structures**) are studied using monitored data. Results indicate that the structural condition is deficient and further special inspection is required at the monitored bridge,

although the structure itself is relatively new, i.e., 12 years old at the time of the study. The deficiency would be related to the skew effect, which may have not been appropriately treated in design. The demonstrated implementation and methodology can be considered as the general basis for monitoring of common workhorse structures.

This field is actively studied and other examples can be found in [18, 19].

4.2 Monitoring for preventive maintenance

Monitoring technology has been applied to improve maintenance at several bridges.

Figure 21. Monitored bridge.

Figure 23. Monitored bridges piers: connecting bridge to Shin–Kitakyushu Airport.

The Shin–Kitakyushu Airport has been constructed on land reclaimed from the sea, 2 km off the east coast of Fukuoka Prefecture, Japan. The airport is connected to the coast by an access bridge having a total of 25 reinforced concrete piers. The bridge is designed taking into consideration 100 years of design service life. However, 22 piers have been constructed in the marine environment as shown in Figure 23, where the piers suffer severe chloride attack, and their durability during the service life is difficult without any maintenance action. Therefore, for constructing a durable bridge, a scheme of durability design of the piers is established on the premise of maintenance action starting from the beginning of the service during its life [20]. The preventive maintenance procedure is shown in Figure 24.

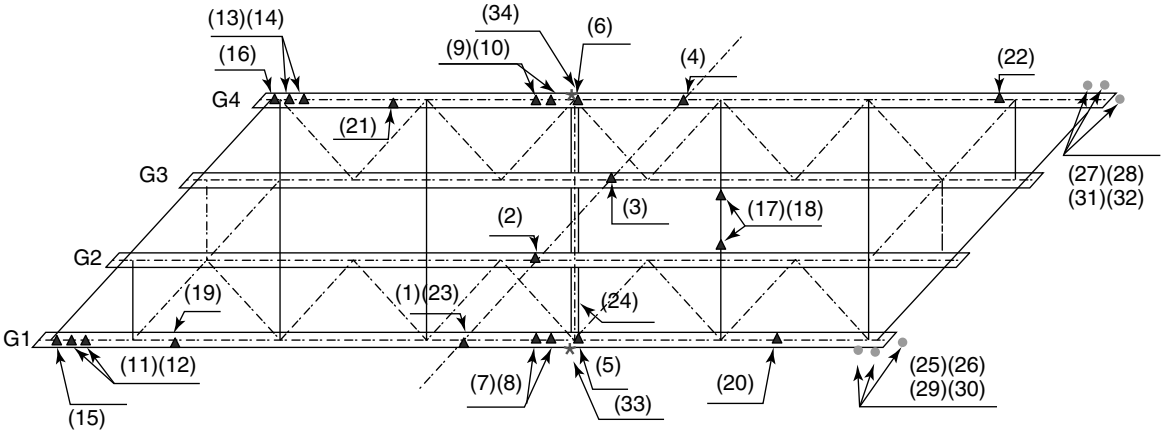


Figure 22. Sensor location. ▲: Strain gauge, 24; ●: displacement sensor, 8; ⋆: thermometer, 2.

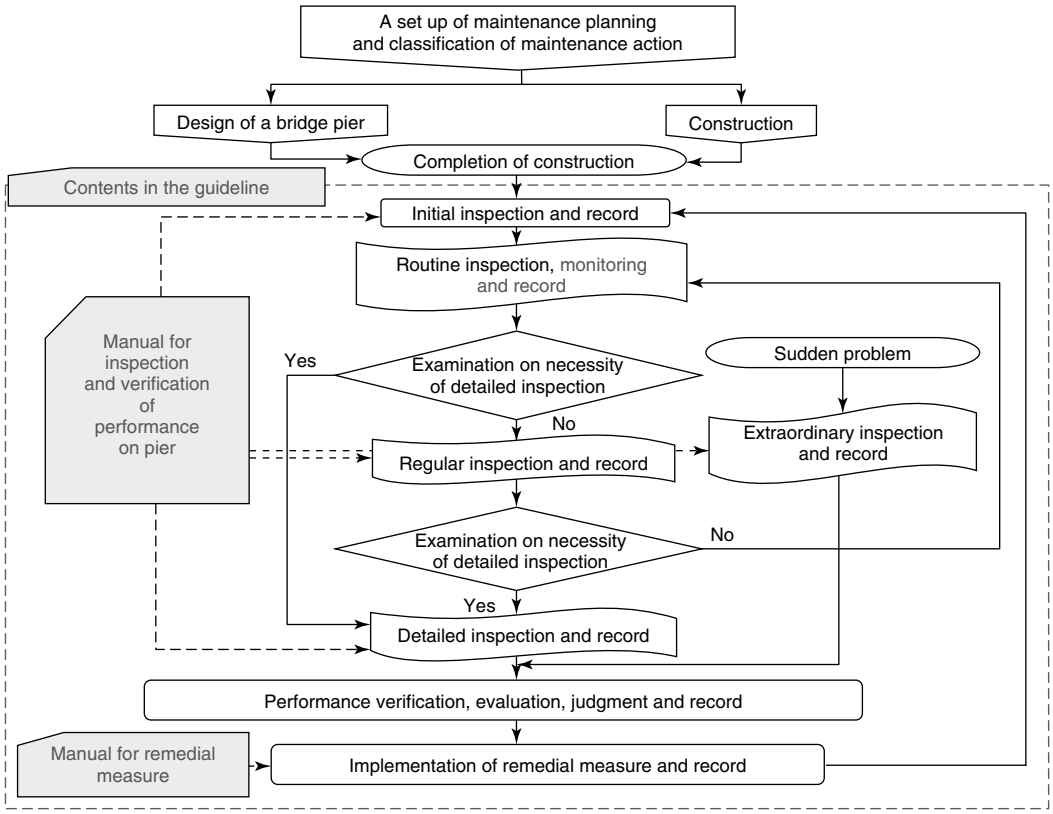


Figure 24. Maintenance procedure.

In the maintenance action, monitoring of chloride penetration into concrete and corrodibility of steel reinforcement is included as continuous inspections for the deterioration of the piers. The outline of the chloride-sensing mechanism is described in Figure 25. The probe body is made of polymer cement mortar cylindrical in shape, around which four thin steel wires are wound over small grooves [21]. Once chloride penetrates into the concrete from the external environment and its content in concrete around the wire reaches the critical value, the wire is depassivated and corrosion starts. After the wire corrodes, it breaks because of its minute cross section.

This chloride-sensing probe utilizes the corrosion of thin wires, which deteriorate faster. A similar sensor is also applied to the fatigue problem. Figure 26 shows the fatigue sensor installed at a railway bridge [22]. The sensor is made of notched thin plate, and by propagation of the crack from the notch, accumulated damage can be estimated.

4.3 Advanced routine inspection

Because transportation infrastructure, such as highways and railways, forms a huge network, bridges are the essential links within. Performance of transportation network relies on the overall performance of these bridges, and efficient inspection is required so that the performance be kept. Conventionally, the structural state is evaluated by qualitative visual inspection. For more efficient inspection for the entire network, and to obtain quantitative information on the infrastructure, vehicle-based monitoring systems are developed for highways and railways. The developed system is called *vehicle intelligent monitoring system* (VIMS), which consists of an accelerometer and a global positioning system (GPS) receiver mounted on a patrol or service vehicle as shown Figure 27 [23]. By using patrol/service vehicles, it is possible to cover the entire network with much higher density and frequency compared with the conventional inspection

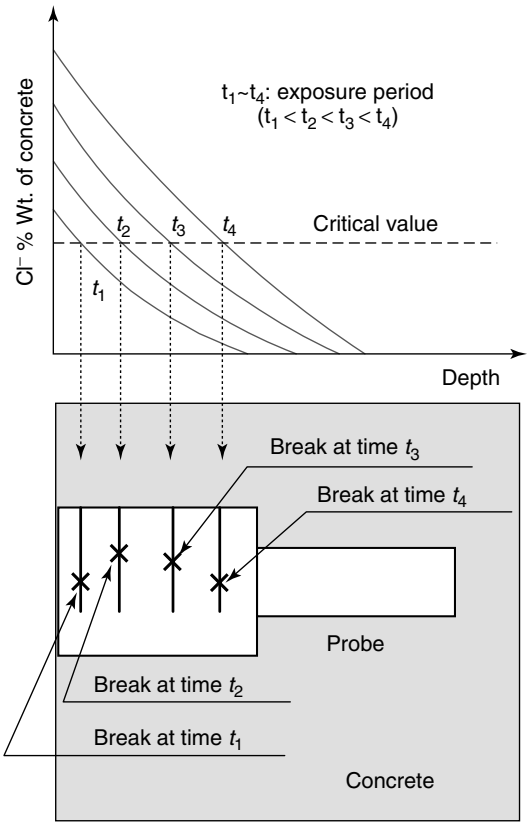


Figure 25. Chloride sensor probe.

with designated personnel and vehicles. Monitoring systems to measure floor vibration of the vehicles are developed so that installation can be made without modification of the vehicles, and to reduce cost of sensors and systems.

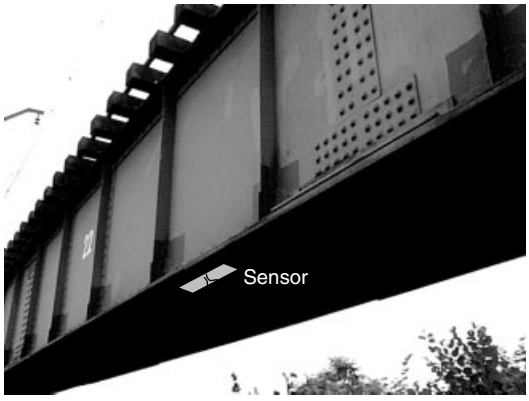


Figure 26. Fatigue sensor installed at a railway bridge.

This system has been applied to Tokyo Metropolitan Expressway, and its effectiveness in road maintenance, i.e., frequent monitoring of the conditions of road pavement and expansion joints, is confirmed [24].

Also, the system is applied to railways, and possibility of utilization of ordinary service vehicles for rail-track monitoring has been studied [25]. The developed system is applied to an actual railway and measurements at different points of time are compared [26]. By looking at the rms acceleration responses, it is found that the responses at the same location are stable with repetitive measurements. It is also observed that repair and deterioration of track beds cause significant change in measurement as shown in Figure 28. Between the two dates indicated in the figure, a major typhoon attacked the area, and the tracks were heavily damaged and repaired.

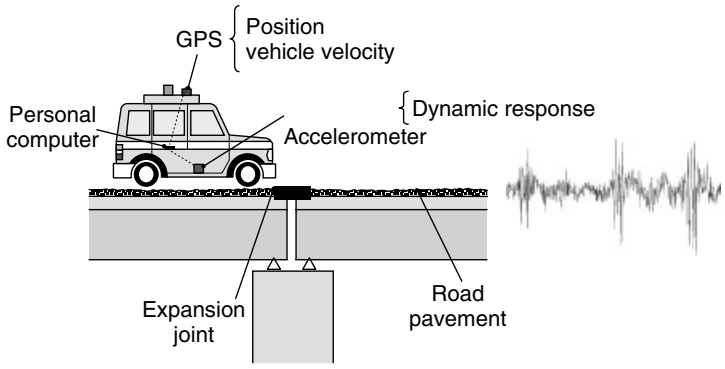


Figure 27. Vehicle intelligent monitoring system.

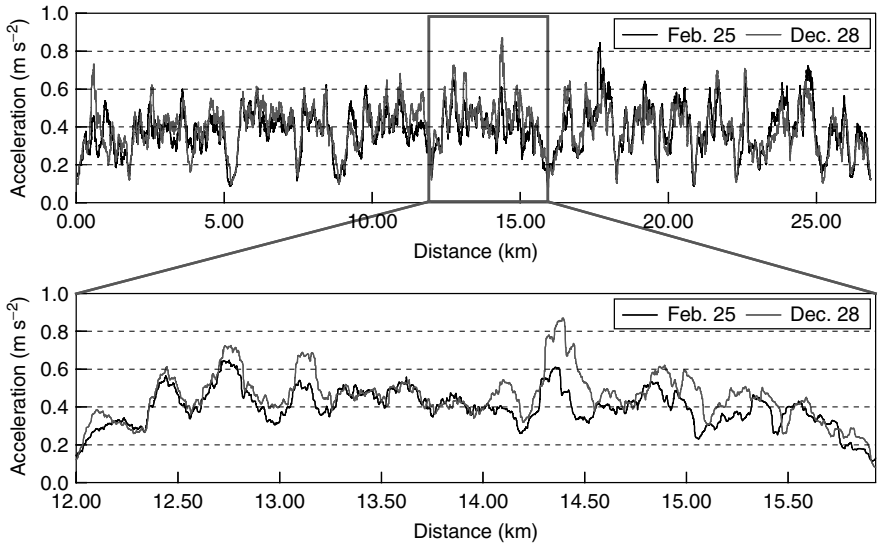


Figure 28. Detected progressive damage.

5 MONITORING FOR RISK AND VULNERABILITY

In this section, monitoring for safety is described from two viewpoints: (i) failure detection; and (ii) vulnerability evaluation (*see* **Risk Monitoring of Civil Structures**).

5.1 Failure detection

When the failure mode of interest is identified or specified, it is not difficult to implement a monitoring system. Several such systems are developed and implemented for safety of railway operations.

Figure 29 shows an unseating sensor. When unseating of a bridge girder occurs, the sensor breaks and an electrical signal is sent to the headquarters to suspend operations. The system is installed to overpass bridges above the highway, where collision risk is supposed to be high. Figure 30 shows clinometers installed at bridges to detect scour-induced collapse

Figure 29. Unseating sensor.

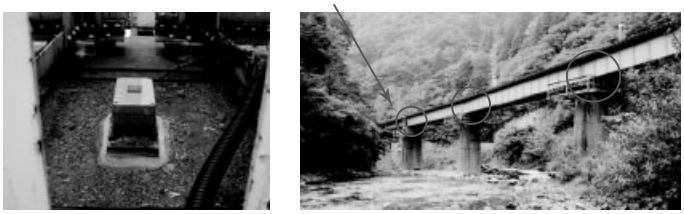


Figure 30. Clinometric type scour-monitoring device. [Reproduced from Ref. 27.

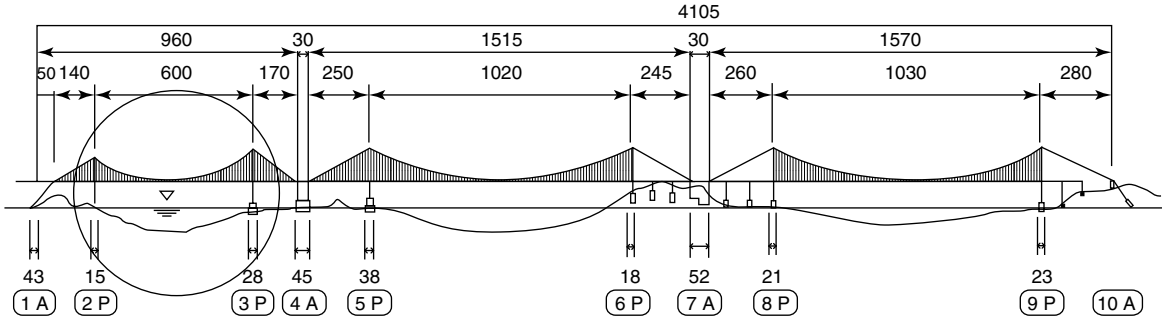


Figure 31. First Kurushima Kaikyo Bridge.

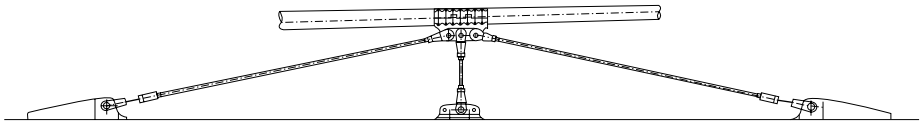


Figure 32. Center stay.

of bridge piers [27]. An alarm is sent when the inclination reaches the maintenance limits of track irregularity. These monitoring systems are developed for specific failure mode and risk, for the purpose of safe operation.

A natural extension of this kind of monitoring system would be the universal sensor system to detect damage and failure, which is the major research area in structural health monitoring.

At the time of the 2001 Geiyo Earthquake, with an intensity of M 6.7 intensity, the center stay rod at the first Kurushima Kaikyo Bridge failed [28]. The outlines of the bridge and the center stay are shown in Figures 31 and 32. Figures 33 and 34 show the failure.

Observed seismic ground motion is applied to dynamic three-dimensional finite element analysis, and it was verified that the failed center stay rods performed as they had been designed to perform. Also, analysis indicates that the failure may have occurred either at 12 or 15 s and other members would not have been damaged. In this way, reanalysis of observed data can provide damage or failure information. Further research and development would be required for damage or failure detection to be made on line in real time.

Figure 33. Observed failure.

5.2 Vulnerability evaluation

Extreme events such as earthquake or scour or accidents due to human error occur in an uncertain manner. Preventive action should be taken in accordance with the probability of the hazard risk and also the extent of vulnerability of the structures. The first step, environmental monitoring, and the next one, vulnerability detection, fall into the core domain of structural health monitoring.

An example of this category of inspection is the impact-testing method, developed by the Railway Technical Research Institute, and is widely used in Japanese railways for the inspection of bridge

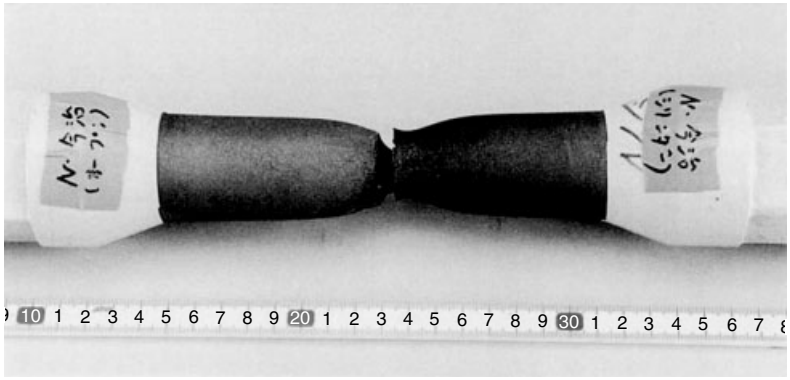


Figure 34. Failed cross section.

substructures [29]. The fundamental procedure is as follows:

1. Measure the free vibration response of the pier excited by impact as shown in Figure 35.
2. Identify the fundamental natural frequency from observed the power spectral density (PSD). This process is manually done by reading the peaks of the PSD.
3. Calculate the integrity index α given by

$$\alpha = \frac{\text{(measured natural frequency)}}{\text{(standard value of natural frequency: } F)} \quad (1)$$

4. Judge the integrity according to the standard given in Table 1. The value of F is provided by regression equations obtained from statistical analysis of the past testing record of railway bridges. For example, the regression equation for a substructure with a pile foundation is given by

$$F = -9.9 \log H_d + 0.005\ W_h + 14.9 \quad (2)$$

where, H_d is the height of the pier from the ground surface (meters), and W_h is the weight of the girder (tonforce), and F, the standard value of natural frequency (Hertz).

The procedure is quantitative, but based purely on heuristic experience. Here, natural frequency is used as the indicator of structural integrity or vulnerability. In design, identical structures have identical natural frequencies, but in reality, natural frequency scatters even for the same design. In this way, the measurement-based index provides at least some information on the relative vulnerability of the structure.

In this direction, trials are being made to use structural response itself as the indicator of vulnerability [30]. A conceptual scheme is shown in Figure 36,

Figure 35. Impact testing.

Table 1. Standard for structural integrity

Integrity index	Integrity criteria	Treatment
Below 0.70	*A* *A*1	Detailed inspection, consideration for countermeasure
Below 0.85	*A*2	Monitor progress of defects such as inclination or scour
Above 0.86	*B*, *C*, *S*	Can be considered sound

*A*1, advanced damage and performance degradation; *A*2, advanced damage, threat to performance degradation; *B*, would reach *A* if progressive; *C*, minor damage; *S*, no damage.

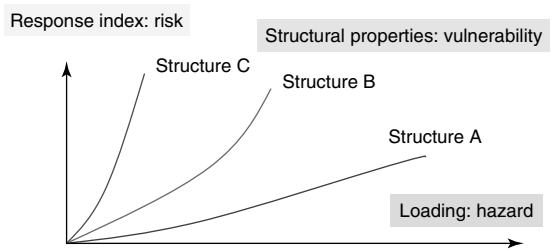


Figure 36. Fundamental concept of vulnerability detection.

where the response index is plotted with respect to loading index, and different response characteristics for a similar structure provides relative vulnerability. The concept is being developed for railway substructures subject to train loading [31] and seismic vulnerability evaluation of buildings [32]. These studies are intended to evaluate overall performance of structures with certain indicators. This approach could be considered "macroscopic" monitoring, in contrast with the "microscopic" monitoring to detect damage or failure. This approach is considered advantageous from a practical point of view since it can readily be related to existing inspection schemes based on visual condition rating.

To date, this approach is based solely on heuristics or statistics. Rationalization and theorization of the method would be the step for its generalization.

6 CONCLUSIONS

In this article, Japanese development of health monitoring of bridges is reviewed from the geographical and socioeconomical background to specific applications. These backgrounds are categorized as (i) environment and natural disasters, (ii) stock management, and (iii) risk and safety. In other words, severe environment, limited resource for maintenance, and increased requirement for safety are the key factors. Structural health monitoring can be considered as a key technology to meet these conflicting demands.

Owing to pronounced natural-disaster risk and severe environment for deterioration, environment and loading have been monitored at various bridges for decades. At several bridges, these measured data are also analyzed in the context of structural monitoring. Although these measurements were intended to measure loading, but not necessarily to measure structural condition, analyses indicate usefulness of these measurements for evaluation of structural integrity.

Structural health monitoring technology is also being developed to improve efficiency of stock management. Monitoring technologies for common workhorse bridges, preventive maintenance, and improvement of routine inspection are being developed and applied to the real world in this domain.

For improved safety, several monitoring technologies for risk and vulnerability are implemented. When the threatening failure mode can be identified, specific monitoring devices, such as unseating sensors or clinometers for scour are installed, and the information is used for operation control. Also, reanalysis of seismic failure implies usefulness of measurement for on-line damage detection, although further research studies would be needed for this purpose. As an example of vulnerability monitoring, the impact vibration testing method to evaluate integrity of structures is introduced. This technique can be used to compare relative safety, and to prioritize the improvement action. Structural health monitoring technologies in this domain can be categorized into (i) microscopic monitoring, where damage detection and localization are main interest; and (ii) macroscopic monitoring, where holistic structural integrity and its comparison are the main focus. The first one is the conventional mainstream of structural health monitoring, while the latter is currently attracting interest especially from practical point of view to connect health monitoring and existing inspection.

REFERENCES

[1] Shinozuka M. *The Hanshin-Awaji Earthquake of January 17, 1995 Performance of Lifelines*. National Center for Earthquake Engineering Research, State University of New York: Buffalo, 1995.

[2] Fujino Y, Abe M. Characterization of risk, hazard and vulnerability in natural disasters. *Proceedings of the 10th International Conference on Applications of Statistics and Probability in Civil Engineering*, University of Tokyo, Japan, 2007.

[3] National Research Institute for Earth Science and Disaster Prevention, Kyoshin-Network, http://www.k-net.bosai.go.jp/ 2007.

[4] Shimizu Y. A new technology for earthquake disaster prevention. *Japan Society of Civil Engineers* 2000 **38**:38–41.

[5] Japan Meteorological Agency, Earthquake Early Warning, http://www.jma.go.jp/jma/en/Activities/eew.html 2007.

[6] Nakamura Y. UrEDAS, the earthquake warning system: today and tomorrow. In *Earthquake Early Warning Systems*, Gasparini P (ed). Springer-Verlag, 2007, pp. 249–281.

[7] Kitagawa M. Technology of the Akashi Kaikyo bridge. *Structural Control and Health Monitoring* 2004 **11**(2):75–90.

[8] Yasuda M, Kitagawa M, Moritani T, Fukunaga S. Seismic design and behavior during the Hyogo-ken Nanbu earthquake of the Akashi Kaikyo bridge. *Proceedings of the 12th World Conference on Earthquake Engineering*, Auckland, 2000.

[9] Fujino Y, Abe M. Challenge to bridge management in the US. *Journal of the Japan Society of Civil Engineers (in Japanese)* 2007 **92**(6):70–73.

[10] Abe M, Fujino Y. Reanalysis of life cycle costs in management of steel railway bridges. *Proceedings of the International Symposium on Integrated Life-Cycle Design and Management of Infrastructure*. Shanghai, 2007.

[11] Cabinet Office of Japan, *Social Capital of Japan*, 2002.

[12] Statistics Bureau, *Historical Statistics of Japan*. Ministry of Internal Affairs and Communications of Japan, http://www.stat.go.jp/english/data/chouki/index.htm 2007.

[13] Ministry of Land, Infrastructure and Transport, *White Paper on Land, Infrastructure and Transport*, 2002.

[14] Chaudhary MTA, Abe M, Fujino Y, Yoshida J. System identification of two base-isolated bridges using seismic records. *Journal of Structural Engineering* 2000 **126**:1187–1195.

[15] Siringoringo DM, Fujino Y. System identification applied to long-span cable-supported bridges using seismic records. *Earthquake Engineering and Structural Dynamics* 2008 **37**(3):361–386.

[16] Nagayama T, Abe M, Fujino Y, Ikeda K. Structural identification of a nonproportionally damped system and its application to a full-scale suspension bridge. *Journal of Structural Engineering* 2005 **131**:1536–1545.

[17] Xia Y, Fujino Y, Abe M, Murakoshi J. Short-term and long-term health monitoring experience of a short highway bridge: case study. *Journal of Bridge Structures* 2005 **1**(1):43–53.

[18] Sasaki E, Miki C, Tohmori M, Ishikawa Y, Miyazaki S. Proposal of a remote bridge monitoring system for damage detection. *Proceedings of the Third International Conference on Urban Earthquake Engineering*, Tokyo Institute of Technology, Japan, 2006; pp. 417–424.

[19] Mutsuyoshi T, Nishimura T, Kato Y, Uomoto T. Structural performance monitoring of reinforced concrete bridges. *Proceedings of the 4th International Symposium on New Technologies for Urban Safety of Mega Cities in Asia*, Singapore, 2005.

[20] Takewaka K. Maintenance plan for 100 years of service life on a newly constructed structures. *Proceedings of the International Workshop on Service Life of Concrete Structures—Concept and Design*, Sapporo, 2005.

[21] Takewaka K, Hoang NX, Leelalerkiat V, Yamamoto S. A non-destructive and quantitative monitoring system for penetration of chloride into concrete structure. *Proceedings of the 8th East Asia-Pacific Conference on Structural Engineering and Construction*, Singapore, 2001.

[22] Abe M, Komon K, Narumoto A, Sugidate M, Mori T, Miki C. Monitoring of railway bridges in Japan. In *Proceedings of the SPIE—Nondestructive Evaluation of Highways, Utilities, and Pipelines IV*, Aktan AE, Gosselin SR (eds). SPIE, 2000; Vol. 3395, pp. 245–252.

[23] Kono H, Abe M, Fujino Y. Development of vehicle intelligent monitoring system (VIMS). *Proceedings of the Annual Conference of the Japan Society of Civil Engineers (in Japanese)*, Hokkaido University, Japan, 2002; Vol. 57, pp. 481–482.

[24] Fujino Y, Kitagawa K, Furukawa T, Ishii H. Development of vehicle intelligent monitoring system (VIMS). In *Proceedings of the SPIE—Smart Structures and Materials 2005: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Tomizuka M (ed). SPIE, 2005; Vol. 5765, pp. 148–157.

[25] Shimozono T, Abe M, Fujino Y, Koshiba A, Shikama T. Development of a track monitoring system using vehicle vibration. *Proceedings of the Annual Conference of the Japan Society of Civil Engineers (in Japanese)*, Tokushima University, Japan, 2003; Vol. 58, pp. 5–6.

[26] Fujino Y. A study of train intelligent measurement system using acceleration of train. In *Proceedings of*

the SPIE—Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, Tomizuka M, Yun CB, Giurgiutiu V (eds). SPIE, 2007; Vol. 65291H.

[27] Kobayashi N, Kitsunai S, Shimamura M. Scour monitoring of railway bridge piers via inclination detection. *Proceedings of the 1st International Conference on Scour of Foundations*, Texas, 2002; pp. 910–917.

[28] Fuchida M, Koashi H, Mohri T, Furuya K. Behavior of Kurushima Kaikyo bridge in response to Geiyo Earthquake. *Proceedings of the 3rd International Suspension Bridge Operators' Conference*, Hyogo, 2002.

[29] Haya H, Nishimura A, Sawada R, Koda M. Comparison of impact vibration test with microtremor measurement for spread foundation piers. *Quarterly Report of Railway Technical Research Institute*, 1995; Vol. 36, No. 2.

[30] Suzuki O, Abe M, Shimamura M, Matsunuma M. Health monitoring system for railway bridge piers. *Proceedings of the 3rd International Conference on Structural Health Monitoring and Intelligent Infrastructure*, Vancouver, 2007.

[31] Abe M, Shimamura M, Matsunuma M. Bridge substructure monitoring using live load induced vibration. *TRB 86th Annual Meeting*, Washington, DC, 2007, #07-1289.

[32] Fujino Y. Development of a practical monitoring system of urban infrastructure toward mitigation of disaster and accidents. *SMSST7—the World Forum on Smart Materials and Smart Structures Technology*, Chongqing and Nanjing, 2007.

Chapter 126

Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge

Edwin Reynders and Guido De Roeck

Department of Civil Engineering, Katholieke Universiteit Leuven, Leuven, Belgium

1 INTRODUCTION

The Z24 bridge was located in the canton Bern near Solothurn, Switzerland. The bridge was part of the road connection between the villages of Koppigen and Utzenstorf, overpassing the A1 highway between Bern and Zürich. It was a classical post-tensioned concrete two-cell box girder bridge with a main span of 30 m and two side spans of 14 m (Figures 1–2). The bridge was built as a freestanding frame with the approaches backfilled later. Both abutments consisted of triple concrete columns connected with concrete hinges to the girder. Both intermediate supports were concrete piers clamped into the girder. An extension of the bridge girder at the approaches provided a sliding slab. All supports were rotated with respect to the longitudinal axis that yielded a skew bridge. The bridge, which dated from 1963, was demolished at the end of 1998, because a new railway adjacent to the highway required a new bridge with a larger side span.

Before complete demolition, the bridge was subjected to a long-term continuous monitoring test and several progressive damage tests in the framework of the Brite-EuRam project CT96 0277 SIMCES [1]:

- A *long-term continuous monitoring test* took place during the year before demolition. The aim was to quantify the environmental variability of the bridge dynamics.
- *Progressive damage tests* took place over a month, shortly before complete demolition. The aim was to prove experimentally that realistic damage has a measurable influence on bridge dynamics. The

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

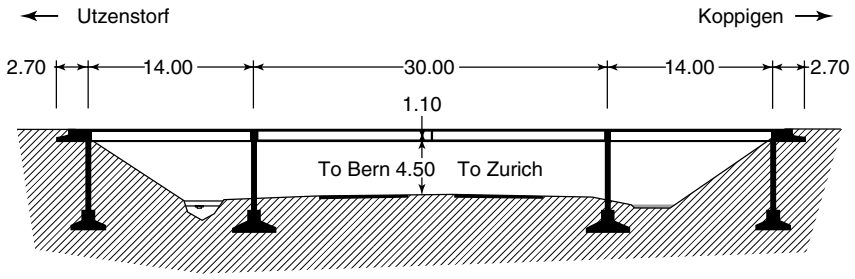


Figure 1. Side view of the Z24 bridge. Distances are in meters.

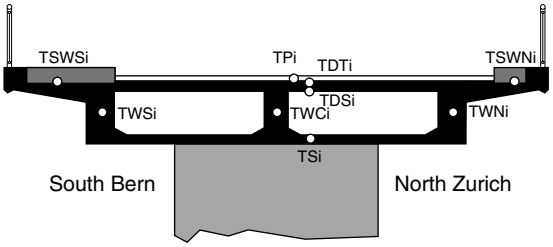


Figure 2. Cross section of the girder, showing the locations where the temperature was monitored.

progressive damage tests were alternated with short-term monitoring tests while the continuous monitoring system was still running during these tests.

The Z24 bridge project was unique in the sense that a long-term continuous monitoring test was combined with realistic short-term progressive damage tests. The measurement data have been used for two benchmarks:

- The shaker, ambient, and drop weight vibration data from the third reference measurement on the Z24 bridge (scenario 8, Table 1) were presented as a benchmark study for system identification methods for operational modal analysis at the IMAC XIX conference in 2001 (Section 3).
- The data from the long-term continuous monitoring tests as well as the data from the progressive damage tests were presented as a benchmark study for algorithms for structural health monitoring and damage identification in the framework of the European Cost Action F3 (Sections 4 and 5).

The benchmark data are still publicly available, and some recently developed methods for modal analysis, damage identification, and health monitoring have been tested on the data.

This article is organized as follows. The different tests are described in Section 2. To provide some insight into the dynamic behavior of the bridge, the operational modal analysis results are discussed in Section 3. Sections 4 and 5 contain a literature review of results obtained from the continuous monitoring data and the progressive damage test data, respectively.

2 TEST DESCRIPTIONS

In this section, a brief overview of the performed tests is given. A profound overview is provided by Krämer *et al.* [2].

2.1 Long-term Continuous Monitoring Test

Since the aim of this test was to quantify the environmental variability of the bridge dynamics, all environmental variables that were considered to be of possible importance for the bridge dynamics have been monitored.

Sensors to measure air temperature, air humidity, rain true or false, wind speed, and wind direction were installed at the bridge, resulting in five sensors for the atmospheric conditions.

A sensor consisting of two inductive loops was installed to detect the presence of vehicles on the bridge.

Since temperature was known to have a key influence on the dynamics of civil engineering structures, the bridge's thermal state was monitored in detail. At the middle of the three spans, the temperature was

Table 1. Progressive damage tests: overview of damage scenarios

No.	Date (1998)	Scenario	Description/simulation of real damage cause
1	04.08	First reference measurement	Healthy structure
2	09.08	Second reference measurement	After installation of lowering system
3	10.08	Lowering of pier, 20 mm	Settlement of subsoil, erosion
4	12.08	Lowering of pier, 40 mm	
5	17.08	Lowering of pier, 80 mm	
6	18.08	Lowering of pier, 95 mm	
7	19.08	Tilt of foundation	Settlement of subsoil, erosion
8	20.08	Third reference measurement	After lifting of the bridge to its initial position
9	25.08	Spalling of concrete, 24 m^2	Vehicle impact, carbonization, and subsequent corrosion of reinforcement
10	26.08	Spalling of concrete, 12 m^2	
11	27.08	Landslide at abutment	Heavy rainfall, erosion
12	31.08	Failure of concrete hinge	Chloride attack, corrosion
13	02.09	Failure of anchor heads I	Corrosion, overstress
14	03.09	Failure of anchor heads II	
15	07.09	Rupture of tendons I	Erroneous or forgotten injection of tendon tubes, chloride influence
16	08.09	Rupture of tendons II	
17	09.09	Rupture of tendons III	

Reproduced from Ref. 16.

measured at eight points on the girder: at the center of the north (TWN), central (TWC), and south (TWS) web; below the north (TSWN) and south (TSWS) sidewalk; at the top (TDT) and soffit (TDS) of the deck, and at the soffit (TS) of the girder (Figure 2). Since the girder was a continuous beam with the intermediate piers clamped into it, the angular deflection of the girder at these piers and the elongation of the midspan were measured. The soil temperature near each of the concrete columns at the approaches was monitored, as well as that near the north, central, and south parts of the intermediate piers (12 sensors in total).

Although the original blueprints of the Z24 bridge indicated that the asphalt layer should have a thickness of 5 cm, the drilling of access holes for the installation of the temperature sensors on the girder revealed a cover of 16–18 cm of asphalt. Therefore, the temperature of the pavement (TP) was measured at the middle of the three spans (Figure 2).

To monitor the bridge dynamics, 16 accelerations have been measured on the bridge at different points and in different directions.

Every hour, 10 scans of environmental data, sampled at 48 sensors, and 8 averages of 8192 acceleration samples, taken at 16 sensors, were collected and stored to a hard disk after compression. The construction works at the new bridge that replaced the Z24, caused the loss of six temperature sensors and damage to one accelerometer. Although the type of accelerometers that had been used was specially designed for long-term use, some showed a considerable drift and a few of them failed during operation.

2.2 Progressive Damage Tests

In order for the damage tests to be significant, it was made sure that (i) they were relevant for the safety of the bridge and (ii) the simulated damage occurred frequently, a condition that was checked in the literature [3, 4] and by questioning Swiss bridge owners. Since the A1 highway was never closed to traffic, some damage scenarios that meet these criteria could not be applied without reducing the safety of the traffic, which was considered of paramount importance. The traffic on the Z24 bridge was diverted to the A36 highway. Table 1 gives a complete overview of all progressive damage tests that were performed. Some of them are illustrated in Figure 3.

Before and after each applied damage scenario, the bridge was subjected to a forced and an ambient

(a) (b)

Figure 3. Settlement system used for damage scenarios 3–6 (a) and cutting of tendons for damage scenarios 15–17 (b).

operational vibration test. With a measurement grid consisting of a regular 3×45 grid on top of the bridge deck and a 2×8 grid on each of the two pillars, 291 degrees of freedom have been measured: all displacements on the pillars, and mainly vertical and lateral displacements on the bridge deck. Because of the limited number of accelerometers and acquisition channels, the data were collected in nine setups using five reference channels. The forced excitation was applied by two vertical shakers of EMPA, Switzerland, placed on the bridge deck. A 1-kN shaker was placed on the middle span and a 0.5-kN shaker was placed at the Koppigen side span. The shaker input signals were generated using an inverse fast Fourier transform (FFT) algorithm, resulting in a fairly flat force spectrum between 3 and 30 Hz. After scenario 8, a drop weight test was also performed. The applied shaker and drop weight forces were periodic with eight periods. A total of 65 536 samples was collected at a sampling rate of 100 Hz, using an antialiasing filter with a 30-Hz cutoff frequency.

3 SYSTEM IDENTIFICATION AND OPERATIONAL MODAL ANALYSIS

From recorded force and acceleration or acceleration-only data, the modal parameters can be determined. Since the ambient forces such as wind excitation or traffic under the bridge could not be excluded during the vibration measurements, all modal tests can be considered as operational modal analysis tests, with or without the use of artificial (exogenous) forces. The shaker, ambient, and drop weight vibration data from the third reference measurement (scenario 8, Table 1) were presented as a benchmark study for system-identification methods for operational modal analysis at the IMAC XIX conference in 2001. Peeters and Ventura compare results obtained by seven different research teams [5].

In addition, new modal parameter estimation techniques have been validated on the benchmark data, such as a parametric and nonparametric setup assembly approach followed by maximum likelihood estimation, developed by Parloo *et al.* [6], and an iterative single degree of freedom (SDOF) technique proposed by Allen and Ginsberg [7]. A very complete set of modal parameters was obtained using the reference-based combined deterministic-stochastic subspace identification (CSI/ref) technique on the shaker data [8]. These recent results are briefly summarized to provide insight into the dynamic behavior of the Z24 bridge.

Figures 4 and 5 show the identified bending modes and lateral modes, respectively. The fact that the lateral modes, which are (almost) exclusively excited by the ambient forces, are identified, delivers an experimental proof for the capability of the CSI/ref method to identify modes that are excited by artificial (measured) excitation and those that are excited by ambient (unmeasured) excitation.

The capability of the CSI/ref method to identify closely spaced modes is clearly illustrated in Figure 6. Because of the skewness of the supports of the bridge, these modes are a combination of bending and torsion.

Figure 7 shows the higher torsion modes that were identified. The fact that even mode 14 at 37.25 Hz could be identified, while the cutoff frequency of the analog antialiasing filter had been set to 30 Hz,

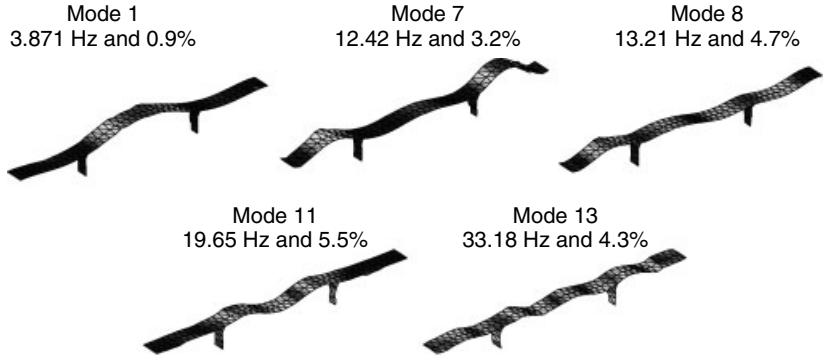


Figure 4. Modal analysis: identified bending modes.

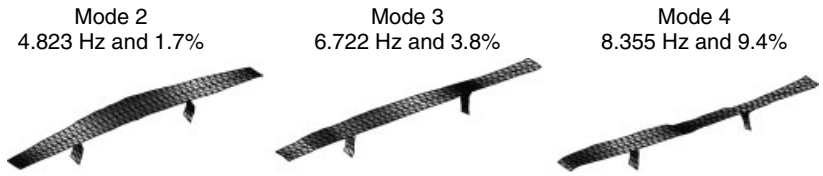


Figure 5. Modal analysis: identified lateral modes.

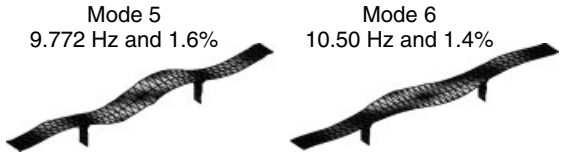


Figure 6. Modal analysis: two closely spaced mixed torsion/bending modes.

indicates that the CSI/ref method is able to identify modes that are very weakly present in the data.

4 CONTINUOUS MONITORING RESULTS

The continuous monitoring data have been investigated by Peeters and De Roeck [9].

First, a modal analysis was performed for each of the 5652 sets of acceleration data recorded every hour. Hereto, the reference-based stochastic subspace identification method was used [10]. Since visual inspection of the stabilization diagram is not an option for continuous monitoring, the selection of modes from the stabilization diagram was automated, resulting in four modes for which the eigenfrequencies could be identified with reasonable accuracy.

Second, the obtained eigenfrequencies were plotted as a function of the recorded environmental variables. Only the temperatures were found to have a clear influence on the eigenfrequencies. As is demonstrated in Figure 8, a plot of the first and second eigenfrequencies versus TP1 and TDS2, respectively, reveals a typical bilinear behavior. This can probably be explained by the change of the Young's modulus

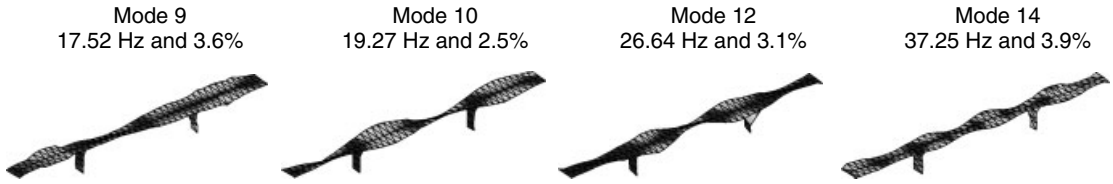


Figure 7. Modal analysis: higher torsion modes.

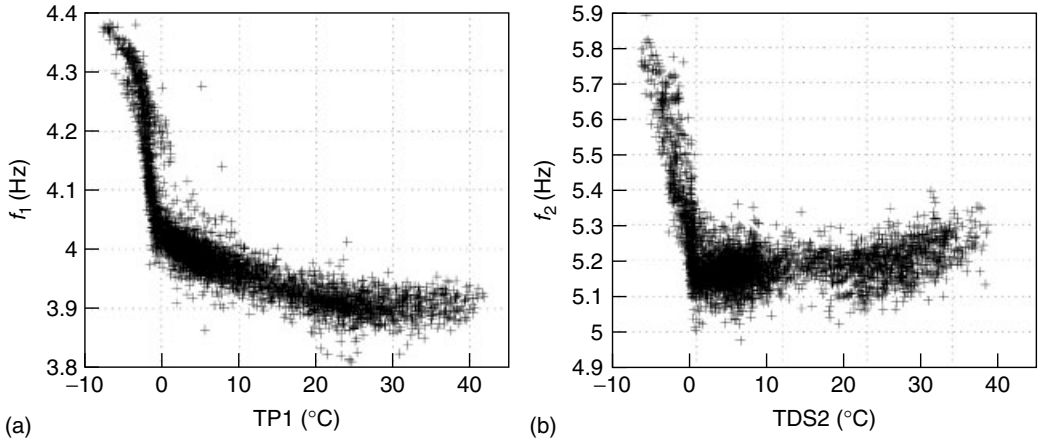


Figure 8. (a) First eigenfrequency versus pavement temperature at the Utzenstorf side span (TP1) and (b) second eigenfrequency versus deck soffit temperature at midspan (TDS2).

of the asphalt layer with temperature. Above 0 °C, the Young's modulus of asphalt is low and varies little with temperature (a constant value of 10 GPa can be taken), whereas below 0 °C, it increases dramatically (to 50 GPa at −10 °C) [11].

Third, a numerical environmental model for the bridge was constructed. The purpose was to remove the environmental influence from the measured eigenfrequencies so as to use them for damage detection (level 1 damage identification [12]). To this end, the number of temperature variables was first reduced from 22 (one variable for each undamaged temperature sensor) to 6 by calculating the correlations between them and grouping variables for which the absolute value of the correlation exceeded 0.99. Because of the lack of a physical model for the temperature influence, a black box ARX model [13] was fitted to each input (temperature) and output (eigenfrequency). A dynamic ARX model performed much better than simple static regression because the thermal inertia of the bridge turned out to be important. It should be mentioned that the ARX model was created with the data for positive temperatures only, since an ARX model is a linear model and so cannot represent the observed bilinear temperature–eigenfrequency behavior. As a consequence, its predictions are only valid for temperatures above 0 °C.

Figure 9 shows the prediction error of the ARX model constructed from the first eigenfrequency as a function of the temperature at the top of the bridge deck at midspan. The vertical line denotes the border between the estimation data that were used to construct the ARX model and the validation data. From day 277 on, the error between the first eigenfrequency as predicted by the ARX model and the measured eigenfrequency becomes significantly large, indicating damage. This date corresponds to damage scenario 5 (pier settlement of 80 mm).

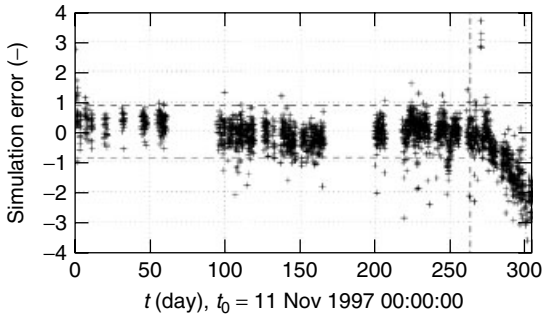


Figure 9. Prediction errors for the first eigenfrequency as a function of TDT2. The horizontal lines denote the 95% confidence bounds.

Mevel *et al.* [14] apply an output-only subspace-based structural damage detection technique to the continuous monitoring data of three accelerometers. Although the environmental influence on the data was denied, damage scenario 5 could be detected. Apparently, the method is more sensitive to damage than to environmental changes.

5 PROGRESSIVE DAMAGE TESTING RESULTS

The data from the progressive damage tests have been investigated by several authors.

Abdel Wahab and De Roeck [15] propose the "curvature damage factor", which is the mean of the absolute differences in measured modal curvatures between a reference state and a damaged state over different modes, for damage detection and localization (level 2 damage identification [12]). Applying it on damage scenarios 2, 5, and 6 (Table 1) using modes 1 and 5 (Figures 4–6), they are able to detect and localize the damage for the 95-mm pier settlement case but not for the 80-mm settlement case. It should be noted that the damage in the immediate vicinity of the settled pier could not be detected because the mode-shape curvature is very small at that location in both damaged and undamaged conditions.

Maeck and De Roeck [16] apply the direct stiffness calculation (DSC) [17] for the detection, localization, and quantification (level 3 damage identification) of damage in beamlike structures on scenarios 2, 4–6, and 8. The DSC is based on the relation that the bending and torsion stiffness in each section of a structure can be written as the quotient of the modal bending and torsion moment, respectively, to the corresponding modal curvature. Although the pier settlement of 40 mm cannot be detected, a maximum bending stiffness decrease of 17 and 30% is detected at midspan near the Koppigen pier for the settlement of 80 and 95 mm, respectively, using mode 1. A torsion stiffness decrease of about 27% is found at the Koppigen pier for the settlement of 80 mm using mode 5. An interesting side result is that for scenario 8, no stiffness decrease is detected, thereby indicating that when the cracks in the concrete closed again due to the lifting of the Koppigen pier, the dynamic stiffness approached its value for the undamaged state.

Mevel *et al.* [14] use a subspace-based damage detection and localization technique to investigate scenarios 2, 3, and 8. While it is claimed that both pier settlements of 20 and 80 mm can be detected (level 1 damage identification), the localization (level 2) fails, because the proposed method is not able to locate damage in symmetric structures and the finite element model that was used for the localization is symmetric.

Kullaa [18] constructs control charts from statistical quality control from eigenfrequencies, eigenfrequencies and reference degree of freedom (DOF) mode shapes, and damping ratios of four modes, obtained from the nine setups for damage scenarios 2, 4, and 5. Using the eigenfrequency and mode-shape data, both pier settlements of 40 and 80 mm can be detected (level 1 damage identification). With the eigenfrequency data, only the settlement of 80 mm can be detected. From the damping ratios, no useful information is extracted.

Parloo [19] considers the analytical sensitivity of the eigenfrequencies of modes 1, 2, 5, 6, and 7 to the stiffness change between two points of the bridge deck. Solving the inverse problem yields the stiffness changes as a function of the observed changes in eigenfrequencies. This method needs mode shapes that are scaled to the mass matrix, but only for the reference configuration, not for the damage configuration. Since the inverse problem is underdetermined, an iteratively weighted pseudoinverse is used, which tries to reduce the identified damage pattern to an as small as possible number of locations. In this way, the method is suitable for identifying local damage. Using a coarse grid, the author is able to locate the pier settlements of 40, 80, and 95 mm correctly and to quantify the damage. Using a denser grid, the localization fails because of the symmetry of the structure, but the detection remains successful. An interesting result obtained with the coarse grid is that, relative to the first reference measurement (scenario 1), damage is localized near the Koppigen pier for the second reference measurement (scenario 2). This indicates that the installation of the settlement system probably caused local damage near the pier, so that the settlement of 20 mm could not be detected with scenario 2 as a reference. However, with scenario 1 as reference, the damage caused by the 20-mm pier settlement could be localized successfully.

Teughels and De Roeck [20] use finite element model updating for detecting, localizing, and quantifying (level 3 damage identification) the damage induced by the pier settlement of 95 mm (scenario 6). An finite element model consisting of three-dimensional beam elements is updated to the reference configuration of scenario 2 using seven damage functions [21] for the bending stiffness, seven damage functions for the torsion stiffness, and two soil

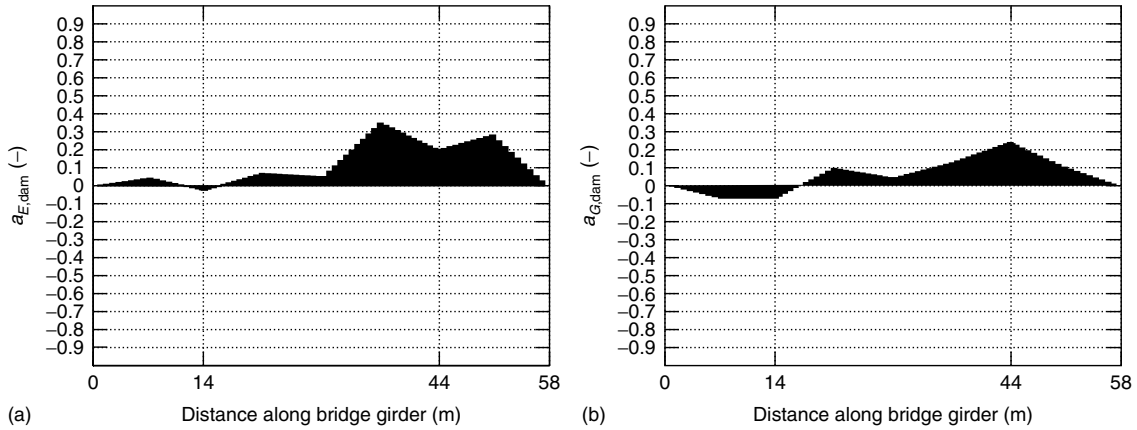


Figure 10. Relative differences in bending stiffness EI (a) and torsion stiffness GI_t (b) for the finite element models updated to scenarios 2 and 6. $\mathrm{EI}_{\mathrm{dam}} = \mathrm{EI}_{\mathrm{ref}}(1 - a_{E,\mathrm{dam}})$ and $\mathrm{GI}_{t,\mathrm{dam}} = \mathrm{GI}_{t,\mathrm{ref}}(1 - a_{G,\mathrm{dam}})$.

springs, resulting in a good match for the eigenfrequencies and mode shapes of modes 1, 2, 5, 6, and 9, for which the updating was performed. Using this updated finite element model for scenario 2, a second updating was performed using the damage functions and the eigenfrequencies and mode shapes of modes 1, 5, 6, and 9, resulting in an updated finite element model for scenario 6. The relative differences in bending and torsion stiffness for both updated finite element models is shown in Figure 10. The maximum bending stiffness decrease of around 35% is in good agreement with the decrease of 30% obtained by Maeck and De Roeck.

6 CONCLUSIONS

One of the goals of the SIMCES project was to deliver a full-scale validation test for the damage-detection methodology based on vibration monitoring. From the tests on the Swiss bridge Z24, the following conclusions can be drawn:

- Reliable modal information can be obtained by output-only dynamic measurements, i.e., accelerations due to ambient influences, e.g., traffic under or on top of the bridge. Closing the bridge to apply controlled force excitation is not necessary.
- The variations between eigenfrequencies obtained by different system identification algorithms are rather small. For the mode shapes, more substantial differences occur. Very good results are obtained by applying the CSI/ref technique.
- Changes in environmental conditions, mainly temperature, lead to changes in eigenfrequencies. With the order of magnitude being similar to that of structural damage, it is important to filter (eliminate) the environmental influence on beforehand. As a consequence, it is advisable to monitor the temperature as well. The relation between eigenfrequencies and temperature can be obtained by monitoring the intact bridge over a period of at least one year. For bridge Z24, the decrease of the first eigenfrequency for an overall temperature increase from $0\,^{\circ}\mathrm{C}$ to $30\,^{\circ}\mathrm{C}$ is about 3%. When the temperature correction is taken into account, the uncertainty interval reduces to $0,7\%$ corresponding to the 95% confidence limits.
- The applied damage scenarios cause eigenfrequency shifts up to 7%. Damage has a selective influence on the eigenmodes: especially those eigenmodes are affected where damage occurs at zones with high modal curvatures. This provides another way to make the distinction between environmental and structural changes.
- Only damage scenarios that produce stiffness reductions could be identified. For instance, this was the case for the support settlement. A loss of prestress will only result in a measurable change in eigenfrequencies if it is accompanied by originating cracks.

- Mode shapes, although less accurately determined than eigenfrequencies, can provide useful information about local changes, e.g., of support stiffness.

ACKNOWLEDGMENTS

This research has been carried out in the Brite-EuRam Programme CT96 0277 SIMCES with a financial contribution of the European Commission. Partners in the project were

- K.U. Leuven (Department of Civil Engineering, Afdeling Bouwmechanica),
- Aalborg University (Institut for Bybninbsteknik),
- EMPA (Swiss Federal Laboratories for Materials Testing and Research, Section Concrete Structures),
- LMS (Leuven Measurement and Systems International N.V.; Engineering and Modeling),
- WS Atkins Consultants Ltd (Science and Technology),
- Sineco Spa (Ufficio Promozione e Sviluppo),
- Technische Universität Graz (Structural Concrete Institute).

REFERENCES

[1] De Roeck G. The state-of-the-art of damage detection by vibration monitoring: the simces experience. *Journal of Structural Control* 2003 **10**:127–143.

[2] Krämer C, de Smet CAM, De Roeck G. Z24 bridge damage detection tests. *Proceedings of the IMAC XVII conference*. Kissimmee, FL, 1999; pp. 1023–1029.

[3] Bundesministerium für Verkehr, *Schäden an Brücken und anderen Ingenieurbauwerken: Ursachen und Erkentnisse, Dokumentation 1982*. Verkehrsblatt-Verlag Borgmann: Dortmund, 1982.

[4] Bundesministerium für Verkehr, *Schäden an Brücken und anderen Ingenieurbauwerken: Ursachen und Erkentnisse, Dokumentation 1994*. Verkehrsblatt-Verlag Borgmann: Dortmund, 1994.

[5] Peeters B, Ventura C. Comparative study of modal analysis techniques for bridge dynamic characteristics. *Mechanical Systems and Signal Processing* 2003 **17**(5):965–988.

[6] Parloo E, Guillaume P, Cauberghe B. Maximum likelihood identification of non-stationary operational data. *Journal of Sound and Vibration* 2003 **268**:971–991.

[7] Allen MS, Ginsberg JH. A global, single-input-multi-output (SIMO) implementation of the algorithm of mode isolation and application to analytical and experimental data. *Mechanical Systems and Signal Processing* 2006 **20**:1090–1111.

[8] Reynders E, De Roeck G. Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis. *Mechanical Systems and Signal Processing* 2008 **22**(3): 617–637.

[9] Peeters B, De Roeck G. One-year monitoring of the Z24-bridge: environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics* 2001 **30**:149–171.

[10] Peeters B, De Roeck G. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing* 1999 **13**(6):855–878.

[11] Watson DK, Rajapakse RKND. Seasonal variation in material properties of a flexible pavement. *Canadian Journal of Civil Engineering* 2000 **27**(1):44–54.

[12] Rytter A. *Vibration based Inspection of Civil Engineering Structures*, PhD thesis, Aalborg University, 1993.

[13] Ljung L. *System Identification*. Prentice Hall: Upper Saddle River, NJ, 1999.

[14] Mevel L, Goursat M, Basseville M. Stochastic subspace-based structural identification and damage detection and localisation–application to the Z24 bridge benchmark. *Mechanical Systems and Signal Processing* 2003 **17**(1):143–151.

[15] Abdel Wahab MM, De Roeck G. Damage detection in bridges using modal curvatures: application to a real damage scenario. *Journal of Sound and Vibration* 1999 **226**(2):217–235.

[16] Maeck J, De Roeck G. Damage assessment using vibration analysis on the Z24-bridge. *Mechanical Systems and Signal Processing* 2003 **17**(1):133–142.

[17] Maeck J, De Roeck G. Dynamic bending and torsion stiffness derivation from modal curvatures and torsion rates. *Journal of Sound and Vibration* 1999 **225**(1):153–170.

[18] Kullaa J. Damage detection of the Z24 bridge using control charts. *Mechanical Systems and Signal Processing* 2003 **17**(1):163–170.

[19] Parloo E. *Application of Frequency-Domain System Identification Techniques in the Field of Operational*

Modal Analysis, PhD thesis. Vrije Universiteit Brussel, 2003.

[20] Teughels A, De Roeck G. Structural damage identification of the highway bridge Z24 by FE model updating. *Journal of Sound and Vibration* 2004 **278**(3):589–610.

[21] Teughels A, Maeck J, De Roeck G. Damage assessment by finite element model updating using damage functions. *Computers and Structures* 2002 **80**(25):1869–1879.

Chapter 127

Continuous Monitoring of the Øresund Bridge: Data Acquisition and Operational Modal Analysis

Bart Peeters
LMS International, Leuven, Belgium

1 INTRODUCTION

The number of civil engineering structures that are equipped with monitoring systems is rapidly increasing. Typical examples are long-span cable-stayed and suspension bridges, which represent a large capital investment and where the use of a permanent monitoring system is easily justified and often recommended by insurance companies. Such a monitoring system can serve several purposes [1]:

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

- **Design verification**
It is verified that the structural static and dynamic response does not exceed the design values.

- **Event recording**
Important load (wind, traffic) or response (strains, accelerations) quantities are recorded for archival reasons or to take decisions about the serviceableness of a bridge when preset thresholds are exceeded; for instance, at too large wind speeds, it may be dangerous to use the bridge.

- **Health monitoring**
The recorded data can be used to derive experimental models. Information on the structural health can be obtained by tracking the evolution of these experimental models or by confronting experimental data with analytical models.

Numerous examples of permanently monitored bridges are readily found in the *Proceedings of the International Workshop on Structural Health Monitoring* [2], *Proceedings of the International Operational Modal Analysis Conference* [3], and *Structural Health Monitoring (SHM) related proceedings of SPIE* [4].

This article presents a state-of-the-art monitoring system that was installed on a state-of-the-art bridge. In Section 2, the Øresund Bridge is introduced. Section 3 presents the monitoring system itself and its normal mode of operation. In Section 4, the vibration signals of the cables, deck, and towers are analyzed. These analyses of dynamic data were done off-line, using data captured by the permanent system, but are not part of the standard analysis procedures of the system.

2 THE ØRESUND BRIDGE

Since July 2000, Sweden and Denmark have been connected through the Øresund fixed link consisting of 8 km of bridge and 4 km of tunnel, joined by a 4-km-long artificial island (Figure 1). The bridge has quite a unique two-level design, with a four-lane motorway placed above a two-track railway (Figure 2). The bridge consists of 49 approach spans (7 spans of 120 m, 42 spans of 140 m) and a

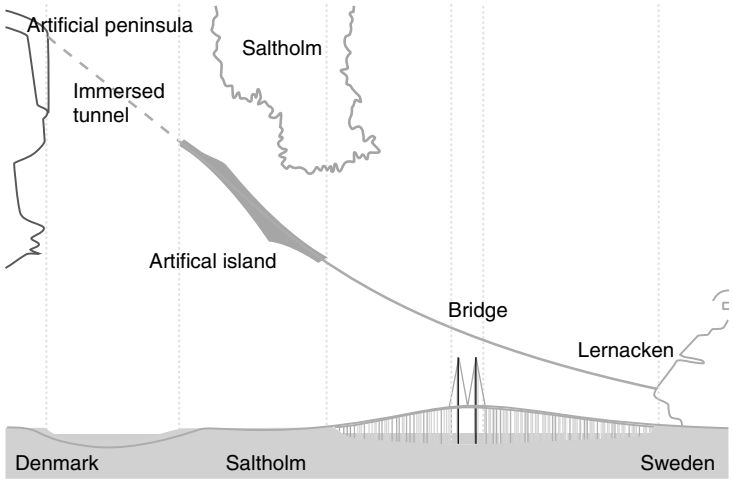


Figure 1. The Øresund fixed link. [Reproduced with permission from Øresundsbro Konsortiet.]

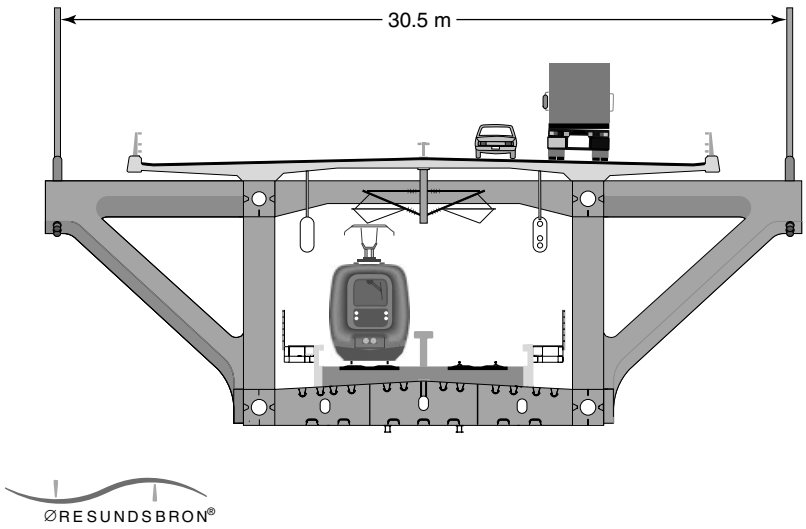


Figure 2. Cross section of the cable-stayed bridge spans. [Reproduced with permission from Øresundsbro Konsortiet.]

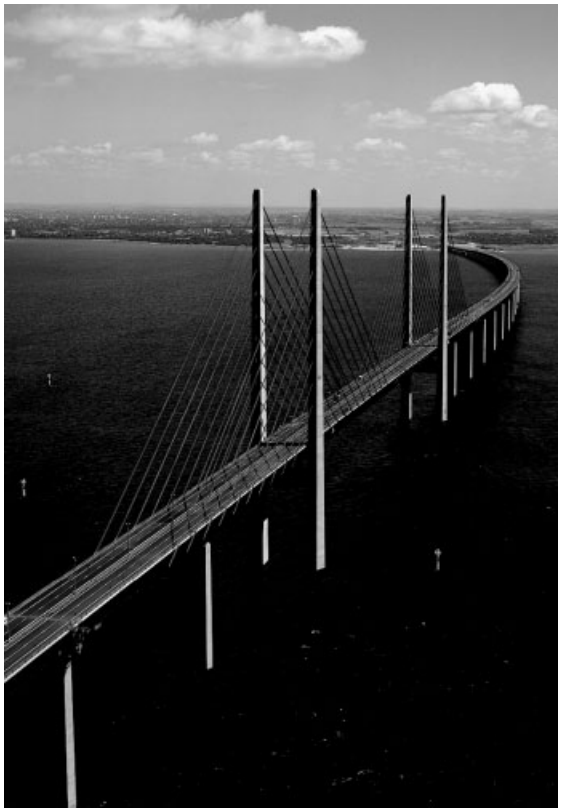


Figure 3. Cable-stayed part of the Øresund Bridge. [Reproduced with permission from Øresundsbro Konsortiet.]

Figure 4. Pylons and stay cables. [Reproduced with permission from Øresundsbro Konsortiet.]

cable-stayed component with two side spans on either side (160 and 141 m) and a main span of 490 m over the navigational channel (Figure 3).

Ten pairs of cables on either side connect the pylons of the two H-shaped towers with the bridge deck (Figure 4). The tops of the pylons are 204 m above sea level and the minimum headroom under the main span is 57 m. The monitoring system discussed in next section is installed on the cable-stayed part of the bridge.

3 THE CONTINUOUS MONITORING SYSTEM

The bridge owner was concerned about the stay-cable oscillations under heavy wind conditions, as well as the deformation of the bridge when trains or heavy trucks pass over it. Therefore, GeoSIG installed a new type of monitoring system (called the *CR-4 Central Recorder*) that could acquire both dynamic and static data.

Eighty-five dynamic channels have been installed at the Øresund Bridge; these channels permanently record at a sample rate of 100 Hz signals from 22 triaxial accelerometers and 19 strain gauges. The static-measurement channels are connected to 12 temperature sensors at different points of the bridge and to two weather stations, one at the top of a pylon, and the other, at road level. Static information, such as minimum, maximum, and mean values, is extracted from the dynamic channels for long-term analysis. The CR-4 acquisition system is placed in a technical room at one of the pylons. There is a telephone connection to the general control room of the Øresund link, 3 km away from the pylon, allowing automatic data retrieval.

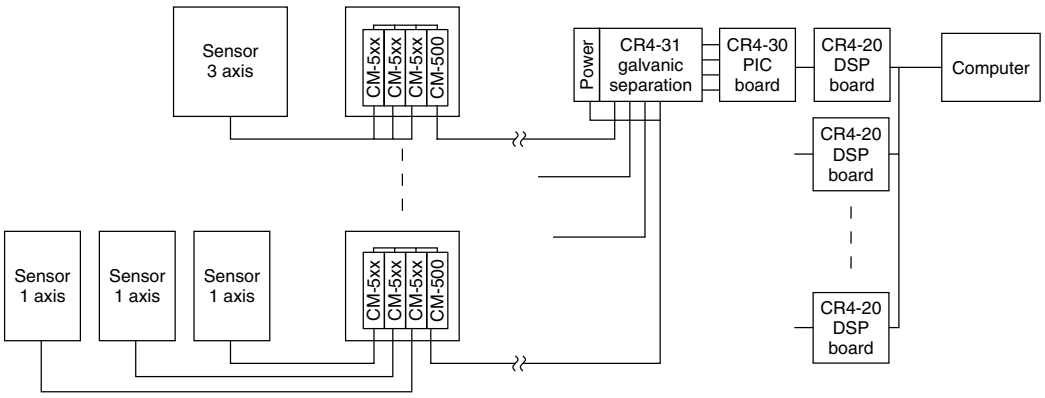


Figure 5. Dynamic data-acquisition architecture.

3.1 Data-acquisition system

The CR-4 system allows different data collection strategies:

- Dynamic acquisition with analog-to-digital (A/D) conversion in the recorder handles up to 33 dynamic channels, with a maximum sample rate of 1000 Hz.
- Dynamic acquisition with A/D conversion near the sensor handles up to 132 dynamic channels, with a maximum sample rate of 200 Hz.
- Static acquisition with A/D conversion near the sensor handles up to 44 chains, each with up to 26 static junction boxes (see below), each of which has up to six channels. So as many as 6864 static channels are possible.

For the Øresund Bridge, a combination of the second and third options has been chosen. The acquisition near the sensor is done by means of a junction box with acquisition modules. Modules differ as a function of the type of sensor that is connected. On the bridge, modules have been used for strain gauges, for voltage inputs like accelerometer and weather station signals, for powering the sensors and generating the sensor test pulse output, and for Pt-100 temperature sensors. Not used in this project, but also available, is a module for linear variable displacement transducers (LVDTs). The junction box is connected to the main cabinet of the CR-4. Power can be supplied locally or, in this case, from the cabinet. The signals from the junction box are digitally transferred to the cabinet through an RS-485 link to avoid any loss of power.

The block diagram of Figure 5 represents the architecture of the dynamic measurement chain: from the sensors on the left to the computer on the right. The analog signal goes from the sensor to the CM-5xx acquisition module where it is converted on request to a digital signal and transferred to the CM-500 module. The CM-500 module makes a package from the collected samples and sends it to the CR4-30 module through the CR4-31 protection board. The CR4-31 is a galvanic separation of the CR-4 and the computer from the external part of the system. The CR4-30 contains four PIC boards, each PIC controlling one CM-500 in dynamic mode and up to 26 chained ones in static mode. The data package of each channel is stored in the PIC and is sent on request to the CR4-20 DSP board where the data is stored in ring buffers. Finally, every second, the computer (CR-4 software) collects 1-s packages of the DSP boards. The computer does signal analysis and treatment, and according to the user-specified trigger mode, the signals are stored in an event file.

The static architecture is different in that several junction boxes can be connected to the same wire. This is not possible in the dynamic mode because of the high baud rate that must be guaranteed. To avoid unnecessary power loss and voltage drops in the powering cables, the CM-500 only powers the sensor if a sample is requested. A start-up routine assures constant values.

A 12-V/115-Ah battery assures power for 15 h in case of power loss. This is sufficient for local conditions, as power failures do not often occur and, if they occur, it would only be for a short period. For

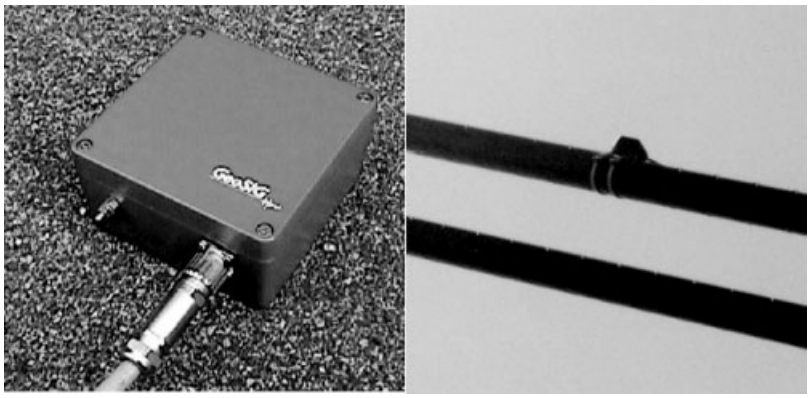


Figure 6. GeoSIG triaxial accelerometer AC-53 mounted on a cable.

accurate timing, a GPS receiver can be connected to the CR-4 that updates the computer time.

3.2 Sensors

The Øresund installation contains four types of sensors. They are specified below together with a description of the type of information that is extracted from the quantities that they measure.

- There are 22 triaxial AC-53, 2*g* full-scale force balance accelerometers (Figure 6). Most of these accelerometers (16) are mounted on the stay cables to measure the cable vibrations. The tops of the two east pylons, as well as four locations along the deck, are also equipped with accelerometers. These sensors allow monitoring the cable vibrations under heavy wind load and the bridge response to railway and road traffic.
- Twelve of the 19 strain gauges LV3400VS0 are mounted on three steel outriggers of the cables, one on either side. Two are mounted on the rail level in the concrete and five are mounted on the lower side of the bridge. These sensors mainly observe torsions due to heavy wind and railway traffic.
- Twelve thermometers Pt-100 are mounted at different locations, most of them at the pylons. These sensors measure temperatures, which are correlated with the strain gauge measurements.
- Of the two weather stations measuring wind speed, wind direction, air humidity, and air temperature, one is mounted on the top of a pylon and the other at road level (Figure 7). The wind measurements serve as a reference for the stay-cable vibrations. The air humidity

Figure 7. Weather station at the deck level of the bridge.

and temperature complete the meteorological information.

4 DATA ANALYSIS

4.1 Standard data analysis procedures

In the daily use of the system, both dynamic and static data are continuously acquired. The dynamic data acquisition is governed by triggering: if the signal of selected data channels exceeds some specified level, the data are logged to an event file, which can be viewed and analyzed later. At the same time, an alarm signal is sent to the traffic control center alerting responsible persons about strong vibrations of the bridge. Static data are logged to another file periodically. There is also a monitoring mode, which indicates the operating status of all dynamic data channels, status of the trigger, and other parameters. The waveform signal of any data channel can be viewed in near real-time mode.

4.2 Detailed dynamic data analysis

In this section, a 5-min recording of the acceleration channels of the CR-4 monitoring system is thoroughly analyzed. Although the type of analysis performed here (operational modal analysis, OMA) is not part of the standard procedures of the Øresund Bridge, it is instrumental in seeing what kind of information can be extracted from the cable, deck, and tower vibrations.

4.2.1 Operational modal analysis

The aim is to identify an experimental dynamic model of the bridge. In laboratory situations, such a model can be obtained by artificially exciting a structure and measuring the responses. Measurement functions (so-called frequency response functions) that relate the input to the output serve as input for *experimental modal analysis* (EMA) methods to achieve this goal. Obviously, the data recorded by the monitoring system are the so-called operational data: bridge responses are measured under dynamic wind or traffic loading without being able to measure all these forces exactly. Nevertheless, it is still possible to derive an experimental dynamic model of the structure from response measurements alone. Hereto, a technique called *operational modal analysis*, OMA, is used. The use of vibration measurements and modal analysis for structural health monitoring (SHM) is also discussed in **Modal–Vibration-based Damage Identification**; **Ambient Vibration Monitoring**.

In the past, OMA in civil engineering was mainly restricted to *peak picking*. The method is named after its key step: the identification of the eigenfrequencies as the peaks of power spectrum plots. However, there exist more advanced methods that better exploit the data and lead to higher-quality models. These methods became accessible through user-friendly commercial software implementations. *Stochastic subspace identification* is, for instance, one such method. In this method, a so-called stochastic state-space model is identified from output correlations [5] or directly from measured output data [6]. The "outputs" in our case are the measured bridge acceleration responses. The first application of stochastic subspace identification to bridge vibration data dates from 1995 [7]. The potential to use stochastic subspace identification for modal analysis applications in general has been described in [8–10]. It is such a stochastic subspace identification method that is used in this article.

Recently, a new frequency-domain method, called *PolyMAX* [11, 12], received considerable attention. Originally developed for use with classical input–output data, it was extended so that it can also cope with output-only data [13, 14]. The main added value of PolyMAX is that it yields very clear stabilization diagrams (Section 4.2.2) so that it has the potential to run autonomously [15]. This is very relevant if the data is almost continuously streaming, as is the case for permanent monitoring systems. PolyMAX is also used in this article.

4.2.2 Cable vibrations

The simplest and least expensive method to measure the cable forces is the measurement of the eigenfrequencies of that cable. Evidently, this is an indirect measurement: the tension force is derived from eigenfrequencies, which are derived from accelerations. Figures 8 and 9 show 5-min recordings of typical accelerations of two-stay cables. The parts (a) of the

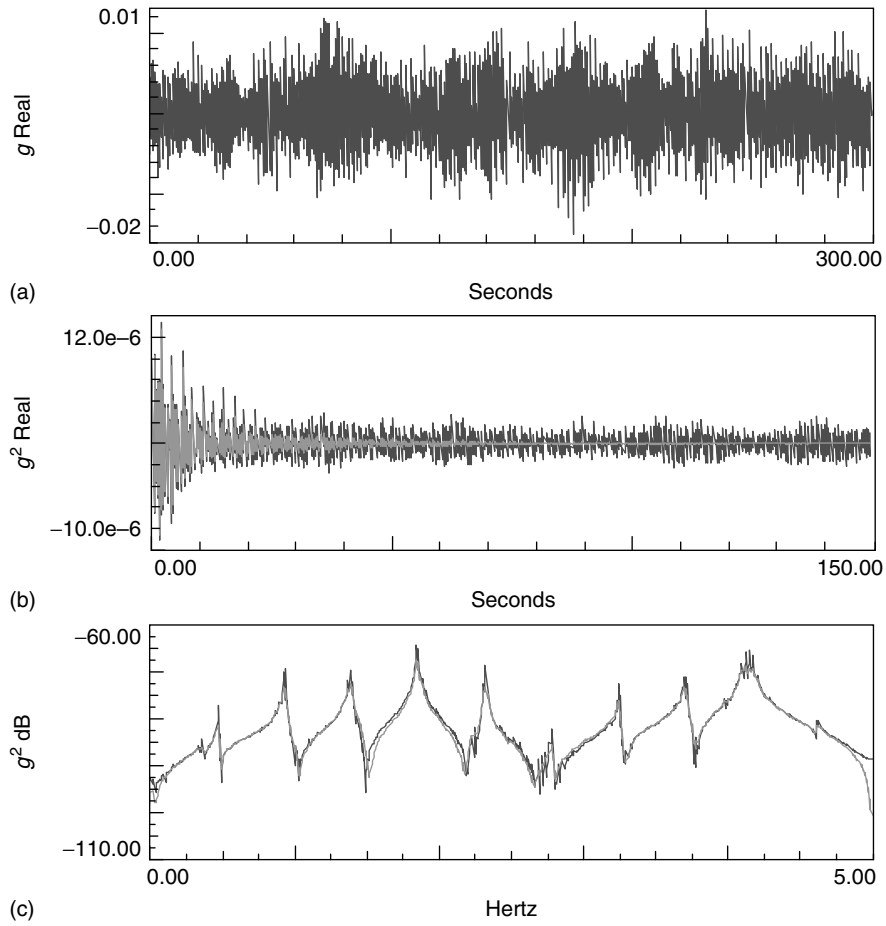


Figure 8. Out-of-plane acceleration of the longest cable from the East–South pylon to mid-span. (a) Time history, (b) output correlations, and (c) half-spectrum magnitude. The effect of the exponential window on the correlations and spectra is clearly visible: no window (black)—exponential window of 1% (gray).

figures show the time histories. The autocorrelations R_i, shown in (b), are estimated from the time data y_k according to

$$R_i = \frac{1}{N} \sum_{k=0}^{N-1} y_{k+i} y_k \qquad (1)$$

where N is the number of time samples used to compute the correlations and k, i both are sample indices. From the plots, it is clear that the correlations have an impulse–response-like behavior. This can also be theoretically proven provided that the structure is excited by white noise. The autocorrelations of the cables are subsequently used as basic functions in stochastic subspace identification to extract the eigenfrequencies. Figures 8(c) and 9(c) show the magnitudes of the discrete Fourier transform of the positive correlation lags. These are estimates of the so-called half spectra, which are the basic functions used in the operational variant of PolyMAX. As in any parametric estimation method, both subspace identification and PolyMAX enable the use of so-called stabilization diagrams that allow for an easy and objective selection of the cable eigenfrequencies. An example of such a diagram is shown in Figure 10.

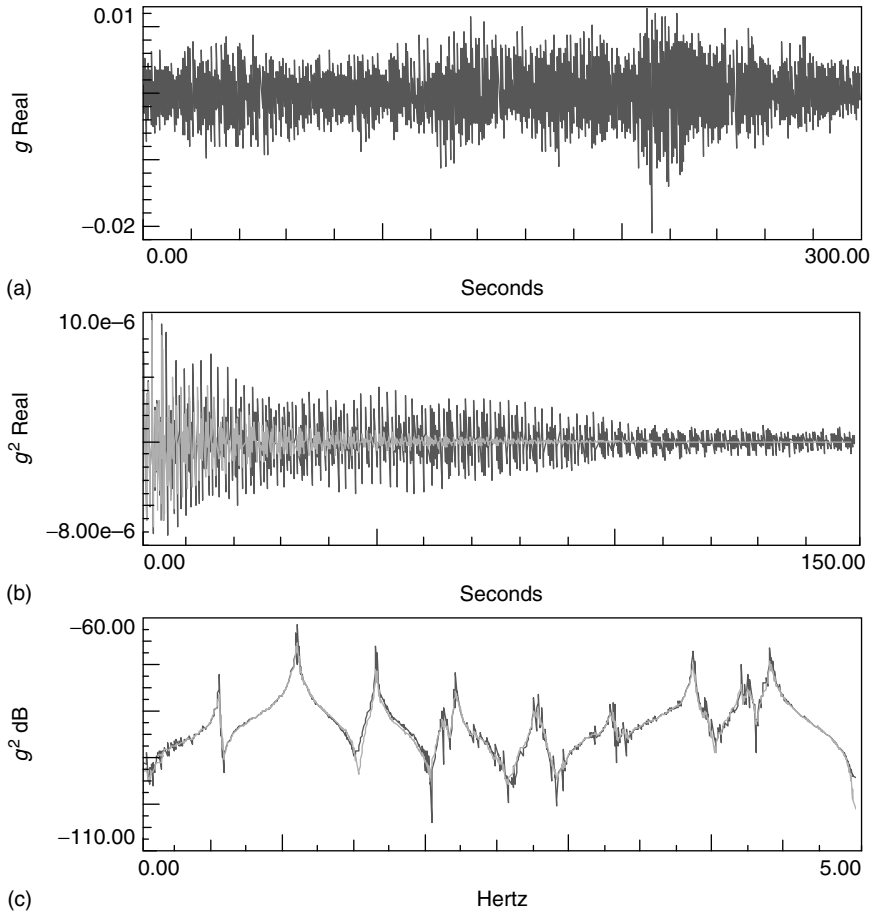


Figure 9. Out-of-plane acceleration of the third longest cable from East–North pylon to side span. (a) Time history, (b) output correlations, and (c) half-spectrum magnitude. The effect of the exponential window on the correlations and spectra is clearly visible: no window (black)—exponential window of 1% (gray).

Finally, Figure 11 compares a measured spectrum with a spectrum that is synthesized from the estimated modal parameters. The good correspondence indicates that all modes were extracted from the data.

By repeating this analysis a number of times, eigenfrequencies are extracted from the accelerations of 10 cables—the 5 longest cables connecting the East–South (ES) pylon to the main span and the 5 longest cables connecting the East–North (EN) pylon to the side span. These frequencies are presented in Table 1. From Figures 8 and 9 and Table 1, the set of cable frequencies seems to be composed of a fundamental frequency f_1 and its higher harmonics $f_n = n\, f_1$. A stay cable is assumed to satisfy the taut string theory with the following relation between frequencies and cable tension forces:

$$f_n^{\mathrm{S}} = n\,\frac{1}{2L}\sqrt{\frac{H}{m}} \tag{2}$$

where f_n^{S} (Hz) is the nth harmonic; L (m) is the cable length; H (N) is the cable force; and m (kg m^{-1}) is the cable mass per unit length. In this article, an estimate of the fundamental frequency is obtained from all harmonics by applying least squares (LS). Following this, the cable forces are

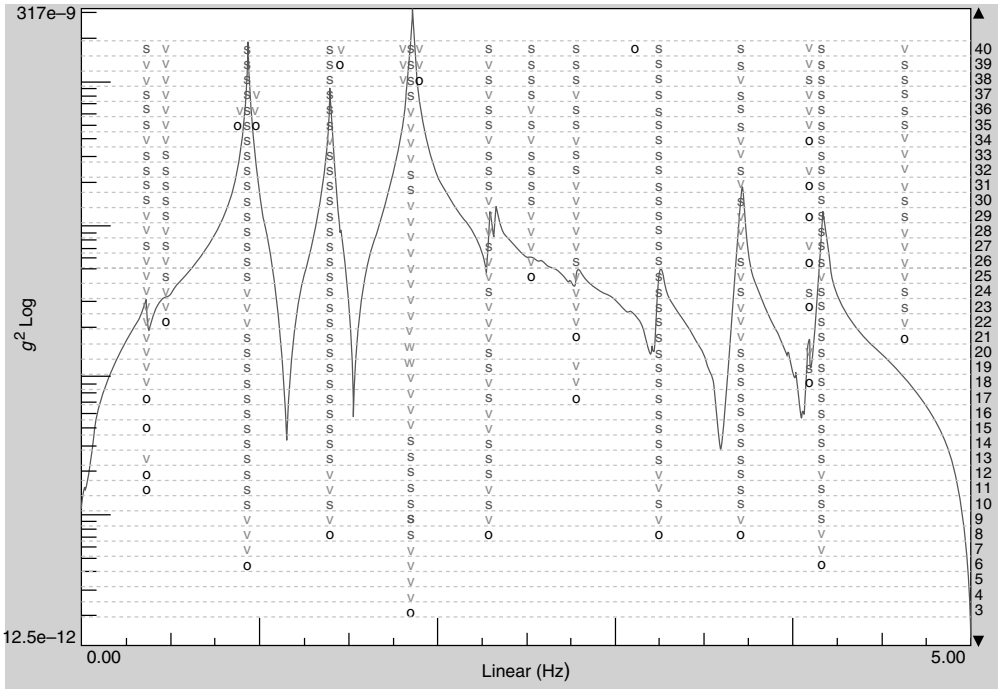


Figure 10. Stabilization diagram obtained by applying operational PolyMAX to the acceleration data of Figure 9.

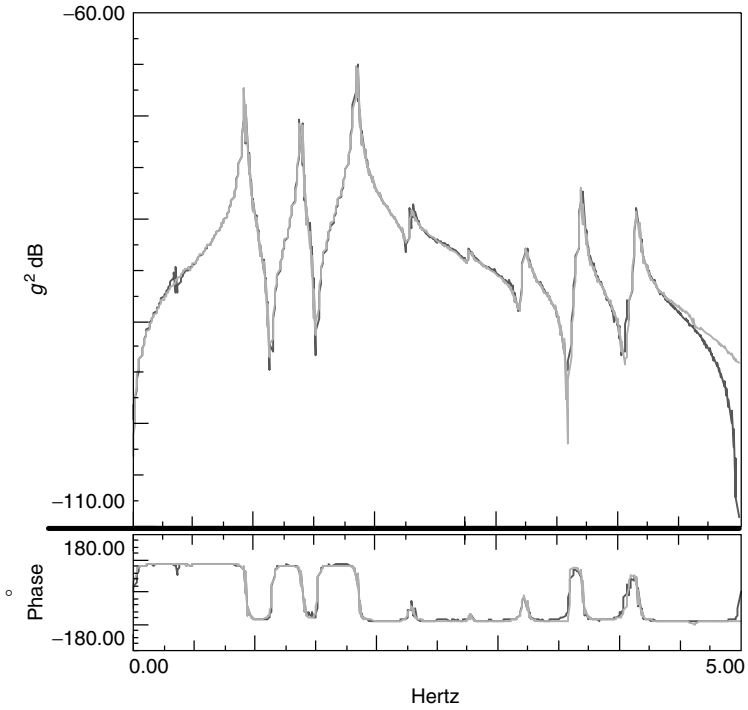


Figure 11. Measured (black) versus synthesized PolyMAX spectra (gray).

Table 1. Determination of cable tension forces from out-of-plane acceleration measurements. Not for all cables it was possible to extract the same number of harmonics

Pylon cable number		ES 1	ES 2	ES 3	ES 4	ES 5	EN 5	EN 4	EN 3	EN 2	EN 1
L (m)		262	239	216	192	169	169	192	216	239	262
m (kg m^{-1})		91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2
Cable eigenfrequencies (Hz)	f_1	0.473	0.508	0.563	0.631	0.728	0.743	0.638	0.560	0.504	0.457
	f_2	0.929	0.995	1.115	1.248	1.445	1.471	1.263	1.106	0.985	0.894
	f_3	1.386	1.498	1.677	1.870	2.166	2.198	1.894	1.662	1.480	1.350
	f_4	1.851	1.990	2.201	2.496	2.878	2.928	2.528	2.218	1.966	1.789
	f_5	2.323	2.493	2.772	—	3.583	—	—	2.784	2.499	2.261
	f_6	2.779	2.991	3.325	3.741	4.302	4.386	3.792	3.307	—	2.712
	f_7	3.242	3.494	3.930	4.351	5.005	—	4.394	3.868	—	3.126
	f_8	3.706	3.974	4.457	4.956	5.727	—	4.999	4.414	3.919	3.586
	f_9	4.171	4.466	4.962	5.547	—	—	5.621	4.960	4.408	4.069
	f_{10}	4.621	4.973	—	—	—	—	—	—	4.911	—
Least squares estimate of f_1^S (Hz)		0.463	0.498	0.555	0.620	0.717	0.732	0.627	0.552	0.491	0.450
Cable force H (kN) based on LS frequency		5368	5157	5248	5169	5352	5581	5288	5194	5030	5069
Dimensionless parameter ε		309	315	283	331	284	208	316	293	312	208
Equivalent taut string fundamental frequency f_1^S (Hz)		0.458	0.493	0.550	0.615	0.710	0.723	0.621	0.547	0.486	0.443
Cable force H (kN) (with bending stiffness)		5246	5059	5140	5102	5249	5444	5207	5091	4930	4911
Bending stiffness EI (kN m^2)		3781	2920	3000	1727	1855	3596	1938	2770	2896	7786
Cable force difference (%)		2.3	1.9	2.1	1.3	2.0	2.5	1.6	2.0	2.0	3.2

computed according to equation (2) and are as shown in Table 1.

However, the taut string theory does not hold exactly for stay cables, which, inevitably, have a certain bending stiffness, which becomes more important as the cables become shorter. In [16], the following equation has been derived to compute the out-of-plane eigenfrequencies f_n^{EI} of a stay cable:

$$\frac{f_n^{\mathrm{EI}}}{f_n^{\mathrm{S}}} = 1 + \frac{2}{\varepsilon} + \frac{4 + n^2\pi^2/2}{\varepsilon^2} \tag{3}$$

where f_n^{S} are the frequencies if the bending stiffness is zero (the taut string frequencies as given by equation (2)) and ε is a dimensionless parameter related to the bending stiffness EI (Nm^2):

$$\varepsilon = L\sqrt{\frac{H}{\mathrm{EI}}} \tag{4}$$

In [17], it is shown how cable forces and bending stiffness can be estimated from the identified eigenfrequencies by the following procedure:

1. Apply nonlinear LS to estimate f_1^{S} and ε from the measured frequencies f_n, which are assumed to behave like f_n^{EI} in equation (3).
2. From the cable length L and mass per unit length m, and f_1^{S} and ε estimated in the previous step, it is straightforward to estimate the cable force H from equation (2) and the bending stiffness EI from equation (4).

The estimated values are shown in Table 1. One cable consists of 70 tendons, each consisting of seven wires of 5-mm diameter. The mass per unit length is estimated at $91.2\,\mathrm{kg\,m^{-1}}$. From the positions of the anchor points on the tower and the bridge, the lengths of the cables are calculated. As seen from Table 1, the dimensionless bending stiffness parameter ε is about 200–300. This is within the limits for normal stay cables: $70 < \varepsilon < 600$. The frequency f_1^{S}, and so the cable force, is determined more accurately than the dimensionless parameter ε, and so the bending stiffness. The resulting bending stiffnesses have a comparable magnitude, except for cable 1 of the EN pylon (last column of Table 1).

Figure 12 compares the two cable force estimates of the different cables: one estimate according to

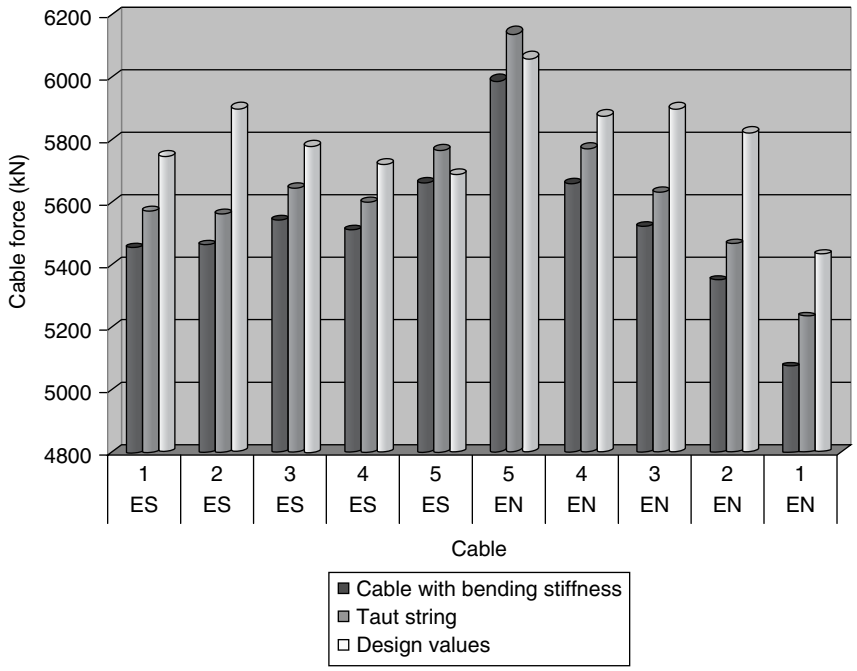


Figure 12. Cable forces estimated from vibration measurements.

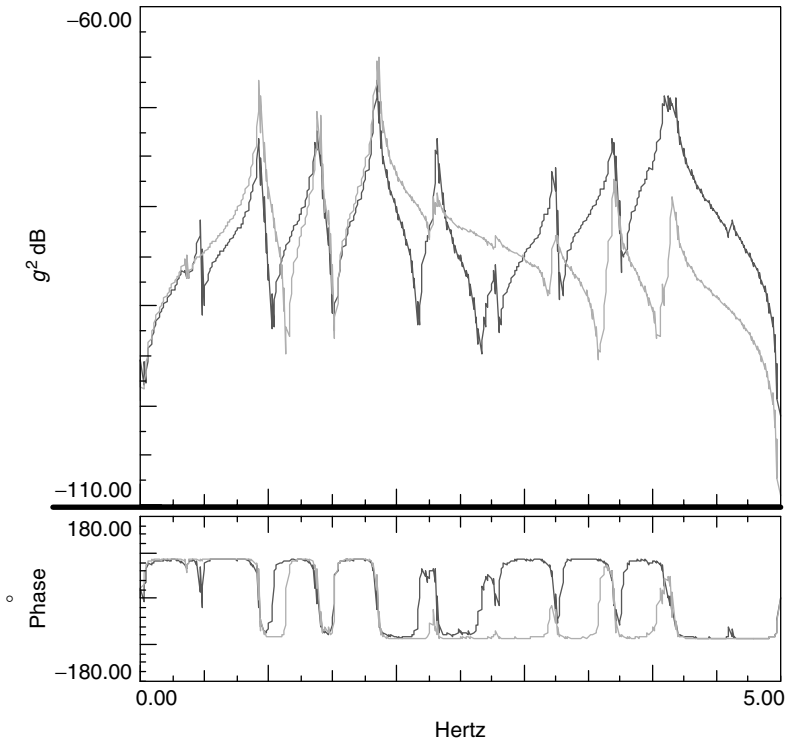


Figure 13. Out-of-plane acceleration of the longest cable from the East–South pylon to midspan. Comparison of two datasets with an interval of more than 1 year.

the taut string theory using an LS estimate of the fundamental frequency; the other taking the bending stiffness into account. The difference between the estimates of the cable forces using the two methods ranges from 1.3 to 3.2%. The tension force differences between cables can go up to 11%, regardless of the estimation method. Also shown on this graph are the design values for the cable forces.

Finally, Figure 13 compares the spectra from the same cable, but using data measured with an interval of more than 1 year (August 2002 vs November 2003). Owing to different excitation conditions (wind speed, direction, distribution), the spectra are not equal, but the location of resonance peaks is still about the same. A more detailed (modal) analysis revealed that the cable frequency differences were in the order of 0.4–0.8%.

The analysis presented in this section shows that it is possible to monitor the cable forces from the accelerations that are recorded by the continuous monitoring system.

4.2.3 Deck and tower vibrations

Figures 14 and 15 show 5-min recordings of a vertical acceleration at the main span deck and a transversal acceleration at the top of the ES pylon. Again, the time histories, the autocorrelations, and the half spectra are shown. The recordings contain the passage of a train, which caused an increase in vibration levels as can be seen from Figure 14. From the pylon accelerations (Figure 15), it becomes difficult to observe the train passage because the pylon vibrations are mainly caused by the wind. The cable vibrations (Figures 8 and 9) are apparently not influenced by the load because of the train. Stochastic subspace identification was applied to determine the deck and tower modal parameters. They are represented in Table 2. As background information and as an assurance that the measured eigenfrequency is plausible, it is interesting to compare these results with the ones obtained from other bridge tests as reported in the literature. For example, in Figure 16 such comparison is made

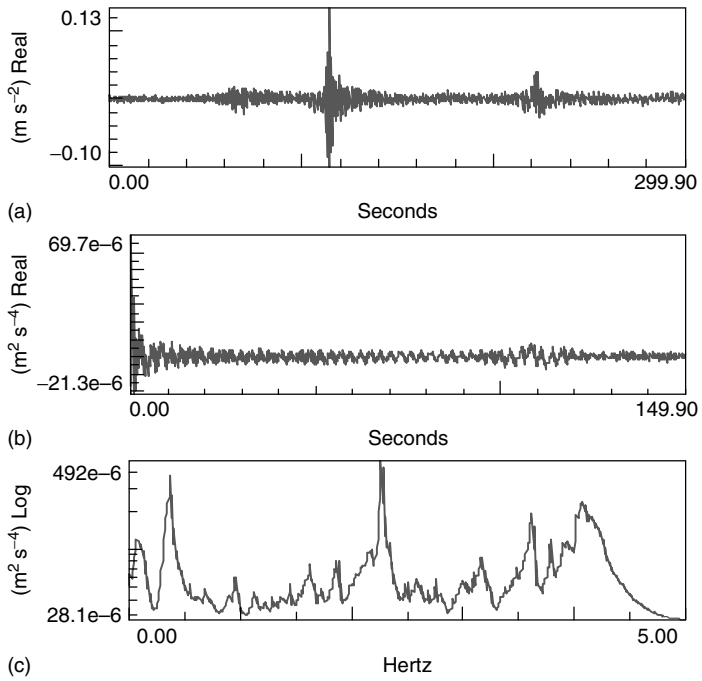


Figure 14. Vertical acceleration at the main span deck. (a) Time history, (b) output correlations, and (c) half-spectrum magnitude.

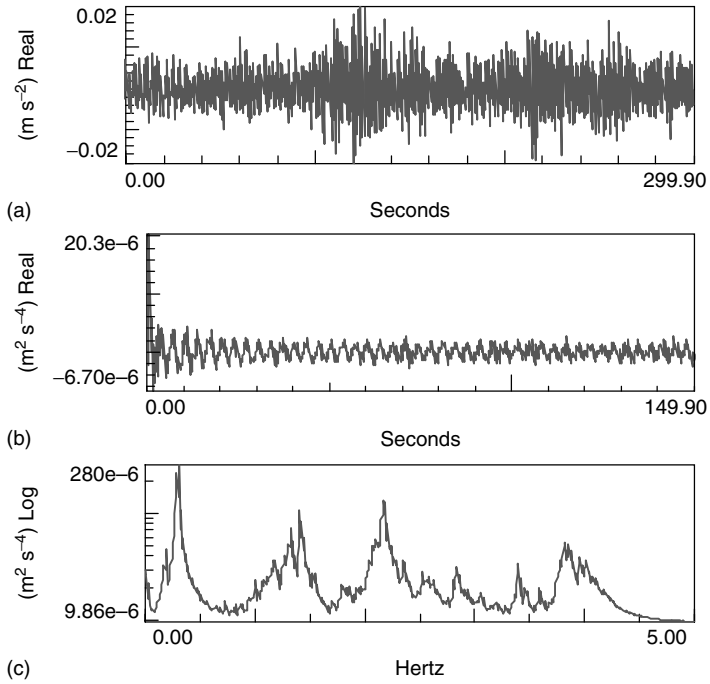


Figure 15. Transversal acceleration at the top of the East−South pylon. (a) Time history, (b) output correlations, and (c) half-spectrum magnitude.

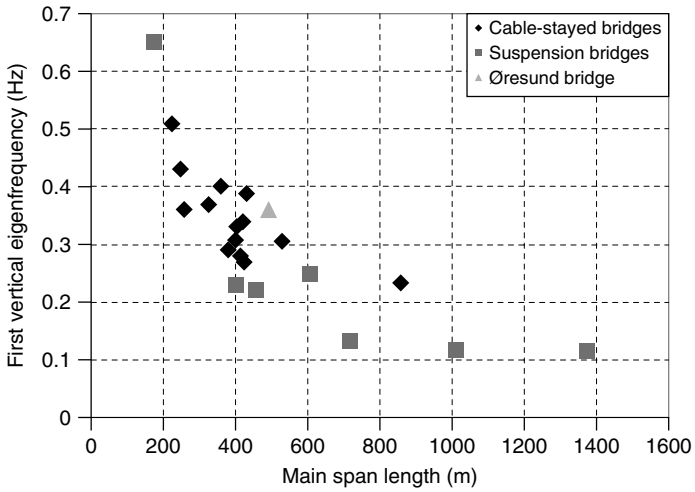


Figure 16. First vertical deck bending eigenfrequency as a function of main span length for cable-stayed and suspension bridges.

Table 2. Deck and tower modal parameters

Description	Frequency (Hz)	Damping ratios (%)
Deck plus tower transversal	0.252	0.74
Tower transversal, pylons in phase	0.294	1.83
Tower transversal, pylons out of phase	0.300	0.82
Deck first vertical bending plus tower longitudinal	0.368	0.65
Tower longitudinal, pylons in phase	0.540	0.40

on the basis of a plot of the first vertical deck bending eigenfrequency as a function of the main span length.

In the present case of the Øresund Bridge, no detailed experimental mode-shape information could be obtained because of the limited amount of accelerometers at the deck and towers. However, OMA methods such as the ones described in Section 4.2.1 are also able to generate accurate mode shapes that can be correlated with numerically predicted mode shapes using, for instance, finite element models. In [18], it is, for instance, demonstrated that detailed mode shapes of a cable-stayed bridge (in this case, the Guadiana Bridge) can be obtained by using ambient vibration data and OMA. In this case, the accelerations at many more deck and tower locations had been measured with nonpermanent equipment. Some of the Guadiana Bridge mode shapes, identified with the PolyMAX method, are represented in Figure 17.

In a continuous monitoring and modal analysis process, the bridge eigenfrequencies could be used to assess the health of the structure. This is, for instance, shown in **Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge** for the Z24-Bridge in Switzerland. The robustness of vibration-based SHM methods is significantly enhanced by measuring and modeling environmental influences such as temperature on the dynamic properties of the structure (*see* **The Influence of Environmental Factors**) [19], and by using appropriate statistical techniques to interpret the data correctly (*see* **Model-based Statistical Signal Processing for Change and Damage Detection**).

5 CONCLUSIONS

In this article, the continuous monitoring system of the Øresund Bridge was presented. It was also shown how OMA, applied to the dynamic data captured by the system, provides useful information about the health of the bridge. Measuring and analyzing cable vibrations allow monitoring the cable tension.

Figure 17. Guadiana Bridge mode shapes extracted from ambient vibration data using PolyMAX.

Measuring and analyzing deck and tower vibrations allow assessing the health of these structural parts. The analyses of dynamic data presented in this article were performed off-line, using data captured by the permanent system, but which are not part of the standard analysis procedures of the system. The main purpose of the article is to demonstrate the possibilities that advanced data processing techniques offer for SHM.

ACKNOWLEDGMENTS

The authors would like to thank Mr Svensson from Øresundsbro Konsortiet (www.oeresundsbron.com) for granting them the right to use monitoring data and photographs of the Øresund Bridge.

REFERENCES

[1] Cunha Á, Caetano E, Calçada R, De Roeck G, Peeters B. Dynamic measurements on bridges: design, rehabilitation and monitoring. *Bridge Engineering* 2003 **BE156**(3):135–148.

[2] *Proceedings of the 5th International Workshop on Structural Health Monitoring*. Stanford University, Stanford, CA, 2005.

[3] *Proceedings of IOMAC 2007 the 2nd International Operational Modal Analysis Conference*. Copenhagen, 1–2 May 2007.

[4] *Proceedings of SPIE, Health Monitoring and Management of Civil Infrastructure Systems*. Newport Beach, CA, 2001; Vol. 4337.

[5] Akaike H. Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control* 1974 **19**:667–674.

[6] Van Overschee P, De Moor B. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers: Dordrecht, 1996.

[7] Peeters B, De Roeck G, Pollet T, Schueremans L. Stochastic subspace techniques applied to parameter identification of civil engineering structures. *Proceedings of the International Conference MV2 on New Advances in Modal Synthesis of Large Structures, Non-Linear, Damped and Non-Deterministic Cases*. Lyon, September 1995; pp. 151–162.

[8] Benveniste A, Fuchs J-J. Single sample modal identification of a nonstationary stochastic process. *IEEE Transactions on Automatic Control* 1985 **AC-30**(1):66–74.

[9] Hermans L, Van der Auweraer H. Modal testing and analysis of structures under operational conditions: industrial applications. *Mechanical Systems and Signal Processing* 1999 **13**(2):193–216.

[10] Peeters B, De Roeck G. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing* 1999 **13**(6):855–878.

[11] Guillaume P, Verboven P, Vanlanduit S, Van der Auweraer H, Peeters B. A poly-reference implementation of the least-squares complex frequency-domain estimator. *Proceedings of IMAC 21 the International Modal Analysis Conference*. Kissimmee, FL, 2003.

[12] Peeters B, Van der Auweraer H, Guillaume P, Leuridan J. The PolyMAX frequency-domain method: a new standard for modal parameter estimation? *Shock and Vibration* 2004 **11**:395–409. Special issue dedicated to Professor Bruno Piombo.

[13] Peeters B, Van der Auweraer H. PolyMAX: a revolution in operational modal analysis. *Proceedings of IOMAC 2005 the 1st International Operational Modal Analysis Conference*. Copenhagen, 26–27 April 2005.

[14] Peeters B, Van der Auweraer H, Vanhollebeke F, Guillaume P. Operational modal analysis for estimating the dynamic properties of a stadium structure during a football game. *Shock and Vibration* 2007 **14**(4):283–303, Special issue: assembly structures under crowd-dynamic excitation.

[15] Van der Auweraer H, Peeters B. Discriminating physical poles from mathematical poles in high order systems: use and automation of the stabilization diagram. *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*. Como, 18–20 May 2004.

[16] Mehrabi AB, Tabatabai H. Unified finite difference formulation for free vibration of cables. *ASCE Journal of Structural Engineering* 1998 **124**(11):1313–1322.

[17] Peeters B, Couvreur G, Razinkov O, Kündig C, Van der Auweraer H, De Roeck G. Continuous monitoring of the Øresund bridge: system and data analysis. *Proceedings of IMAC 21 the International Modal Analysis Conference*. Kissimmee, FL, February 2003.

[18] Magalhães F, Caetano E, Cunha Á. Assessment of dynamic properties of Guadiana cable-stayed bridge based on different output-only identification techniques. *Proceedings of EVACES 2005 the International Conference on Experimental Vibration Analysis for Civil Engineering Structures*. Bordeaux, October 2005.

[19] Peeters B, De Roeck G. One-year monitoring of the Z24-bridge: environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics* 2001 **30**(2):149–171.

Chapter 128

Condition Compensation in Frequency Analyses—a Basis for Damage Detection

Robert Veit-Egerer

VCE – Vienna Consulting Engineers, Vienna, Austria

1 INTRODUCTION

Recent publications have raised doubts as to whether damage in structures can be detected by the application of frequency analyses. In fact, very often, temperature changes show larger reactions in spectra than any smaller damage. Beside temperature, other environmental influences such as radiation from sunshine create changes in the structural systems that have to be considered.

This article demonstrates the capabilities of new compensation methods in frequency analyses. When the environmental conditions are monitored together with structural response, a proper reaction can be predicted. The described compensation process, in general, deals with the following sources of input:

- temperature (daily and annual cycles)
- compensation of live load (moving vehicles, etc.)
- influence of wind loads
- bearing friction
- restoring forces
- change of boundary conditions
- impact energy
- instrumentation.

After elimination of all operational and environmental factors, stable frequencies are achieved.

$$\boldsymbol{f}_{\text{total}} = \boldsymbol{f}_0 + \sum_{i=\text{factors}} \boldsymbol{f}_i \tag{1}$$

where $\boldsymbol{f}_{\text{total}}$ is the vector of total (measuredlike) frequencies, $\boldsymbol{f}_0$ the vector of structural (ownlike) frequencies, $\boldsymbol{f}_i$ the vector of influenced (measured) frequencies, and i: corresponds to the considered operational and environmental factors.

Any deviation therefrom can be interpreted as damage or extraordinary event. This procedure opens new possibilities for structural management and lifetime prediction. This article describes investigation of two sources (temperature and live load) that are assumed to be the major ones.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 COMPENSATION OF TEMPERATURE

2.1 Introduction

The following analysis is based on investigations on the Europabrücke—a well-known Austrian steel bridge near Innsbruck, opened in 1963—which is part of one of the main alpine north–south routes for urban and freight traffic (Figure 1).

A long-term preoccupation of VCE with BRIMOS® (*Bri*dge *Mo*nitoring *S*ystem) on the Europabrücke (since 1997) led to the installation of a permanent monitoring system in 2003 [1].

It consists of 24 channels for measuring the accelerations of the main span, the pier, and the cantilever, the dilatation of the abutment, the wind speed and direction, and the temperatures at several locations (Figure 2). The bridge's reference sensor (3-D-forced balance accelerometer) is installed within the main span, at a distance of 0.4 times of the span's length from pier II. At this base point, global stiffness and its dependence to several environmental influences are assessed (sampling rate = 100 Hz, file length = 330 s).

By evaluating the results (frequency spectra) of several measurements, telescoping them together, and viewing them from above (so-called trend cards), the visuals obtained are as shown in Figure 3. These exemplarily show the main span's relevant vertical stiffness patterns of a particular day with a distinctive progression of temperature.

For the sake of completeness, the corresponding frequency spectra themselves are shown in Figure 4, again over a period of this day. An individual procedure has been developed in [2], which contains some measurement preconditioning (offset elimination and bandpass filtering). To enable a more stabilized, automatically performed peak, picking in different ranges of frequency, the response spectra are smoothed in the course of frequency assessment.

The permanent monitoring system exhibits the remarkable loading impact, as the bridge is currently stressed by more than 30 000 motor vehicles per day (approximately 20% of them are freight traffic). By applying the previously described method to the reference sensor's measurement data for the whole day, a progression of stiffness, which consist of 281 single peaks, is obtained (Figure 5a) and represents randomly occurring ambient and forced vibration conditions (scatter).

The complementary relation between stiffness and the air temperature itself (registered at the bridge's base point directly above pier II) is obvious and can be interpreted as a long sinusoidal wave of the main span in the vertical direction. In the course of the described procedure, it should be considered whether to omit the described bandpass filtering and replace it by a further optimized smoothing of the frequency spectra for reasons of stabilizing the accuracy of peak picking.

2.2 Stiffness versus temperature

To describe the verified phenomenon mechanically, attention has to be focused on temperature dependence of the roadbed's asphalt layer, as the change of steel characteristics under varying climate conditions

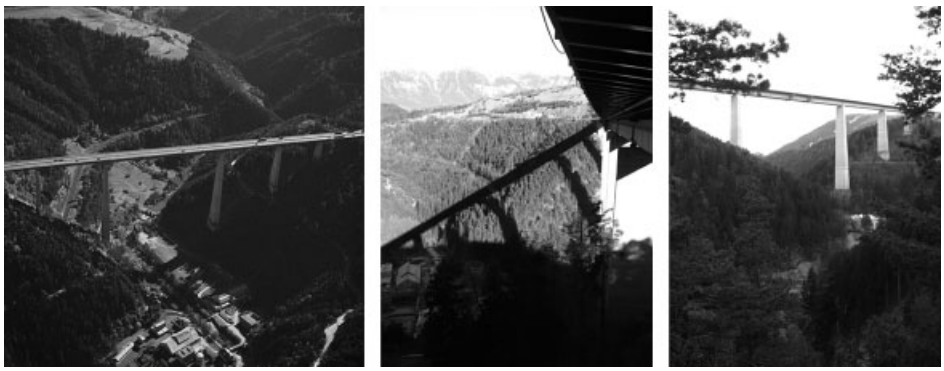


Figure 1. Europabrücke—overview.

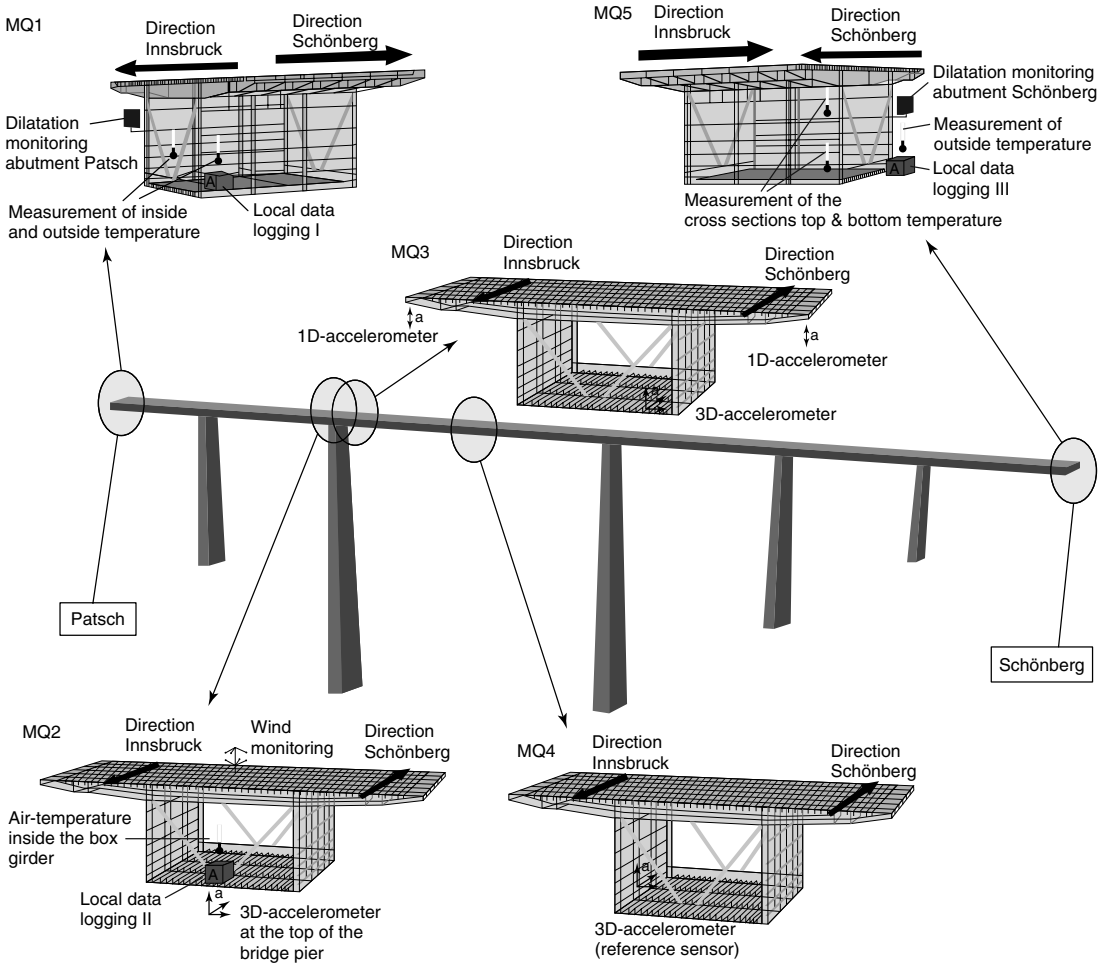


Figure 2. The permanent monitoring system and its several measurement sections.

is negligible. In the first step, the characteristic relationship of the temperature dependence of dynamic Young's modulus is used [3].

Owing to this relation, a temperature-sensitive asphalt layer is implemented into the cross sections of the global structural analysis model, which leads to a distinctive progression of the midspan's flexural rigidity (Figure 6b).

$$f_i = \frac{\lambda_i^2}{2 \cdot \pi \cdot L^2} \left(\frac{EI}{m} \right)^{1/2} \tag{2}$$

According to the widely known equation (2), the frequency of vibration for a bridge beam is proportional to the square root of the moment of inertia [4]. The eigenfrequency f_i is defined as a function depending on the bending stiffness EI of the cross section, the length L of the structural member, the mass m per meter, and the type of the boundary conditions (λ). The utilization of the present equation without considering axial forces is permissible, as the acting forces in the present case are relatively small—furthermore, probable changes of the effective axial forces occur in quite a short timeframe, which minimizes their effect on the dynamic response in the course of frequency analysis.

For further investigations, a curve in terms of frequency needs to be generated (Figure 7b) for the next step, when a temperature-based stiffness path is eliminated from the overall trend (Figure 7a–c).

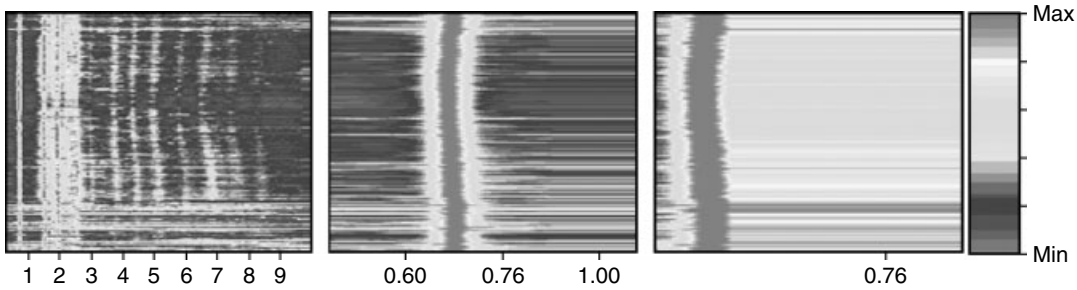


Figure 3. Trend of stiffness during one day: 0.30–10/0.30–1.10/0.60–0.80 Hz.

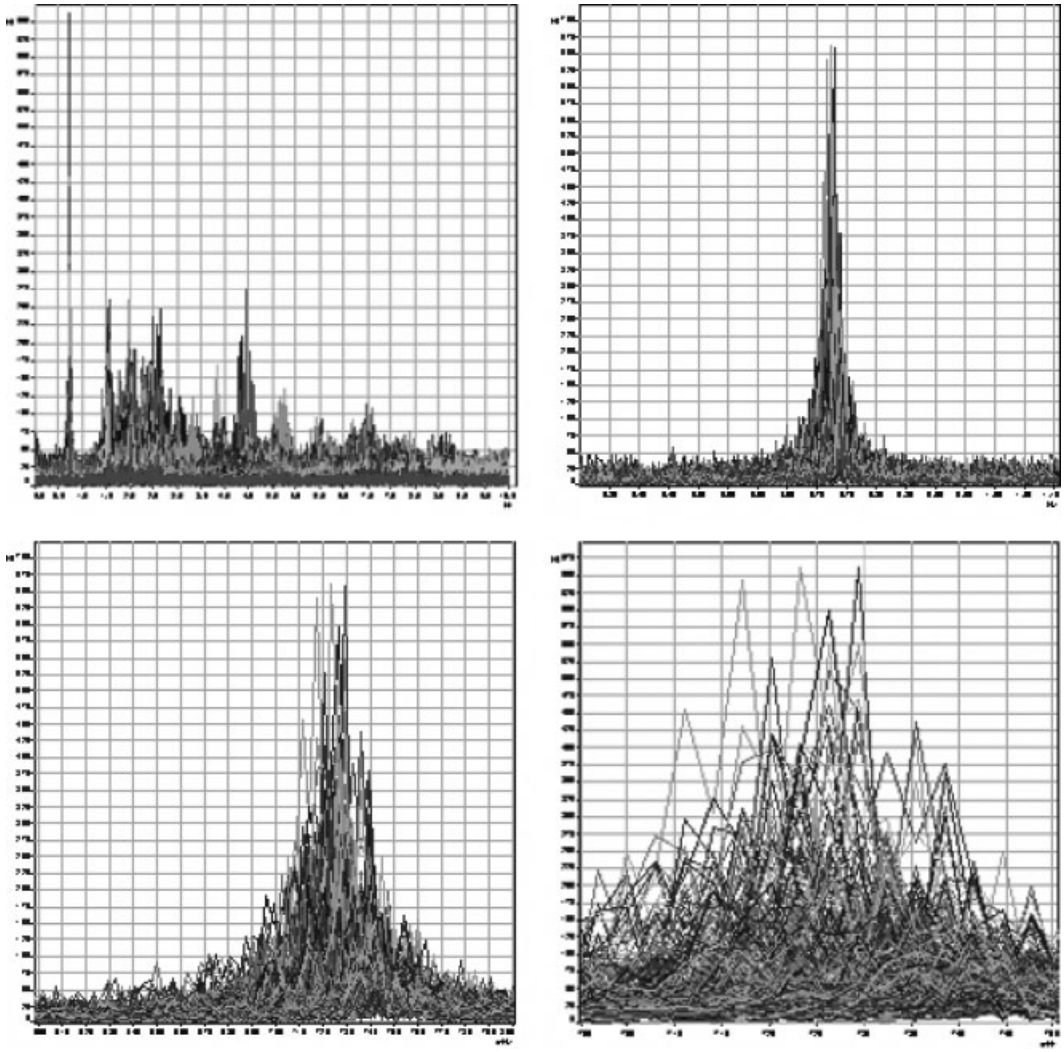


Figure 4. The front views of the trendcard for one day: 0.30–10/0.30–1.10/0.60–0.80/0.68–0.74 Hz.

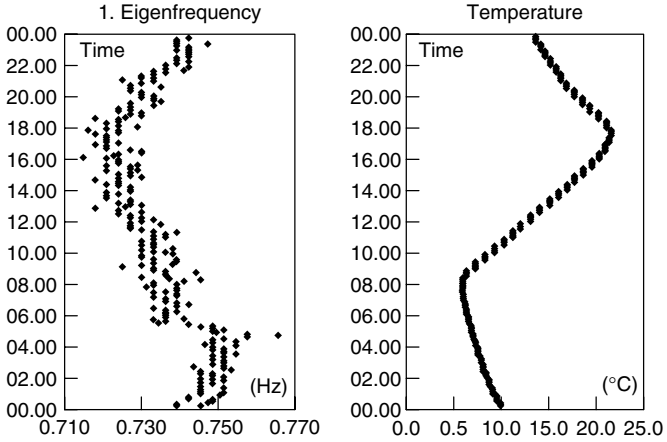


Figure 5. Pattern of first eigenfrequency and its obvious dependency on temperature.

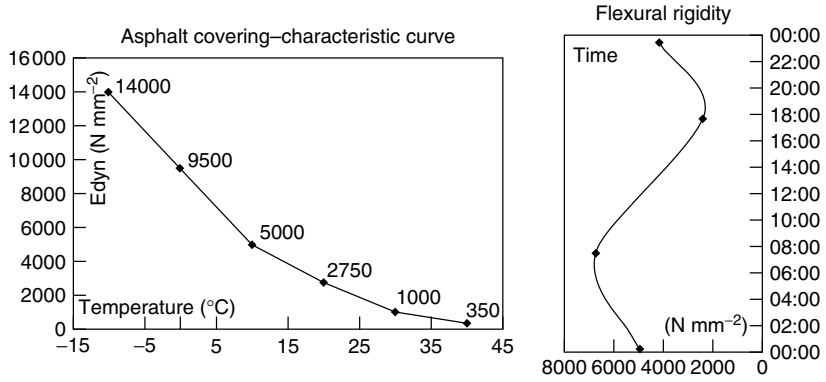


Figure 6. Progression of the asphalt layer's flexural rigidity in dependence of its temperature.

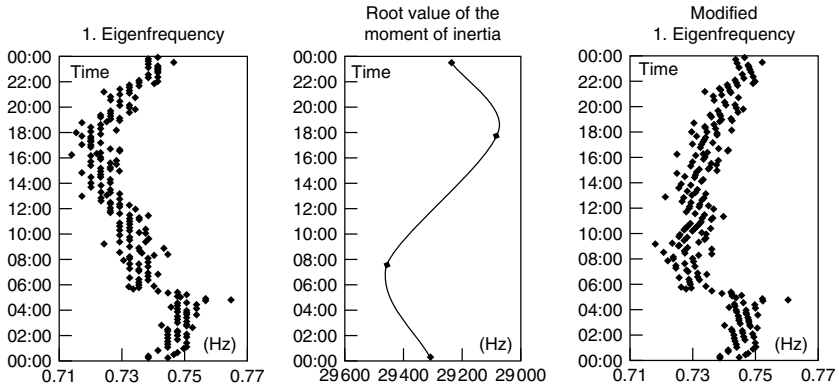


Figure 7. Pattern of first eigenfrequency before and after compensation of temperature.

The obtained trend very clearly shows the remaining impact of freight traffic itself, which strongly affects the timeframe between 5 a.m. and 10 p.m., when trucks are allowed to pass the bridge and cause two characteristic offsets during the course of the day.

3 COMPENSATION OF ADDITIONAL (MOVING) MASSES

The modified trend of the main span's stiffness already includes a number of characteristics of the prevailing freight traffic progression (Figures 8 and 9). Unfortunately, traffic data from the competent authorities are available only per hour. For some introductive exploration on approximate additional mass compensation, further steps need to be undertaken.

The frequency of vibration, based on equation (2) again, is inversely proportional to the square root of the mass. This means that live loads cause increase in effective mass, which leads to hourly calculated factors to modify the fluctuating frequency. Owing to that relation, the scattered trend of frequency is straightened in dependence of modal contribution of trucks per hour (Figure 10).

In fact, the present configuration of the permanent monitoring system provides the possibility to develop a more sophisticated and more reliable, strictly measurement data–based method. Forced balance accelerometers located at a defined distance along the cantilever's outer edges—in both directions of traffic—enable the verification of recurring truck passages and their related velocity and tonnage without any disturbance of traffic. A dynamic freight traffic registration system was developed (a certain pattern-recognition procedure introduced in [5]), which utilizes accelerometer-based, reproduced

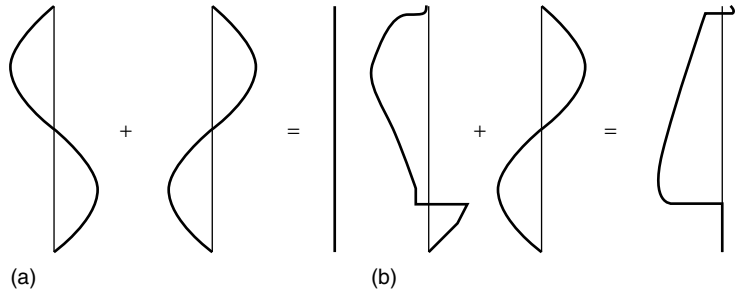


Figure 8. Comparison of expected (a) and actual (b) consequences of temperature-compensated natural frequency patterns.

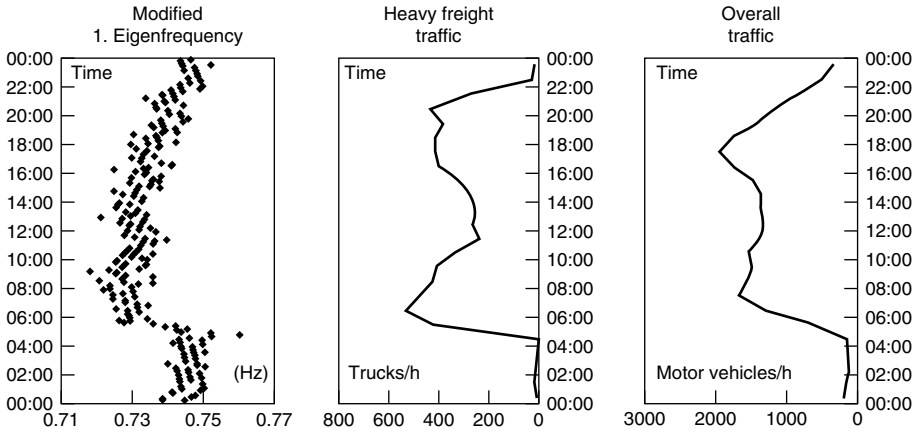


Figure 9. Modified pattern of stiffness—strongly affected by traffic loading (moving additional masses).

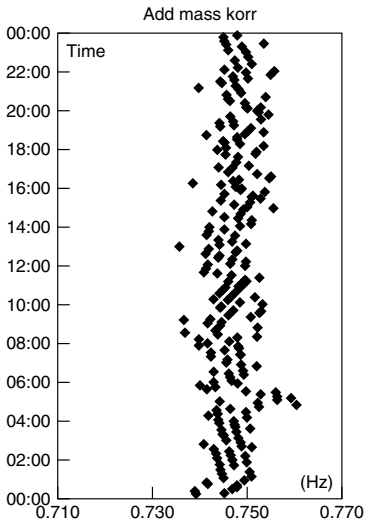


Figure 10. Stiffness pattern after approximate compensation of additional masses.

cantilever deformations. By this method, the moving loads within each measurement file—passing the main span simultaneously—could be identified and would lead to a shifting of the single peak in the frequency-response spectrum, which represents the registered time history in each measurement file.

4 CONCLUSIONS

As permanent monitoring systems produce huge amount of data, they have to be processed systematically to exploit the information fully. For easy handling, it is proposed that statistically based threshold levels are calculated. To do so, continuous monitoring systems—providing information about changed modal parameters under "normal" operational conditions—can be used to trigger warning and alarm levels with regard to damage assessment.

The acquisition and elaboration of the quantities that are provided by the installed instrumentation allow setting up a structural behavior model that is considered as the "regular model" (baseline model). The periodic elaboration of the acquired measurements and the comparison with the baseline model allows to point out indicators of potential structural damages. The availability of periodic surveys of the cause quantities allows, moreover, setting up statistical models of the structural behavior, where

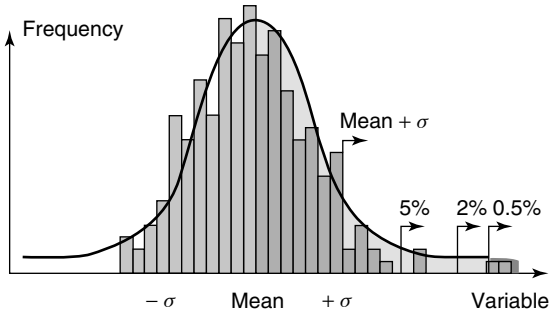


Figure 11. Histogram and best fit distribution-based determination of threshold values.

the structural response is statistically correlated to the trend of the cause quantities. Meanwhile, these models allow a control in time of the structural response by pointing out the "weight" of the cause quantities. The definition of alert threshold levels from analysis of historical database (e.g., extreme-value analysis) may be illustrated as in Figures 11 and 12.

5 OUTLOOK

The discussed approach—benefited by permanent monitoring—allows to reach the conducted goals, even if the experience already gained with the applied methods is still approximate in character. Therefore, the article solves the problem more qualitatively, rather than quantitatively, so far. The main reason is the fact that every condition compensation process based on permanent monitoring has necessarily to be a tailor-made one—depending on the observed bridge object. The polynomial functions—describing the occurring temperature—have always to be derived from measurements for further analysis (implementation of temperature-sensitive asphalt layer into the finite element bridge model—leading to distinctive characteristics of span's flexural rigidity). Although air temperature is used instead of structural elements, the results are very promising. This approach represents an innovation in stiffness assessment appropriate for long-term application. The goal of generating frequency progressions over time, without major environmental and operational impact, has come within reach. The procedure will be optimized in

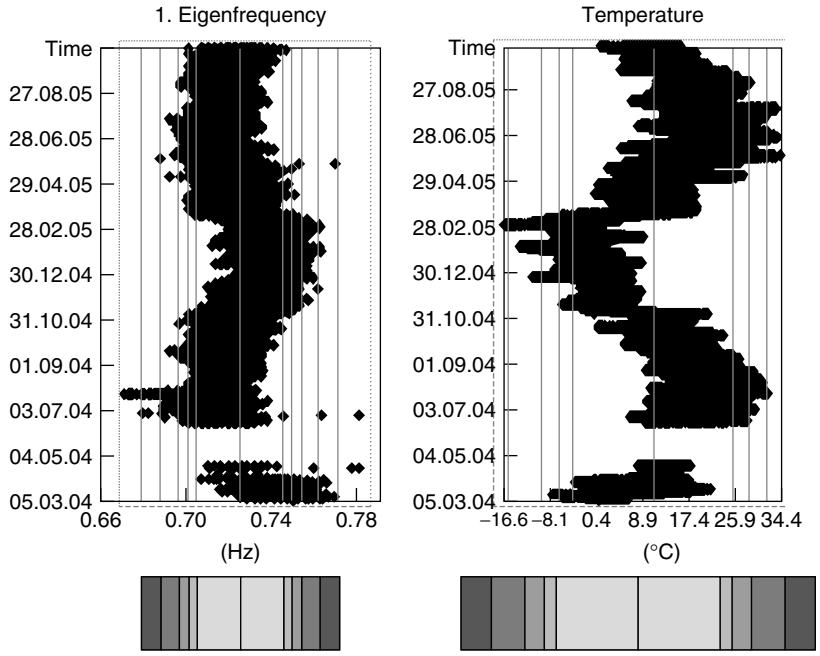


Figure 12. Response for 18 months using threshold levels (statistical time history) with 5/2.75/2.5/1/0.135% probability of exceedance.

progressive stages, as soon as the cantilever-sensor-based approach that has already been introduced is implemented.

RELATED ARTICLES

Free and Forced Vibration Models

Civil Infrastructure Load Models for Structural Health Monitoring

The Influence of Environmental Factors

Loads and Temperature Effects on a Bridge

REFERENCES

[1] Wenzel H, Pichler D. *Ambient Vibration Monitoring*. John Wiley & Sons: Chichester, 2005, ISBN 0470024305.

[2] IMC. *FAMOS Version 5.0, Reference Manual*. Berlin, 2005.

[3] Willberg U. *Asphaltschichten auf Hydraulisch Gebundenen Tragschichten—Untersuchungen Zum Tragverhalten*, Doctoral Thesis. Munich, Technische Universität München, 2001, pp. 10–11.

[4] Blevins RD. *Formulas for Natural Frequency and Mode Shape*. Van Regional Company: New York, 1979.

[5] Veit R, Wenzel H. Measurement based performance prediction of the Europabrücke against traffic loading. *Proceedings of the 16th European Conference of Fracture ECF16*. Alexandroupolis, July 2006, ISBN 13-978-1-4020-4972-9.

Chapter 129

Modal Testing of the Vasco da Gama Bridge, Portugal

Elsa Caetano and Álvaro Cunha
Faculty of Engineering, University of Porto, Porto, Portugal

1 INTRODUCTION

The safety condition of large bridges is normally verified at the end of construction on the basis of static and dynamic tests. Static load tests employ heavy road or railway actions applied in different combinations, reproducing exceptional loading scenarios, in order to ensure that the required strength has been achieved. The structural response is measured in terms of strains or displacements and compared with the results from numerical models. The purpose of dynamic testing is to identify the relevant natural frequencies, vibration modes, and damping ratios in order to calibrate and validate numerical models that have been used during the design to simulate the structural behavior under dynamic excitations such as wind and earthquakes.

The dynamic tests performed on the Vasco da Gama Bridge took place in March 1998 shortly before the opening of the bridge [1]. They comprehended response measurements developed under natural excitation (ambient vibration test (AVT)) and under the releasing of a load applied at the central span of the main cable-stayed bridge (free vibration test (FVT)). The use of high-sensitivity sensors and of an efficient wireless technique allowed the identification of a significant number of vibration modes with high amplitude and spatial resolutions based only on ambient vibration measurements. It also permitted the identification of installed force in some stay cables on the basis of laser measurements and on the indirect evaluation of cable frequencies.

Conventional identification techniques initially employed to extract modal parameters from the recorded time series were subsequently extended to

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Figure 1. Vasco da Gama Bridge.

more recent and powerful output-only identification algorithms, which are of interest to improve the quality of model estimates and automatic implementation in the context of long-term monitoring of a structure.

2 THE VASCO DA GAMA BRIDGE

The Vasco da Gama Bridge is a crossing of the river Tagus located in the city of Lisbon, Portugal, and has a total length of 17.3 km (Figure 1), involving three interchanges, a 5-km-long section on land and a continuous 12.3-km-long bridge. This bridge includes a main cable-stayed part over the main navigational channel with a total length of 829.2 m, formed by a central span of 420 m and three lateral spans on each side ($62 + 70.6 + 72$ m) (Figure 2). Two lateral prestressed concrete girders 2.6-m high connected by a cast *in situ* slab 0.25-m thick and by transversal steel I-girders every 4.42 m, form the 31-m-wide deck. The bridge deck is continuous along the total length and is fully suspended at a height of 52.5 m above the river by two vertical planes of 48 stays connected to each tower. The two H-shaped towers are 147 m high,

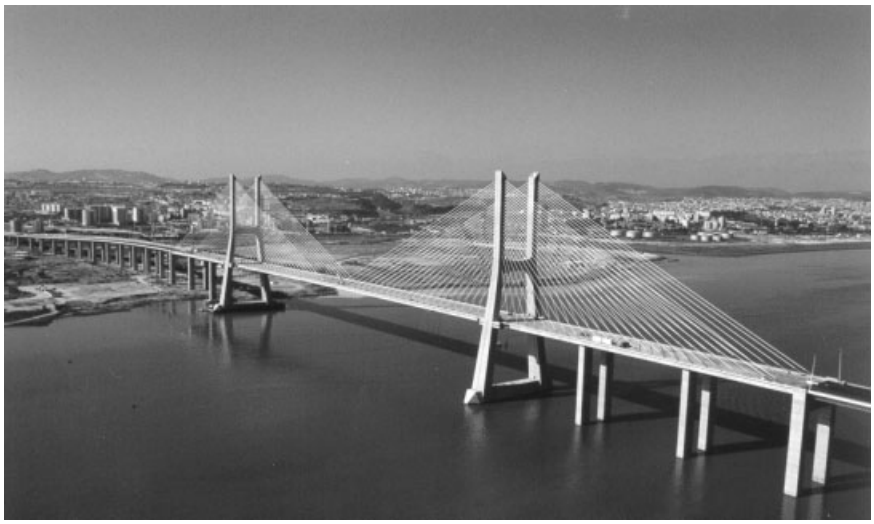


Figure 2. Vasco da Gama Bridge: cable-stayed part.

above a massive zone at their base used as protection against ship collisions.

The stay cables consist of bundles of parallel self-protected strands covered by a high density polyethylene (HDPE) sheath. Specific protection against vibration was adopted, namely, by inclusion of a double helical rib in the cable cover for prevention of rain–wind vibration, and by use of damper devices installed close to the deck anchorage inside the steel guide pipe of the cables. Given the active seismic location of the bridge site, specific measures were taken in the design, namely, the adoption of a full suspension deck from flexible towers to minimize the seismic forces and the introduction of a set of hysteretic steel dampers connecting the pylons and the deck to limit the displacements. Under service loads, the transverse dampers work within the elastic range, acting as elastic supports, whereas the longitudinal dampers allow low-speed displacements. In case of an earthquake, the steel hysteresis is used to dissipate energy.

3 INSTRUMENTATION

The identification of the major natural frequencies and vibration modes of the cable-stayed bridge and viaducts within a very limited time period would be compromised if a conventional acquisition system based on a set of accelerometers and a central unit was employed, due to the limited number of available accelerometers and to the huge requirements in terms of electrical cable length and manipulation to cover a dense mesh of points necessary to characterize higher order vibration modes. An efficient technique was employed instead, which consisted in the performance of the measurements with a total of six triaxial 16-bit strong motion recorders. These devices included force-balance type accelerometers and individual acquisition card and memory units. The operation of these recorders was programmed and synchronized by a laptop, which allowed the performance of measurements without electrical cables (Figure 3). Therefore, it was possible to freely move the sensors along the deck and towers and record the bridge response at a dense mesh of points.

A noncontact measurement system was tested for the instrumentation of the cables, which was based on a laser Doppler velocimeter. Although the purpose of the test was not to identify installed cable force at all stays, such a requirement would be extremely demanding if conventional accelerometers had been employed, due to the need to successively attach the accelerometer to each cable (Figure 4a). Using the laser velocimeter, the sensor head is placed underneath the cable on the deck (Figure 4b) and the relative velocity between the deck and the cable is measured from the analysis of frequency change of the reflected laser beam according to the well-known Doppler effect. The output of this sensor can be processed by a conventional Fourier analyzer. Although the precision of the laser sensor may be lower than that of a conventional accelerometer, the fact that stay cables evidence vibration amplitudes that are typically at least one order of magnitude higher than the deck means that the laser sensor can still provide accurate measurements of cable response. The average power spectrum densities (PSD) represented in Figure 5 obtained for one stay cable using an accelerometer and the laser sensor shows the high quality of the collected signal and the usefulness of this sensor to measure cable frequencies.

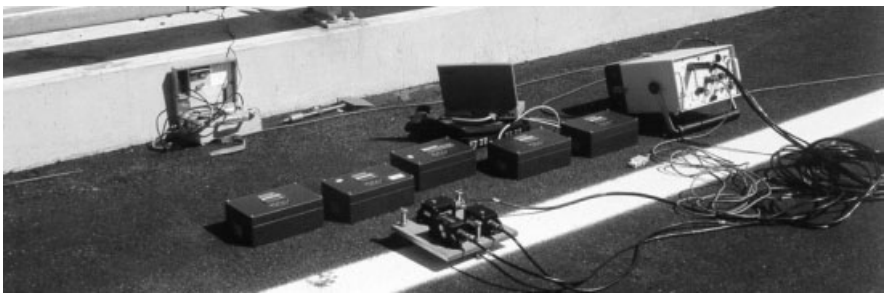
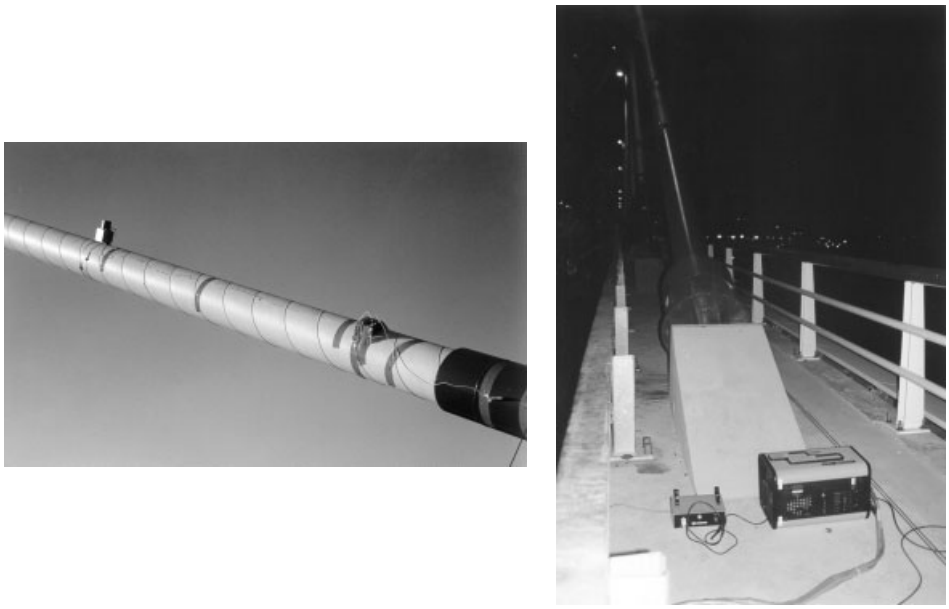


Figure 3. Strong motion recorders used in dynamic testing.



(a) (b)

Figure 4. Stay-cable instrumentation: (a) conventional accelerometer and (b) laser Doppler velocimeter.

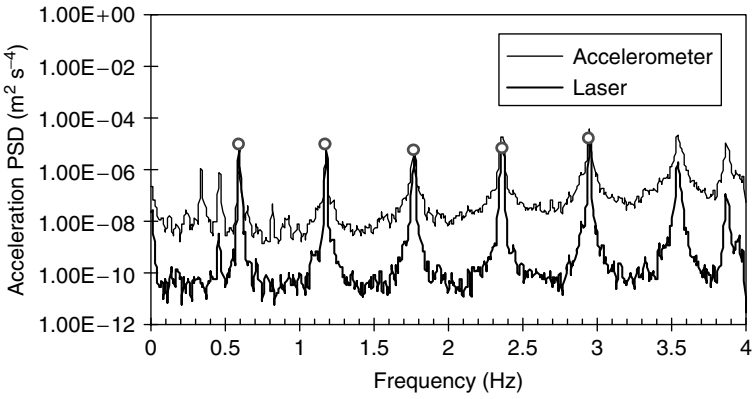


Figure 5. Comparison of acceleration average power spectrum densities (PSD) of the ambient response of a stay cable measured with a conventional accelerometer and a laser sensor.

4 AMBIENT VIBRATION TEST

The AVT was conducted considering two strong motion recorders fixed at reference positions (upstream and downstream), and moving the other four recorders along the successive measurement sections.

For the cable-stayed bridge, the chosen reference sections were located about one-third of the central span on the North side, on marks 10U and 10D of Figure 6. The other four recorders scanned the bridge deck and the towers using a total of 29 measurement sections. The sampling frequency was 50 Hz. The expected interesting frequency range was very low (0–1 Hz). Therefore, a measurement time of 16 min was chosen for each setup, which was sufficiently long to capture various periods of the low-frequency modes. Data recording started every 20 min, meaning

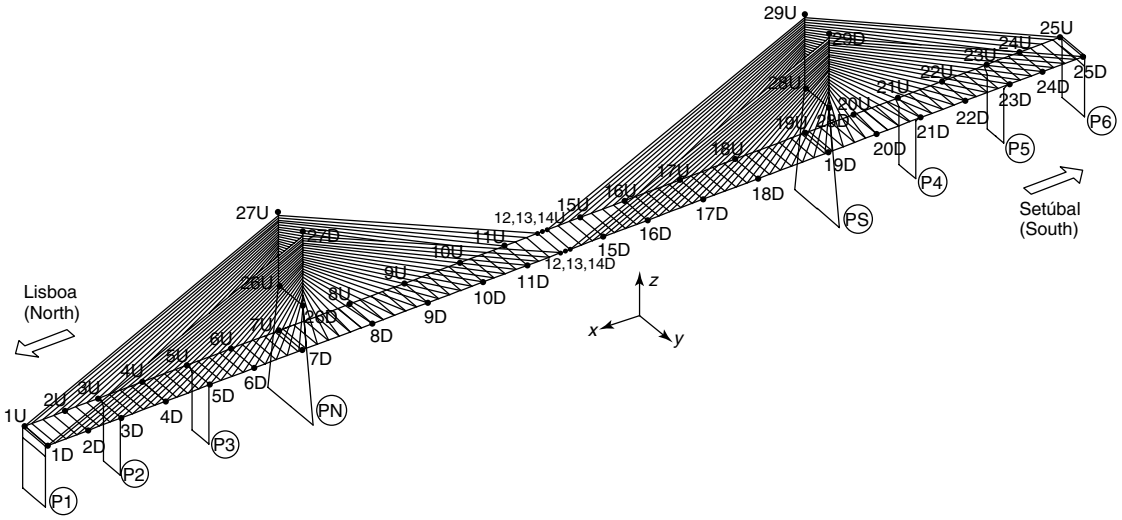


Figure 6. Measurement locations on the Vasco da Gama cable-stayed bridge.

that 4 min were left for moving the four recorders to the adjacent measurement sections. A total of 2.5 days was required to record the triaxial response at the 29 pairs of measurement locations represented in Figure 6.

The excitation source was wind, the speed varying between 1 and 22 $\mathrm{m\,s^{-1}}$ during the complete AVT campaign. This resulted in large differences of bridge acceleration magnitudes, whose root mean square values ranged between 0.03 and 0.4 mg for the horizontal directions, and, between 0.06 and 1.3 mg for the vertical directions, with the consequence of quality differences of the acquired data.

5 FREE VIBRATION TEST

Since there is no reliable analytical model available for the evaluation of damping ratios and, considering the large influence on the bridge response to wind and earthquake excitation, experimental identification of those parameters is extremely important.

Damping-ratio estimates obtained from ambient vibration measurements exhibit a large scatter, as a consequence both of variation of damping with amplitude of vibration and wind speed and of measurement/processing errors. For higher amplitudes of vibration, the scatter reduces and the accuracy of measurements increases. The easiest way of generating higher than ambient vibrations on a full-scale structure is by application of impulsive loads. On the Vasco da Gama Bridge, the impulsive excitation was obtained by suspending a mass of 60 t from a point on the bridge deck, close to location 10U (Figure 6), and suddenly releasing it (Figure 7). The resulting free vibrations were then recorded over 16 min at measurement sections 10, 13, and 16 (Figure 6). This test was performed under low wind speeds (less than 2.5 $\mathrm{m\,s^{-1}}$ were measured), so that the identified damping ratios represented the real structural damping ratios, with no added aerodynamic component.

6 MODAL PARAMETER IDENTIFICATION AND FINITE ELEMENT CORRELATION

6.1 Operational modal analysis

The extraction of modal parameters from the acquired data was initially made on the basis of peak picking method [2]. Subsequently, more powerful output-only modal parameter estimation methods were applied, the stochastic subspace identification (SSI) and the frequency-domain decomposition methods [3, 4], implemented in the software packages MACEC [5] and ARTeMIS [6].

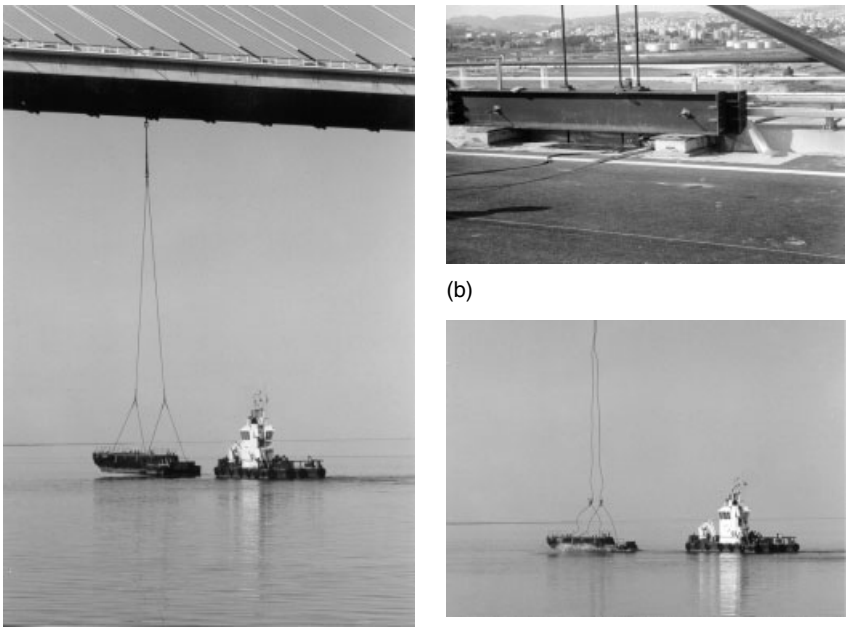


Figure 7. Free vibration test (a) eccentrically suspended 60-t barge, (b) starting cut of hanging Dywidag bar, and (c) release of barge.

Peak picking is the simplest and the most popular method used in civil engineering to estimate the modal parameters of a structure subjected to ambient loading. The method is based on the identification of the natural frequencies from the peaks of power spectra estimates. Mode-shape components are then identified from the analysis of the amplitude and phase of transfer functions relating the ambient response at a reference section with the response at the other measurement points.

Although this method relies on the assumptions of low damping and well-separated eigenfrequencies, some refinements have been introduced [7], which allow more complex applications. The coherence function between two channels tends to go to 1 in the vicinity of the resonance frequencies, because of the high signal-to-noise ratio at these frequencies. Consequently, inspecting the coherence function can assist in selecting the eigenfrequencies. Also, the phase angles of the cross spectra are helpful: if real modes are expected, the phase angles should be either 0° or 180° at the resonance frequencies. An additional improvement consists in separating bending and torsional effects by simple sum/difference of simultaneous records collected at the opposite sides of the deck. However, a violation of the basic assumptions (low damping and well-separated modes) can lead to erroneous results. In fact, the method identifies operational deflection shapes instead of mode shapes and for closely spaced modes such an operational deflection shape will be the superposition of multiple modes. Other disadvantages are that the selection of the eigenfrequencies can become a subjective task if the spectrum peaks are not very clear and that the method does not yield any damping estimates.

The frequency-domain decomposition (FDD) method is an interesting alternative that can be understood as an extension of the peak picking method, enabling the treatment of closely spaced modes and eventually the extraction of damping estimates. This method employs a process of noise reduction of a cross-power matrix of the output measurements based on the singular value decomposition. The obtained singular values are related with the natural frequencies and damping ratios, while singular vectors represent the corresponding mode shapes [8]. Another alternative to the peak picking method is the SSI method, which identifies a so-called stochastic

state space model from measured output data or output covariances. This model is a good representation of a structure excited by unknown forces, which are assumed white-noise signals. After the identification of the state space model, the modal parameters are obtained from corresponding matrices [5]. The three methods referred to above are the most widely used output-only methods in civil engineering applications at the current state of art and were applied to the Vasco da Gama data for comparison of performances.

6.2 Ambient vibration data

The extraction of modal parameters from the ambient vibration data was performed first using the conventional peak picking method. Figure 8 shows average normalized power spectrum density (ANPSD) and average normalized cross power spectrum density (ANCPSD), respectively, associated to vertical and transversal accelerations along the bridge deck, whereas Figure 9 shows normalized cross power spectrum densities (NCPSD) and corresponding coherences relating the ambient response at sections 10 and 16 (Figure 6). The natural frequencies identified on the basis of the peak picking technique are summarized in Table 1, as well as the calculated frequencies obtained at the design stage by EEG (Europe Études Gecti, Villeurbanne, France), using the finite element program Hercules. Figure 10 shows some of the most significant mode shapes of the deck, and the corresponding numerical modes, as well as some modal components identified using the FVT data.

The analysis of Table 1 and Figure 10 shows the excellent correlation between identified and numerical modal parameters, even though the bridge is characterized by a large number of very closely spaced in-frequency vibration modes. The success of the peak picking method in this application stems from the preliminary separation of bending and torsional components of the signal and from the high-frequency resolution of the analysis. The major differences between experimental and numerical data are associated with identified multiple modes, characterized by natural frequencies in the

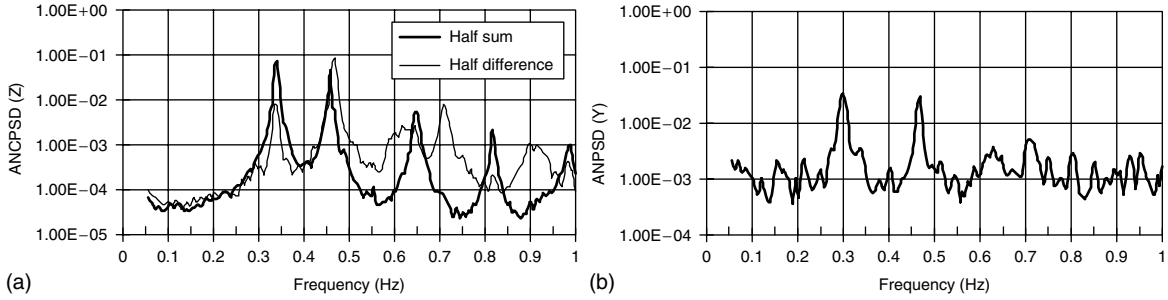


Figure 8. Average normalized spectra associated to (a) vertical acceleration (half-sum and half-difference signals, upstream−downstream) and (b) transversal acceleration (half-sum signal).

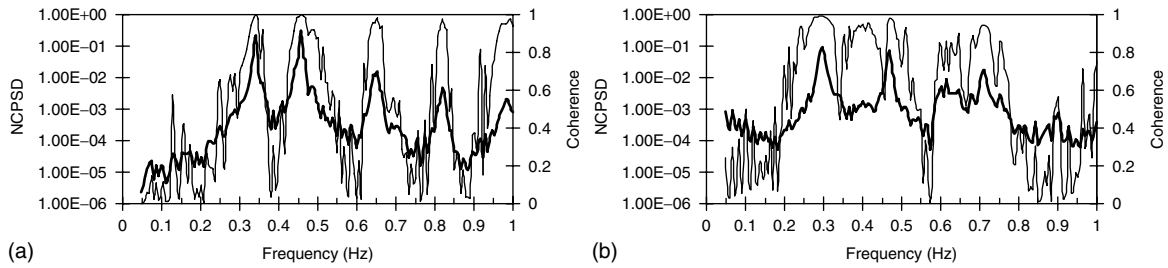


Figure 9. NCPSD spectra (amplitude) of the half-sum signal of (a) vertical acceleration and of (b) transversal acceleration at sections 10 and 16, and corresponding coherences.

Table 1. Identified and calculated natural frequencies

Calculated frequencies (Hz)	Identified frequencies (Hz)	Type of mode of vibration
0.2624	0.298	First transversal bending (BT1)
0.3185	0.341	First vertical bending (BV1)
0.4287	0.437	Second vertical bending (BV2)
0.4386	0.471	First torsion + transversal bending (T1)
0.6268	0.572[(a)]/0.590[(a)]/0.599[(a)]/0.619[(a)]/0.624[(a)]	Second torsion + transversal bending
0.6077	0.651	Third vertical bending (BV3)
0.6268	0.693[(a)]/0.707[(a)]/0.718[(a)]/0.755[(a)]	Second torsion + transversal bending (T2)
0.7600	0.817[(a)]	Fourth vertical bending (BV4)
[(b)]	0.895[(a)]/0.917[(a)]	Third torsion (T3)
[(b)]	0.985	Fifth vertical bending (BV5)
[(b)]	1.129[(a)]	Fourth vertical bending

[(a)] Multiple modes, low signal level.
[(b)] Unknown.

range 0.693–0.755 Hz and with common deck modal configuration, but involving different cable motions. These modes reflect cable–structure interaction [2] and have not been identified numerically, since the cable dynamics was not included in the analysis.

Having the purpose of analyzing and comparing the efficiency of the more powerful identification algorithms based on the FDD and SSI methods, a reanalysis of the ambient vibration data of the Vasco da Gama cable-stayed bridge was done [3, 4], based on the implementations of the SSI in the MACEC software package [5] and on the FDD and SSI implementations in the commercial package ARTeMIS [6]. Tables 2 and 3 show comparisons between natural frequencies identified using both methods and the peak picking method, also summarizing the estimates of modal damping ratios achieved by application of the SSI method. Figure 11 shows a three-dimensional representation of the fundamental identified modes based on the MACEC implementation.

The analysis of Tables 2 and 3 shows that the three methods lead to very consistent estimates, especially in terms of identified natural frequencies. The SSI and FDD methods provide excellent results for closely spaced modes, without special combination (bending/torsion) of the original signals. The identification of multiple modes involving different cables is, however, more difficult. This fact is due to the variable wind speed along the test, responsible for different excitation of cables along the different setups. As for damping estimates, it is seen that similar order of magnitude is obtained using different implementations of the SSI method. The scatter of estimates is, however, very large and is inherent to the identification procedure. It results also from the variation of damping along the test due to aerodynamic variable contribution and amplitude variation of structural damping for very low amplitudes of vibration.

6.3 Free vibration data

The eccentrical release of the 60-t mass from a point in the one-third of the central span of the bridge produced a free decay oscillation with maximum displacement amplitude of 25 mm (30-mg peak acceleration), which attained the ambient oscillation level about 6 min after release of the load.

The identification of natural frequencies from the collected records was made, in a first instance, by inspection of the peaks of fast Fourier transforms (FFTs) of the acceleration time series (Figures 12 and 13). With regard to the mode shapes, these were identified by digital filtering around each of the natural frequencies identified, and by comparing the amplitudes and phases of the filtered signals at different points of measurement. Figure 10 shows the modal components identified by this procedure, which are clearly in good agreement both with the mode shapes obtained by the AVT and with the modal configurations calculated numerically. The identification of the modal damping ratios was done on the

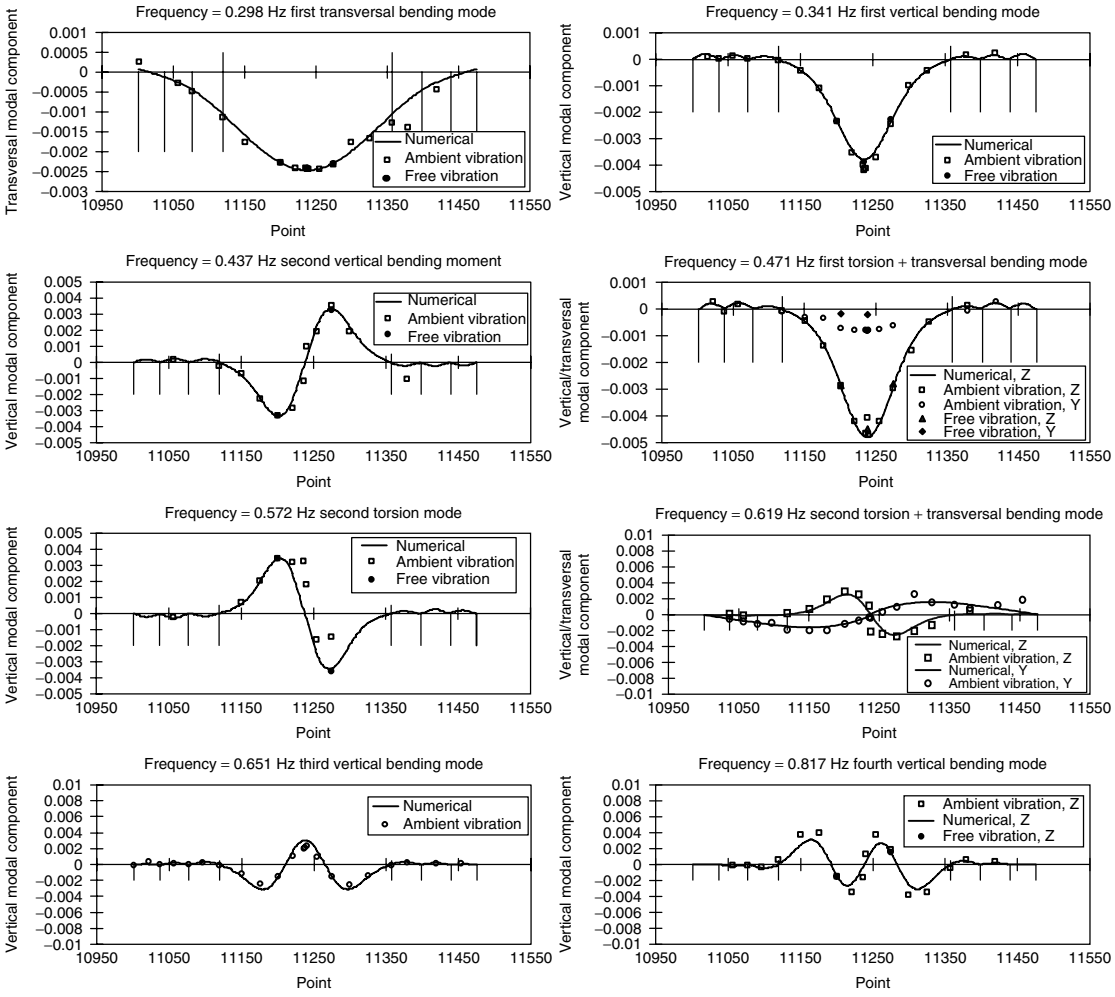


Figure 10. Some of the most relevant identified mode shapes (deck).

basis of the decay of the envelope of the filtered signals obtained as exemplified in Figure 14.

The SSI method was also applied to the set of free vibration records [5]. Table 4 shows a comparison between natural frequencies identified using both the peak picking technique and the SSI method. This table also shows the modal damping ratio estimates achieved. It can be concluded that natural frequency estimates obtained from peak picking method are almost identical to those resulting from the SSI method, except for the third torsional mode (T3), whose peak was unclear for identification with peak picking method. Damping ratios estimated with SSI method are also very similar to the ones estimated from the free decay analysis of the band-pass filtered records.

6.4 Discussion of AVT and FVT results

Considering the different levels of response involved in AVT and FVT, it is of interest to compare the modal parameters obtained from those types of tests. Table 5 systematizes natural frequencies and damping ratios identified for the two situations. The mode shapes are compared by computing the (squared) correlation between the components of the modal vectors, which have been measured in both AVT

Table 2. Ambient vibration modal parameters (AVT) obtained with peak picking (PP) and the MACEC implementation of SSI

Type of mode	AVT/PP	AVT SSI (MACEC)			
	f (Hz)	f (Hz)	σ_f (%)[a]	ξ (%)	σ_ξ (%)[a]
BT1	0.298	0.302	1.66	1.47	61
BV1	0.341	0.339	0.29	0.52	39
BV2	0.437	0.458	0.22	0.44	31
T1	0.471	0.468	0.21	0.43	22
	0.572–0.624	—	—	—	—
BV3	0.651	0.649	0.46	0.72	45
T2	0.693–0.755	0.711	0.56	1.09	50
BV4	0.817	0.817	0.37	0.44	17
T3	0.895	0.917	0.00	—	—
BV5	0.985	0.987	0.51	0.74	23

BT1, first transversal bending; BV1, first vertical bending; BV2, second vertical bending; BV3, third vertical bending; BV4, fourth vertical bending; BV5, fifth vertical bending; T1, first torsion; T2, second torsion; T3, third torsion.
[a] Standard deviations based on the setups used.

Table 3. Ambient vibration modal parameters (AVT) obtained with peak picking (PP) and the ARTeMIS implementation of FDD and SSI

Type of mode	AVT/PP	FDD (ARTeMIS)		SSI (ARTeMIS)			
	f (Hz)	f (Hz)	σ_f (%)[a]	f (Hz)	σ_f (%)[a]	ξ (%)	σ_ξ (%)[a]
BT1	0.298	0.303	1.65	0.303	1.56	1.25	44
BV1	0.341	0.339	0.69	0.339	0.27	0.33	73
BV2	0.437	0.458	0.36	0.458	0.14	0.26	64
T1	0.471	0.470	0.40	0.469	0.30	0.29	58
	0.572–0.624	0.593; 0.620	0.68; 1.29	0.596; 0.627	0.78; 0.98	0.80; 0.84	81; 67
BV3	0.651	0.649	0.46	0.650	0.32	0.60	58
T2	0.693–0.755	0.712	0.74	0.714	0.76	0.89	46
BV4	0.817	0.818	0.30	0.818	0.26	4.52	450
T3	0.895	0.899	0.55	0.900	0.75	0.74	57
BV5	0.985	0.987	0.45	0.988	0.47	1.11	233

BT1, first transversal bending; BV1, first vertical bending; BV2, second vertical bending; BV3, third vertical bending; BV4, fourth vertical bending; BV5, fifth vertical bending; T1, first torsion; T2, second torsion; T3, third torsion.
[a] Standard deviations based on the setups used.

and FVT. This squared correlation is called *modal assurance criterion* (*MAC*) in modal analysis and is represented in Figure 15, showing that the AVT and FVT mode shapes are very similar. From Table 5, it is clear that the natural frequencies and damping ratios are also very much alike. In general, damping ratios are identified with larger uncertainty. It is known that they vary with the magnitude of vibrations and that an aerodynamic component may be present in the AVT due to the relatively high wind speeds measured during some of the setups. It appears that the FVT natural frequencies are systematically slightly lower than their AVT counterparts.

7 SURVEILLANCE AND STRUCTURAL MONITORING

As an important infrastructure in the access to the city of Lisbon, the Vasco da Gama Bridge is currently

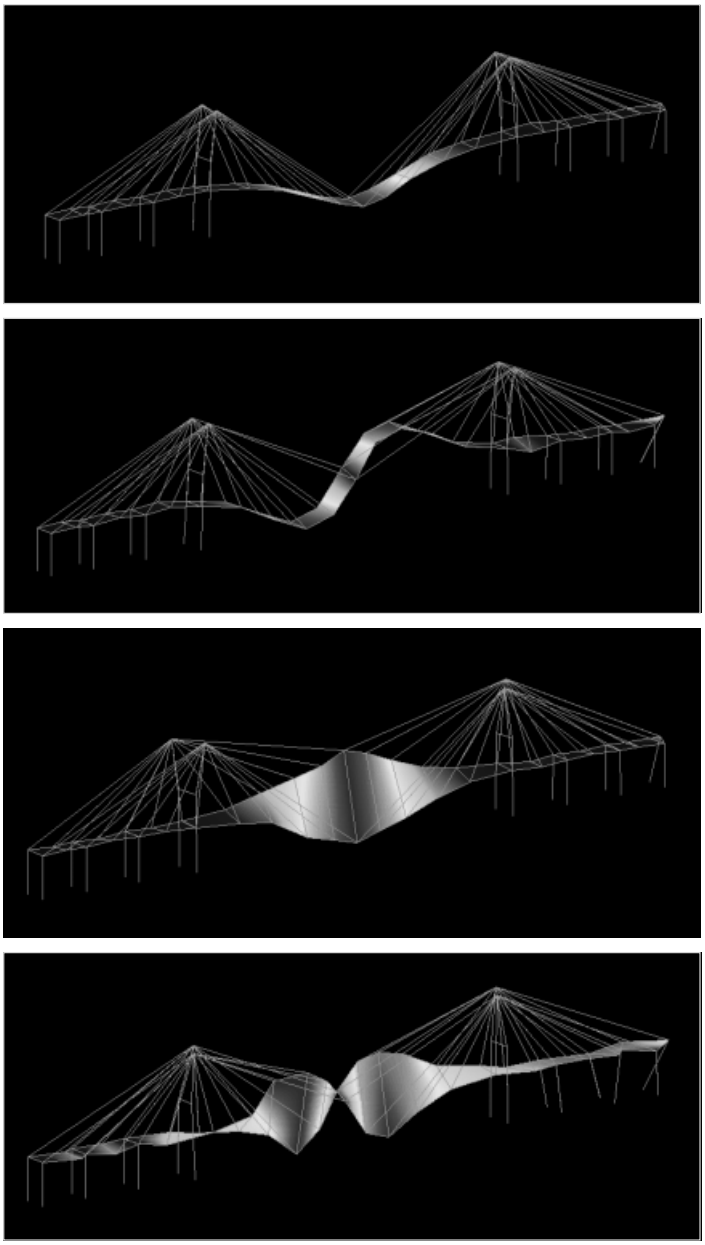


Figure 11. 3-D representation of some of the identified mode shapes using MACEC (first and second vertical bending modes and first and second torsional modes).

the object of a surveillance and monitoring program [9]. The purposes are (i) the guarantee of users' safety; (ii) the verification of the variation of structural behavior according to the predicted behavior; (iii) the characterization of the effects associated with exceptional events like earthquakes, ship collision, or high winds; and (iv) the identification of cable retensioning needs according to a predefined plan. To accomplish these objectives, the bridge has been fully instrumented, including, in particular, a set of

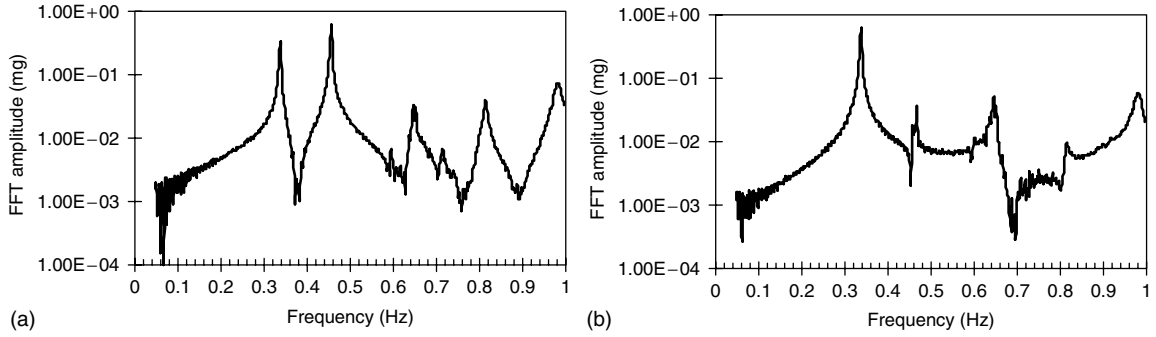


Figure 12. Amplitude of the FFT of the half-sum signal of vertical acceleration (upstream−downstream) at (a) one-third span North and (b) one-half span.

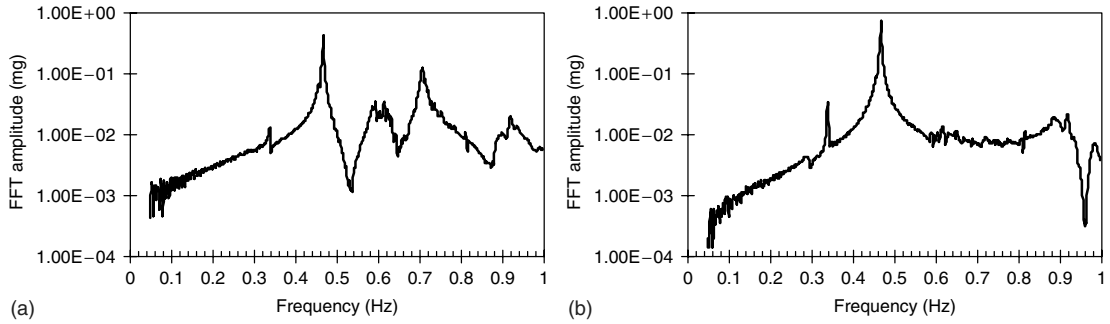


Figure 13. Amplitude of the FFT of the half-difference signal of vertical acceleration (upstream−downstream) at (a) one-third span North and (b) one-half span.

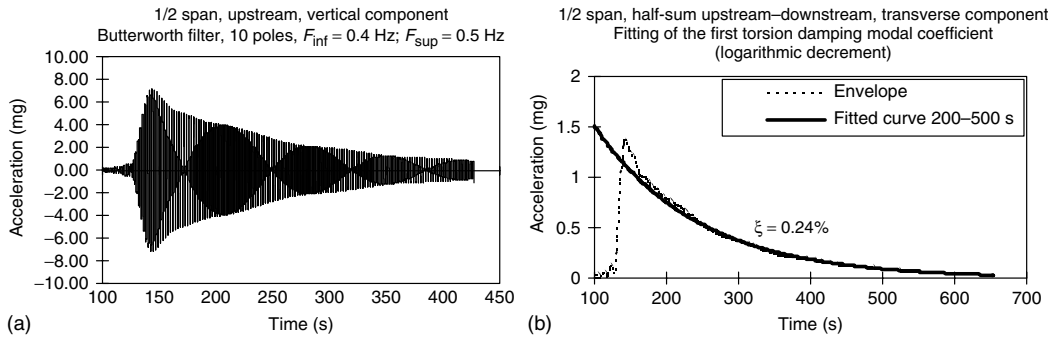


Figure 14. Identification of the modal damping ratio associated to the natural frequency 0.467 Hz. Analysis based on the measured response at one-half span upstream (a) Band-pass filtered free decay record; (b) Measured and fitted envelope of response (in module).

electrical sensors for measurement of joint openings, temperatures, wind, strains, and accelerations. All observations are permanently referred to a zero-state condition that corresponds to the measurements performed at the end of construction of the bridge, including the ones referred above. Records are saved in case of exceedance of predefined limits. Otherwise, complete measurement records are obtained every six years. Alert levels have been stipulated, which require different levels of intervention, like closing the bridge to traffic, conducting visual inspections, or developing particular retrofit actions. It is relevant to mention

Table 4. Free vibration modal parameters (FVT) obtained with peak picking (PP) and stochastic subspace identification (SSI)

Type of mode	FVT PP		FVT SSI	
	f (Hz)	ξ (%)	f (Hz)	ξ (%)
BT1	0.295	1.23	0.293	1.12
BV1	0.338	0.21	0.337	0.39
BV2	0.456	0.23	0.455	0.31
T1	0.467	0.24	0.466	0.27
	0.591	0.34	0.590	1.30
BV3	0.647	0.37	0.647	0.56
T2	0.707	0.78	0.705	0.74
BV4	0.814	0.48	0.814	0.50
T3	—	—	0.917	0.48
BV5	0.982	0.74	0.982	0.76

BT1, first transversal bending; BV1, first vertical bending; BV2, second vertical bending; BV3, third vertical bending; BV4, fourth vertical bending; BV5, fifth vertical bending; T1, first torsion; T2, second torsion; T3, third torsion.

Table 5. Comparison between AVT and FVT modal parameters obtained with stochastic subspace identification

Type of mode	AVT SSI		FVT SSI	
	f (Hz)	ξ (%)	f (Hz)	ξ (%)
BT1	0.302	1.47	0.293	1.12
BV1	0.339	0.52	0.337	0.39
BV2	0.458	0.44	0.455	0.31
T1	0.468	0.43	0.466	0.27
	—	—	0.590	1.30
BV3	0.649	0.72	0.647	0.56
T2	0.711	1.09	0.705	0.74
BV4	0.817	0.44	0.814	0.50
T3	—	—	0.917	0.48
BV5	0.987	0.74	0.982	0.76

BT1, first transversal bending; BV1, first vertical bending; BV2, second vertical bending; BV3, third vertical bending; BV4, fourth vertical bending; BV5, fifth vertical bending; T1, first torsion; T2, second torsion; T3, third torsion.

that until the current moment, no particular threshold has been attained demanding unexpected correction measures.

8 CONCLUSIONS

Experimental modal identification and finite-element correlation of the Vasco da Gama cable-stayed bridge have been conducted on the basis of the performance of AVT and FVT. The high-quality database created was processed using both the conventional peak picking technique and the modern frequency-domain decomposition and SSI methods. The results obtained allow drawing, in particular, the following conclusions:

- The levels of vibration of the cable-stayed bridge under ambient excitation are very low, even for significant wind speeds.
- The measurement system used in the AVT and FVT, based on the use of independent triaxial strong motion recorders conveniently programmed and synchronized by a portable PC, revealed to be a very efficient and comfortable solution, avoiding the use of several hundred meters of electrical cables and permitting the integral data acquisition in a relatively short period of time.
- A large number of modes could be identified in a low-frequency range (0–1 Hz) from low-level ambient vibrations using sensors dedicated to civil engineering applications, collecting long records of response and using appropriate signal processing and modal analysis techniques.
- The FDD and SSI application confirmed earlier results obtained with peak picking.
- The FVT was very useful as a complementary test that permitted not only to check the previous modal identification from ambient data but also to accurately identify modal damping ratios, whose knowledge is particularly relevant in terms of the study of the aerodynamic stability of the bridge.
- The SSI method allows the identification of damping ratios from ambient vibration data. These ratios present a significant scatter and seem to be systematically slightly higher than the corresponding estimates obtained from the FVT, reflecting fluctuations due to aerodynamic variability. Therefore, although it is understood that a lesser accuracy exists in the identification of damping ratios based on ambient vibration data and that this type of tests cannot replace FVTs; information obtained is of extreme usefulness, especially when the latter tests cannot be accomplished.
- There is, in general, an excellent correlation between modal parameters identified and the corresponding parameters calculated on the basis of the 3-D finite element model developed at the

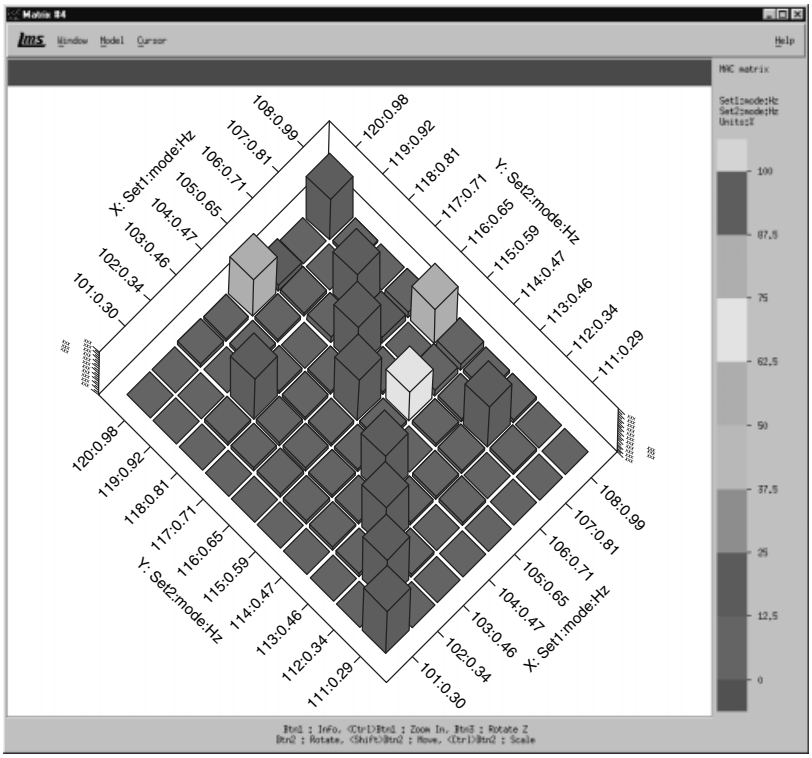


Figure 15. MAC values (mode-shape correlations) between 8 AVT modes and 10 FVT modes.

design stage, though some small differences can be found, as it is the case of the multiple modes identified associated to the second torsion + transversal bending numerical mode, related with local stay-cable frequencies, or the third torsion mode, in which no transversal bending component was experimentally detected.

The Vasco da Gama Bridge has been the object of a surveillance and monitoring program implemented by Lusoponte. This system comprehends extensive instrumentation that permanently collects data that are compared to a zero-state condition, for possible detection of degradation and characterization of the effects of exceptional events. Until the current moment no particular threshold has been attained demanding unexpected correction measures.

ACKNOWLEDGMENTS

The authors acknowledge the financial support obtained from the Portuguese Scientific Foundation FCT, under the terms of the research projects PBIC/CEG/2349/95 and PRAXIS/ECM/13251/98, as well as the collaboration provided by NOVAPONTE and LNEC.

RELATED ARTICLES

Free and Forced Vibration Models
Modal–Vibration-based Damage Identification
Statistical Time Series Methods for SHM
Ambient Vibration Monitoring
Long-term Monitoring of Dynamic Loads on the Brandenburg Gate
Continuous Monitoring of the Øresund Bridge: Data Acquisition and Operational Modal Analysis
Condition Compensation in Frequency Analyses—a Basis for Damage Detection
Multiple-model Structural Identification
Suspended Roof of Braga Sports Stadium, Portugal

REFERENCES

[1] Cunha A, Caetano E, Delgado R. Dynamic tests on a large cable-stayed bridge. An efficient approach. *Journal of Bridge Engineering, ASCE* 2001 **6**(1):54–68.

[2] Caetano E. *Dynamics of Cable-stayed Bridges: Experimental Assessment of Cable-Structure Interaction*, Ph.D. Thesis. Faculty of Engineering of the University of Porto: Portugal, 2001.

[3] Peeters B, De Roeck G, Caetano E, Cunha A. Dynamic study of the Vasco da Gama Bridge. *International Conference on Noise and Vibration Engineering, ISMA 2002*. Leuven, 2002.

[4] Cunha A, Caetano E, Brincker R, Andersen P. Identification from the natural response of Vasco da Gama Bridge. *XXII International Modal Analysis Conference IMAC*. Deaborn, MI, 2004.

[5] Peeters B. *System Identification and Damage Detection in Civil Engineering*, Ph.D. Thesis. Katholieke Universiteit Leuven, 2000.

[6] SVS. *ARTeMIS Extractor Pro, Release 3.41*. Structural Vibration Solutions: Aalborg, 1999–2004.

[7] Felber A. *Development of a Hybrid Bridge Evaluation System*, Ph.D. Thesis. University of British Columbia (UBC), 1993.

[8] Brincker R, Zhang L, Andersen P. Modal identification from ambient responses using frequency domain decomposition. *XVIII International Modal Analysis Conference IMAC*. San Antonio, TX, 2000.

[9] LUSOPONTE, *Structural and Topometric Monitoring of the Vasco da Gama Crossing. Manual of Inspection and Monitoring of the Vasco da Gama Bridge*, 1998 (in Portuguese).

Chapter 130
Multiple-model Structural Identification

Ian F. C. Smith
Institute of Structural Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

1 INTRODUCTION

While there are other good applications, multiple-model structural identification methods are useful principally for complex civil structures in uncertain environments. In such situations, determining the number of possible causes of anomalous behavior is combinatorial. This aspect has important implications. For example, it is impossible to place sensors so that they can measure directly every type of damage and deterioration at every location on every element. Most measurements are consequently indirect. Behavior models are thus essential to perform diagnoses, to design and plan possible repairs, and then to continue to manage structures throughout the rest of their service lives. Finding good models to support such decision making requires development of appropriate methodologies for structural identification. Such development is one of the most important scientific challenges within the field of structural health monitoring.

Behavior models are rarely accurate. Errors due to factors such as idealizations inherent in models of mechanical principles, numerical errors, inaccurate boundary conditions, and wrong material constants create situations where high model accuracy is unlikely in most practical situations. In addition, measurements have errors. These errors can be several times larger than the accuracies reported by sensor manufacturers owing to factors such as installation difficulties and connection losses. Therefore, engineers cannot assume the same level of accuracy as is observed in laboratory installations.

Service lives of civil infrastructure often exceed 100 years. Monitoring and decision making are thus iterative processes. Once measurements are made and the data is interpreted, additional measurements using new sensors may be warranted. Engineers need to make the most of existing information to

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

plan downstream tasks such as further measurement, preventative maintenance, and structural replacement.

Conditions of indirect measurements, an unavoidable presence of errors in modeling and in measurement, as well as monitoring/interpretation/decision-making cycles over decades create many challenges for structural health monitoring. This article proposes that appropriate selection of the best identification order for the task at hand provides good support for engineers involved in structural management.

2 CURRENT PRACTICE

It is common practice to assume that the service behavior of a structure can be modeled using the same model that was used in design. Since this assumption requires no verification of characteristics such as support conditions, geometry, and damage on the as-built structure, it is clearly the most attractive in terms of time and money. The predictions of this model are then compared with measurements.

When agreement between measurement data and model predictions is bad, measurements are used to "calibrate" the model. This involves selecting a small number of model parameters that may have values that are different from those used in design. Once these parameters are selected, various procedures are used to find their values for which the measurements best match the model predictions. Such a strategy is identification order 0 (Table 1). This approach is commonly used, for example, to interpret measurement data from load tests on concrete bridges where values for the product of Young's modulus and the moment of inertia (EI) are determined [1]. This strategy has also been used to find coefficients of stiffness matrices for vibration model calibration, as described in [2–4], for example.

Approaches more sophisticated than identification order 0 include the determination of more than parameter values within one general design model. Most work to date involves manual selection of a few likely behavior models using engineering experience. Factors such as values of member stiffness, support conditions, and in-span hinges are varied to obtain several models. Again model predictions are compared with measurements to update parameters. The model that best fits all types of measurements at all locations is identified to be the most likely model. An example of this approach is given in [5]. This is identification order 1 in Table 1.

The two approaches described above are the lower two orders of structural identification (Table 1). Order 0 modeling involves the use of the same general model that was used for design; only parameter values are changed. Order 1 modeling includes evaluation of several manually selected models. Order 0 and order 1 modeling involve attempts to match sensor measurements with model predictions directly. If formulated as an optimization problem, the difference between measurement and prediction is minimized.

Recent work has proposed automatic generation of candidate models. This builds on work performed

Table 1. Orders of structural identification

Order	Generation of candidate models	Determination of model parameter values	Criteria for selection of model classes and parameter values	Support for further measurement
0	No generation—engineers use design model only	Calibration with load test	Predictions = measurements	Weak
1	Several models selected manually	Optimization using measurement data	Predictions = measurements	Weak
2	Automatic generation	Optimization using measurement data	(Predictions − measurements) < maximum error threshold	Fair
3	Automatic generation	Optimization using measurement data	(Predictions − measurements) < error threshold that is fixed probabilistically	Fair
4	Automatic generation	Optimization using measurement data with data mining for feature extraction	(Predictions − measurements) < error threshold that is fixed probabilistically	Good

in the field of model-based diagnosis over the past 30 years [6–10]. Several proposals have been developed and tested for applications in areas such as electrical circuits and photocopy machines. A promising approach for automatic model generation is compositional modeling [10, 11]. Compositional modeling is a framework for constructing adequate device models by composing model fragments that are selected from a model fragment library. Model fragments partially describe components and physical phenomena. A complete model is created by combining a set of fragments that are compatible. For modeling the behavior of structures, fragments represent support conditions, material properties, geometric properties, nodes, elements, and loading. Assumptions are explicitly represented in model fragments so that the model composition module only generates valid models that are compatible with the assumptions chosen by the users.

Model composition makes it possible to search for models containing varying numbers of degrees of freedom. There is no need to formulate an optimization problem in which the number of variables is fixed *a priori*. Models are automatically generated by combining model fragments and are analyzed by, for example, the finite element method to compare their predictions with measurements.

An important aspect of the methodology is the use of stochastic global search algorithms for the selection of populations of candidate models whose predictions match measurements [12]. Other optimization techniques that make use of gradients and sensitivity equations are not used because multiple local minima are common in these search spaces. This type of search is common for all identification orders higher than order 1.

3 ERROR THRESHOLDS

Errors play a major role in the structural-identification process. Errors from different sources may compensate each other such that predictions of wrong models match measurements [12, 13]. Rather than search for exact matches, in such situations, it is only possible to identify those models whose predicted values, when subtracted by the measurement data, fall within error thresholds. When absolute values of these differences are taken, models that generate values below these thresholds are taken to be candidate models.

Modeling and measurement errors have been investigated by a small number of researchers in the field of structural identification. Banan *et al.* [14] stated that the selection of an appropriate model is difficult; it is problem dependent, and usually requires the intuition and judgment of an expert in modeling. For example, mathematical models may not be able to exactly capture variations in cross-sectional properties, existing deformations, residual stresses, stress concentrations, and variations in connection stiffness. Sanayei *et al.* [15], and Arya and Sanayei [16] emphasized that errors in parameter estimates may arise from many sources, the most significant of which are measurement errors and modeling errors. Measurement errors can result from equipment as well as on-site installation faults [15]. A statistical evaluation of the performance of a system-identification methodology must account for modeling and measurement errors.

Modeling error (e_{mod}) is the difference between the predicted response of a given model and that of an ideal model that accurately represents behavior. Modeling error propagation is graphically depicted in Figure 1. Modeling error has three constituents: e_1, e_2, and e_3 [17]. The component e_1 is the error due to discrepancy between the behavior of the mathematical model and that of the real structure. Component e_2 is introduced during numerical computation of the solution of partial differential equations. Component e_3 is the error arising from inaccurate assumptions made during simulation. Such a definition of modeling errors by subdividing it into sources is similar to the delineation of errors in physical system modeling [13].

Component e_3 is further separated into two parts: e_{3a} and e_{3b}. The error part, e_{3a}, arises from assumptions made when using the model (typically assumptions related to boundary conditions such as support characteristics and connection stiffness). The error part, e_{3b}, arises from errors in values of model parameters such as moment of inertia and Young's modulus. While it might be impossible to separate the components in practice, it is still important to distinguish between these errors since the only error source that is usually recognized by traditional model-calibration techniques is e_{3b}.

Measurement error (e_{meas}) is the difference between the real and measured quantities in a measurement. Measurement errors result from equipment as well

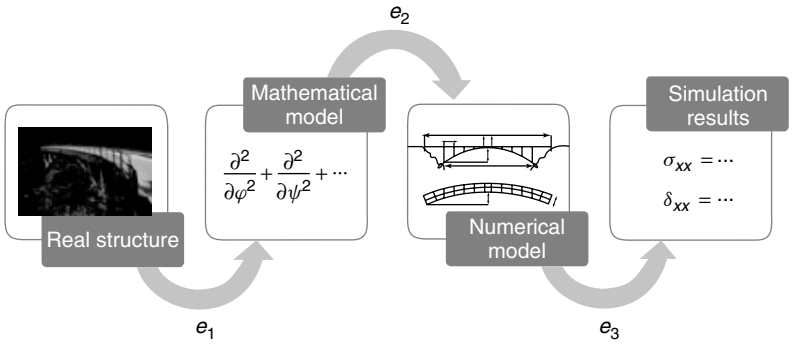


Figure 1. Three sources of modeling errors. e_1, formulation of mathematical models of structure; e_2, representing mathematical models numerically; e_3, simulating numerical models on computers.

as on-site installation faults [15]. In addition to sensor precision values reported by manufacturers, the stability and robustness (for example, with respect to temperature) and the effects of location characteristics (for example, connection losses) also account for measurement error.

The model generation task requires an objective function that accounts for the errors in order to generate a set of candidate models. In Robert-Nicoud *et al.* [12], the objective function is formulated as follows. If x_a is the real value of a behavior quantity such as deflection, x_{meas} is the measured value, and x_c is the value computed using a model, the following relationships have been obtained for a single measurement.

$$x_a = x_{meas} + e_{meas} \tag{1}$$

$$x_a = x_c + e_1 + e_2 + e_3 \tag{2}$$

Model-calibration procedures minimize the absolute value of the difference between x_{meas} and x_c. The difference between x_{meas} and x_c is known as the residue q. Rearranging the terms in equations (1) and (2),

$$q = |x_{meas} - x_c| = |e_1 + e_2 + e_3 - e_{meas}| \tag{3}$$

Thus, model-calibration techniques minimize the quantity $(|e_1 + e_2 + e_3 - e_{meas}|)$. This is equivalent to inaccurately assuming that this quantity is always zero. The objective function that is minimized during the optimization routine is the root mean square error (RMSE), which is calculated as

$$RMSE = \sqrt{\frac{\sum q_i^2}{n}} \tag{4}$$

where $q_i = |x_{i,meas} - x_{i,c}|$ = difference between the value measured at the ith measurement point and the predicted value computed using the model. Any model that gives an *RMSE* value less than a threshold value is considered to be a candidate model. The threshold is computed using an *approximate* estimate of modeling and measurement errors. From equation (3), since errors could be positive or negative,

$$q \le |x_c| + |x_{meas}| \le |e_1| + |e_2| + |e_3| + |e_{meas}| \tag{5}$$

$$\text{Residue, } q \le \text{threshold} = e_{mod}^{est} + e_{meas}^{est} \tag{6}$$

e_{mod}^{est} and e_{meas}^{est} are estimates of the upper bound for modeling errors and measurement errors, respectively. For quantifying threshold, e_{mod}^{est} has been assumed to have a value of $\pm 4\%$ (from finite element simulations) and e_{meas}^{est} was taken (liberally) to be only the precision of the sensor [12]. This is identification order 2 in Table 1. For identification order 2, it is assumed that all candidate models can be generated exhaustively.

4 PROBABILISTIC SUPPORT

While there has been much work on statistics and structural identification [18–20], it is of interest to examine the reliability of structural identification

explicitly. This is done by determining the probability that the correct model is present in the set of candidate models. Absolute reliability (100%) requires that (i) all possible models be considered in the set of models, (ii) there are sufficient measurement data to filter out wrong models, and (iii) all errors are zero.

Estimating the reliability of structural identification involves calculation of a threshold range of errors given a statistical tolerance limit. Many structures can be evaluated using the assumption that, through use of good stochastic search algorithms and high tolerance limits, all possible candidate models are generated. Another assumption made is that enough measurement data is available to filter out wrong models. When these assumptions are not possible, evaluations of reliability that are described in this section provide upper-bound values.

The formulation described earlier for evaluating candidate models is improved by combining errors using statistical methods [21]. In this section, errors are no longer considered to be ranges. Instead, they are values having a statistical distribution. Modeling error e_1 is difficult to quantify. It is problem dependent and can be minimized using intuition and judgment along with modeling expertise [14]. Assuming an ideal situation, $e_1 = 0$. It is assumed that all other errors follow a normal distribution.

Consider x^i_{meas} as the measured value at the ith measurement location and e^i_{meas} as the measurement error at that location. Similarly, x^i_{pred} is the predicted value at the ith measurement location, and ($e^i_{\text{pred}} = e_1 + e_2 + e_3$) is the total modeling error. In the absence of errors, predictions from a candidate model exactly match the measurements. Since errors are present, this is represented in mathematical terms as

$$x^i_{\text{meas}} + e^i_{\text{meas}} = x^i_{\text{pred}} + e^i_{\text{pred}} \tag{7}$$

$$\Delta x^i = x^i_{\text{meas}} - x^i_{\text{pred}} = e^i_{\text{pred}} - e^i_{\text{meas}} \tag{8}$$

Modeling error is defined by a variable e_{pred} with mean μ_{pred} and standard deviation σ_{pred}, and measurement error is defined by a variable e_{meas} with mean μ_{meas} and standard deviation σ_{meas}. Assume that e_{pred} remains the same for one modeling problem. Since values of measurement error depend on sensor type and location characteristics, e_{meas} changes for each measurement location. Many quantities of engineering interest that are not extreme loads generally follow the normal distribution [22]. Assuming both e_{pred} and e_{meas} to be Gaussian distributions, the combined error is defined by a cumulative distribution function Z with mean μ_z and standard deviation σ_z, such that

$$\mu_z = \mu_{\text{pred}} - \mu_{\text{meas}} \tag{9}$$

$$\sigma_z = \sqrt{\sigma^2_{\text{pred}} + \sigma^2_{\text{meas}}} \tag{10}$$

Following from equation (9), the threshold values for a certain reliability of identification (p_{reqd}) are given by

$$r^i_1 \le \left(x^i_{\text{meas}} - x^i_{\text{pred}}\right) \le r^i_2 \tag{11}$$

such that

$$P\left(r^i_1 \le Z \le r^i_2\right) = p_{\text{reqd}} \tag{12}$$

and

$$r_1 = \mu_z - c \quad \text{and} \quad r_2 = \mu_z + c \tag{13}$$

where c is the value that is determined from the required statistical tolerance limit, p_{reqd}.

A function f_i is defined as

$$f_i = \begin{cases} 0 & \text{if} \quad r^i_1 \le \Delta x^i \le r^i_2 \\ \left(\Delta x^i - r^i_1\right)^2 & \text{if} \quad \Delta x^i < r^i_1 \\ \left(\Delta x^i - r^i_2\right)^2 & \text{if} \quad \Delta x^i > r^i_2 \end{cases} \tag{14}$$

where superscript i refers to the ith measurement location.

The new objective function is then defined as

$$E = \sqrt{\sum_1^n \frac{f_i}{n}} = 0 \tag{15}$$

Equation (15) is used to generate candidate models. Differences between measurements and predictions at measurement locations are thus compared with threshold values. The new objective function, E, includes values of errors at each measurement location and provides a probabilistic basis for the reliability of candidate models. This is identification order 3 in Table 1.

5 DATA MINING

Data-mining methods, such as clustering techniques, aid in eliminating incorrect models from candidate model sets and thus they contribute to rapid convergence to the correct model. Integration of data-mining methods is classified as identification order 4 in Table 1. An example of a methodology [23] that combines principal component analysis (PCA) [24] and K-means clustering [25] is used for illustration below.

5.1 Principal component analysis (PCA)

PCA is a method for linearly transforming data in parameter space to a new and uncorrelated feature space [24]. In the machine-learning community, PCA is usually used as a preprocessing technique, for example, before supervised learning. PCA is also used for visualization. It is difficult to visualize clusters when clustering techniques such as K-means are applied to model sets of dimensionality greater than three. PCA finds a set of principal components (PCs) that are sorted such that the first few components explain most of the variability in the model sets. By plotting the first two PCs instead of two randomly chosen parameters, clusters are easier to visualize.

The first step in evaluating the PCs of a data set is the construction of the covariance matrix $\boldsymbol{S}$. The formula for evaluating $\boldsymbol{S}$ is given below.

$$\boldsymbol{S}_{ij} = \mathrm{cov}(x, y) = \sum_{k=1}^{N} \left(x_k - \overline{x}\right)\left(y_k - \overline{y}\right) \qquad (16)$$

$\boldsymbol{S}_{ij}$ represents element at the ith row and jth column of $\boldsymbol{S}$. x and y are the ith and jth parameter and $\overline{x}$ and $\overline{y}$ are their respective means. N is the number of samples. After constructing $\boldsymbol{S}$, its eigenvectors and eigenvalues are found (details are in [24]). The PCs are obtained through sorting eigenvectors in decreasing order of their eigenvalues.

5.2 K-means clustering

K-means [25] is a widely used clustering algorithm that is simple to understand and implement. However, it is useful only when applied and interpreted correctly. The K-means algorithm divides the data into K clusters according to a given distance measure. Although the Euclidean distance is often chosen as the distance measure, other metrics may be more appropriate in certain cases. The algorithm iterates over K clusters to minimize their intracluster distances, shown as the measure J in equation (17):

$$J = \sum_{j=1}^{K} \sum_{x_i \in D_j} \left\| x_i - c_j \right\|^2 \qquad (17)$$

K is the number of clusters, x_i is the ith data point, and c_j is the centroid of jth cluster D_j. The K starting centroids are chosen randomly among all data points. The data set is then partitioned according to the minimum squared distance J. The cluster centers are updated by computing the mean of the points belonging to the clusters. The process of partitioning and updating is repeated until a stopping criterion is reached. The stopping criterion is attained if there is no significant change in the values of c_j or J in equation (17) over two consecutive iterations of the algorithm.

The methodology for grouping models into clusters combines PCA and K-means. Model sets in parameter space are transformed using PCA into an uncorrelated feature space. Next the best number of clusters is estimated using a score function [26]. Once the number of clusters is known, K-means algorithm is applied to the data in feature space.

5.3 Clustering for system identification

The goal of multimodel structural identification is to filter out incorrect candidate models in order to converge to the correct model. Additional measurement locations that produce data, which lead to efficient filtering, are found using cluster information. Figure 2 shows a flowchart with the methodology for system identification assuming that there are several monitoring and interpretation cycles [27]. Engineers provide the structural modeling assumptions that define the parameter set.

The next step, *model generation*, involves creation of a set of candidate models using stochastic search. Measurements, a set of model parameters, and an

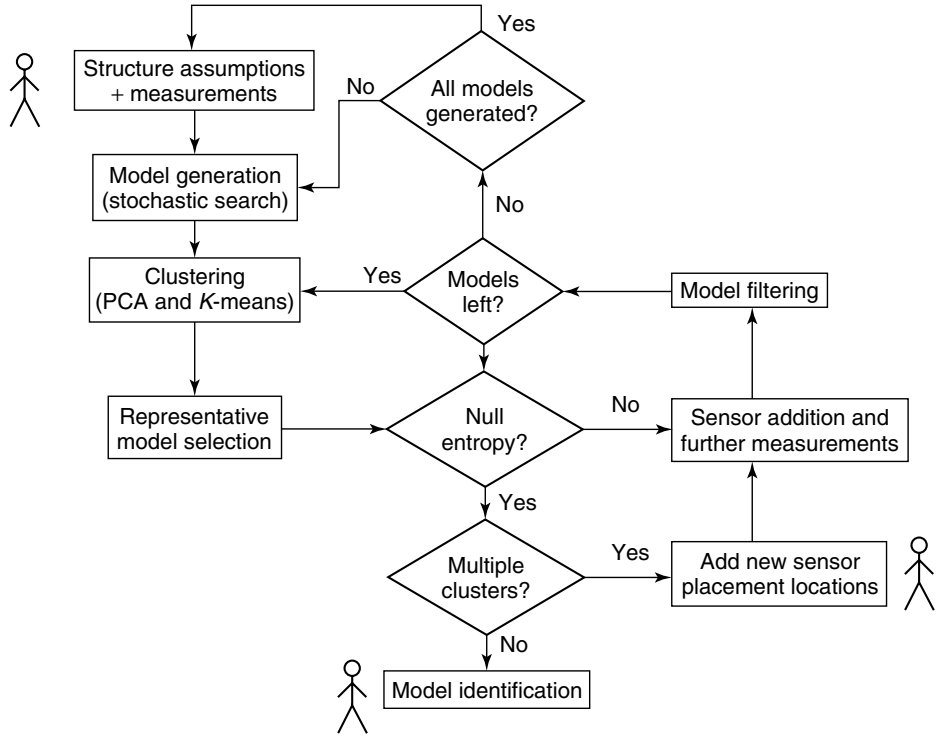


Figure 2. Flowchart showing an example of identification order 4. The human icon means that user interaction is required.

objective function that defines candidate models are needed to generate the set of candidate models.

Once the models have been generated, a clustering algorithm is used to group models through analysis of model predictions. Models are grouped into clusters to (i) facilitate visualization of the model space and (ii) reduce the number of models given to the engineer (the centroid of the cluster is a possible representative model for the entire cluster). Visualization of clusters is improved through the use of PCA. As described earlier, PCA is first applied to models before the K-means algorithm is used.

In the *representative model selection* step, a few models representing each cluster are selected. Only models that are close to the center of the cluster are selected. In this study, 5% of the total number of models in each cluster are taken to be representative models.

After clustering, entropy [28–30] is used as a measure of model separability to identify the next measurement location. Entropy H_s of model sets at measurement location s is given by the following equation:

$$H_s = -\sum_i P_i \cdot \log_2 P_i \tag{18}$$

P_i is the probability of the ith interval in the prediction distribution at location s. P_i is calculated as the ratio between the number of models that have predictions within the ith interval and the total number of models. If model sets have high values of entropy, more candidate models can be filtered.

Additional sensors provide further measurement data. If the entropy of predictions is zero at all measurement locations, data are checked for multiple clusters. If there is evidence of multiple clusters, the current set of measurement locations is incapable of further filtering models. At this point, the engineer needs to provide other measurement types and locations to find the correct model. If there is only one cluster and the entropy is zero, the center of remaining models is given to the engineer as the correct model for the structure (*model identification* step).

During the *sensor addition and further measurements* step, the entropies of predictions of selected representative models are used to find the best position for the next sensor. The location with the highest entropy is chosen. Further measurement of the structure follows.

In the *model filtering step*, sensor measurement at the new location is compared for every candidate model. Candidate models that do not predict the measurement are eliminated from the current set of models. If models are left, then the next step is *clustering*. However, if no models are left, then it is likely that all candidate models are not yet generated by the model generation module. While it may be possible to generate all models for a simple problem at first, it may be difficult to generate all possible candidate models in a complex structure. In this case, the *model generation* phase is revisited. On the other hand, if all models have been generated, then some assumptions related to modeling the structure could be incorrect. Therefore, *structure assumptions* have to be checked and modified by the engineer.

6 SUPPORT FOR DECISIONS RELATED TO FURTHER MEASUREMENT

The activity of filtering candidate models provides opportunities for systematic and rational decision making related to further measurement. In this context, measurement activities can be seen as a way to reduce the number of candidate models to unity. In the last column in Table 1 a qualitative measure of this support is given for each order. If the number of candidate models is small, there is only weak support for choosing what and where to measure next. This support becomes fair when many tens of candidate models are under consideration (orders 2 and 3) [30] and strong when clustering is used in a general framework (order 4) as described in the previous section.

7 CHALLENGES

Technological advances over the past decade have shifted the focus of challenges in structural health monitoring to subfields of structural identification. Challenges have been discussed by Aktan [31] for modal analysis. Taking a more general view, modern challenges related to all types of structural identification are as follows:

- Data management and interpretation (not sensor and computing technologies) have become the dominant causes of the structural-identification bottleneck.
- Measurement interpretation is an inverse engineering task that may have many possible solutions. While models used at the design stage are usually appropriate for design, they are rarely useful for interpreting measurements on existing structures.
- The presence of errors in modeling and in measurements further increases the number of possible solutions. Single-value optimization methods, traditionally used in model updating strategies, are often inappropriate.
- Engineers lack a generally accepted and systematic methodology for selecting sensor type, quantity, accuracy, and position.
- When confronted with decision making related to interpreting and filtering multiple candidate models, engineers can be overwhelmed by the complexity of coping with large decision spaces.

While recent results, such as those reported in the papers listed in the References section, have made progress, these challenges require important long-term multidisciplinary efforts before nonspecialists are able to benefit from the new methodologies that are used in the higher identification orders.

8 CONCLUSIONS

Management of complex civil engineering structures is supported by multiple-model structural identification. Five identification orders classify methodologies. Engineers should select the most appropriate order according to their resources and the levels of decision support that are required for specific infrastructure-management tasks.

ACKNOWLEDGMENTS

The author would like to thank B. Raphael, P. Kripakaran, Y. Robert-Nicoud, Sandro Saitta, and Suraj

Ravindran for their contributions to multimodel structural identification. Some of the work described in this article was funded by the Swiss National Science Foundation.

REFERENCES

[1] Burdet O. Load testing and monitoring of Swiss bridges. *Bulletin d'information Comité Européen du Béton, Safety and Performance Concepts*. Swiss Federal Institute of Technology: Lausanne, 1993; Vol. 219.

[2] Friswell M, Motterhead J. *Finite Element Model Updating in Structural Dynamics*. Kluwer Academic Publishers, 1995.

[3] Doebling SW, Farrar CR, Prime MB. A summary review of vibration-based damage identification methods. *The Shock and Vibration Digest* 1998 **30**(2):91–105.

[4] Brownjohn JMW, Moyo P, Omenzetter P, Lu Y. Assessment of highway bridge upgrading by dynamic testing and finite-element model updating. *Journal of Bridge Engineering* 2003 **8**(3):162–172.

[5] Robert-Nicoud Y, Raphael B, Burdet O, Smith IFC. Model identification of bridges using measurement data. *Computer-Aided Civil and Infrastructure Engineering* 2005 **20**(2):118–131.

[6] de Kleer J, Williams BC. Diagnosing multiple faults. *Artificial Intelligence* 1987 **32**:97–130.

[7] Hamscher W, Console L, de Kleer J. *Readings in Model-Based Diagnosis*. Morgan Kaufmann, 1992.

[8] Console L, Friedrich G. Special issue on model-based diagnosis. *Annals of Mathematics and Artificial Intelligence* 1994 **11**(1–4):1–524.

[9] Struss P. Knowledge-based diagnosis: an important challenge and touchstone for AI. *Proceedings of the 10th European Conference on Artificial Intelligence*. John Wiley & Sons, August 1992, pp. 863–874.

[10] Falkenhainer B, Forbus KD. Compositional modelling: finding the right model for the job. *Artificial Intelligence* 1991 **51**:95–143.

[11] Raphael B, Smith I. Finding the right model for bridge diagnosis. *Artificial Intelligence in Structural Engineering, Computer Science*, *Lecture Notes in Artificial Intelligence 1454*. Springer: Heidelberg, 1998, pp. 308–319.

[12] Robert-Nicoud Y, Raphael B, Smith IFC. System identification through model composition and stochastic search. *Journal of Computing in Civil Engineering* 2005 **19**(3):239–247.

[13] Mahadevan S, Rebba R. Inclusion of model errors in reliability-based optimization. *Journal of Mechanical Design* 2006 **128**:936–944.

[14] Banan MR, Banan MR, Hjelmstad KD. Parameter estimation of structures from static response. II: numerical simulation studies. *Journal of Structural Engineering, ASCE* 1994 **120**(11):3256–3283.

[15] Sanayei M, Imbaro G, McClain JAS, Brown LC. Structural model updating using experimental static measurements. *Journal of Structural Engineering, ASCE* 1997 **123**(6):792–798.

[16] Arya B, Sanayei M. Structural parameter estimation accuracy in the presence of modeling error. *Proceedings of the International Conference on Applications of Statistics and Probability*. Sydney, 12–15 December 1999.

[17] Raphael B, Smith IFC. *Fundamentals of Computer-Aided Engineering*. John Wiley & Sons, 2003, p. 306.

[18] Beck JL, Katafygiosis LS. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics, ASCE* 1998 **124**(4):455–461.

[19] Beck JL. Statistical system identification of structures. *Proceedings of the 5th International Conference on Structural Safety and Reliability*, San Francisco, 1989; pp. 1395–1402.

[20] Sohn H, Law KH. A Bayesian probabilistic approach for structure damage detection. *Earthquake Engineering and Structural Dynamics* 1997 **26**:1259–1281.

[21] Ravindran S, Kripakaran P, Smith IFC. Evaluating reliability of multiple-model system identification. *14th EG-ICE Workshop*. Maribor, 2007.

[22] Jordan I. *Decisions Under Uncertainty—Probabilistic Analysis for Engineering Decisions*. Cambridge University Press, 2005.

[23] Saitta S, Raphael B, Smith IFC. Data mining techniques for improving the reliability of system identification. *Advanced Engineering Informatics* 2005 **19**(4):289–298.

[24] Jolliffe I. *Principal Component Analysis*. Springer, 2002.

[25] Webb A. *Statistical Pattern Recognition*. John Wiley & Sons, 2002.

[26] Saitta S, Raphael B, Smith IFC. A bounded index for cluster validity. *International Conference on Machine Learning and Data Mining*. Leipzig, 2007.

[27] Kripakaran P, Saitta S, Ravindran S, Smith IFC. System identification: data mining to explore multiple models. *3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure*. Vancouver, 2007.

[28] Saitta S, Raphael B, Smith IFC. Rational design of measurement systems using information science. *IABSE Conference*. Budapest, 2006; pp 118–119.

[29] Shannon C, Weaver W. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[30] Robert-Nicoud Y, Raphael B, Smith IFC. Configuration of measurement systems using Shannon's entropy function. *Computers and Structures* 2005 **83**(8–9):599–612.

[31] Aktan AE, Ciloglu SK, Grimmelsman KA, Pan Q, Catbas FN. Opportunities and challenges in health monitoring of constructed systems by modal analysis. *International Conference on Experimental Vibration Analysis for Civil Engineering Structures*. Bordeaux, 2005.

Chapter 131

Construction Process Monitoring at the New Berlin Main Station

Rosemarie Helmerich

Division VIII.2 Non-destructive Damage Assessment and Environmental Measurement Methods, BAM Federal Institute for Materials Research and Testing, Berlin, Germany

1 INTRODUCTION

The six dead-end train stations of the first railway concept in Berlin, developed in the nineteenth century, were almost completely destroyed in World War II (Figure 1). The Lehrter Bahnhof, covered by a true arch roof, being one of the six dead-end railway stations, was destroyed too. After the reunification of Germany and the city of Berlin, the German railways (DB AG) developed a new concept with the New Berlin Main Station as a crossing of trans-European high-speed trains. The optimum location for a main station in the city center was found to be near the suburban train station Lehrter Bahnhof, which was rebuilt after the war.

The architects Gerkan, Marg, and partners from Hamburg developed the architectural design for this most modern and large train station. Professor Schlaich, Bergermann, and partners from Stuttgart are responsible for the structural analysis and design.

In the center of Berlin, the new main station with a wide-spanning modern glass roof has been under construction since 2001 (Figure 2). In 2006, the station was taken into service as an important high-speed train crossing of the two international train corridors from Paris to Moscow and from Rome to Stockholm. The German railways, DB Station and Service AG, represented by DB Projektbau, owned the structure during construction.

1.1 Description of the structure

The New Berlin Main Station is a complex structure. The supports of the interesting glass roof do not lead the load directly to foundations and soil, but use other static systems as the viaducts carrying platforms and tracks. The whole bridge viaduct has a length of about 680 m, and the width of the tracks in the station varies between 33 and 68 m [1]. Two of four reinforced concrete viaducts, carrying the east–west

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Figure 1. The former Lehrter Bahnhof was one of the six dead-end stations forming the railway system in the nineteenth century.

Figure 2. New modern main station Lehrter Bahnhof with flat glass arch roofs during construction in 2003.

directed tracks for the international railway traffic, support the relatively flat glass roof that is up to 60 m wide. The railway station is located in a curve. That is the reason that each of the almost 8000 glass plates has a different geometry.

Partially prestressed concrete bridges in the east–west viaduct are crossing the north–south international passenger train line and have a larger span compared with the other bridges of the station. The prestressed concrete bridges carry the platforms with spans of 18, 21 and 18 meters. The platforms for the north–south and underground line are at about 15 m below the ground level. The partly prestressed, very slender massive concrete bridges of the east–west viaduct are at about 10 m above the ground level. To get sunlight to the underground level, the view through the central station from the top level to the bottom platforms is open, without any through floors. For this reason, 23-m-high steel columns carry the massive prestressed east–west concrete bridges. The steel columns are composed of four single tubes each and form a fork-shaped support for the partly prestressed concrete bridges at their tops.

1.2 Monitoring needs

The steel structure of the glass roof was calculated following the rules of the German national standards DIN 18 800 using partial safety factors according to the modern concept with limit states [1]. During the 1990s, the calculation of the massive concrete bridges followed the rules of the former edition of the German railway standards. The railways still use the concept with allowable stresses and global safety factors. In the level of the supports, a safety factor was applied to compensate for any shortfall in the different safety concepts, the traditional calculation of the concrete superstructure below this level, and the calculation of the glass roof above that followed the new limit states concept. Although additional adapting factors were introduced by the steel specialists at the Technical University Aachen, the Federal supervising authorities for German railways (Eisenbahnbundesamt) required a monitoring system to follow differential displacements between neighboring glass roof supports along the sensitive outer bridges. According to these requirements, the vertical level of the structure must not alter between adjacent glass roof supports by more than 10 mm. A monitoring system should survey the limits of differential vertical displacements. A geometric benchmark system shall be connected to the ends of the monitoring system to have a link between the absolute vertical level and relative displacement data.

2 MONITORING CONCEPT

2.1 State of the art

Monitoring of critical parameters, mainly referred to as *structural health monitoring*, originates from

the airplane and space industries. Continuous data acquisition of critical parameters allows survey of critical areas during changing loading conditions. On-line data availability gives early warning if the data exceed limits to providers of the systems or to the owners. In the late 1990s, it was quite a new concept to apply these ideas to displacements and strains of structures in civil and infrastructure engineering. Furthermore, appropriate long-term stable sensors were not available for all complicated tasks. Sensor development and data acquisition systems are quite cost intensive. Advanced sensor development, but as simple as possible, makes the idea of continuous survey affordable to their application in civil engineering structures. Finally, on-line monitoring for display of data to researchers and owners of the structure was a completely new and higher level of structural survey [2].

2.2 The objective of the monitoring system

A group of scientists at the Federal Institute for Materials Research and Testing, Berlin, BAM, developed a concept for a continuous monitoring system to quickly measure the most relevant data needed for a reliable interpretation of possible changes in the structural condition during construction[2]. The aim was to obtain immediate on-line information about significant differences of displacements and strains in chosen cross sections. The significant period for monitoring was the construction process until the opening of the station in 2006. During the construction process—different load cases from excavation, flooding of the excavated pit, erecting of new structural parts, and dismantling of old structures in the neighborhood—the different load cases cause a continuous change in the structural performance. In some cross sections, the changing loading conditions may cause critical performance scenarios. Therefore, the concrete was cast after the sensor cable tubes were already located in the scaffolding (Figure 3).

The following global load cases were expected:

- loading the steel columns after removing the scaffolding of the concrete bridges;
- surveying the changes during prestressing process in the middle bridges;

Figure 3. The cable tubes for the sensors were fixed in the structure before the concrete (see arrow).

- unloading the immediate vicinity due to dismantling the old suburban train viaduct;
- unloading the surrounding by excavating the north–south tunnel;
- loading the surrounding by flooding the excavation pit with ground water;
- loading the surrounding by casting foundations for the underground track in the north–south tunnel;
- unloading the ground beside the viaducts by removing the ground water;
- completing the track on the bridges and in the underground level;
- completing the framing of buildings.

The function of the chosen measurement system and its long-term stability are contemporarily verified and validated in laboratory monitoring on beams with comparable loading conditions. The information is made available in real time, using the Internet. The Internet collects data of both the sites, at the station and in the laboratory. The software design was specified and developed for this application at BAM.

2.3 Data handling and data transfer

Both the systems, model beams in the laboratory and the sensors at the station, deliver all data also to the central computer, located at the Federal Institute for materials research and testing. Scheme for data acquisition, preprocessing, and transfer is shown in Figure 4. The fiber-optic sensors use the commercial data acquisition amplification and processing. All the other sensors are connected to an amplifier "centipede" (Hottinger) and have a different time basis. Both the systems are connected to the local

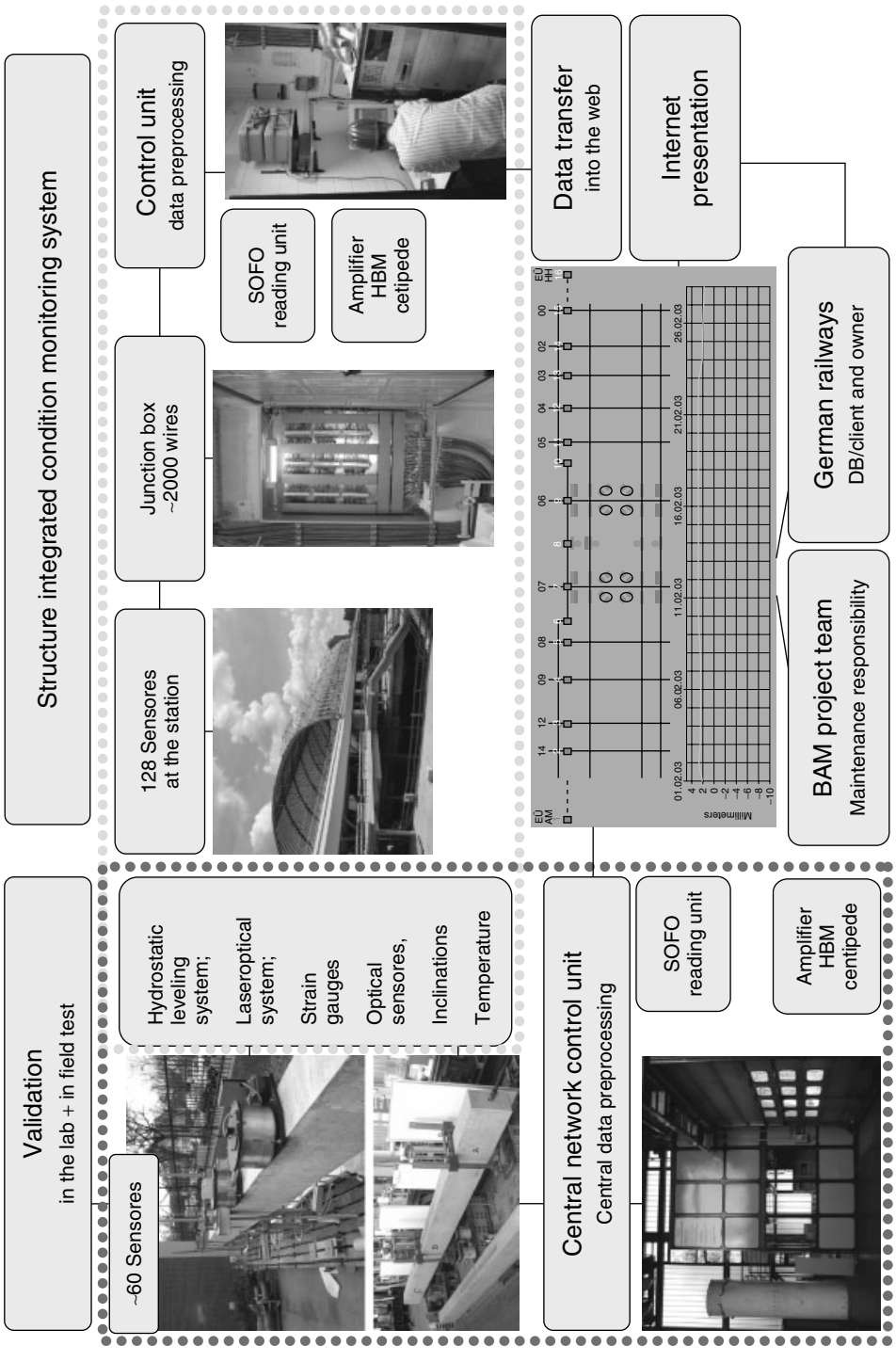


Figure 4. Scheme of the data acquisition, preprocessing, transfer via Internet, and accessibility of the webpage to researchers and end users.

computer and after processing to a unified Internet representation. For central data processing and maintenance of the monitoring system, a central network control unit was arranged at BAM in cooperation with the Engineering laboratory. The central network control unit was the local junction box for the model beams, one located in the laboratory and the other outside the building under environmental conditions. For more information about the model beams, see [3].

2.4 Monitoring system

The continuous survey of vertical movements, strains, inclinations, and temperatures makes it possible to react very fast on early warnings to avoid possible damage. A procedure was proposed to the Federal railway authorities on how to react with corrective decisions and measures in time, i.e., raising or lowering the bridges at supporting points, depending on measurement results. Geodetic measurement of vertical displacements at the supports of the glass roof and at the head of the columns cannot be repeated at sufficiently short intervals owing to the time-consuming procedure and limited accessibility of measurement points [4]. The monitoring system consists of two main elements, installed on both the places, as well at the new main station as in the laboratory beams, to validate the function of the system:

1. monitoring of the vertical displacement at the supports of the glass roof and on the prestressed middle bridge (Figure 5);
2. strain measurement in the partial prestressed outer bridges in the structure of the main crossing (Figure 6).

Figure 5. View of the monitored bridge of the main station during the construction phase. dark: laser displacement sensors; white: hydrostatic leveling system.

As the Federal railway authority required, a redundant strain measurement system, including interference-free fiber-optic sensors, for the monitoring of the partly prestressed bridges at least during the construction period of about 5 years is needed.

The data were collected together with temperature as the environmental parameter, which has an influence on the system. About 2000 single wires in cables are collected in hollow tubes, hidden in the concrete structure, and lead to junction box in a central room at the station. All components of the equipment, such as the electric distributing switchboard and the main computer for data acquisition, are located in the central unit at the station.

The long-term control of changes in the structure was made possible from the beginning of the construction period in 2002. Initially, data were taken manually, before the communication system started working reliably. The positions of sensors in

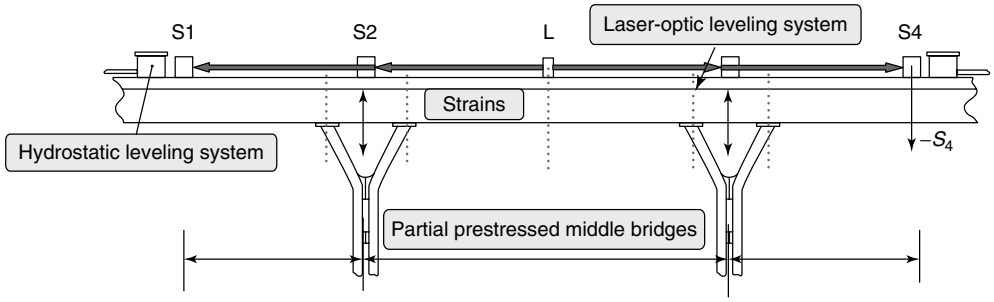


Figure 6. Strain measurement in the partial prestressed bridges and sketch of the measurement system for vertical displacements consisting of laser displacement and hydrostatic leveling components. [Reproduced with permission from Harald Kohlhoff.]

the partial prestressed middle bridges were already prepared during sheeting of the concrete construction to minimize the risk for the structures during construction and demolishing work in the vicinity of the building.

To restrict the number of sensors at the station, it was decided to monitor only the two outer bridges that have additional static loading from the roof structure. To get as much reliable and redundant information as possible, different types of sensors were installed.

2.5 The basic elements and types of sensors

The concept of the field test consists of 128 sensors. Table 1 gives an overview on the sensor types, their location, and measurement uncertainty.

To increase the redundancy of the strain measurement data, different sensor types were installed. The concrete strain measurement—only in the partial prestressed bridges—is performed by means of electric strain gauges, fiber-optic sensors, and mechanic contact strain measurement (type Pfender, BAM). Figure 6 shows the cross sections with strain monitoring in the middle and at the supports.

Contact strain measurement delivered absolute elongation (strain) data in the first phase of monitoring. Unfortunately, the measurement points were removed during the following completion works by sandblasting. Figure 7 shows the cross sections with the strain measurement in the prestressed bridges 12 and 15.

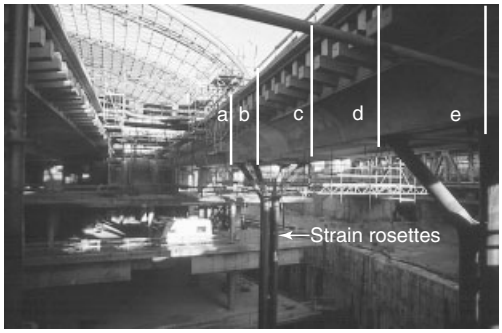


Figure 7. Cross sections with strain measurement points in the prestressed middle bridge and in the upper part of the steel columns. Sensors a–e: fiber-optic sensors and electric strain gauges. [Reproduced with permission from Ref. 5. © IABSE, 2003.]

3 MEASUREMENT SYSTEMS AND ADVANCED SENSORS

3.1 Strain measurement

Strains are measured in

- five cross sections of the partly prestressed middle bridges (north and south) to get information about the near-surface strains, nonsymmetric movement, and/or inclination;

Table 1. Sensors, their location, and measurement uncertainty

Type of sensor	Number	Location	Measurement uncertainty
Strain gauges	40	Bridges 12 and 15	$\pm 3\,\mu m\,m^{-1}$
Fiber-optical sensors		Bridges 12 and 15	
Strain gauge rosettes	16 (+16)	Columns under bridges 12 and 15	$\pm 3\,\mu m\,m^{-1}$
Hydrostatic leveling buoyancy cylinders	30 (+4)	North and south viaduct	Distance: approximately 500 m ±0.3 mm
Laser sensors	10	Bridges 12 and 15	Measurement range ±40 mm ±0.2 mm and add. ±0.1 per 10-m beam length
Temperature sensors	12	North and south viaducts	±0.2 K
Inclination sensors	10	North and south viaducts	Measurement range $\pm 3^{\circ}$ $\pm 0.03^{\circ}$

- the upper part of the steel columns to get information about moments and inclinations.

In the outer concrete bridges, in line with the bridges carrying the roof, fiber-optic sensors, electric strain gauges, and a mechanical strain measurement system (type Pfender) were installed. Since concrete may have local inhomogeneity or cracks, a comparison between locally measured strains by means of electric strain gauges with strain measurement on a longer base line is advisable. For this purpose, fiber-optic sensors are used. For comparability, strain gauges and fiber-optical sensors are located in the same slits and have the same gauging axis.

Like all other data, the strains are measured four times a day. For their near-surface position, the influence of temperature changes is expected to be relatively high compared to other influences as, e.g., symmetric traffic load. The system does not measure the strain distribution inside the massive concrete cross section since the number of sensors was limited.

3.1.1 Electric strain gauges

Prestressed concrete bridges

In five cross sections in the middle of the structure and at the supports of the steel columns, strain gauges (TML) are applied in slits. Electric strain gauges are embedded together with the fiber-optic sensors in slits, relatively close to the upper and lower surface in five cross sections of the prestressed bridges crossing the platforms of the north–south track. All slits were closed after all sensors were installed and the function was validated. Temperature influences the strains very much. In few measurement points, temperature sensors are applied together with the strain gauges.

Steel columns

Two ~25-m high steel columns, each composed of four single columns (see Figure 7), support the prestressed massive concrete bridges. In these cross sections at the supports above the "arms" of the composed steel columns, the maximum strain should occur (Figure 7). The composed steel columns have a hinge bearing on the ground level. Strain rosettes are positioned in the upper part to get information about the loading, inclination, and settlements. For temperature compensation, additional strain sensors were added to each pair of electric strain sensors perpendicular to their axis.

3.1.2 Fiber-optic sensors

Fiber-optic sensors have the advantage of not being sensitive to electromagnetic influences, e.g., from high-voltage cables. The applied, commercially available fiber-optic sensors (SOFO-Smartec) with a length of 0.50–3.50 m work on interferometer principle. In a tube two standard optical fibers are placed and, one of them, the sensing fiber, is fixed at defined points and the other, the interference fiber, is loosely placed in the tube. The second fiber is used to compensate for temperature influences. In real time, the measured data are collected, amplified, stored, and preprocessed by calculation included in the software package. The fiber-optic sensors are connected to a demodulator, which is a one-channel device using a multiplexer for multichannel operation. The fiber-optic sensors have their own time base in a separate receiver [4].

3.2 Vertical displacement measurement

3.2.1 Introduction

The vertical displacement measurement at the outer bridges 12 and 15 consists of 16 sensors each, in a chain of hydrostatic-leveling sensors in the arched sections and the laser-based vertical measurement above prestressed bridges crossing the north-south track. Figures 5 and 8 show the scheme for the location of the measurement points at the supports of the roof on the north and south bridges.

Changes in the vertical level between two neighboring supports should not exceed 1 cm. The relative displacement is measured continuously four times a day. In some places, inclinations perpendicular to the axle of the bridges and the temperature are measured. Geometrical imperfections resulting from displacements and inclinations cause additional loadings in the bridge structures. If the limit value would be attained, then a vertical lifting or lowering of the bridges is needed using the vertical adjustability mechanism.

In the middle part of the bridges between the two roofs, the architects denied installation of the relative huge hydrostatic cylinders for aesthetical reasons. Both the hydrostatic leveling system and the laser-based optical system are installed at the outer bridges and appear in the Internet presentation as one chain.

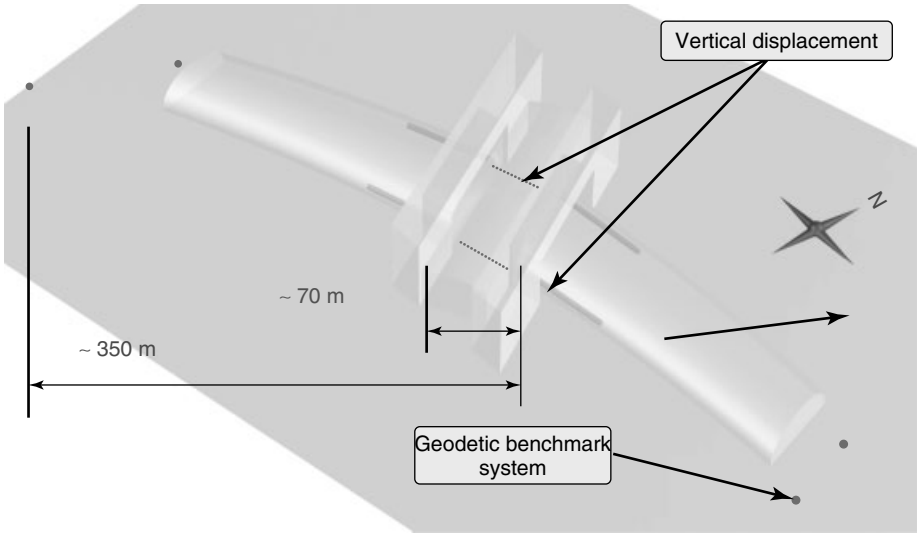


Figure 8. Length of the monitored supports of the outer bridges and connections to the geodetical benchmarks. [Reproduced with permission from Harald Kohlhoff.]

3.2.2 Laser-optic leveling system

On this place, the BAM combines the hydrostatic leveling with an elegant small laser-based sensor system. One laser source emits a vertical laser in the middle of the bridge. Prisms divide and switch the vertical laser beam into horizontal direction to the left and right side along the bridge axis. The widespread lasers meet laterally displaced photodiode chains. Each chain consists of 64 photodiodes; the activated diode is a measure for vertical displacement (Figure 9).

3.2.3 Hydrostatic leveling system

The traditional hydrostatic leveling was improved to obtain a long-term stable chain of sensors to reduce creep and drift of the force transducer. The principle of measuring the displacement of a buoyancy cylinder has been selected for avoiding measurement errors due to influences by partial heating of the sensors, e.g., if exposed to direct sunlight. Raising the liquid level by pumping liquid into the system permits the buoyancy body to separate from the force transducer, as well as checking the operation of the hydraulic system (Figure 10) [6]. This allows to obtain a zero-value in system maintenance. The hydrostatic leveling system automatically records the level of the liquid as a measure about relative vertical displacements. The data is measured as load cell. A benchmark point is connected to the hydrostatic leveling system within the next weeks. The connection to this "fix point" makes it possible to connect the relative settlements with a geodetic measurement, performed on behalf of the German railways (DB AG). The principle of the hydrostatic leveling system was improved to obtain long-term stability with regard to reducing creep and drift of the force transducer.

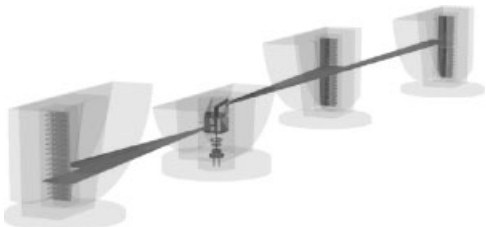


Figure 9. Laser measurement system upgraded by Knapp and Kohlhoff. [Reproduced with permission from Harald Kohlhoff.]

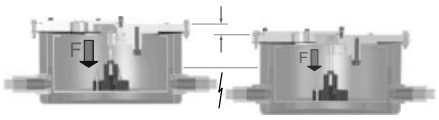


Figure 10. Hydrostatic leveling system type Kohlhoff (patent [7]). [Reproduced with permission from Ref. 7.]

3.3 Maintenance and long-term stability

The chains of hydrostatic sensors on both the bridges have been collecting data since May 2002. During that summer, the data was transferred periodically. Sensors used for construction monitoring must be robust. Construction work and finishing may damage sensors or cables. The ongoing construction work prevented the connection of the measurement points to fixed points. The continuous transfer of data for online presentation on the internet website was available from October 2002 (Figure 11).

During the 5 years of monitoring, the combined system for the vertical displacement, consisting of the hydrostatic leveling and the laser-based leveling system, was maintained several times. The advanced hydrostatic leveling allows to validate the measured data by pumping the fuel into an external pot. This procedure unloads the load cells, after which a real zero-value can be read. Alternatively, mechanical measurement of the fuel level is possible to compare the on-line with real data. In case of dysfunction, some of the sensors can be replaced by new ones. Others may be in non-accessible positions. During harsh conditions of construction processes, accidents can lead to disturbance of the measurement chain. Since the access is limited in the station that is in service, maintenance is almost impossible.

4 RESULTS

The monitoring system was installed before the beginning of the extreme loading conditions, during construction activities in the vicinity of the east–west viaduct.

In 2003, the trains were shifted from the old track to the new track through the new station. Before the first train was shifted, the German Railways made a proof load test, with measurements of displacements. It was possible to increase the data acquisition rate for the BAM monitoring system from only four data per measurement point to about 80 Hz to get information about the strains under traffic. Fiber-optical sensors and the leveling system were not used for these measurements, since the configuration for these systems was appropriate for long-term measurements with only a few data collected per day.

4.1 Relative settlements at the roof supports

The main objective of the monitoring system was controlling the differences between the displacement of neighboring roof supports. During demolition of the old station (Figure 12), excavation of the north–south tunnel, and flooding of the excavation pit (Figure 13), the east–west viaducts carrying the

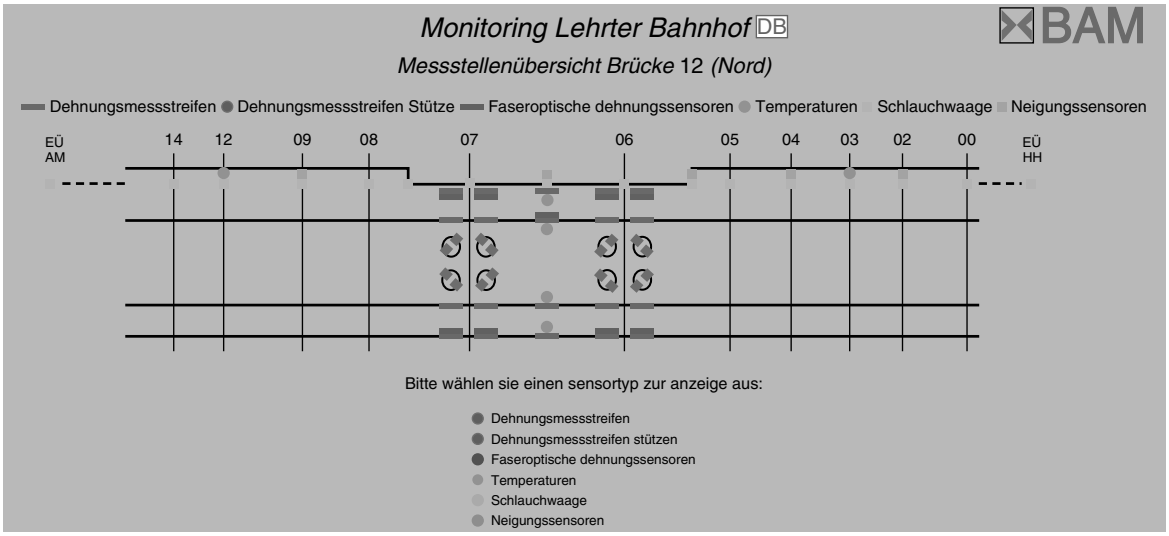


Figure 11. Website with the position of the sensors in the north bridge (bridge number 12).

glass roof were exposed to load differences. Figure 14 shows the differential in settlements and heaving with relative displacements during the loading and unloading phases in the vicinity of the station from May 2002 to July 2003. The positive result was, that over the whole construction, required displacement limits were not exceeded, not even in the year 2003, during the hectic period of dismantling of the old structure and excavating the north-south tunnel (Figure 13).

Figure 12. Load case: demolition of the old station.

Figure 13. Load case: flooding of the excavated pit with ground water.

Figure 15 shows the differences in relative vertical displacements between neighboring roof supports obtained from the chain of the hydrostatic leveling and laser-optic system, for 1 month, as an example. For this purpose, the relative displacement is compared to a mean value of all sensors, including both the leveling systems, by using data displayed from the website.

4.2 Strain measurement

The monitoring of strains is of special interest in partially prestressed bridges. As in the measurement of relative displacements, the measured data have been reliably collected, amplified, stored, and preprocessed by calculation in real time for 5 years. Although both the systems, the electric strain gauges and the fiber-optic sensors, have a different time basis, both the data sets are given in a unified Internet representation.

The analysis of data showed, e.g., in the first year a maximum strain difference of 585 μm. For a maximum temperature difference of 40 K, it results in a mean value of about $10\,\mu\mathrm{m\,m^{-1}\,K^{-1}}$ for the gradient.

The measured values are influenced by many factors, such as shrinking, creeping, nonlinear temperature distribution in the massive concrete cross section, and the fact that a real zero-value cannot be measured again.

The measurement began immediately after the structure was erected, partially prestressed, and the shrinkage was already in process. These values were estimated from calculation and from evaluation of the model beams. Figure 16 shows the course of strains measured with fiber-optic sensors in four points of the prestressed bridge no. 15 for 5 years. The curves with

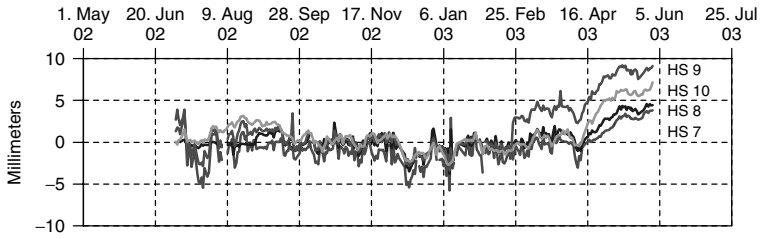


Figure 14. Settlements and heaves during demolition of the old station, excavation of the north–south train line, and flooding of the excavated pit.

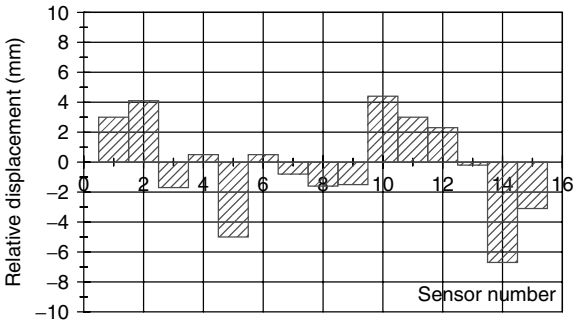


Figure 15. Example for relative displacements for 1 month in March 7, 2003 by means of the hydrostatic and laser measurement at the south bridge.

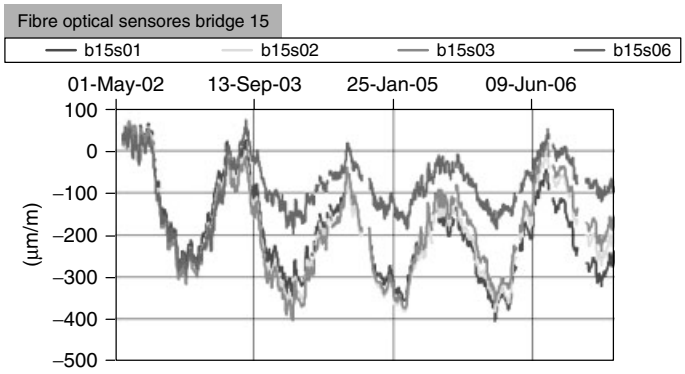


Figure 16. Continuous strain data measured with fiber-optical sensors for 5 years in the partial prestressed bridges. The compression during winter was lower since 2006 because the middle roof was closed and the temperature influence was reduced.

the daily midnight temperatures in three points of the same bridge show that the influence of temperature on the main strain is dominant (see Figure 17). Also it shows, of course, that the temperature and, as a result, also the strains did not reach the negative temperatures, as before. Two reasons are responsible for this result: the warm winter 2006/2007 and the finalizing of the roof. In 2006, the middle roof was built in such a way that the bridges are now located inside and are not exposed to harsh environmental conditions anymore. The strains are lower, which means the structure is safer.

4.3 Long-term stability and maintenance

Continuous control of the stable function of all sensors and the load cells is needed. During first year in service, only three sensors failed (less than 3%), a temperature sensor and two electric strain gauges. Since all slits with the strain sensors are closed, no access is possible anymore.

It can always happen during construction processes that scaffoldings destroy sensors or application of secondary elements hit a cable, e.g., by drilling. Repair and maintenance of systems of this logistic extend require experienced specialists. That is why the access to all sensors, especially to sensors for vertical displacement, is advisable. Reduced maintenance may affect the long-term stability and service of the monitoring system.

5 CONCLUSIONS AND OUTLOOK

The presented monitoring system at the Berlin Main station has been reliably working through all construction phases. The system provided continuous

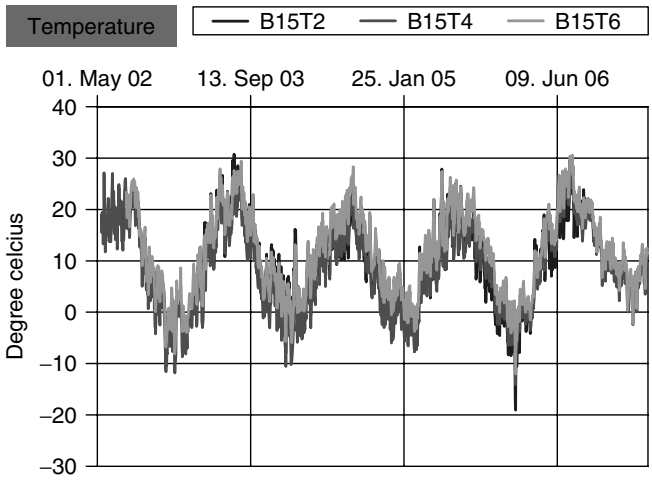


Figure 17. Continuous temperature measurement for 5 years (only midnight values).

on-line data for sensitive parameters of the glass roof. The limits, stated by the railway authorities, as for relative vertical displacements between neighboring glass roof supports, have not been exceeded. The sensors have been working stably for 5 years. The system needs continuous maintenance by experienced specialists.

Measured data are immediately preprocessed and have been presented on line via Internet. Thus, the researchers and the client, the German railways, DB AG, always have the opportunity to identify possible changes of critical parameters at any time.

ACKNOWLEDGMENTS

The German railways (DB) were financing this monitoring system. We thank DB for the confidence in the implementation of advanced techniques and new sensor prototypes, which have never been used before. We thank numerous colleagues from the Federal Institute for Materials Research and Testing, who carried out the research and enabled the realization of the project. We acknowledge especially Klaus Brandes and Wolfgang Habel, who developed the idea and concept to continuously monitor the construction process during all phases. Klaus-Dieter Werner designed and realized the software for the on-line data presentation, and Hans-Joachim Peschke cared for the data preprocessing and transfer. Without the upgraded sensors for the widespread multipoint laser measurement with several lateral displaced measurement points from Juergen Knapp and the patented hydrostatic leveling system by Harald Kohlhoff, the data would not have been so convincing. All systems worked reliably and stably. Joachim Niemann made continuous efforts to keep the system running for over more than 5 years.

REFERENCES

[1] Albrecht G, Klähne T, Stucke W. Aspects of the structural examination of the project Lehrter Bahnhof, Berlin, in German: Aspekte der bautechnischen Prüfung des Bauvorhabens Lehrter Bahnhof, Berlin. *Stahlbau* 2002 **71**:Heft 12: S-890–S-903.

[2] Knapp J, Brandes K, Werner K-D. Optical monitoring system for settlements and inclinations. *Proceedings, IMEKO 2000*. Wien, 2000.

[3] Ullner R, Helmerich R, Knapp J. Laboratory model test for monitoring the new main station of Berlin, Lehrter Bahnhof. *Proceedings of the 1st International Conference on Reliability and Diagnostics of Transport Structures and Means*, ISBN 80 7194-464-5. Pardubice, September 2002.

[4] Habel W, Kohlhoff H, Knapp J, Helmerich R, Hänichen H (DB Projekt Verkehrsbau GmbH Berlin), Inaudi D (Smartec SA, Manno/CH). Monitoring system for long-term evaluation of prestressed

railway bridges in the new Lehrter Bahnhof in Berlin. *Proceedings of the 3rd World Conference on Structural Control*. Como, 7–12 April 2002; Vol. 2, S-713–S-719.

[5] Helmerich R, Kohlhoff H, Werner K-D, Niemann J. Structural condition monitoring of a high-speed train station. Keynote presentation at *IABSE Conference*, ISBN 3-85748-109-9. Antwerp, 2003.

[6] Niemann J, Habel WR, Hille F. Complex monitoring system for long-term evaluation of prestressed bridges in the new Lehrter Bahnhof in Berlin. *Proceedings of the 2nd International Conference on Reliability and Diagnostics of Transport Structures and Means*, ISBN 80 7194-769-5. Pardubice, July 2005.

[7] Kohlhoff H. *Hydrostatic Levelling System, Type BAM*, Patent No. 10203231, April 2003.

Chapter 133
SHM of a Tall Building

James M. W. Brownjohn[1] and Tso-Chien Pan[2]

[1] *Department of Civil and Structural Engineering, University of Sheffield, Sheffield, UK*
[2] *Nanyang Technological University, Singapore*

1 INTRODUCTION AND CHRONOLOGY

Structural health monitoring (SHM) is frequently associated with damage detection, often using vibration-based techniques. In fact, damage detection represents only a fraction of the spectrum of SHM technologies, which as a whole comprise characterization of structural performance through response measurements, with the aim of learning about the structure *and* the loads that drive it. Misunderstanding the nature of loads on a structure is often a cause of structural "failure"; examples include unstable behavior of bridges due to wind or pedestrian loads, scour of bridge piers, and various forms of seismic-induced structural failure. Hence, SHM focused on characterizing loads is highly relevant, particularly if it can be used to feed information on loading characteristics back into future designs. That is the philosophy of the exercise reported here, in which a new tall building was monitored with the principal aim of providing information on the unknown characteristics of loading.

Singapore is not known for strong winds, nor is it in a seismic area. However, powerful storms have been known to remove aircraft hanger roofs and flip shipping containers over, while the most powerful earthquakes on the planet occur within a few hundred kilometers of a vulnerable financial center that has seen rapid growth in construction of high-rise structures. Local design codes borrowed from the United Kingdom have little relevance yet there is no better guidance, and hence the need to

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

provide information about wind and seismic loads by monitoring performance of a tall building.

In 1993, Shimizu Corporation instrumented the 18th story of a new 65-story 280-m office tower under construction in Singapore with an array of stress and strain gauges and hired Nanyang Technological University researchers to report on the instrument readings. Given the access to the structure, it was also possible to use a simple acceleration recorder to track the parameters of the two lateral fundamental modes of vibration during construction.

When the building was completed but unoccupied at the end of 1995, a full-scale vibration survey with four high-resolution accelerometers was conducted. The results of the operational modal analysis on the data were used to update a finite element model of the structure for interpreting future performance data.

In 1996, a monitoring system was installed, comprising biaxial accelerometers at the basement and roof levels (total of four channels) as well as a pair of anemometers perched on the building parapet.

For several years, the system recorded wind and vibration response, including data on major seismic events in neighboring Indonesia and on the different classes of wind loading. In order to resolve questions about the mixture of static and dynamic response of the building under local wind conditions, a pair of global positioning system (GPS) antennae was installed in order to identify absolute wind-induced displacements.

The system operated until 2005, and since the most useful data obtained concerned seismic response, a simpler four-channel acceleration recording system was installed later. Performance data obtained from the system operation up to 2005 have been useful for preliminary developments of locally oriented design guidance for wind and seismic loads.

2 BUILDING CONFIGURATION

Figure 1 shows a perspective view of the Republic Plaza building, and Figure 2 shows a typical cross section at story 18. The structure is built around a reinforced concrete (RC) central core with almost a square profile of side 22 m that extends for the full height of the building. The perimeter of the building, occupying a 45-m square footprint, comprises 16 steel tube columns around 1-m diameter that connect with the core via a horizontal framing system and are concrete-filled up to the 49th story.

Figure 1. Perspective view of completed building.

Two double-story mechanical equipment floors are located at storys 28 and 47, where the building profile tapers and outriggers are installed for controlling wind-induced drift. Apparent axes of symmetry, labeled A and B in the horizontal and vertical axes of Figure 2, are used for reference in describing the performance measurements.

The whole structure sits on a very rigid foundation system that includes caissons founded up to a depth of 62 m in marine (boulder) clay.

3 TRACKING NATURAL FREQUENCIES DURING CONSTRUCTION AND AMBIENT VIBRATION SURVEY

During the latter period of building construction from June 1994, a portable acceleration recorder was used

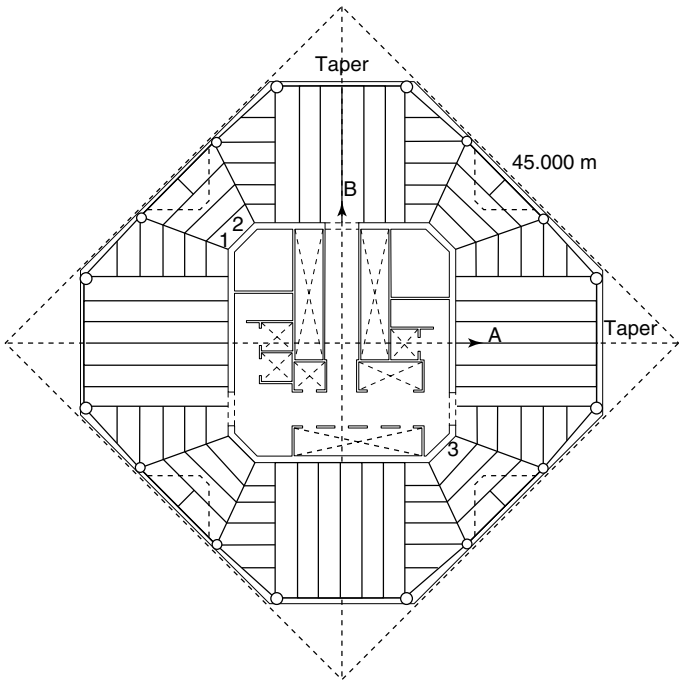


Figure 2. Typical cross section at story 18. Coordinate system for measurements is labeled A and B.

to record horizontal accelerations in the natural A and B axes at the same time as static stress and strain gauges were read. Pairs of first-mode frequency values were obtained from curve fitting to autospectra of the signals showing doubling of both periods up to the point at the end of 1995 when a formal ambient vibration survey (AVS) [1] was undertaken.

The variation, during construction, of period for fundamental modes in each direction, i.e., A1 and B1 is shown in Figure 3 together with mass (above story 18) and construction progress, as indicated by highest story of completed office slab. It is clear that while direction B developed into the stiffer of the two directions, due to structural arrangements inside the core wall, frequencies were originally identical.

The procedure for the AVS in late 1995 involved estimating frequency and damping values by fitting curves of single degree of freedom oscillator response functions around autopower spectral peaks for many data sets acquired over 5 days of measurements at different locations in the building. Then phase angles and amplitudes of cross-power spectra of rooftop response were compared with the response at every second floor level down to the basement at the

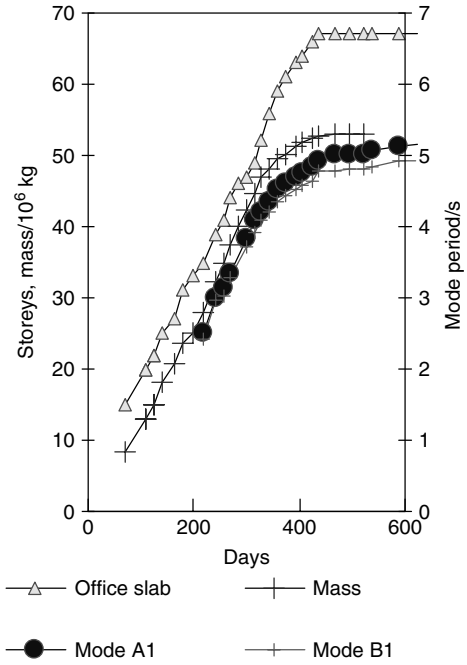


Figure 3. Variation of fundamental mode periods during construction.

estimated modal frequencies to map out mode shapes (in fact operating deflection shapes).

The measurements showed that modes did not divide exactly into the natural A and B directions, rather that the principal axes of movement were rotated unknown angles with respect to these obvious symmetry axes. Also, there was evidence of significant torsional response even in the translational modes, so a set of four measurements were made, one at each of storys 18, 32, 46, and 65, to identify the unknown angles and mode shapes in horizontal planes using an arrangement of the set of four accelerometers.

A complete set of 12 modes was identified. Modes are numbered A1 through A4 for modes aligned closest to the A direction, B1 through B4 for modes aligned closest to the B direction, and T1 through T4 for modes of almost pure torsion. Modes A1, A2, A3, and T1 are illustrated in Figure 4 together with an autospectrum of A-direction story 65 acceleration obtained during the AVS. The modal parameter estimates are given in Table 1 from the AVS as well as from a period in early 1997 when the building was occupied.

The modal survey data were subsequently used for validating a finite element model constructed

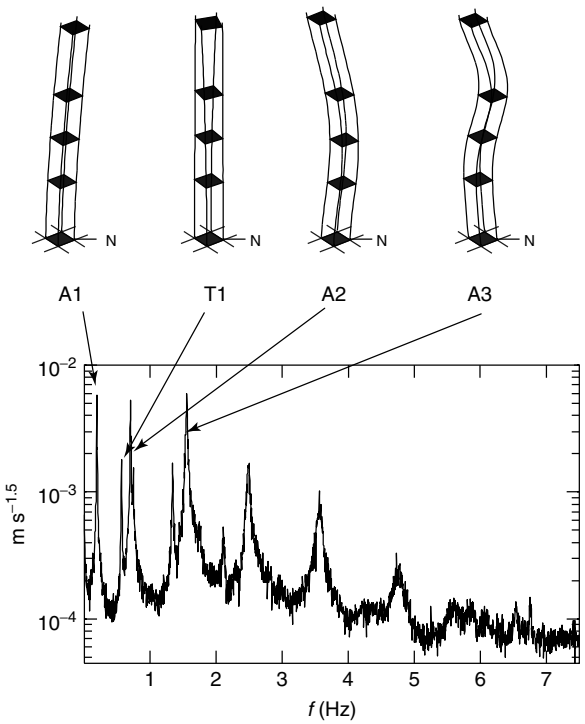


Figure 4. Vibration autospectra at story 65 and measured modes.

Table 1. Modal parameters for Republic Plaza: unoccupied in 1995, occupied in 1997

A mode	Frequency (Hz)		ζ (%)	B mode	Frequency (Hz)		ζ (%)	T mode	Frequency (Hz)		ζ (%)
	1995	1997	1997		1995	1997	1997		1995	1997	1997
A1	0.192	0.184	0.66	B1	0.201	0.194	0.70	T1	0.564	0.528	0.53
A2	0.702	0.676	0.85	B2	0.749	0.726	0.52	T2	1.341	1.258	1.25
A3	1.553	1.490	0.87	B3	1.739	1.690	0.77	T3	2.309	2.205	1.65
A4	2.486	2.403	0.74	B4	3.004	2.904	1.67	T4	3.329	3.140	1.59

using the SAP2000 finite element code [2] by comparing predicted and measured characteristics and carrying out systematic modifications of the model.

4 WIND AND ACCELERATION RECORDING SYSTEM

Long-term monitoring of Republic Plaza followed the experience with a wind and acceleration response recording system installed previously in a 26-story apartment block in Singapore [3] that operated until the end of 1995, when the equipment was removed for use in the Republic Plaza AVS. This preliminary study of the apartment block provided strong evidence that earthquakes occurring in Indonesia were inducing significant and alarming but (so far) nondestructive vibration response in residential structures in Singapore. Because the purpose was to study wind-induced response, the ground motions could only be inferred from the recorded rooftop response. The Republic Plaza study provided the opportunity to record foundation level acceleration measurements using cables installed during construction. Thus, it was intended that over the long term, synchronous roof and foundation level acceleration recordings could be used to capture ground motions and their effects on the building.

Hence, from October 1996 until January 2005, acceleration signals were recorded and analyzed with a few interruptions due to hardware issues. Initially, two accelerometers were used and while a second pair was being procured the two-channel system was left to record continuously for 2 weeks. The aim was to check expected response levels for optimizing system dynamic range, but during the 2 weeks the response due to a strong (Ms6.3) Indonesian earthquake at a relatively close epicentral distance of 700 km was recorded, being the first recorded time series of building structural response in Singapore.

Sensitive quartz-flexure force-balance accelerometers having noise threshold as low as 1 μg were used. Acceleration signals were digitized with a 12-bit analog-to-digital converters and saved in frames of 4096 samples acquired at 7.5 Hz. In fact, signals were oversampled in short frames after analog low-pass filtering and decimated in order to benefit from the sharper cutoff characteristic of digital filters in the acquisition software. The system used double-buffering for simultaneous acquisition (in one buffer) and processing (in the other). The processing involved calculation of FFTs and various statistical properties of the signals so that for every frame of approximately 9-min duration a set of parameters describing mean, variance, and narrow-band root mean square (RMS) corresponding to known vibration modes were stored.

Data frames were saved together with trigger indicators for interesting events. Acceleration trigger conditions, applying only to story 65 accelerometers, included strong broadband response as well as response in lower vibration modes measured using narrow-band RMS.

Trigger levels were also set relative to a moving average of response so that weak signals could be captured against the weaker background of building response to both wind and internal machinery at night. After a short learning period we observed that while wind generated strongest acceleration response in the fundamental vibration mode (A1 and B1), distant earthquakes invariably caused the strongest response in the second mode (A2 and B2). Hence triggering on second-mode RMS was found to be very effective for capturing local effects of distant earthquakes [4].

From almost 5 years of data up to July 2001, daily maximum RMS acceleration values due to wind excluding known earthquake response were obtained and used to obtain statistical distributions of extreme values, presented as Gumbel plots in Figure 5. These plots confirm the relative weakness of A direction and that, even allowing for statistical peak factors (ratio of peak to RMS) around 3.0 for a 10-min record, the serviceability criteria of 98 mm s^{-2} (0.01 g) established for the building by the architect are comfortably higher than the acceleration response expected for a 50-year return period.

Even the strongest ever recorded response to wind, with 15 mm s^{-2}-peak amplitude and peak gusts between 20 and 30 m s^{-1}, was dwarfed by the acceleration response due to the magnitude 8.0 Bengukulu earthquake that occurred in June 2000. Apart from the December 2004 Aceh earthquake, this was the strongest earthquake in the region during the monitoring and it generated the strongest recorded signals, probably clipping at the system range of 50 mm s^{-2}. A summary of characteristics and effects of recorded tremors up to 2000 is given in [4].

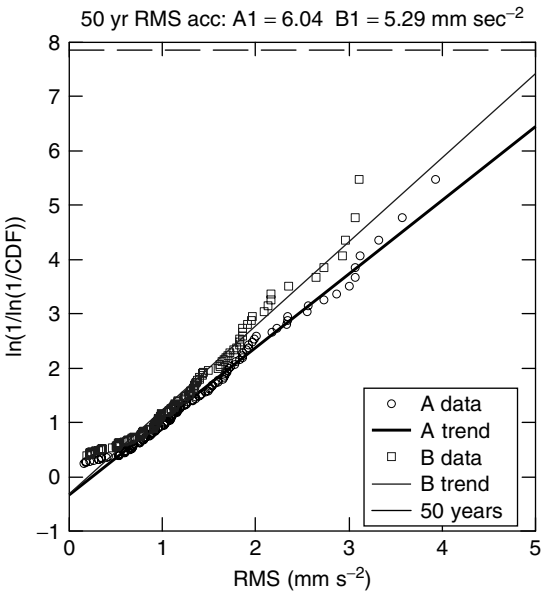


Figure 5. Gumbel plots of story 65 first-mode RMS accelerations.

From an ensemble of large amplitude response time histories collected during strong winds, we established that for unity normalization at the roof, mode-shape amplitudes for the first three modes are 0.006, −0.010, and 0.016, showing the foundation to be extremely rigid. Hence we believe that the recorded basement signals can be taken as representative of local ground movement, with the triggering algorithm working very well to detect weak earthquake events. Moreover, the transmissibility function between basement and roof, (Figure 6) allowed for identification of modal frequency and damping through classical modal analysis procedures such as circle fitting, confirming the results given in Table 1.

Besides discrete events, we tracked modal parameters over the 10 years of monitoring and found a gradual degradation of stiffness; for translational modes A2, B2, A3, and B3 the frequencies drop on average 0.65% per annum.

5 RECOVERY OF DISPLACEMENTS FROM ACCELEROMETER SIGNALS

With the development of a Singapore-specific wind loading code at the time of the project, Republic Plaza provided an opportunity to calibrate the candidate code provisions.

As with a UK exercise [5] in which an empty apartment block was used to study total static and dynamic response to wind load, and the subsequent Chicago Project [6], Republic Plaza provided a convenient form of wind sensor. Wind loading comprises static and dynamic components, and modern wind loading codes deal with the effect via a dynamic magnification factor multiplying the static effect of mean wind and including the effect of both broadband turbulence and response in first-mode resonance. Calculations based on the Australian wind code [7] give a dynamic

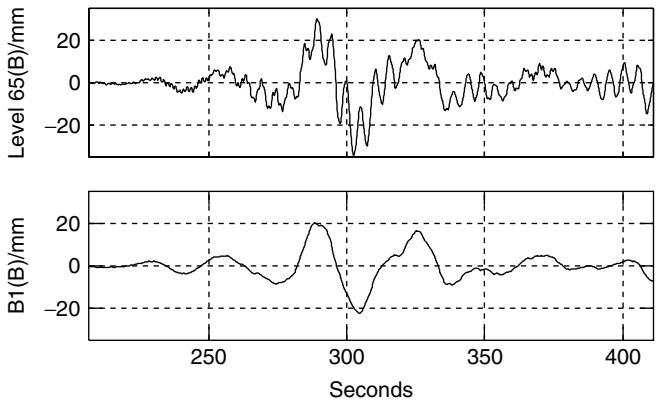


Figure 6. B-direction displacements at basement B1 and story 65 recovered by integrating accelerations resulting from "great earthquake" near Singapore in 2000.

amplification factor close to two, but this depends on the character of the wind, particularly the turbulence, which is relatively high for Singapore.

Apart from issues of occupant comfort during first-mode sway due to wind turbulence, it is the total wind force that concerns designers. The resonant component of base shear can be found by combining known mode shape and mass distribution with the rooftop acceleration signals, but the static component depends on a combination of the stiffness (known from the validated finite element model) with measured absolute mean deflections.

Absolute static displacements cannot be obtained even from the best accelerometers because double-integration of acceleration to displacement is subject to contamination by instrument noise at very low frequencies, translating to high displacements.

Figure 7 shows B-direction displacements obtained by integration of accelerometer signals from the 2004 Bengukulu earthquake and aftershock, using a cutoff frequency of 0.02 Hz. Apart from the obvious first-mode contribution of story 65 response, the two signals are identical indicating a rigid body motion of the entire building. Using a lower cutoff frequency in the integration introduces low-frequency ripples in different channels at different times due to noise, so the result also fixes a reasonable lower cutoff frequency for general application of the integration process and effectively rules it out for the very slow quasi-static wind-induced response.

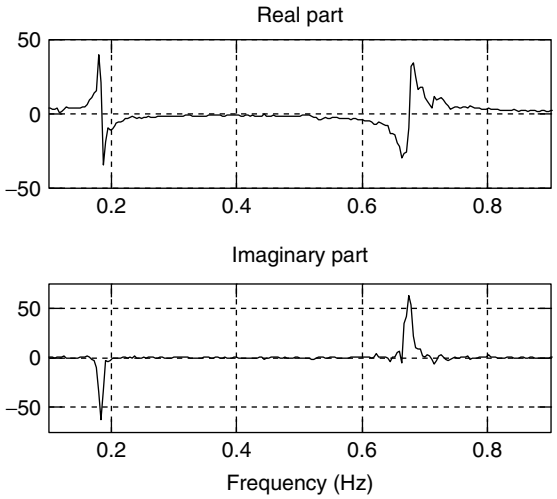


Figure 7. Transmissibility function between level 65 and basement response during an earthquake.

An alternative method was investigated, involving the relationship of tilt to deflection at the roof and interpreting DC acceleration signals as a component of gravity due to rotation of accelerometer sensing axis from horizontal. Even allowing for compensated thermal bias, displacements thus recovered did not correlate with any measured "effects", e.g., wind and differential solar heating, and hence the technique was abandoned.

6 GLOBAL POSITIONING SYSTEM

With a proven track record in recording dynamic structural deflections in suspension bridges [8], the GPS offered possibilities of absolute position measurement for resonant, dynamic nonresonant, and static responses. Hence, in 1999 the acceleration and wind recording system was upgraded to include a dual-rover real-time kinematic (RTK) differential GPS.

For a single GPS antenna/receiver (rover), estimates of position are subject to errors that depend on modification to signal transit time from the satellite due to atmospheric and other effects. If a fixed base or reference (base) station is located nearby, given the known fixed location, the "differential" errors can be identified and used to adjust the rover position estimate. When the errors are transmitted from base station to rover and incorporated in rover position estimates in real time, position fixes accurate to the order of a centimeter or better may be possible. RTK operation uses software embedded in the receivers to correct for the errors. More recent GPS capabilities allow for direct supply of corrections without the need for a base station using corrections supplied over the internet from a collective "virtual" base station.

The system at Republic Plaza used RTK solutions output at 1 Hz for each sample as NMEA (ASCII) text data. Text data is problematic to fuse with signals from analog sensors and hence the NMEA data were converted to analog signals and fed into the existing acquisition system with the analog signals.

The rover antennae were positioned flush with the level of the parapet, and the base station was located on a low-rise building 10 km away. The antenna configuration was not ideal as it was not possible to

use larger, more expensive, and more accurate choke ring antennae.

The initial GPS data appeared very noisy and not to correlate with other signals, and hence we found it hard to believe what the data represented. Validating the GPS data was a major issue, as the signals were subject to various forms of error such as multipath, cycle-slip, random noise, and systematic noise. Also, the total movements of the building, expected to be of the order of ± 0.1 m, would comprise components of dynamic and static response to wind as well as static response to temperature changes in and around the building. We obtained direct evidence that the system was working first by physically moving the antenna during a recording and then by studying the signal during strong winds and/or earthquakes that generated first-mode deflections of at least 1-cm amplitude.

Two periods in early 2003 and early 2004 generated useful GPS data during the relatively strong seasonal monsoon winds. However, the signals were affected by dropouts occurring at slightly irregular intervals of approximately 30 min. A good example of typical data is shown in Figure 8, representing a storm generating first-mode response with an amplitude 10 mm s^{-2}. The RTK data shown were corroborated by the second rover and indicate a steady drift at the roof level of the building with a kind of ratcheting, together with dynamic response.

Surprisingly, the best validation of the GPS performance was provided by the December 26, 2004 earthquake off the coast of Sumatra, Indonesia. Figure 9

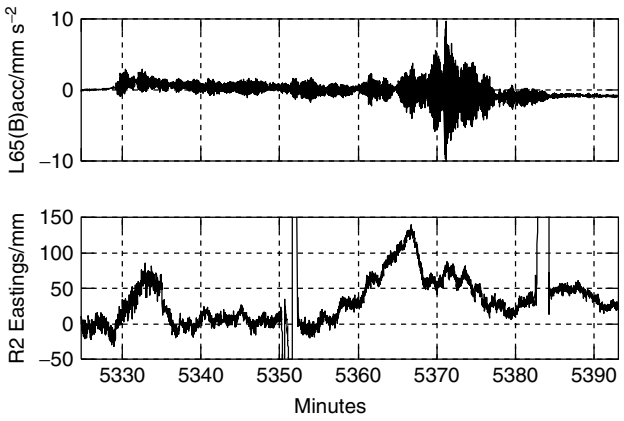


Figure 8. Sampled time series (1 Hz) of story 65 accelerations and RTK displacements during a storm.

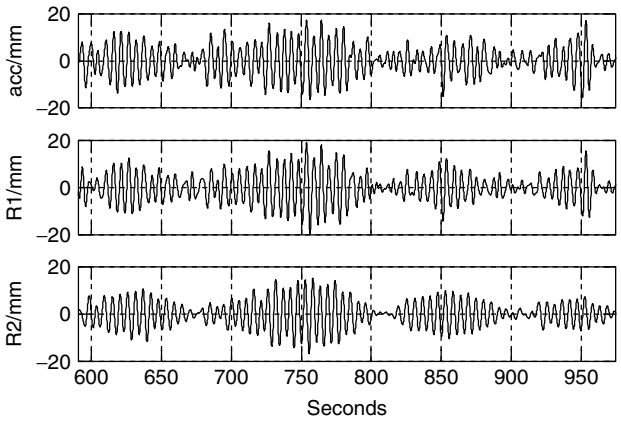


Figure 9. Evidence of GPS performance: correspondence of A-direction relative displacements recovered from accelerometer (acc) compared with RTK-GPS displacements (R1 and R2) during Aceh earthquake on December 26, 2004.

shows story 65 displacement in A direction relative to basement recovered from the acceleration signals, and also displacements obtained directly using RTK from the two rovers. The RTK displacements are the A-direction component of the vector sum of Eastings and Northings, band-pass filtered between 0.1 and 0.3 Hz.

While story 65 acceleration can be integrated to show the type of large low-frequency displacements visible in Figure 7, RTK displacements do not, probably because they were relative to the base station, which moved to the same extent as the ground at Republic Plaza.

7 OBSERVATIONS ON STRUCTURAL AND LOADING MECHANISMS

From the character of the mode shapes as well as the transmissibility function, we deduced that the foundation is extremely stiff, allowing the building to work as an excellent tremor detector, using the second-mode response at roof level as a sensitive trigger. That said, it is not absolutely certain to what extent the building basement movement represents the "free field" motion. Even so, the rigid body motion observed at long periods (20 s to the limit of integration at about 50 s) almost certainly represents a larger scale ground motion in the greater area of Singapore, and the fact that the differential GPS does not show the low-frequency motion supports this.

Over the 10 years of monitoring, the dynamic characteristics in the lower modes changed only slightly, with a detectable gentle downward trend in the higher mode frequencies. The effect is open to interpretation and could result from changes in mass and/or stiffness.

One of the biggest surprises that emerged from the monitoring was that, for Singapore, dynamic loads due to distant tremors are far in excess of the dynamic loads due to wind [9]. It also appears that this also applies when total (dynamic plus static) wind loads are considered. As the basement seismic response spectra show a concentration of energy between 0.5 and 1.0 Hz, there are implications for the majority of tall buildings on Singapore Island having this range of first-mode frequency. The reasoning for this is that Republic Plaza responds in second mode, where (seismic) modal participation factor is relatively low, approximately 25% of first-mode participation factor, depending on the structure. Hence, low-rise buildings will attract proportionally much more seismic load.

It has been possible to draw conclusions about the nature of the loading and response because a validated finite element model has been available. Our recommendation is that dynamic testing and creation of a validated numerical structural model are prerequisites for an effective SHM system.

REFERENCES

[1] Brownjohn JMW, Pan T-C, Cheong HK. Dynamic response of republic Plaza, Singapore. *The Structural Engineer* 1998 **76**(11):221–226.

[2] Brownjohn JMW, Pan T-C, Deng XY. Correlating dynamic characteristics from field measurements and numerical analysis of a high-rise building. *Earthquake Engineering and Structural Dynamics* 2000 **29**(4):523–543.

[3] Brownjohn JMW, Ang CK. Full-scale dynamic response of a high rise building subject to lateral loading. *ASCE Journal of Performance of Constructed Facilities* 1998 **12**(1):33–40.

[4] Brownjohn JMW, Pan T-C. Response of a tall building to long distance earthquakes. *Earthquake Engineering and Structural Dynamics* 2001 **30**: 709–729.

[5] Littler JD, Ellis BR. Interim findings from full-scale measurements on hume point. *Journal of Wind Engineering and Industrial Aerodynamics* 1990 **36**:1181–1190.

[6] Kijewski-Correa T, *et al.* Validating the wind-induced response of tall buildings: a synopsis of the Chicago full-scale monitoring program. *Journal of Structural Engineering, ASCE* 2006 **132**(10):1509–1523.

[7] AS/NZS 1170.2. *Structural Design Actions, Part 2: Wind Actions*. Standards Australia, 2002.

[8] Ashkenazi V, Roberts GW. Experimental monitoring of the Humber bridge using GPS. *Civil Engineering, Proceedings, Institution of Civil Engineers* 1997 **120**:177–182.

[9] Brownjohn JMW. Lateral loading and response for a tall building in the non-seismic doldrums. *Engineering Structures* 2005 **27**:1801–1812.

Chapter 134

Dynamic Response of Buildings of the Cultural Heritage

Paolo Clemente and Giacomo Buffarini

ENEA, Casaccia Research Centre, Rome, Italy

1 INTRODUCTION

The dynamic behavior of historical buildings is quite complex owing to several reasons, such as the complexity of the geometrical characteristics, the nonlinear behavior of the material, and the noneffectiveness of the connection between the masonry walls and the decks. As a result, identification of the dynamic characteristics and the seismic analysis are very hard and we cannot easily generalize the results obtained on a specific building to other structures. Besides, the numerical analysis by means of finite element models, which are usually very helpful to analyze the expected behavior and to interpret the experimental results, contains large uncertainties: the structural size of the various elements (walls, floors, etc.) cannot be evaluated with the needed accuracy; the material characteristics, such as the tension–strain relationship and the strength, are not known; structure and materials often exhibit inelastic behavior; rigid floors are often missing; these are substituted by masonry vaults or by wooden floors, whose effectiveness in connecting walls is uncertain; neither the depth of the foundation nor their geometry and material properties are known; often the depth is variable as well as the soil characteristics; buildings are often connected to other constructions so that their behavior is very complex. Actually, a complete analysis should include clear identification of the structural systems, their dynamic characteristics, and the knowledge of the mechanical properties of the materials used to construct the original buildings. Furthermore, in many cases, the building under investigation may have been altered repeatedly over time. This can be a critical issue because, for example, the building may be founded on older buildings that got buried over. Therefore, for such kind of structures, the experimental analysis is often the only way to improve our understanding about their dynamic behavior.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Instrumentation of structures to measure their motion under various loading conditions is a widely used practice in engineering. Temporary arrays are generally used for dynamic characterization of structures, which can be excited by means of ambient or forced vibrations, obtained by means of a vibrodine, impulse loading, explosions, etc. In this case, velocimeters are to be preferred, because their deployment is often easier and because they are more sensitive than accelerometers, so also vibrations of very low amplitude, such as ambient vibrations, can be recorded. Accelerometric sensors, which have to be fixed to the structure, are to be preferred in the cases of strong motion recording, due to their larger full-scale value, and long-term monitoring of structures. Obviously, for a fixed instrumentation, the choice of the locations of sensors and the ways for the cables is influenced very much by the interference of the instrumentation with the normal use of the building and the optimization of the ways of the cables and the accessibility of the locations, which must allow the safe installation of sensors. These problems can be often ignored for temporary arrays.

The experimental study of a structure should be organized in two steps. The dynamic characterization should be first performed, in order to have a first glance at the dynamic properties of the structure, such as resonance frequencies, modal shapes, and damping; few sensors can be temporarily deployed in different configurations also to define the optimum deployment. Then the permanent array is designed on the basis of the experimental results obtained from the temporary deployment. A numerical model should accompany both the experimental phases [1, 2].

Two case studies are shown in the following. The first one is a typical masonry bell tower and the second is a historical building. Both of them were damaged by earthquakes.

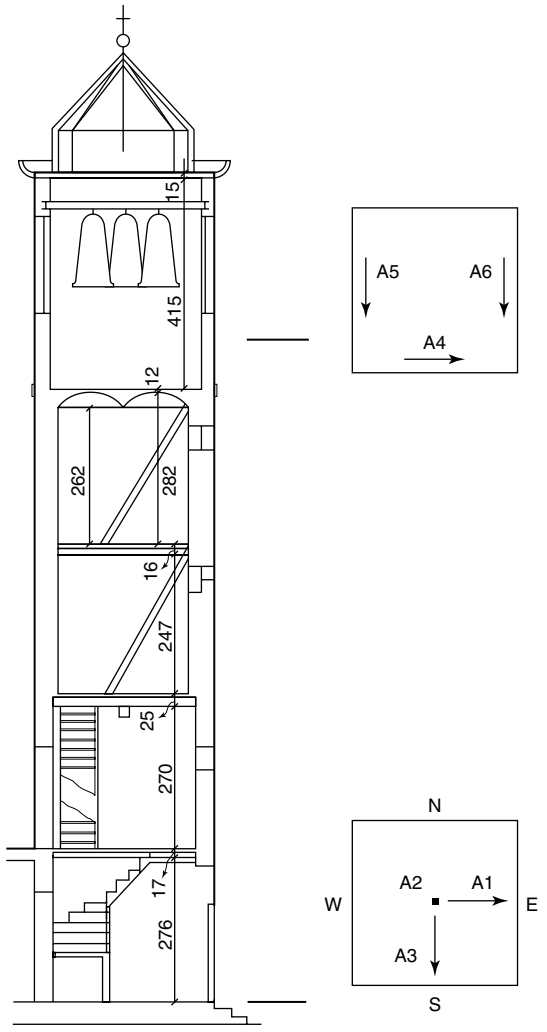


Figure 1. Bell tower at Trignano: accelerometric network.

2 BELL TOWER OF S. GIORGIO CHURCH IN TRIGNANO

The original structure of the medieval Bell Tower of S. Giorgio Church in Trignano, 18.5-m tall and 3.35 m × 3.00 m at the basement (Figure 1), withstood several changes and additions in the past centuries. Four masonry pillars at the corners, about 40-cm thick, compose the main structure. Very poor masonry walls, whose connections with the pillars are not effective, fill the spaces between the pillars. The first three floors are made of timber, the fourth one was substituted by a two-brick little vault floor, supported by a central steel I-beam. The stairs were composed by wooden and steel flights. The tower is connected to the structure of the church and to other masonry buildings on three sides, up to the height between 6 and 7 m.

The tower was seriously damaged by the Reggio Emilia earthquake ($M_L = 4.8$) of October 15, 1996. The most apparent effect of the earthquake was the opening of a near horizontal crack in the freely

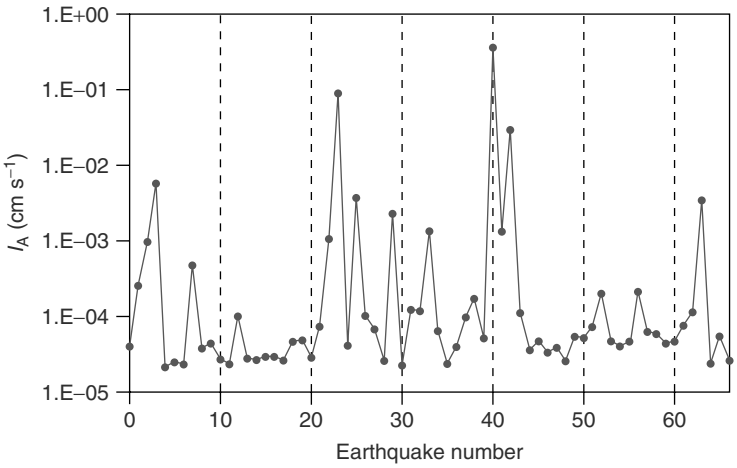


Figure 2. Energy for the recorded earthquakes.

rising part, above the roofs of the adjacent buildings. The crack intersected three of the four sides of the tower. A 3-cm offset was also apparent between the upper and the lower part of east wall, due to a clockwise rotation of the upper part with respect to the lower one.

The dynamic characterization of the tower was performed by means of ambient vibration tests [3]. Forced vibrations were also considered, by recording the vibrations due to the effects of a mass dropped on the ground near the tower. The motion, in terms of modal shapes, was examined by means of the amplitudes of power spectral densities (PSDs). The ambient vibration tests showed a resonant frequency of about 2.7 Hz with prevalent displacements in the N–S direction, and a resonant frequency of about 2.9 Hz with prevalent displacements in the W–E direction. A resonant frequency of 6.9 Hz was associated to a torsional modal shape.

2.1 Analysis of the seismic input at the basement

The fixed instrumentation consisted of a triaxial accelerometric sensor located on the ground floor, which could be assumed as basement, and three uniaxial horizontal accelerometric sensors at the top (Figure 1). Sixty-seven aftershocks were recorded between October and December 1996. The events were classified on the basis of the input energy for the structure, estimated by means of the Arias scalar intensity at the basement $I_A = \pi 2g \cdot \int (a_x^2 + a_y^2 + a_z^2)\,dt$. In Figure 2, I_A is plotted for each earthquake. As one can see, most recorded earthquakes showed values of I_A lower than 1.0E–4 cm s^{-1}; a relevant number of earthquakes had an energy value between 1E–4 and 1E–2 cm s^{-1}. Three events showed energy much higher than the others.

2.2 Dynamic characteristics of the structure

The records obtained under the lower energy earthquakes confirmed the results of the dynamic characterization [4]. Peaks at the already mentioned resonance frequencies are evident. In more detail, the resonance frequency of 2.7 Hz is apparent for CH5 and CH6, while the resonance frequency of 2.9 Hz is present for CH4. The analysis of the phase factor pointed out that the signals at the top and the corresponding ones on the basement are 90° out of phase at these frequencies. The torsional resonant frequency of 6.9 Hz is also present.

The dynamic characteristics, in terms of resonance frequencies, modal shapes, and damping, changed significantly, under earthquakes of higher level energy. In Figure 3, the acceleration time histories relative to strongest event (No. 40) are plotted. The PSDs (Figure 4) show the different frequency content of the top records in comparison to the basement

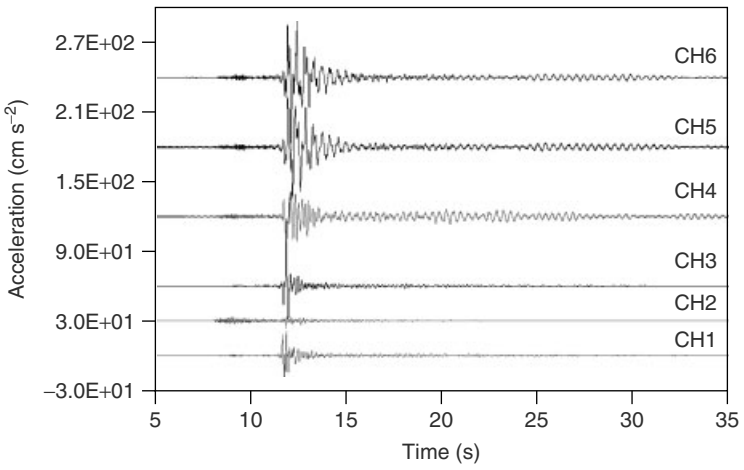


Figure 3. Time histories of the event no. 40.

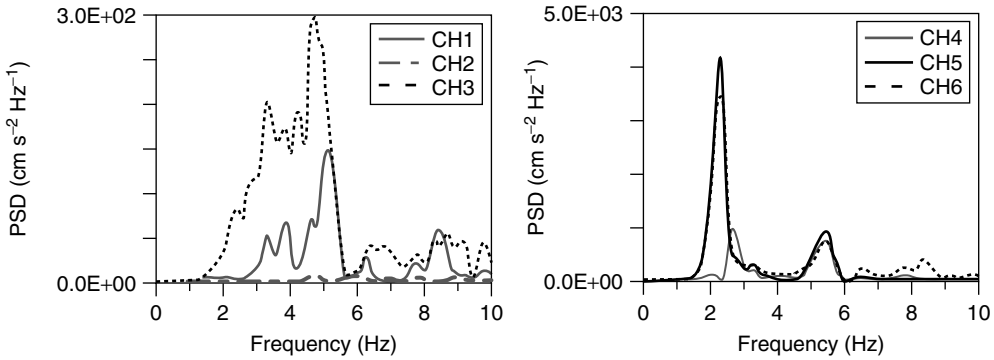


Figure 4. Power spectral densities.

ones, which contain the characteristics of the seismic motion. The reduction of the resonant frequencies is apparent. In fact, the first frequency is equal to 2.25 Hz and is associated to a modal shape with prevalent displacements in the N–S direction. The second frequency is at 2.60 Hz and is relative to a modal shape with prevalent displacements in the W–E direction. It is also evident that, in this case, the energy in the N–S direction is much higher than the energy in the W–E direction. It is also interesting to point out that just the opposite happened in the PSDs of other events. This occurrence is related to the different directivity of the earthquakes.

The resonance frequency of 5.5 Hz is apparent in the cross spectral density (CSD) between sensors CH5 and CH6 (Figure 5). The signals being 180° out of phase, this happens to be the torsional frequency of the tower. The PSD amplitudes of the three records at the top being almost similar, we can conclude that the prevalent motion, associated with this frequency, is a rotation around the vertical center line of the tower.

2.3 Seismic response of the tower

The frequency-domain analysis was performed for all the events. In Figure 6, the resonant frequencies are plotted for each event. Changes in the resonant frequencies for the different earthquakes are apparent. Recalling the classification of the recorded events based on the scalar Arias intensity, I_A, at the basement, we plotted the values of the first three resonance frequencies versus the corresponding I_A (Figure 7). From these plots, the influence of the seismic

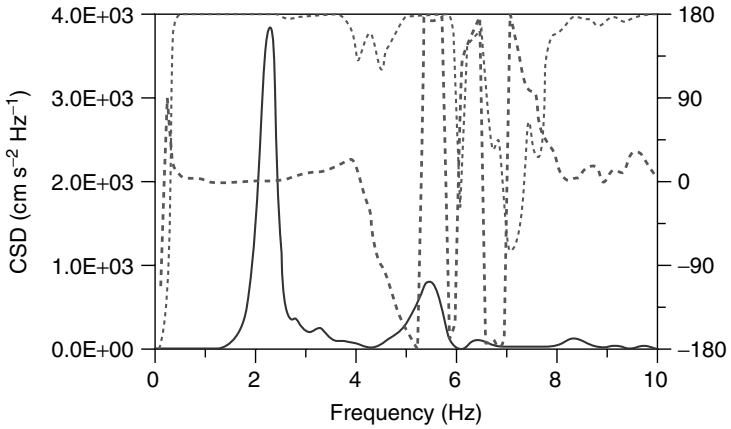


Figure 5. Cross spectrum CH5–CH6.

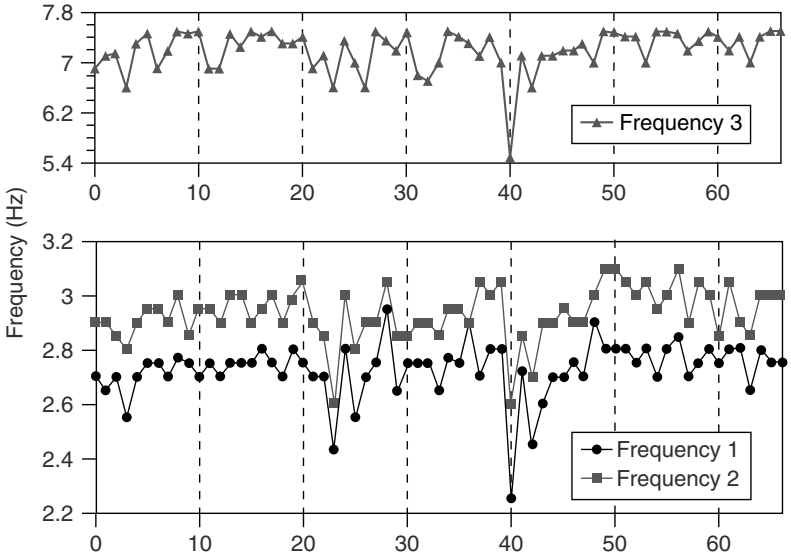


Figure 6. Resonance frequencies for all the recorded events.

input energy on the resonance frequencies is well emphasized. The energy range can be divided into two intervals. In the first one, the resonance frequencies are almost independent of the seismic energy and are very close to those obtained from the characterization tests. In the second one, all the three resonance frequencies decrease almost linearly with $\log(I_A)$. The intersection between the horizontal line, corresponding to the average value of the first resonant frequency found in the lower energy earthquakes, and the straight line that fits the decreasing first frequency values better, gives a numerical evaluation of the boundary between the two intervals. This operation brings the same value $I_{A,0} \approx 5\text{E}{-}4\,\text{cm s}^{-1}$ for the three frequencies, so it can be assumed as the limit value for the seismic energy, which separates the range of the linear behavior of the damaged tower from the nonlinear behavior range. The coherence function assumed lower values for the strongest earthquakes. This occurrence is also a measure of the nonlinearity in the seismic behavior of the tower. It was also noticed that earthquakes of similar intensity

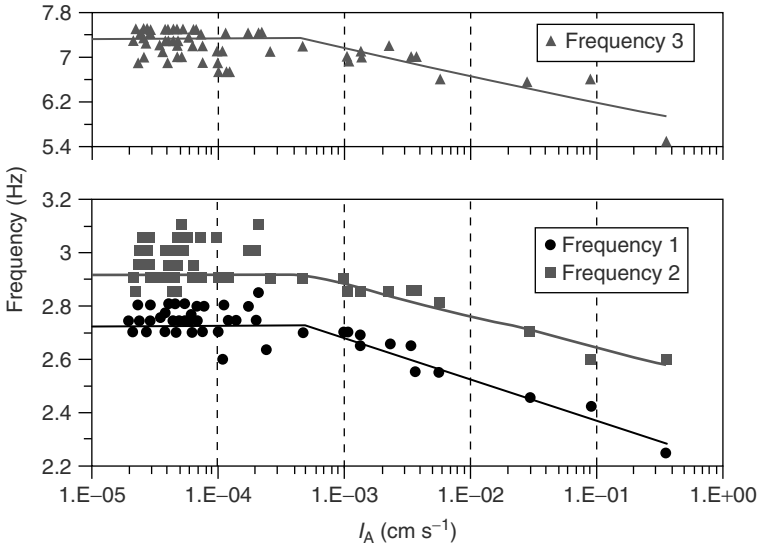


Figure 7. Resonance frequencies against the event energy.

had slightly different effects on the dynamic behavior of the tower. This occurrence demonstrated that even the higher recorded earthquakes did not cause significant additional damages to the tower. In some cases, a different behavior was detected under earthquakes of similar amplitude. This was probably related to the directivity of the earthquakes. In fact, different ratios between the PSDs of the three components were found. Damping was calculated by means of the half-power bandwidth method. It increases with the energy level.

3 CEDRAV BUILDING AT CERRETO DI SPOLETO (ITALY)

The Centre for Anthropological Documentation and Research of Nerina Valley (CEDRAV) was built as a monastery in the fourteenth century on the top of a carbonatic ridge. The building, which is irregular both horizontally and vertically (Figure 8), is composed of a main building and three additional ones connected to it. In the main building, the foundations are not at the same level. The first level is partially embedded into the ground and is mostly founded directly on rock, the second level is partially founded on the rock, and the foundation of the N–W portion of the building is not known in detail. The carrying structure is composed of masonry walls. At the east side, the main building is connected to a small square-shaped structure of three floors, having about half a wall in common. At the north corner, a rectangular building is connected to the main one and at the west side, another rectangular building is connected to the main one by means of a masonry arch. The secondary structures influence the dynamic behavior of the main building very much. In fact, torsional modes with very low damping and coupling with the principal vibration modes of the buildings are generated.

3.1 Experimental tests and data analysis

The building was damaged during the Umbria–Marche seismic sequences of 1997. By observation, it was determined that the building suffered some damage during the two main shocks of September 26 ($M_S = 5.6$ and $M_S = 6.0$, respectively), although the epicenter of the earthquake was 30 km away from the building. Most of the damages were caused by the event of October 14 ($M_S = 5.4$), the epicenter of which was approximately 8–10 km away from the building site.

Following these two earthquakes, the structure was first instrumented using temporary arrays in order

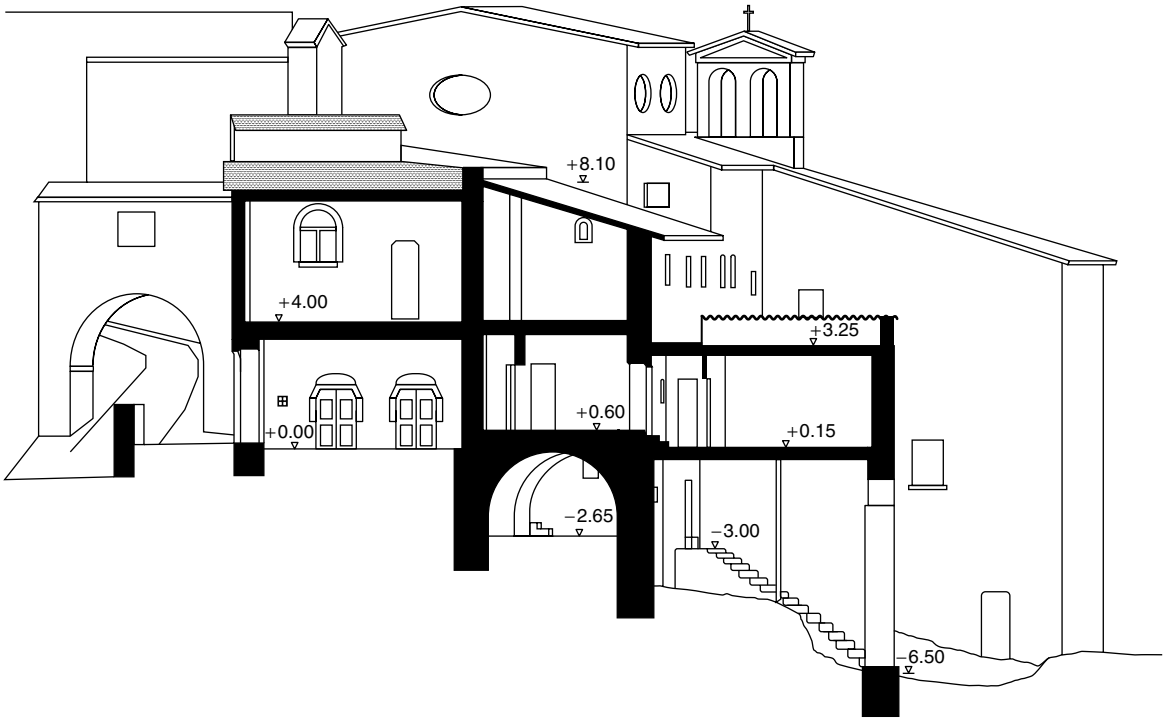


Figure 8. CEDRAV building: vertical section.

to characterize its dynamic properties [5]. Then a permanent accelerometer array was installed. This consisted in three K_2 acquisition systems, each of them having an internal triaxial accelerometric sensor and 27 external FBA11. Schematics showing the locations of the instruments are provided in Figure 9. The permanent deployment recorded several seismic events in about one year.

Collected data were analyzed in the frequency domain. Sample PSDs of the recorded response to a seismic event are provided in Figure 10. Peaks at 9.7 and 10.4 Hz are apparent. The first frequency is associated to a translational mode and the second to a torsional mode. Using system identification, very low damping percentages ($<1\%$) were extracted. The following considerations have to be pointed out: the building behaves like a very complex and rigid system; translational and torsional frequencies are close to one another; mode coupling occurs; and damping percentage is low.

Coupling of the frequencies and low damping are two factors that cause beating effect when shaking is strong enough. In this case, the building is affected by beating effect, even though the shaking level is not high.

3.2 The numerical model

The experimental results were compared with those obtained from a linear finite element analysis, which allowed to interpret the behavior under low energy events very well. Besides, cracks at the test time were very narrow, so no discontinuities were introduced in the model. Walls, floors, and vaults were modeled using four nodes shell elements, having both membrane and bending behavior. The structure is very complex both in plan and in elevation and some uncertainties affected the model. For example, the presence of buried rooms without access was pointed out by the experimental dynamic behavior of the building and the finite element model was consequently modified.

The structure was supposed to be composed by a unique homogeneous and isotropic material, which reproduces the average characteristics of the actual

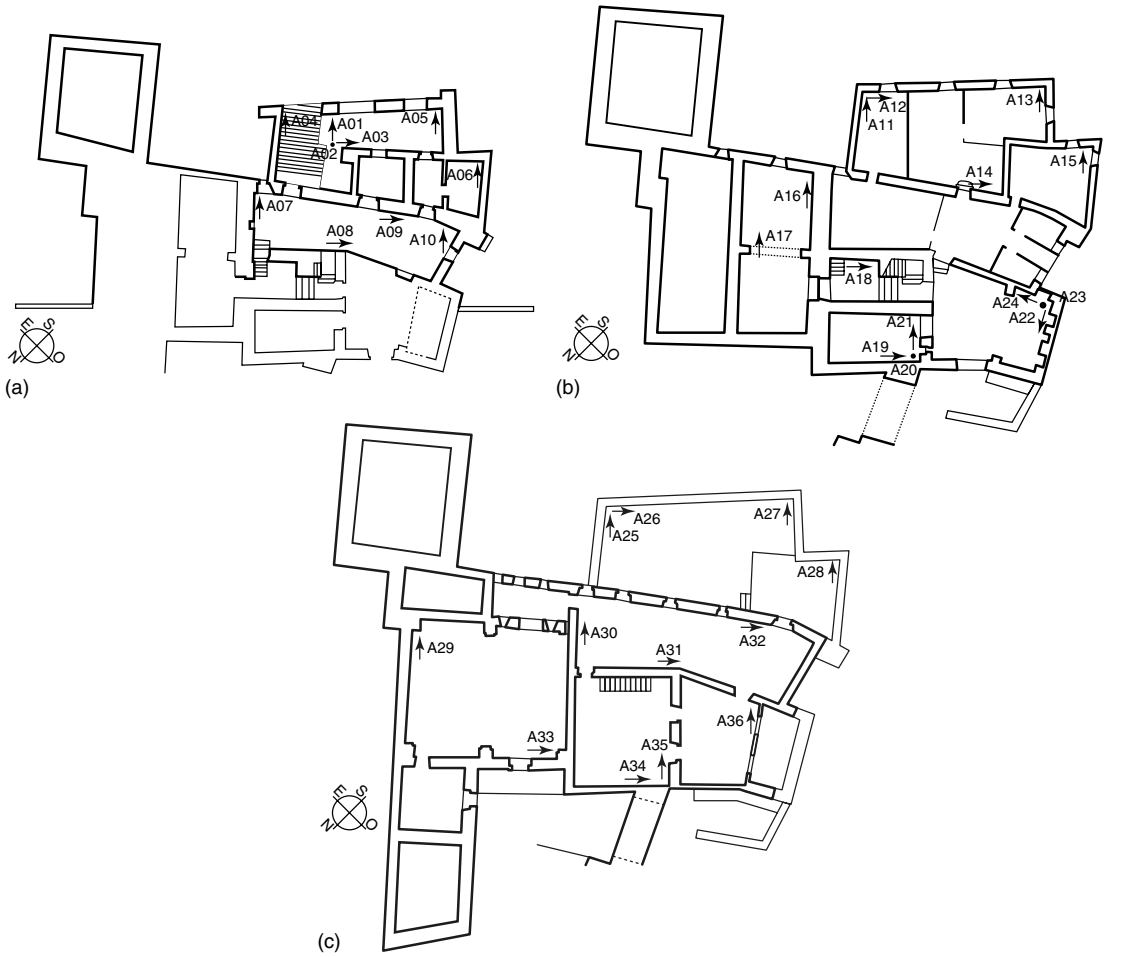


Figure 9. CEDRAV building: plans and sensor locations on the (a) ground floor, (b) first floor, and (c) second floor.

ones. Material characteristics given in the literature were first assumed, and then the Young's modulus was changed so that the first numerical frequency matches the first experimental frequency. In this way, we obtained a quite good correspondence between most of the numerical and experimental frequencies. In the final model $E = 1000\,\mathrm{N\,mm^{-2}}$ for both walls and vaults, while the density was 1800 and $1600\,\mathrm{MNs^2/m^4}$, respectively. The modal analysis gave the frequencies shown in Table 1. As one can see, very close frequencies were found due to the structural complexity. This behavior was shown by the experimental data too. In fact, many substructures behaved also as separated structures, but being linked they influence each other reciprocally. In the following analysis, only the modes

Table 1. Numerical frequencies

No.	1	2	3	4	5	6	7	8	9	10
Hertz	6.35	7.01	7.17	7.61	7.98	8.55	9.26	9.27	9.69	9.71
No.	11	12	13	14	15	16	17	18	19	20
Hertz	10.24	10.41	10.88	11.00	11.24	11.37	11.97	12.31	12.39	12.76

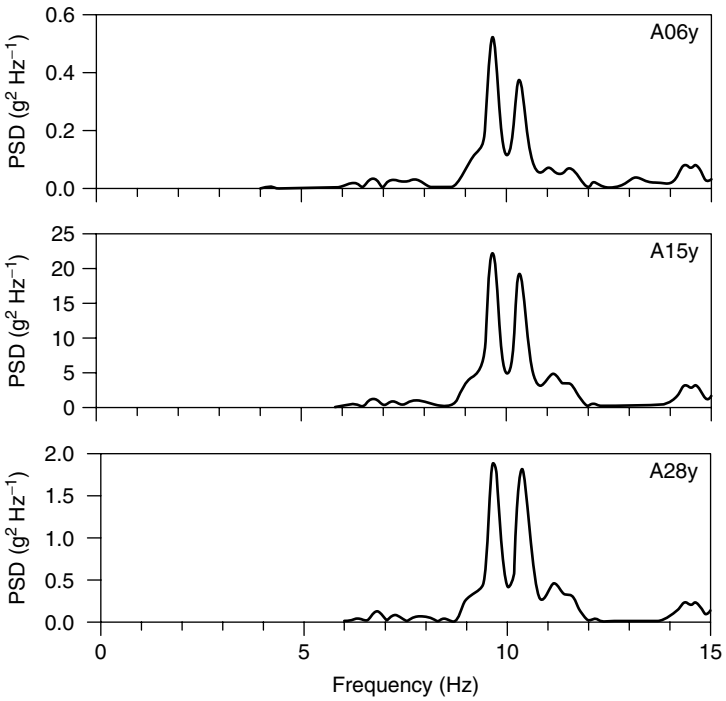


Figure 10. Power spectral densities.

with a significant participation factor were used for validation of the model.

The mathematical model was validated by using the experimental behavior under the higher energy level earthquakes. The time histories recorded on the basement (A01, A02, and A03) were assumed as input accelerations at all the basement joints. The output obtained at sensor locations were compared with those recorded at the same locations, both in time and frequency domain. Modal damping ratio was assumed to be equal to 1% for mode 9 (9.69 Hz) and mode 11 (10.25 Hz), while it was assumed equal to 3% for all the other modes.

In Figure 11, the PSDs of the experimental recorded responses (dashed line) are compared with those from the numerical model (bold line). The correspondence between experimental and numerical graphics is quite good; this occurrence confirms the validity of the hypotheses assumed. In other words, in the frequency range considered (0–10 Hz), the elastic model reproduced the experimental behavior very well. In fact, the same peaks were found in the PDSs, obtained from experimental and numerical

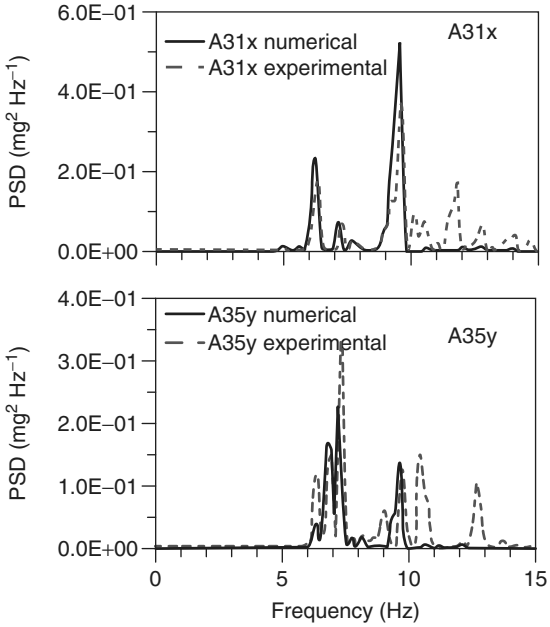


Figure 11. Comparison between numerical and experimental PSDs.

analyses, respectively, often with very similar amplitude.

4 CONCLUSIONS

The dynamic response of buildings of cultural heritage has been investigated in two steps: (i) dynamic characterization tests to evaluate the structural properties and to have preliminary information about the health status of the structure and (ii) strong motion monitoring over a consistent period. The advantage of this second step is that the structure is analyzed under true seismic inputs. The analysis of the experimental results pointed out the main features of the dynamic behavior of the buildings of cultural heritage.

ACKNOWLEDGMENTS

The results of the studies reported here are parts of research projects funded by the European Commission (Bell Tower at Trignano) and the Italian Ministry of Scientific Research (CEDRAV Building).

RELATED ARTICLES

Signal Processing for Damage Detection
Statistical Time Series Methods for SHM
The Character of SHM in Civil Engineering
Ambient Vibration Monitoring

REFERENCES

[1] Clemente P, Rinaldis D. Design of temporary and permanent arrays to assess dynamic parameters in historical and monumental buildings. *Proceedings of North American-Euro-Pacific Workshop Sensing Issues in Civil Structural Health Monitoring (CSHM)*, ISBN I-4020-3660-4 (HB), ISBN I-4020-3661-4 (e-book), ISBN 978-I-4020-3660-6 (HB), ISBN 978-I-4020-3661-3 (e-book), November 2004. Springer, 2005, pp. 107–116.

[2] Rinaldis D, Clemente P, De Stefano A. Design of seismic arrays for structural systems. In *Structural Health Monitoring of Intelligent Infrastructure (Proceedings SHMII-2'2005)*, Ou JP, Li H, Duan ZD (eds). Taylor & Francis/Balkema: Leiden, 2005, pp. 1447–1453.

[3] Bongiovanni G, Buffarini G, Clemente P. Dynamic characterisation of two earthquake damaged bell towers. *Proceedings of the 11th European Conference on Earthquake Engineering*. Balkema: Rotterdam, September 1998.

[4] Clemente P, Bongiovanni G, Buffarini G. Experimental analysis of the seismic behaviour of a cracked masonry structure. *Proceedings of the 12th European Conference on Earthquake Engineering*, Paper No. 104. EAEE: London, September 2002.

[5] Rinaldis D, Çelebi M, Buffarini G, Clemente P. Dynamic response and seismic vulnerability of an historical building in Italy. *Proceedings of the 12th World Conference on Earthquake Engineering*, Paper No. 3211. IAEE, August 2004.

Chapter 135

Suspended Roof of Braga Sports Stadium, Portugal

Álvaro Cunha, Filipe Magalhães and Elsa Caetano
Faculty of Engineering, University of Porto, Porto, Portugal

1 INTRODUCTION

The Braga Municipal Sports Stadium (Figure 1) is one of the stadia that were recently constructed in Portugal to host some of the matches of the 2004 European Football Championship. The stadium, designed by Eduardo Souto Moura in conjunction with the consultancy office Afassociados [1], has been considered a masterpiece of architecture. The roof, suspended by cables, is unique and it presented a particular challenge in terms of conception, structural design, and construction.

The innovative characteristics of this roof structure, as well as the resulting flexibility, have motivated extensive studies developed during the design phase by various independent entities, whose purpose was to adequately define the design wind load, to evaluate the corresponding static and dynamic behavior, and to investigate the susceptibility to aeroelastic instabilities. These studies comprehended the development of different numerical models of the structure, and a series of wind-tunnel tests performed on physical models. The Laboratory of Vibrations and Monitoring (VIBEST, www.fe.up.pt/vibest) of the Faculty of Engineering of the University of Porto (FEUP) was consulted in this context, at an early stage of the project, with the aim of developing a static and dynamic study of the roof structure [2]. The developed numerical models were first used in the definition of the geometric and mechanic characteristics of the cables and slabs, and the calculated dynamic properties were later used in the construction of a physical model for wind-tunnel tests. After the stadium had been constructed, VIBEST/FEUP was also consulted to analyze data provided by forced and free vibration tests developed by the contractor; a complete ambient vibration test of the suspended roof was performed to measure the corresponding dynamic

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Figure 1. General views of the Braga Municipal Sports Stadium.

properties and to use them to validate the developed numerical model.

Under these circumstances, it is the purpose of the current work to describe the research developed for the operational modal analysis and finite element correlation of the suspended roof of the new Braga Sports Stadium on the basis of temporary ambient vibration monitoring. This work allowed the experimental validation of a numerical modeling of the dynamic behavior of the suspended roof, which takes into account the geometric nonlinear structural behavior and the progressive application of loads during the construction phase.

The identification of the modal parameters presented a particular challenge, as the flexibility of the roof is associated with very low and closely spaced natural frequencies. Special attention was given for the identification of modal damping ratios, owing to the necessity of analyzing the susceptibility of the suspended roof to buffeting effects. These coefficients were estimated using data provided by free, forced, and ambient vibration tests, and using an improved implementation of the enhanced frequency domain method and the covariance-driven stochastic subspace identification (SSI-COV) method. The comparison of the estimates achieved by application of output-only identification methods with the ones provided by data collected in artificial excitation tests was important to understand the capabilities of each approach.

2 DESCRIPTION OF THE STRUCTURE

The stadium was constructed on the slopes of Monte Castro, and developed as an amphitheater over a wide rural landscape, formed only by two rows of stands, on either side of the pitch, and by a granite massif (Figure 2). The most noticeable element of the stadium is its roof, which is formed by pairs of full-locked coil cables with diameters varying

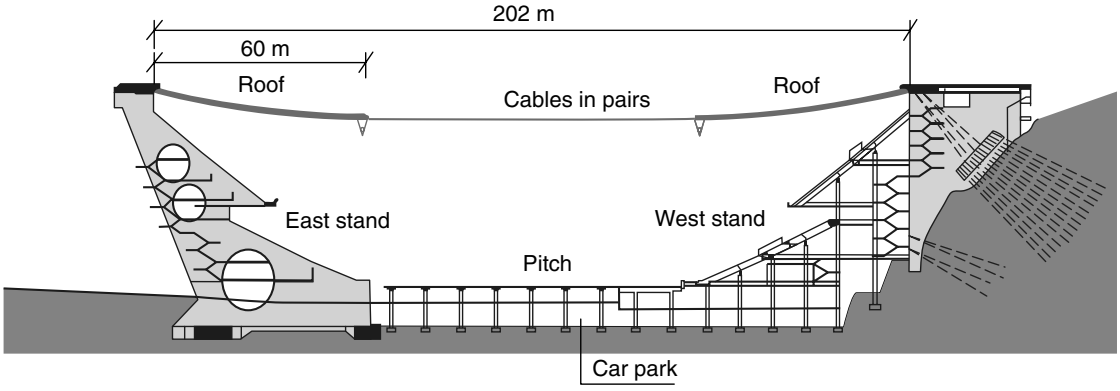


Figure 2. Scheme of the Braga Municipal Sports Stadium.

between 86 and 80 mm, spaced 3.75 m apart from each other, supporting two concrete slabs over the two stands of the stadium. The cables' span is 202 m and the slabs' length is 57.3 m; therefore, the remaining 88.4 m of the central part is free. Rain water is drained from the roof along one side only, the slope being achieved by a variation in the lengths of the cables. The concrete slabs have a thickness of 0.245 m and are connected to the cables only in the normal direction, allowing relative tangential movements. A transversal triangular truss is suspended from the inner border of each slab, acting as a stiffness beam and simultaneously accommodating the floodlights and loudspeakers.

The roof cables are anchored in two large beams at the top of both stands—east and west. The east stand is structurally formed by 50-m-high concrete walls, whose geometry was defined to minimize the unbalanced moments at the level of the foundation, motivated by the combination of the gravitational action of the stand and the high forces transmitted by the roof cables. In the west stand, the concrete walls are anchored in the rock and the roof cables' tension forces are transmitted to the foundation by prestressed tendons embedded in the concrete.

The outstanding characteristics of the structure and the need for a tight control of the corresponding behavior during construction justified the installation of a monitoring system, which comprehends static and dynamic components. The static monitoring system was essential during the construction and is based on a series of load cells installed in the cables anchorages, on embedded instrumentation of the concrete structure (strain gauges, tiltmeters, and thermometers), and on instrumentation of the rock massifs and foundations, with load cells installed in the anchors to the earth and in-place inclinometers. The dynamic monitoring system is important for observing the response of the roof to wind excitation; it is composed of six accelerometers, installed in the inner edges of the concrete slabs, and of cells to measure the wind pressure at various points on the underside and top of the roof slabs.

3 AMBIENT VIBRATION MONITORING AND STRUCTURAL IDENTIFICATION

3.1 Ambient vibration measurements

The temporary ambient vibration monitoring comprehended the measurement of the vertical acceleration at 42 points of the roof, using three strong-motion recorders (Figure 3a), synchronized by GPS and programmed using a laptop. The use of these recorders was very practical, as no electrical cabling was required. In the present test, the use of cables connecting equipment placed on both sides of the suspended roof would have made the test preparation very complicated.

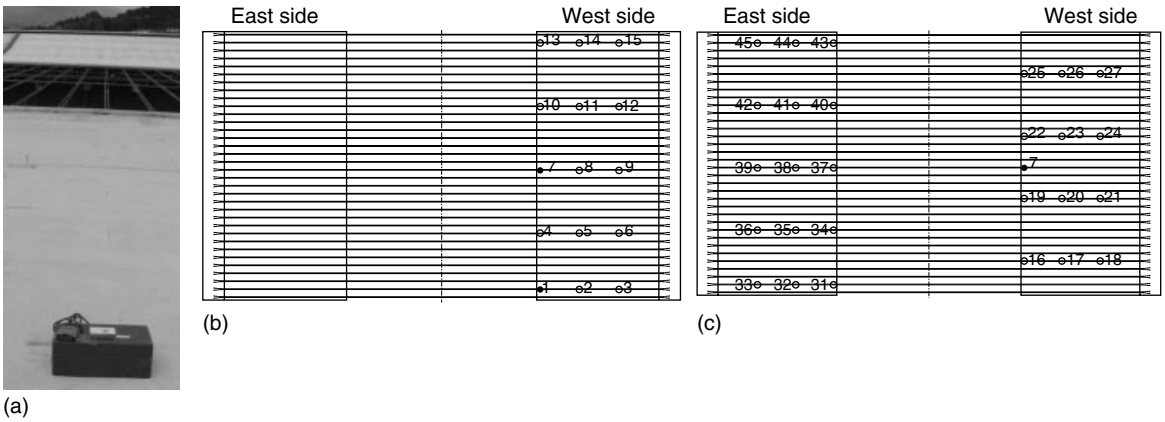


Figure 3. Placement of one of the used seismographs (a) and measurement points of the ambient vibration test: (b) first day (test 1 layout), and (c) second day (test 2 layout). ●, reference points; ○, points measured with moving sensors.

The test was performed over two days. On the first day, the measurements were taken at the points of the west slab represented in Figure 3(b), using 13 setups, while on the second day, the measurements were carried out at the points of the west and east slabs represented in Figure 3(c), using 15 setups.

On the first day of measurements, two reference points were considered (two recorders were permanently placed at points 1 and 7 during all setups). After a preliminary analysis of the data, it was concluded that, for the frequency range of interest (0–1 Hz), all the modes were detected by the reference sensor placed at point 7. Thus, to reduce the duration of the test, it was decided to use just this reference point for the remaining measurement sections. The test developed on the second day provided a set of new responses on the east slab and more response measurements on the west side slab that allowed the improvement of the spatial resolution used in the characterization of the mode shapes.

For each setup, time series of 16 min were collected with a sampling frequency of 100 Hz (minimum sampling frequency allowed by the acquisition system). Figure 4 represents one of the time series collected at point 7 (reference point) and shows the variation of the standard deviation of the time series measured at point 7 during the 28 setups. The amplitude of vibration is essentially dependent on wind, as this was the only significant dynamic excitation of the roof during the test. Therefore, the graph shows that, during the first day of the test (a rainy and windy day), the wind speed was higher and had greater fluctuations than on the second day (a sunny and calm day).

During the tests, the maximum value of recorded vertical acceleration was of about 5 mg, denoting a very low level of oscillation of the roof structure.

3.2 Operational modal identification

3.2.1 Frequency-domain decomposition

The ambient vibration response was initially processed by the Artemis software [3] using the frequency domain decomposition (FDD) method [4]. In this method, the natural frequencies are identified from the peaks of the singular values of the spectral matrices. The mode shapes are identified from the singular vectors of the spectral matrices evaluated at the identified resonance frequencies and associated with the singular values that contain the peaks.

In the present application, the spectral matrices were calculated with a resolution of 0.00488 Hz, which allowed sufficient accuracy in the identification of the natural frequencies. Figure 5 represents the configurations of the most relevant identified mode shapes, showing the good quality of the results and the existence of pairs of closely spaced modes. These are justified by the slope of the roof to drain the rain water, which breaks the symmetry of the structure with respect to the middle axis parallel to the cables. The lower spatial resolution of the measurements performed in the east slab explains the lower quality of the estimates of the last three mode shapes in this slab.

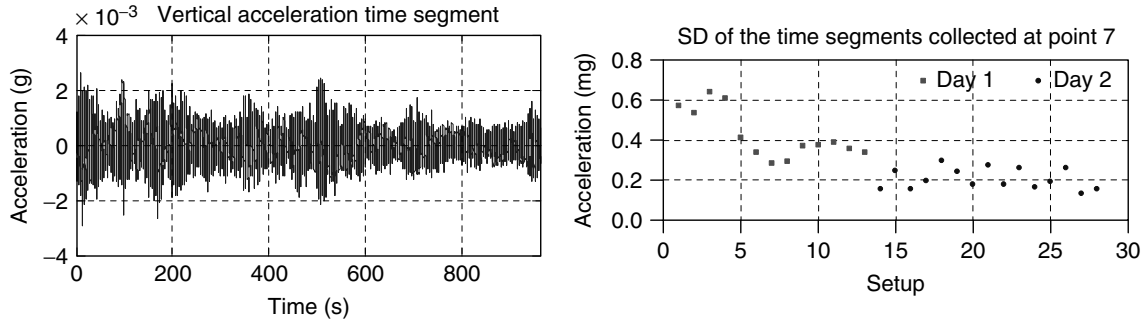


Figure 4. Vertical acceleration time series measured at point 7 during the first setup and variation of the standard deviation of the time series measured at point 7 along the 28 setups.

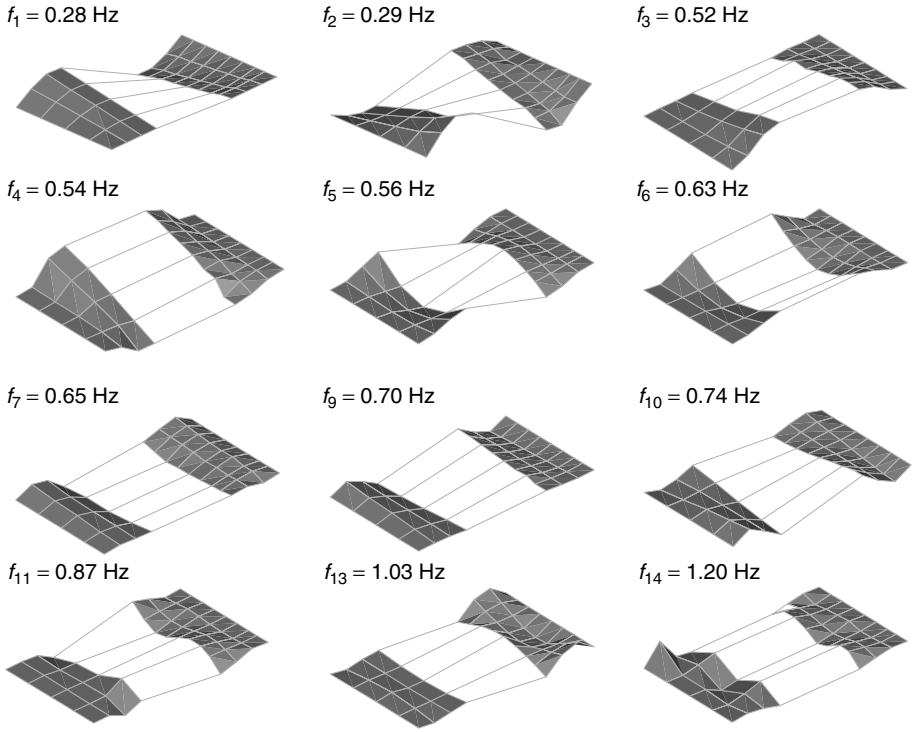


Figure 5. Mode shapes estimated by the FDD method.

The modal damping ratios can be estimated using an enhanced version of the FDD method, the enhanced frequency-domain decomposition (EFDD) method [5]. Usually, when the ambient vibration test is performed using several setups, the modal damping ratios, as well as the natural frequencies, are estimated using independently the spectral matrix of each setup; the final estimates are then obtained, averaging the estimates provided by all the setups.

In the present case, the duration of each setup time series is not sufficiently long to accurately estimate modal damping ratios of modes with such low natural frequencies, especially because very small values of damping are expected.

However, on the first day of tests, the vertical accelerations at points 1 and 7 (points defined as references in Figure 3b) were measured simultaneously during 13 setups, which means that these 13 time series with 16 min were a good basis to evaluate the modal damping ratios.

Beyond that, to increase the accuracy of the estimates, the autocorrelation functions were calculated using an alternative procedure to the one presented in [5] to avoid the circular error (the calculated autocorrelation is a superposition of the desired function and its mirror image [6]). These were obtained using an adaptation of the procedure described in [7], which is also based on a fast Fourier transform (FFT) approach [8].

In the processing of the available data collected at points 1 and 7, time series with 11 min were selected using an overlap of 54%, allowing the accomplishment of averages over 26 record estimates. The singular values calculated from the spectral matrix, with dimension 2×2, are represented in Figure 6. In the same graph, the points chosen to estimate the autospectra (Spi) associated with six different modes are selected (points with modal assurance criterion (MAC) > 0.8). For the other modes, identified in the plot by the peaks, it is not possible to select well-defined autospectra, because of the proximity between resonance peaks. The natural frequencies identified in the graph and summarized in Table 1 are more accurate than the ones presented in Figure 5

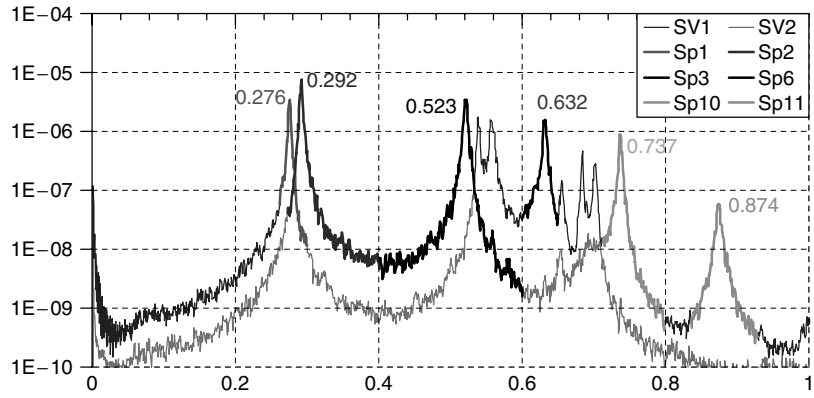


Figure 6. Singular values of the spectral matrix calculated using the time series collected at the reference points (1 and 7) and selection of the autospectra of each mode.

Table 1. Natural frequencies (f.) and modal damping ratios (damp.) identified with the ambient vibration test

Mode	EFDD		SSI-COV			
	f. (Hz)	damp. (%)	f. (Hz)	SD f. (Hz)	damp. (%)	SD damp. (%)
1	0.276	0.58	0.276	0.0008	0.50	0.25
2	0.292	0.52	0.292	0.0007	0.42	0.20
3	0.523	0.47	0.521	0.0014	0.44	0.20
4	0.539	—	0.539	0.0014	0.40	0.33
5	0.556	—	0.558	0.0012	0.47	0.19
6	0.632	0.35	0.632	0.0012	0.54	0.37
7	0.655	—	0.655	0.0019	0.28	0.09
8	0.684	—	0.684	0.0010	0.27	0.19
9	0.703	—	0.702	0.0010	0.26	0.10
10	0.737	0.25	0.737	0.0009	0.26	0.13
11	0.874	0.36	0.874	0.0016	0.41	0.10

because the frequency resolution is higher in this analysis (0.000763 Hz).

Figure 7 shows two of the autocorrelation functions calculated from the identified autospectra. The envelopes of these functions were fitted by exponential decays to estimate the modal damping ratios presented in Table 1.

3.2.2 Stochastic subspace identification

Despite the very good quality of the estimates of natural frequencies and mode shapes obtained by the FDD method, a stochastic subspace identification (SSI) method was also applied with the main objective of comparing estimates of modal damping ratios using different output-only modal identification techniques. For this purpose, the data collected on the first day of tests was used. To reduce the influence of the aerodynamic damping, it would have been better to use data collected during the second day (very low wind velocities). However, the data of the first day has the advantage of having two reference points.

In the present research, the SSI-COV method was applied using MatLab routines developed at the University of Porto [9]. Before application of the identification algorithm, the measured signals were filtered by a low-pass filter with a cutoff frequency of 1 Hz and the application of a decimation of order 20 (reducing the sampling frequency from 100 to 5 Hz).

The correlation functions are the basis of this identification method. These can be estimated using three different procedures: the direct definition (summation

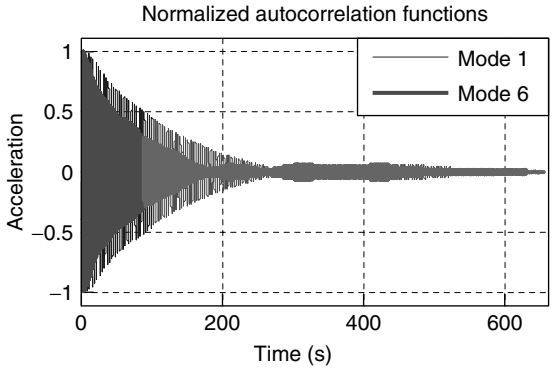


Figure 7. Normalized autocorrelation functions associated with modes 1 and 6.

formula), the FFT, or the random decrement. In the present application, the first alternative was used, which is not very computationally efficient, but does not introduce bias and is easy to program.

The data of the 13 setups was processed independently using the three collected time series and consequently, 13 stabilization diagrams, like the one presented in Figure 8, were constructed. This diagram shows that the dynamic behavior of the structure is well represented by state-space models of order between 20 and 40. The use of models of this relatively low order was only possible because a low-pass filter was applied, which eliminated the contribution of modes with frequencies greater than 1 Hz.

The most suitable model order was selected for each of the 13 stabilization diagrams, and the corresponding natural frequencies and modal damping ratios were calculated. Table 1 presents the mean and standard deviation (SD) of the identified modal parameters. The standard deviations show that the dispersion of the estimates of natural frequencies for the various setups is very low, whereas the estimates of modal damping ratios present significantly higher scatter. This dispersion shows the difficulty of getting reliable damping estimates, which stems not only from the uncertainties of the method but also from the variation of damping with the level of oscillation and with the wind characteristics.

4 FORCED AND FREE VIBRATION TESTS

The identification of modal damping ratios on the prototype structure was developed, in the first instance, at the request of the design office Afassociados [10], on the basis of a set of data collected by the instrumentation installed at the roof structure, during the forced and free vibration tests developed at the commissioning phase. In this section, the corresponding estimated modal damping factors are compared with the ones provided by the output-only identification techniques.

The free vibration test was based on the sudden release of a 5-t mass from a point located close to mark 1 (represented in Figure 3b). The response was collected by the six triaxial force-balance accelerometers of the dynamic monitoring system, which are located at marks 1, 7, 13, 31, 37, and 43 (Figure 3).

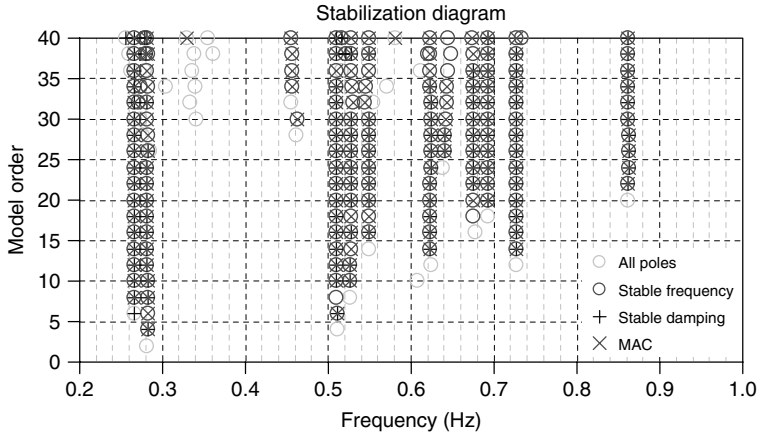


Figure 8. SSI-COV method stabilization diagrams of setup 4.

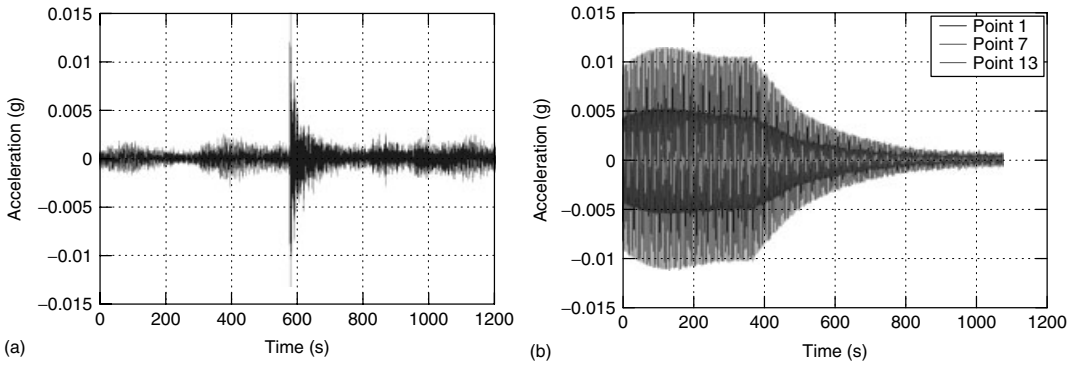


Figure 9. Free decays measured after the application of an impulse (a) and after harmonic excitation of the second mode (b).

Figure 9(a) represents the response measured at point 31. The application of band-pass filters to the measured signals enables the evaluation of modal free vibration responses, which were used to estimate the modal damping ratios presented in the second column of Table 2. This procedure faces two problems that explain the missing values in the table: the low level of excitation of some modes and the difficulty in isolating the contribution of modes with very close natural frequencies. An alternative to this procedure consists in using the measured response to the impulse as input to the SSI-COV method, as responses motivated by impulses are proportional to correlations of responses to white-noise excitations. This technique was used after application of a low-pass filter with a cutoff frequency of 1 Hz, and a decimation to reduce the sampling frequency to 5 Hz. The results obtained are presented in the third column of Table 2. It is relevant to observe that the new estimates are very consistent with the previous ones for the modes where both techniques provided results.

Forced vibration tests were further conducted, based on a harmonic excitation of the roof at resonance, by means of a cable pulled by an electric engine from mark 1 or 7 (Figure 3b). After resonance was attained, the excitation was suppressed and the free vibration response measured by the same six accelerometers. Using this procedure, five modes were excited and so five free decays were measured, like the one represented in Figure 9(b). The fourth column of Table 2 shows the values of the corresponding identified modal damping ratios, which show consistency with previously described estimates.

Table 2. Modal damping coefficients (%) identified with the free vibration tests

Mode	Free vibration filter	Free vibration SSI-COV	Harmonic excitation
1	—	0.29	0.28
2	—	0.37	0.27
3	0.28	0.32	0.22
4	0.25	0.22	—
5	—	0.44	—
6	0.34	0.36	0.43
7	—	0.29	—
8	—	0.11	0.20
9	—	0.18	—
10	0.20	0.18	—
11	—	—	—

The comparison between modal damping coefficients identified using artificial and ambient excitation shows the existence of satisfactory correlation. However, one can notice that relative differences tend to increase at lower frequencies. In effect, the difficulty in identifying modal damping ratios is well known, since they are dependent on the amplitude of vibration and also on the wind characteristics, which can introduce a significant component of aerodynamic damping, as experimentally observed in [11]. The very low damping values of this structure make the comparison even more difficult because very small differences are expressed by significant relative errors.

As for the quality of the estimates provided by the various identification methods, it is relevant to stress that in the present case, the results provided by the FDD method are close to the ones obtained by the SSI

methods, as very long-time series were used and an alternative procedure to estimate the autocorrelation function was introduced. It is still worth mentioning that the application of the standard EFDD method, using independently the time series of each setup (with 16 min.), led to values of modal damping ratios for the first modes of about 1%.

5 FINITE ELEMENT CORRELATION

5.1 Finite element modeling

The analysis of the static and dynamic behavior of the suspended roof was based on a three-dimensional finite element model [12]. This model was formed by a total of 34 cables spaced at 3.75-m intervals, which were idealized as 89 truss elements each. These were linked by shell elements, simulating the slabs, which were only activated after full application of the corresponding weight, and were linked also by transversal truss girders at the ends of the slabs, simulating the lattice structures used to accommodate the floodlights and loudspeakers. An overall slab thickness of 0.245 m was considered. To accommodate thermal deformations without further stressing the cables, sliding between the cables and slabs was allowed. This effect was achieved by definition of different layers of nodes for the cables and slabs, which were constrained to identical vertical displacements. A 1% slope was created along the transversal direction by gradual modification of the lengths of successive cables, for the purpose of water draining.

Considering the structure deformed under permanent loads, natural frequencies and vibration modes were calculated. Table 3 summarizes the first 11 calculated natural frequencies, while the first six mode shapes are represented in Figure 10.

Table 3. Calculated versus identified (SSI-COV) natural frequencies and MAC values

Mode	Frequency (Hz)		MAC
	Calculated	Identified	
1	0.277	0.276	0.98
2	0.305	0.292	0.99
3	0.520	0.521	0.94
4	0.532	0.539	0.94
5	0.574	0.558	0.97
6	0.610	0.632	0.96
7	0.673	0.655	0.84
8	0.678	0.684	0.97
9	0.712	0.702	0.84
10	0.754	0.737	0.97

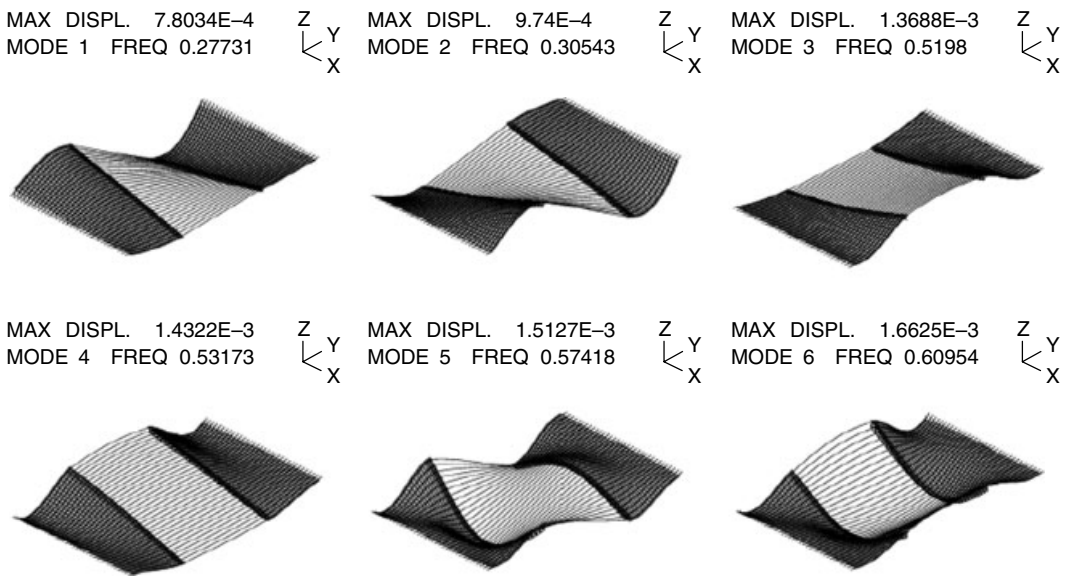


Figure 10. Calculated modal configurations of the first six modes.

5.2 Correlation between experimental and numerical frequencies and mode shapes

The numerical and experimental assessment of the dynamic behavior of the Braga Stadium suspended roof permitted to achieve an excellent correlation between calculated and identified modal parameters (natural frequencies and mode shapes), which evidences the high quality of the complex finite element modeling developed, as well as the accurate and reliable nature of the sophisticated output-only modal identification techniques employed.

The very good agreement obtained in terms of natural frequencies and mode shapes is shown by Table 3. The mode shapes are compared using the (MAC) [13] that present values close to those when the modes are similar.

6 CONCLUSIONS

The most relevant dynamic properties of the new Braga Sports Stadium suspended roof have been assessed experimentally, on the basis of the application of different output-only modal identification techniques, allowing the subsequent validation of the finite element modeling developed to investigate the static and dynamic behavior of this complex structure.

It is shown that, despite the low levels of signal captured during the ambient vibration test and the existence of a large number of modes of vibration in the frequency range 0–1 Hz, the available stochastic modal identification methods can provide very accurate estimates of natural frequencies and mode shapes, which present an excellent correlation with the corresponding calculated values. This agreement validates the sophisticated finite element modeling, which takes into account the geometric nonlinear structural behavior and the progressive application of the loads.

Because of the major importance of damping in the aerodynamic behavior of the suspended roof, special attention was dedicated to the estimation of modal damping ratios either from free vibration or ambient vibration tests. Results achieved show that, even in this challenging structure, with extremely low natural frequencies and damping factors and with closely spaced modes, operational modal analysis can provide reliable estimates of the order of magnitude of modal damping ratios (absolute differences inferior to 0.25%), constituting therefore an interesting alternative to the more costly, but certainly more accurate, procedures based on artificial excitation. An essential aspect in the application of the operational modal analysis is the use of very long-time records, preferably collected over a period of very low and steady wind velocity condition. However, further research is still needed to improve the potential of operational modal analysis in the estimation of modal damping ratios.

ACKNOWLEDGMENTS

The authors acknowledge all the financial support got from the Portuguese Foundation for Science and Technology (FCT) to the present research at the group of vibrations and monitoring (VIBEST) at CEC/FEUP and, in particular, the Ph.D. scholarship (SFRH/BD/24423/2005) provided by FCT.

REFERENCES

[1] Furtado R, Quinaz C, Bastos R. The New Braga Municipal Stadium, Braga, Portugal. *Structural Engineering International* 2005 **15**(2):72–76.

[2] Caetano E, Cunha A. *Numerical Modeling of the Structural Behaviour of the New Braga Stadium Roof*, Technical Report FEUP/VIBEST. FEUP, 2001.

[3] Structural Vibration Solutions. *SVS—Artemis Extractor Pro, Release 3.41*. Aalborg, 1999–2004.

[4] Brincker R, Zhang L, Andersen P. Modal identification from ambient responses using frequency domain decomposition. *IMAC XVIII*, San Antonio, TX, 2000.

[5] Brincker R, Ventura C, Andersen P. Damping estimation by frequency domain decomposition. *IMAC XIX*, Kissimmee, FL, 2001.

[6] Bendat J, Piersol A. *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons: New York, 1980.

[7] Brincker R, Krenk S, Kirkegaard P, Rytter A. Identification of dynamical properties from correlation function estimates. *Bygningsstatiske Meddelelser* 1992 **63**(1):38.

[8] Magalhães F, Caetano E, Cunha A. Operational modal analysis of the braga sports stadium suspended roof. IMAC XXIV, St. Louis, MO, 2006.

[9] Magalhães F. *Stochastic Modal Identification for the Validation of Numerical Models*, Master Thesis (in Portuguese). University of Porto: Porto, 2004.

[10] Magalhães F, Caetano E, Cunha A. *Experimental Identification of Modal Damping Ratios from the New Braga Stadium Roof*, Technical Report. FEUP, 2004.

[11] Macdonald JHG. Daniell WE. Variation of modal parameters of a cable-stayed bridge identified from ambient vibration measurements and FE modelling. *Engineering Structures* 2005 **27**(12):1916–1930.

[12] Caetano E, Cunha A, Magalhães F, Furtado R. Numerical and experimental studies of braga sports stadium suspended roof. *EVACES*. Bordeaux, 2005.

[13] Allemang RJ. Brown DL. A correlation coefficient for modal vector analysis. *IMAC I*. Orlando, FL, 1982.

Chapter 136
Dams

Reto Cantieni
rci dynamics, Duebendorf, Switzerland

1 INTRODUCTION

1.1 Health monitoring of dams

Standard procedures to monitor dams include visual inspection, monitoring of the static deformation behavior using geodetic instruments, plumb lines and clinometers, checking for pore and uplift pressures using piezometers, measuring drainage water as well as measuring temperature and water level. In addition, some dams are equipped with vibration sensors and a local data acquisition facility that can be accessed remotely. In Switzerland, five dams are part of the Swiss National Strong Motion Network [1]. Signals are acquired upon the trigger level of, depending on the local conditions, $0.001, \ldots, 0.005\,\mathrm{g}$, being crossed. From the installation of the system in 1992 until the end of 2004, 502 acceleration time signals from 161 events have been recorded [2]. This yielded insight into the dynamic behavior of the dams in a limited number of points under earthquake excitation. However, to the knowledge of this author, no real attempt to monitor a dam's structural health based on its dynamic characteristics has been realized up to now. There are several reasons for this.

Firstly, the dynamic characteristics of a civil engineering structure are determined through its mass and stiffness. Before being able to relate changes in these characteristics to changes in structural health, any other source of changes in the dynamic characteristics has to be identified and its influence eliminated. For bridges, these "other sources" predominantly include temperature, because this may influence the stiffness of (asphalt) pavement and (flat) foundations. For dams, the situation is worse. Here, the main "other source" is the level of water in the reservoir. This simultaneously influences the mass and stiffness of the dam structure. This is discussed in a later section.

Second, determination of a dam's dynamic characteristics is not easy. Large dams tend to be located in remote areas where accessibility might be a problem. Performing a dynamic test on such a structure usually becomes an expensive challenge.

Although no real health monitoring system making use of changes in dynamic characteristics has been installed in a dam up to now, several steps in this direction have been undertaken in the last 20–30 years.

Step 1. Achieving a set of experimental data reflecting the real behavior of the structure and hence

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

becoming able to update an existing finite element (FE) model of the structure. Within the limits of linearity, further analyses can then be performed on a model known to be as close to reality as possible. This article presents two examples of this kind. Attempts to extend the boundaries of the problem to the surrounding rock and to the reservoir and to study dam–reservoir–soil interaction effects are discussed in [3–6]. Then, the challenge from the view of experimental side is the simultaneous measurement of dam vibrations and hydrodynamic pressure oscillations.

Step 2. Determining the influence of changes in the reservoir water level on the dynamic characteristics of a dam [5, 6]. This article presents one example of this kind.

1.2 Dynamic (dam) testing methods

Today, two basically different test procedures are available: forced vibration testing (FVT) and ambient vibration testing (AVT). FVT is also called *experimental modal analysis* or *traditional modal analysis* in the literature. For AVT, there are even more expressions used: "output-only modal analysis", "natural input modal analysis", and "operational modal analysis". For the sake of simplicity, we will stay with the expressions FVT and AVT here.

With FVT, the structure is excited dynamically using a vibration generator (shaker) and the resulting structural vibrations are measured. The dynamic excitation force is either measured with a load cell or calculated through determination of the inertial forces involved. The structure's vibrations are measured using accelerometers. The number of measurement points usually being much larger than the number of sensors disposable, the measurement is divided into several setups. One setup being finished, the sensors are moved to new positions and the next setup is measured. The force signal acts as a reference, which all response signals are related to. It is then possible to put together, e.g., the mode shapes determined to cover all the measured points. Data processing, i.e., determination of the structure's dynamic characteristics is based on the assumption that the artificial force produced by the vibration generator is the only source of the dam vibrations measured. We see that this assumption is not valid in cases where, e.g., machinery is operating close to the dam. We also see that this is not always a serious problem. With dams, problems might arise when the excitation through wind and/or waves is of the same order of magnitude as the shaker-induced vibrations.

With AVT, the structural vibrations excited by unknown ambient sources are measured. In the case of dams, these ambient forces can be wind, waves, and/or microtremors. Instead of using the force input signal as a reference signal, the response signal measured in (one or more) so-called reference points is used for this purpose. The reference points have to be measured in all setups. Data processing, i.e., determination of the structure's dynamic characteristics, is based on the assumption that the frequency content of the exciting forces is flat, i.e., more or less of the white noise type.

Besides being much cheaper than FVT, AVT has the advantage of being a multiple input multiple output (MIMO) procedure. This means that the ambient forces exciting the structure are simultaneously acting on many points of the structure. Especially for civil engineering structures, FVT is usually a single input multiple output (SIMO) procedure. It can be seen from the examples, discussed later, that using more than one shaker on a structure is not the standard procedure. Choosing the proper position of the driving point, where the vibration generator is located is hence always one of the crucial problems to be solved. The shaker should not sit in a node of the shape of one of the modes of concern. The same problem arises with AVT tests when only one reference point is used. This can, however, be solved easily through using more than one reference point. Performing a preliminary FE analysis is also a good means to prevent problems with the placement of shaker and reference points.

1.3 Short history of dynamic testing of structures

Historically, efficient FVT procedures have been developed much earlier than AVT procedures. The first important step was development of the fast Fourier transform (FFT) algorithm, which allowed easy calculation of a frequency spectrum from a measured time signal in 1965 [8]. Subsequently, further signal processing routines were developed by mechanical engineers dealing with "small" structures easily fitting into a laboratory. In such cases, artificial

excitation, e.g., using a hammer or a small electrodynamic shaker was not a problem. The first International Modal Analysis Conference (IMAC) 1, held in 1982, and its successors reflected the respective development. In the late 1970s, when civil engineers started to make use of the methods developed in the "mechanical world" they tried to apply the well-developed FVT methods as far as possible.

However, there are civil structures like, e.g., suspension bridges or dams, which are not easy to excite artificially to a sufficient degree. This is why AVT soon became a topic for civil engineering structures. Of course, "manual signal processing" using a two-channel frequency analyzer and determining the amplitude and phase relationships between two time signals has been possible since about 1975. However, in the mid 1990s, the time had come for civil engineers to develop the procedures necessary for the efficient processing of an AVT test with a large number of degrees of freedom involved [9]. Today, even more efficient methods operating in the frequency and/or the time domains are available to process the signals of an ambient test [10–12].

1.4 Short history of dynamic dam testing

Probably, first attempts to experimentally determine a dam's dynamic characteristics were undertaken in the late 1970s. In a book edited by Graham Tilly [13], Calciati is cited having tested 10 dams in Italy [14], and a British team of the University of Bristol and the British Research Establishment (BRE), is cited to have tested 6 dams in the United Kingdom and in Switzerland [15, 16]. Flesch of Austrian Arsenal is cited to have performed tests on a dam with the reservoir being full and empty [17, 18]. Fanelli and his colleagues report on extensive tests performed during three years on Talvacchia Dam in Italy [5, 6]. All these tests were performed between 1978 and 1988.

As a consequence of the historical development described earlier, the FVT method using vibration generators was applied in all the above-mentioned cases. Since hammer is not really a well-suited instrument to excite a dam, vibration generators of the unbalanced mass type were built. These are relatively easy to install and they can be operated with a comparatively small amount of energy. The inertial force produced is concentrated into one simply harmonic vibration. However, performing a test covering the typical frequency band $f = 2, \ldots, 20\,\text{Hz}$ was a very time-consuming procedure. Assuming a frequency resolution of $\Delta f = 0.1\,\text{Hz}$, the test consisted of 180 individual tests using a different frequency of excitation (and keeping the response measurement points always in the same place). To keep the amplitude of the dynamic force generated in certain limits, the position of the unbalanced masses had to be adapted from time to time. Usually, this force could not be measured directly. Therefore, modern modal analysis methods could not be applied. It was, however, possible to determine natural frequencies, mode shapes, and damping of earth dams, gravity dams, and arch dams like the Emosson dam in Switzerland with a height of 180 m. With the latter, some problems arose with the wind influencing the dam vibrations [16]. For strong wind conditions, the vibrations excited through wind were of the same order of magnitude as those excited with the shaker.

As a next step, servohydraulic vibration generators were built e.g., in Switzerland by the Swiss Federal Laboratories for Material Testing and Research, EMPA. (A similar device was also built by Arsenal in Austria.) Besides a simply harmonic vibration's these could produce a broadband random-type force signal covering the whole frequency band of concern. The energy available was distributed over a larger frequency band, but the tests were much less time consuming. This allowed increasing the number of measurement points significantly. Instead of using the time slot available for testing with sweeping through the frequency band of interest, it was used for roving the sensors available over the structure in several test setups. As a result, the measurement point grid density and hence the resolution of the mode shapes could be increased significantly. As is shown with the three tests described here, respective limits are dictated by the accessibility of points to be measured only. Tests performed by EMPA on two dams using servohydraulic shakers are described, in more detail, later.

The AVT signal-processing routines having become available in the mid 1990s, ambient tests were subsequently performed on dams. Such a test, again performed by EMPA, is also described later.

Today, dynamic tests on dams are performed relying either on the AVT method [7] or, again, on

the FVT method using unbalanced mass exciters [4]. These two methods are much cheaper to be applied than FVT using servohydraulic vibration generators.

2 VIEUX EMOSSON DAM

2.1 The dam

Vieux Emosson dam is a curved concrete gravity structure located close to the border between Switzerland and France in the Mont Blanc region at a height of 2200 m above sea level. With a length at the crest of 175 m, a maximum height of 42 m, and a concrete volume of 62 500 m^3, it is of moderate size (Figures 1 and 2). The dam consists of 13 monolithic blocks of roughly 13.5 m length each. The block width varies between 4 and 7 m at the dam crest and reaches 20 m at the foundation (Figure 3).

No machinery is located close to the dam. Vieux Emosson dam is simply collecting water throughout the summer and the reservoir is emptied in autumn through letting the water flow freely in the reservoir of Emosson dam, which is located downstream of Vieux Emosson dam. This means that no disturbing dam vibrations generated by machinery occur.

Figure 1. Vieux Emosson dam as seen from the air side.

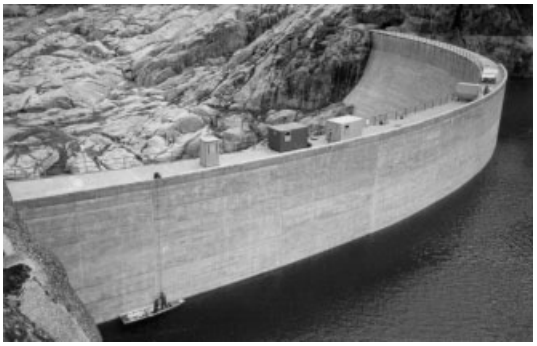


Figure 2. Vieux Emosson dam as seen from the water side.

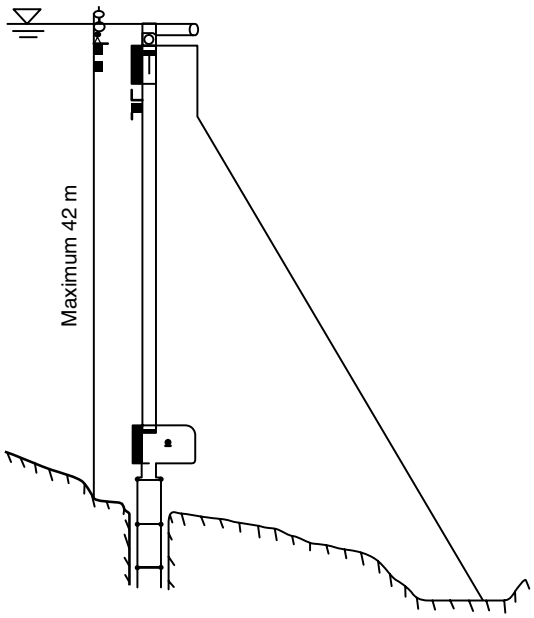


Figure 3. Vieux Emosson dam cross section.

As the dam site is accessible with cars but not with trucks, all the equipment discussed here had to be flown to the dam crest using helicopters of the Swiss Army (Figure 4).

The dam is owned by the Swiss Federal Railways, SBB. The tests were financed as an EMPA research project [19–21].

Figure 4. Helicopter taking over equipment at the base station.

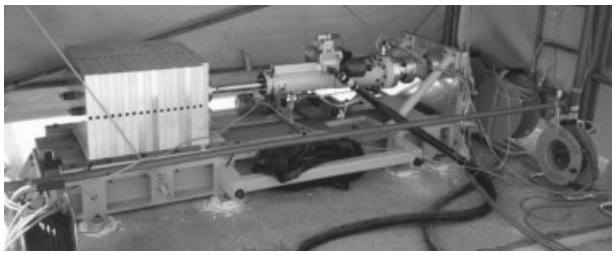


Figure 5. Servohydraulic vibration generator fixed to the crest of Vieux Emosson dam.

2.2 The excitation

The main part of the vibration generator is a 32-kN cylinder with a 1000 kg mass fixed to the piston rod (Figure 5). A load cell is located between the cylinder and its supporting device. The 250-mm-stroke cylinder is equipped with a $63\,l\,min^{-1}$ servo-valve and controlled through an electronic circuitry. The hydraulic power pack driving the cylinder produced $40\,l\,min^{-1}$ of 280 bar oil (Figure 6). In order not to disturb the measurements, the power pack was supported by air springs. A 60-kW diesel generator drove the hydraulic power pack (Figure 7).

Points 10 and 17 were chosen as driving points (where the vibration generator was located). The influence of the shaker location on the results is discussed in [20]. Here, results for only the shaker position 10 are presented.

The driving signal of band-limited burst random type was provided by the modal analysis software package. To achieve a force spectrum being flat in the region of interest, $f = 5, \ldots, 40$ Hz, the signal driving the cylinder in the displacement-controlled mode had to be fine-tuned on site. The peak force amplitude was chosen to 28 kN.

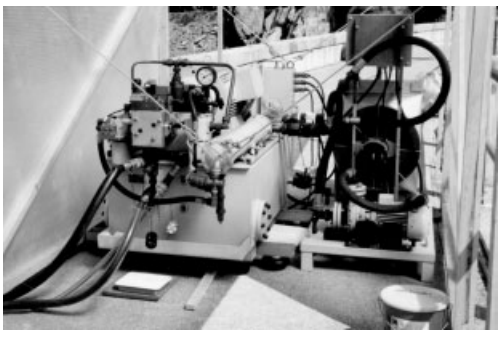


Figure 6. Hydraulic power pack (to the left) and air cooler (to the right).

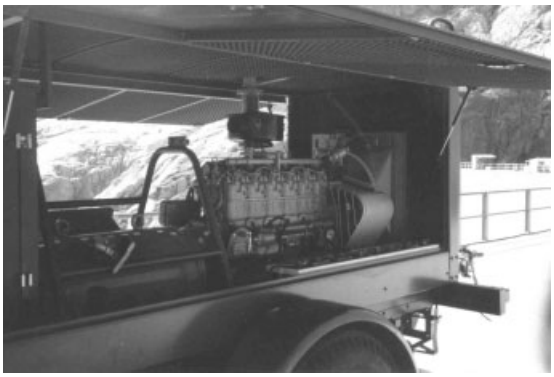


Figure 7. Sixty kilowatt diesel power generator.

2.3 The response

Three accelerometers Brüel&Kjær 8306 mounted orthogonally to each other on a supporting steel plate were used to measure the structural response (*see* **Microelectromechanical Systems (MEMS)**). The sensitivity of this type of instruments is $10\,V\,g^{-1}$ and the resolution is $10^{-6}\,m\,s^{-2}$. As Vieux Emosson dam has an inspection gallery close to the foundation only, the measurement points were located on the dam crest and above the water line (Figures 8 and 9). One of the questions that should have been answered through the tests was, is there any relative movement between two adjacent blocks? As a consequence, measurement points were located on both sides of the joints. The measurement point grid consisted of a total of 53 points (Figure 10).

2.4 Signal acquisition and processing

The forcing signal and six response signals were acquired simultaneously using an 8-channel front

Figure 8. 3-D accelerometer response measurement point fixed to the dam crest.

Figure 9. 3-D accelerometer response measurement point fixed to the dam water face.

end, a computer, and dedicated software. From this, frequency response functions (FRFs) were calculated and the modal parameters subsequently extracted. Figure 11 shows the modal indicator function (MIF), indicating the existence of nine physical modes with a maximum modal participation at mode 2, $f = 9.90$ Hz. Table 1 gives the frequency and damping in percent of critical for the first six modes.

The test revealed that there is some relative movement between two adjacent blocks and that the movement at the dam foundation is not equal to zero. The latter indicates the existence of soil–structure interaction. These facts are discussed in more detail in [19] and [21].

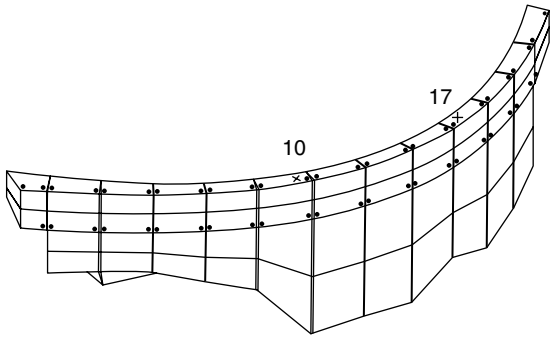


Figure 10. Measurement point grid (dots) and location of the driving points 10 and 17 (crosses).

2.5 Finite element modeling and updating

After completion of the tests, a preliminary FE model was updated using a dedicated software package [21]. Updating included the introduction of a soft layer between the blocks and optimizing the stiffness of the springs reflecting soil–structure interaction. The modal assurance criterion (MAC) value is used to compare the coincidence of two mode shapes. $MAC = 1$ (also given as $MAC = 100\%$) means that two shapes are identical, $MAC = 0$ means that two shapes are orthogonal (completely different from each other). For the final model, MAC values comparing the mode shapes as determined experimentally and as calculated from the FE model are higher than 0.92 for the first five modes. This can be rated as "very good". Figure 12 shows a comparison between the frequencies and shapes of the first four modes as extracted from the experiment and as calculated using the FE model.

3 NORSJÖ DAM

3.1 The dam

Norsjö dam is located in northern Sweden, west of Skellefteå. The dam is a cylindrically shaped reinforced concrete structure with a length at the crest of 169 m, a maximum height of roughly 46 m, and a radius of curvature of $R = 110$ m (Figures 13 and 14). The dam width is 2.5 m at the crest and 5.5 m at the foundation (Figure 15). At the dam downstream side, a reinforced concrete wall of 150 mm width is located at a 0.8 m distance of the main

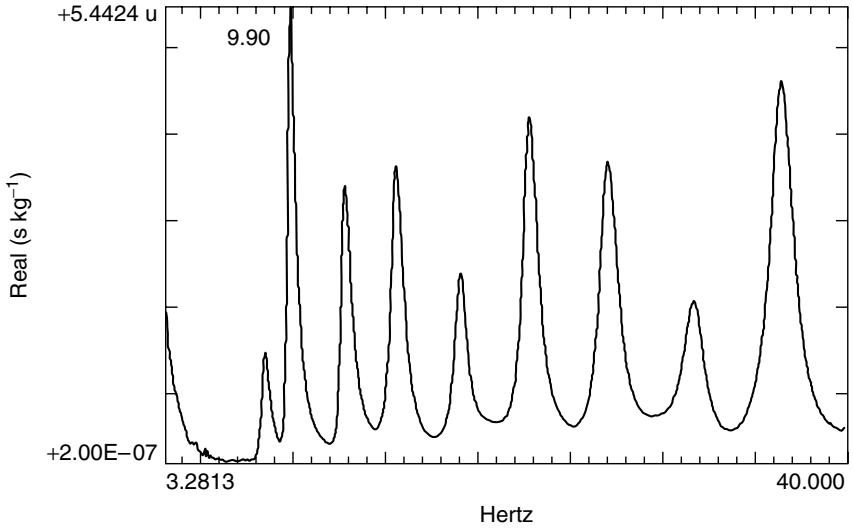


Figure 11. Modal indicator function (MIF) for Vieux Emosson dam.

Table 1. Frequency and damping of the first six modes of Vieux Emosson dam

Mode number	Frequency (Hz)	Damping (%)
1	8.42	3.5
2	9.90	2.6
3	12.80	2.8
4	15.62	2.7
5	19.10	2.4
6	22.86	2.5

structure. This wall and the dam's main structure are connected through a large number of reinforced concrete bars of 150 mm × 150 mm cross section positioned in a grid with a mesh width of 2.88 m. Wooden planks are located on every second row of these concrete bars, which makes the downstream face of the main structure easily accessible over the whole dam height (Figure 16). This wall prevents the downstream face of the dam from freezing and hence from locking the water flow. Norsjö dam is part of a river exploitation system with the turbines and generators located at the dam footing. Although the reservoir surface freezes in winter, the production of electricity continues throughout the year.

Norsjö dam is owned by Vattenfall Utveckling AB. The tests were jointly financed by Vattenfall and an EMPA research project [22, 23].

3.2 The problem

A preliminary FE analysis performed at the Royal Institute of Technology, KTH, Stockholm [24], had shown that the behavior of the structure was strongly dependent on the real state of the boundary conditions: Where is the connection to the surroundings pinned, where is it elastic, and where is it clamped? Having a look at the cross section shown in Figure 15, it becomes clear that it is, e.g., not easy to predict the behavior of the connection between the footing of the dam and the rock. It was, therefore, the major goal of the tests described here to determine an FE model being as close to reality as possible.

3.3 Preliminary ambient test

To get an idea of the dam fundamental natural mode a preliminary test was performed. Two facts resulted from this: (i) the dam fundamental frequency is $f \approx 3.2$ Hz and (ii) operation of the turbines and generators located at the powerhouse at the dam bottom resulted in disturbing peaks in the dam vibration spectra (Figure 17). The very narrow peaks in the spectrum could all be related to either the Kaplan or the Francis turbine located at the powerhouse. This meant that testing was reasonable at times when

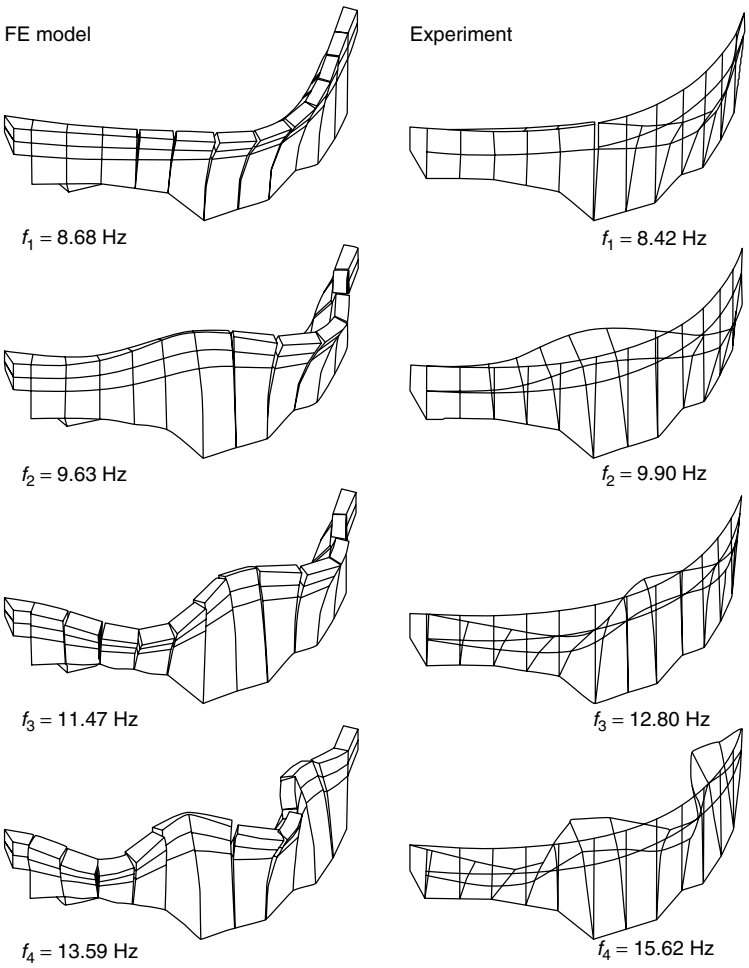


Figure 12. Natural frequencies and mode shapes of the first four Vieux Emosson dam modes.

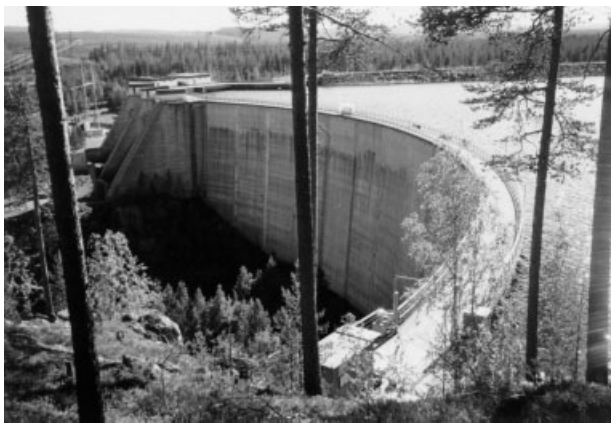


Figure 13. Norsjö dam. To the left, spillways and powerhouse.

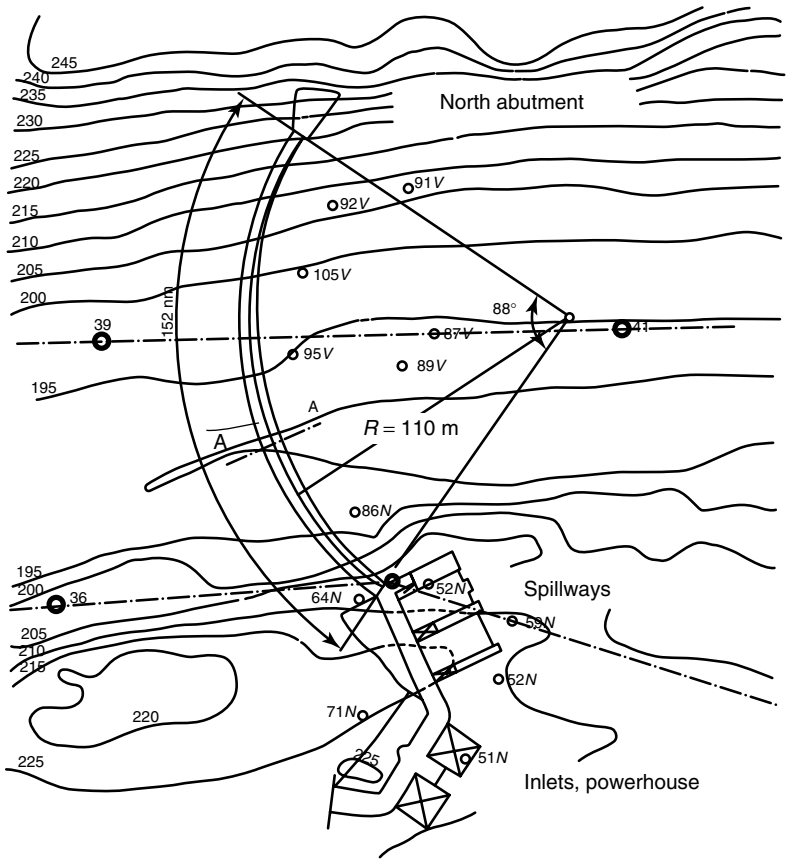


Figure 14. Plan view of the Norsjö Dam and the adjacent structures.

the powerhouse was shut off only. As a consequence, tests were performed during the night and during weekends only and close cooperation with the powerhouse control center was necessary. The consequences of the dam natural frequency being $f \approx 3\,\mathrm{Hz}$ are discussed below.

3.4 The excitation

A similar servohydraulic shaker as for the Vieux Emosson tests was used (Figures 18 and 19). However, as there were no insurmountable problems with the accessibility of the dam here, a hydraulic power pack with an $80\,\mathrm{l\,min^{-1}}$ capacity was chosen instead of the $40\,\mathrm{l\,min^{-1}}$ power pack used for Vieux Emosson. The reason for this was the fact that the Norsjö dam fundamental frequency, $f \approx 3.2\,\mathrm{Hz}$, lies significantly lower than the $f = 8.42\,\mathrm{Hz}$ of Vieux Emosson dam.

The basic problem with any kind of vibration generator using inertial forces is the force produced being directly proportional to the acceleration of the moving mass and hence to the square of the frequency of excitation. This means that it is difficult to generate high forces at low frequencies. Whereas it was acceptable for Vieux Emosson dam to have an excitation spectrum being flat for $f = 5, \ldots, 40\,\mathrm{Hz}$, it was preferred for Norsjö dam to dispose of a force spectrum being flat for $f = 2, \ldots, 40\,\mathrm{Hz}$. This could be achieved using the larger power pack because this allowed to make use of the full capacity of the $63\,\mathrm{l\ min^{-1}}$ servovalve and of the 250-mm-cylinder stroke for low frequencies. The forcing signal was of the continuous random type with a maximum force amplitude of 32 kN (Figure 20).

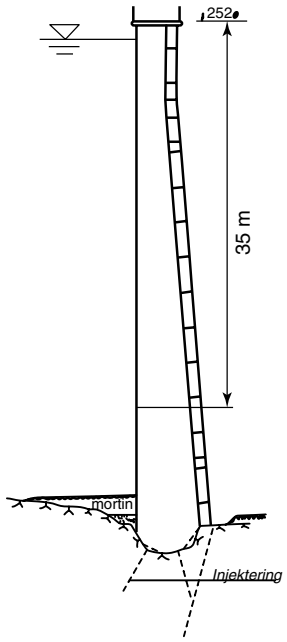


Figure 15. Norsjö dam cross section.

Figure 16. Between water (to the right) and air (to the left).

The standard problem with having enough electric power available to drive the power pack could be solved in a very elegant way here. If we need a 90-kW connection, we can simply plug into the local network (Figure 21).

On the basis of the results of the preliminary FE analysis performed at KTH earlier [24], point 6 was chosen as the driving point (Figure 22). At the end of the first day of testing, this choice was confirmed through analyzing the results for the measurement points on the dam crest. The shaker was obviously not sitting in a node of the shape of a mode of concern.

3.5 The response

The measurement point grid consisted of 227 three-dimensional measurement points distributed over the dam crest and the downstream face (Figure 22). This grid was subsequently extended to the rock foundations and to the spillway and inlet/powerhouse structures, where accessible. The number of measurement points thus increased to 270. The full measurement point grid is reflected in the graphics shown in Figures 30–33.

Three sensor units consisting of three Brüel&Kjær 8306 accelerometers (Figure 23) and a fourth unit consisting of three PCB 393B31 accelerometers (Figure 24) were used simultaneously. Both sensor types have similar specifications (Section 2.3).

3.6 Signal acquisition and processing

The forcing signal and 12 response signals were acquired simultaneously using a 16-channel front end, a computer, and dedicated software (Figure 25). The sampling rate was $s = 100\,\text{Hz}$, the length of a time window was roughly $T = 41\,\text{s}$. No disturbing effects due to nonstationary environmental conditions leading to system nonlinearities were to be observed. Water-level variations in the reservoir were less than 100 mm for the whole week of testing. From the time signals, FRFs averaged over eight times windows were calculated and the modal parameters extracted. Both, the driving point FRF (Figure 26) and the MIF (Figure 27), indicate the existence of 12 natural modes in the frequency range $f = 3, \ldots, 13.5\,\text{Hz}$.

The shape at the crest, the deformation type, the frequency, and the percentage of critical damping of

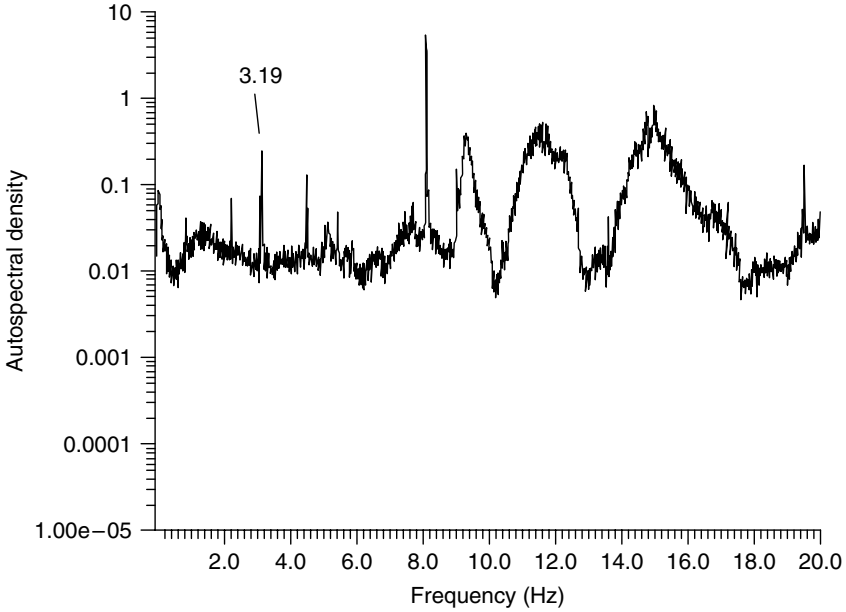


Figure 17. Spectrum of the Norsjö dam vibrations measured with the powerhouse under operating conditions.

Figure 18. The servohydraulic shaker fixed to the Norsjö dam crest. In the background, the air-sprung hydraulic power pack to the left and the air cooler to the right.

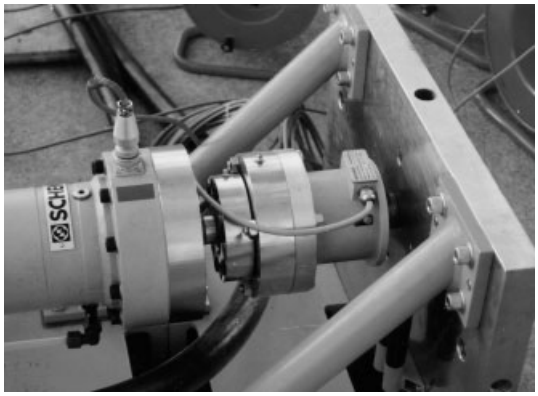


Figure 19. Load cell between the cylinder and the supporting structure.

the 12 modes identified are given in Figure 28. Three basically different shape types can be distinguished: antimetric horizontal bending (AHB), symmetric horizontal bending (SHB), and vertical bending (VB). The horizontal bending modes follow the well-known schedule of an arch-type structure. There is no horizontal bending mode without a node in the crest shape. However, the respective crest shape appears at mode 6, which turns out to be a VB mode (Figure 28).

Unfortunately, it is not possible to discuss all details concerning modes shapes here, in detail. However, to summarize, for all modes, the crest shape indicates the dam being simply supported at the spillway side and completely clamped in at the rock side (Figure 28). The vertical shape of the modes indicates the boundary condition rock/structure at the dam bottom-line to be elastic with being closer to simply supported than to fully clamped in (Figures 30–33).

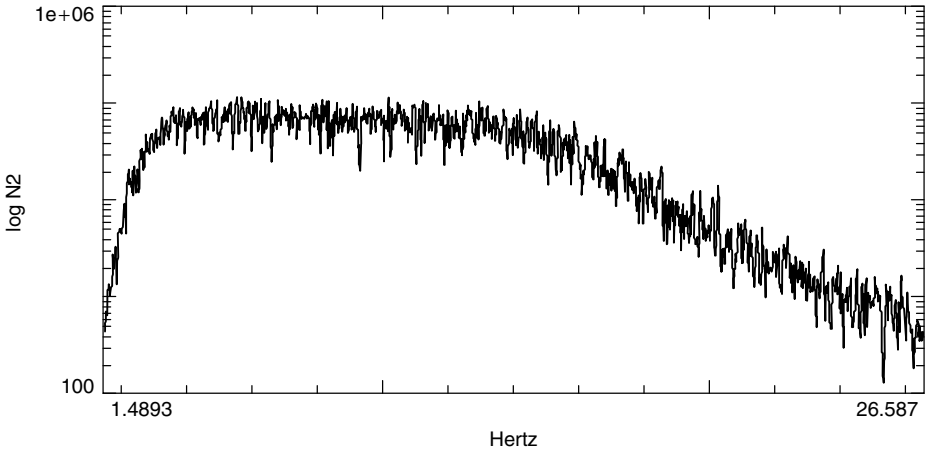


Figure 20. Spectrum of the force introduced into Norsjö dam with the shaker.

Figure 21. Swedish–Swiss 90-kW power connection.

3.7 Finite element modeling and updating

The FE model for Norsjö dam consisted of roughly 1000 solid elements representing the dam's main structure, 1500 plate elements representing the secondary wall on the downstream side, and 500 beam elements for the connections "main structure—secondary wall". The boundary conditions in the structure/rock contact area were modeled using three-dimensional elastic springs. Upon updating using a dedicated software package, a quite nice correlation between the experimental and analytical modes could be achieved (Table 2 and Figure 29). This applies, above all, for the first five modes where not only MAC is between 0.8 (80%) and 0.95 (95%) but also the sequence of the modes is the same for experiment and modeling. For the higher modes, there is some disorder in the mode sequence and MAC is between 0.58 and 0.86. Further discussion in this matter can be found in [22]. In Figures 30–33, the experimental and the analytical mode shapes are displayed overlaid; the wire-framed shape shows the experimental results.

4 MAUVOISIN DAM

4.1 Introduction

The goal of the tests discussed here was to prove that it is also possible to determine the dynamic properties of a large dam using ambient methods. To investigate into the effect of the reservoir water level on the dam dynamic characteristics, the dam was tested seven times. Subsequent continuous monitoring of the dam

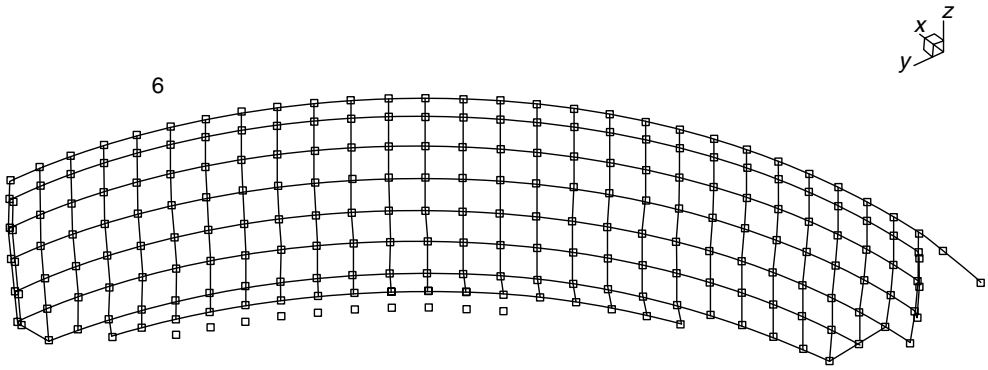


Figure 22. Norsjö Dam: primary measurement point grid as seen from the downstream side. Later on this was extended to also cover the spillways and powerhouse. Point 6 indicates the driving point where the shaker was located.

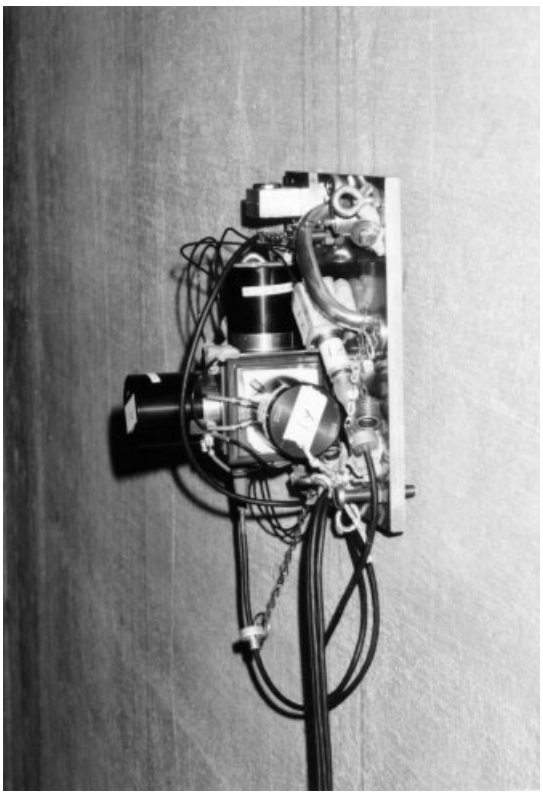


Figure 23. Norsjö Dam: 3-D measurement point with Brüel & Kjær sensors.

during 180 days in 1998/1999 as well as comparison of ambient dam vibrations and vibrations induced by an earthquake occurring during this time is discussed in detail in [25].

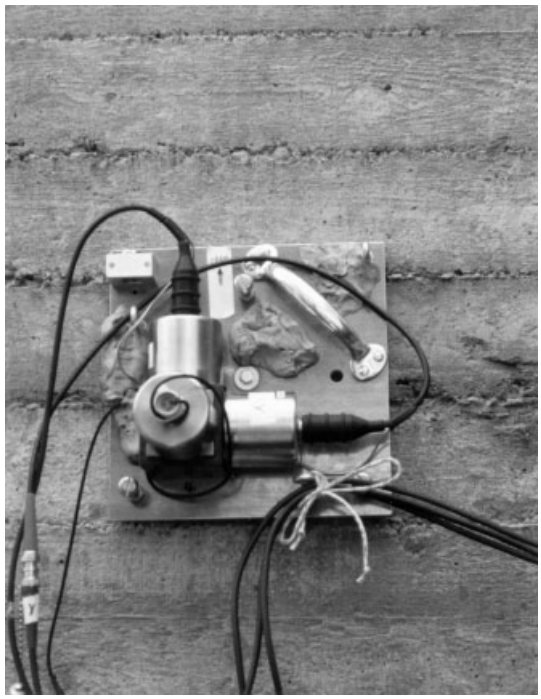


Figure 24. Norsjö Dam: 3-D measurement point with PCB sensors.

4.2 The dam

Mauvoisin dam is a double-curved concrete arch dam with a height of 250 m and a crest length of 520 m (Figure 34). The dam width is 12 m at the crest and 53 m at the foundation. The crest lies at 1976 m above sea level. The dam consists of 2.1 million cubic

Figure 25. The measurement center on Norsjö dam crest.

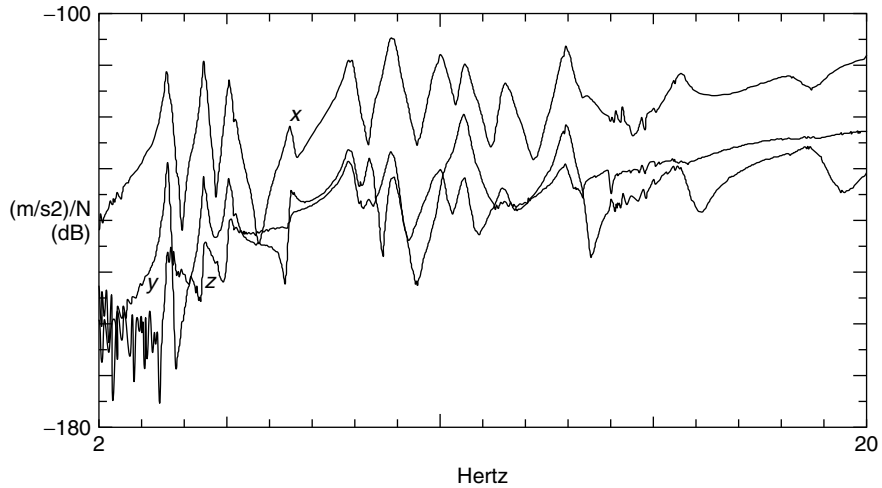


Figure 26. Norsjö Dam: driving point 6 FRF for the x, y, and z directions.

meters of concrete; the reservoir has a volume of 204 million cubic meters. Mauvoisin dam is owned by the Forces Motrices de Mauvoisin SA, Sion, Switzerland. The tests described here were financed through an EMPA research project [26, 27].

When it comes to planning a dynamic test on a dam, one of the first problems to deal with is the accessibility of the measurement points. Usually, small and medium-sized dams like Vieux Emosson dam do not have horizontal inspection galleries between the crest and the foundation. A gallery at the foundation is not very interesting from the point of view of dam dynamics. It is used to connect vertical galleries allowing for checking of the dam static deformations and to check for possible incoming water. The conditions at Norsjö dam were very fortunate allowing installing a very tight grid of measurement points and hence determining the mode shapes

with an extraordinary high resolution. In this respect, Mauvoisin dam is somehow in-between the two.

The nice thing with Mauvoisin dam is that, as a consequence of the dam height having been increased in 1990, there is very large gallery at the old crest level, 1961 m above sea level (Figure 35). In the snow-free seasons, this gallery is accessible with trucks and provides a very nice sheltered area at an interesting dam level. Further galleries are at 1957, 1885, 1837, and 1789 m and at the foundation, 1728 m above sea level. With the exception of the 1837-m gallery, all the galleries are accessible through an elevator at the east side of the dam.

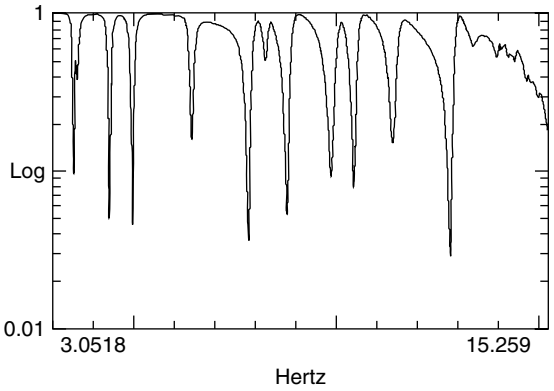


Figure 27. Norsjö Dam: modal indicator function, MIF.

4.3 The test program

Figure 36 shows the relationship between the seven tests performed (phase 1–7) and the reservoir water level.

4.4 The excitation

Although ambient tests were planned, excitation has to be a topic here. Taking a look at the Mauvoisin installation scheme, it becomes clear that several "disturbing vibration generators" exist (Figure 37). Operation of the Chanrion powerhouse located close to the dam crest was easy to identify and to work

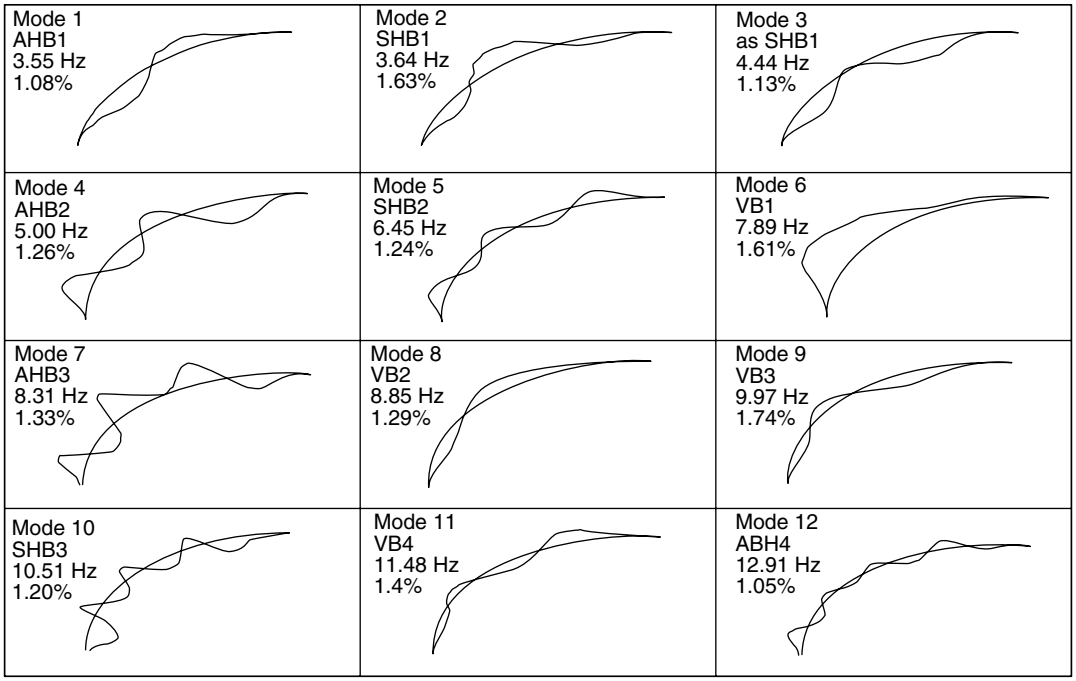


Figure 28. Frequency, damping in percent of critical and crest mode shape for the 12 first modes of Norsjö dam. AHB = antimetric horizontal bending, SHB = symmetric horizontal bending, and VB = vertical bending. The spillway side is to the left, the "rock side" to the right of the shape.

Table 2. Norsjö dam natural frequencies as calculated finite element analysis (FEA) and as measured experimental modal analysis (EMA) and MAC values comparing the mode shapes

Mode pair	FEA number	Frequency (Hz)	EMA number	Frequency (Hz)	MAC (%)
1	1	3.66	1	3.55	87.4
2	2	3.71	2	3.64	79.5
3	3	4.64	3	4.44	95.8
4	4	4.92	4	5.00	94.7
5	5	6.16	5	6.45	89.2
6	6	7.70	7	8.31	58.6
7	9	8.13	6	7.89	83.1
8	10	8.85	8	8.85	86.1
9	11	9.48	10	10.51	72.8
10	12	9.75	9	9.97	77.2
11	13	10.81	11	11.48	74.6
12	14	11.58	12	12.91	69.4

around (Figure 38). This "work around" is discussed in more detail in Section 4.6.

It took some time to find out that operation of the Fionnay power station had quite bad effects on the signals measured. Opening and closing of the valves controlling the water flow to the turbines produced shock waves traveling back to the dam and significantly disturbing the dam's natural vibrations. As these processes could not be easily controlled, tests were performed with the Fionnay power station being out of operation for phases 2–7.

Furthermore, problems arose with changing wind conditions. This is discussed in detail in [26, 27]. Summarizing here, the level of the dam vibrations was significantly influenced by the wind conditions. Therefore, not all the maximum 16 dam natural modes could be identified for all seven tests performed (Section 4.8).

4.5 The response

The dam vibrations were measured using three-dimensional force balanced accelerometers, Kinemetrics FBA-23, and one-dimensional FBA-11. These have a full-scale range of $\pm 0.5\,\mathrm{g}$, a sensitivity of $5\,\mathrm{V\,g^{-1}}$ and a dynamic range of 140 dB for frequencies $f = 0, \ldots, 10\,\mathrm{Hz}$. The 16-channel signal-conditioning hardware consisting of amplifiers and filters is described in detail in [26].

Seven tests were performed between June 1995 and October 1996 (Figure 36). All the measurement points shown in Figure 39 were covered in one of these tests only. For the other tests, the measurement points were located on the "old crest" level only.

This is quite easy to understand, because distributing the cables from the "old crest" level, where the measurement center was located, to the lower galleries was very difficult and very time consuming. It can be seen from Figure 34 that from two of the lower galleries access to a device mounted on the dam air side was possible. Let us call this device an "air ladder". Here, the cables could be led

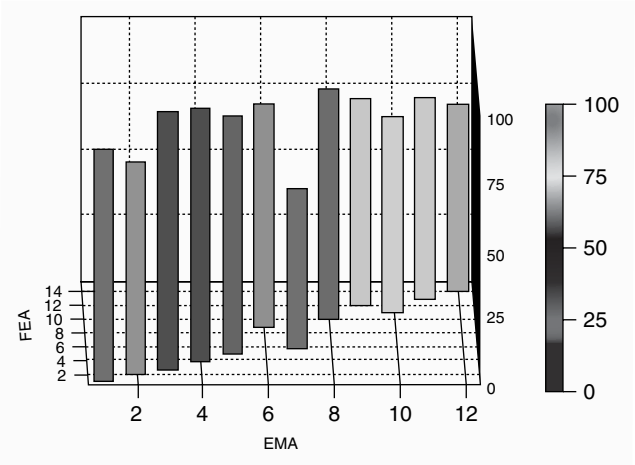


Figure 29. Norsjö Dam: graphical representation of the MAC values comparing the mode shapes as calculated (FEA) and as measured (EMA).

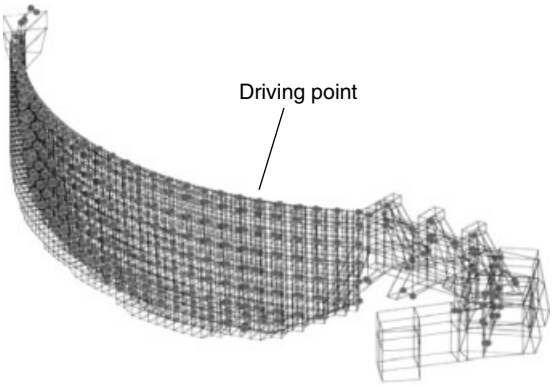


Figure 30. Norsjö Dam: finite element model and measurement point grid overlaid.

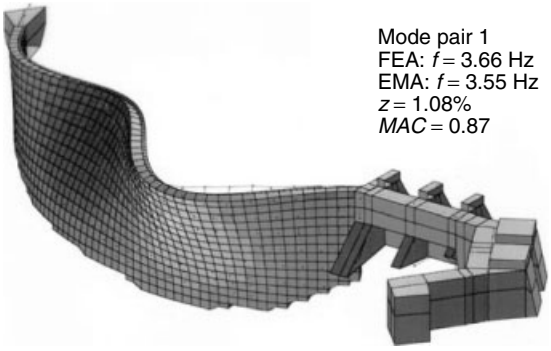


Figure 31. Norsjö dam mode pair 1.

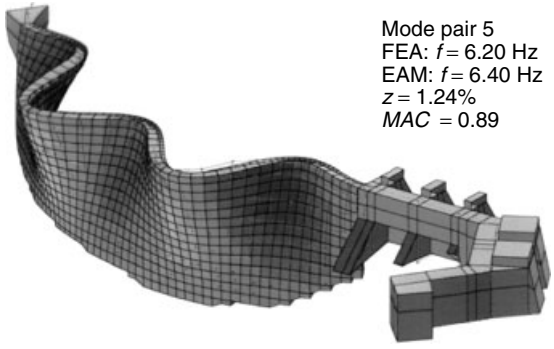


Figure 32. Norsjö dam mode pair 5.

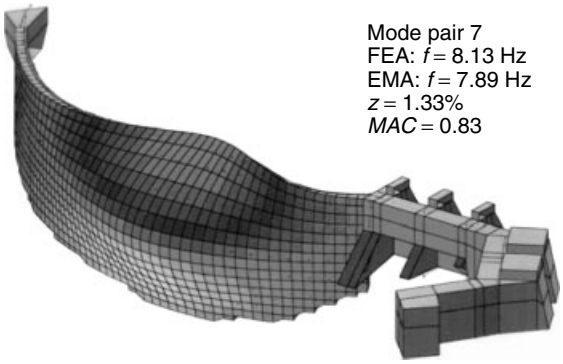


Figure 33. Norsjö dam mode pair 7.

Figure 34. Photo of Mauvoisin dam downstream face. The "air ladders" mentioned in Section 4.5 can be seen.

down to a lower gallery through the air. Access to the gallery without this air ladder was possible through a vertical tunnel only. This was quite a bad experience. Another bad experience was the fact that Mauvoisin dam is one of the favorite places for base jumpers to practice. The opening of their parachutes sounds very much like an explosion and if we are situated on one of these air ladders and do not hear them coming, and their parachutes open some meters from us, we are in for a shock.

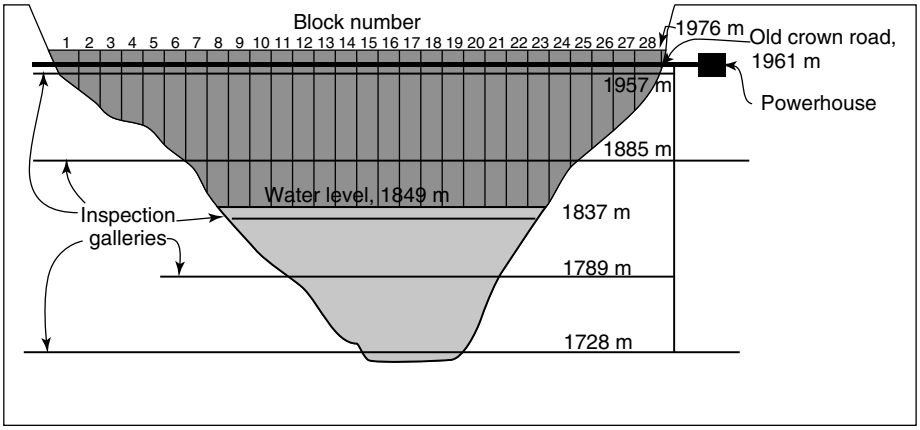


Figure 35. Mauvoisin dam. Topology of blocks and galleries as seen from downstream. The elevator is indicated to the right.

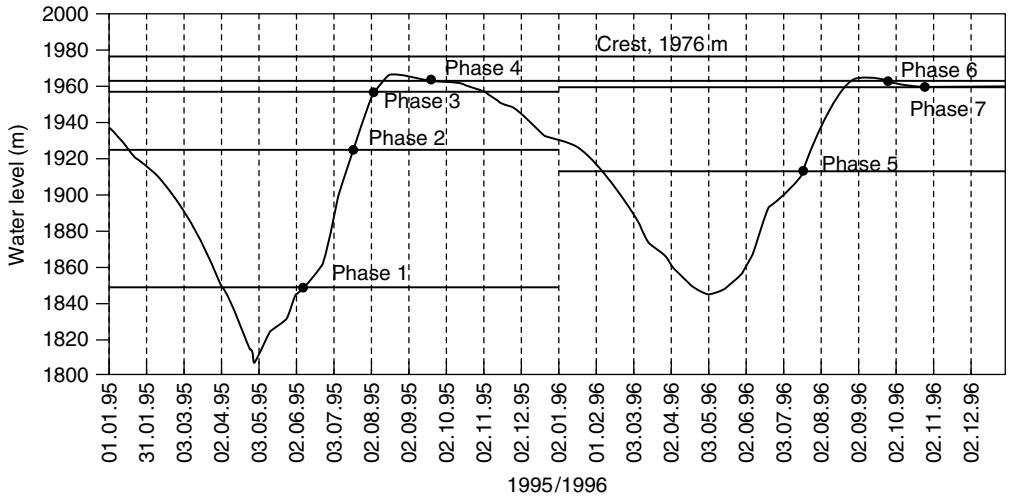


Figure 36. The seven test phases and the respective reservoir water level.

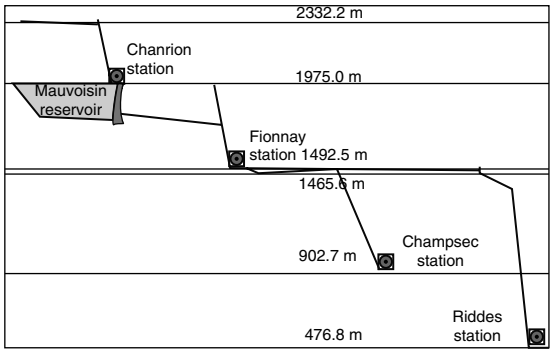


Figure 37. The Mauvoisin exploitation system.

4.6 Signal acquisition and processing

The sampling frequency was $s = 40\,\text{Hz}$, the length of the time window acquired per setup was either $T = 54\,\text{s}$ or $T = 108\,\text{s}$. Figure 40 shows a typical time signal for both a low-level and a high-level ambient excitation state. The amplitude ratio is about 1 : 30.

The signals were processed by Andreas Felber using the software package that he had developed [9]. Today, his method is called *peak picking*. This means that the amplitude and phase relationships

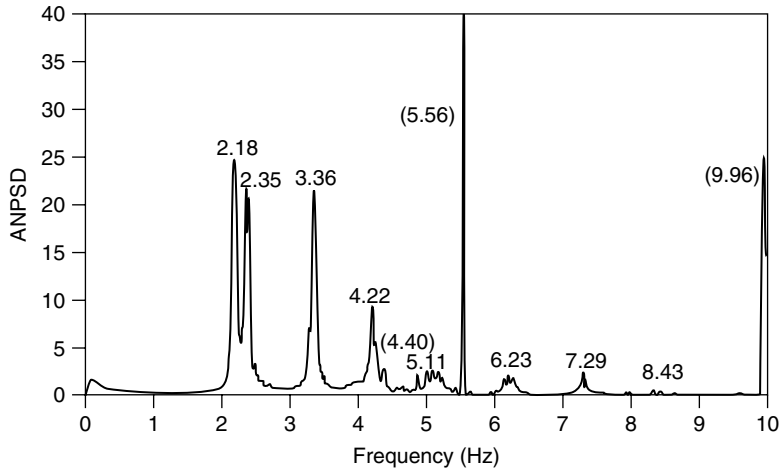


Figure 38. Typical frequency spectrum for the upstream/downstream direction with the Chanrion powerhouse in operation. ANPSD = averaged normalized power spectral density. Frequency values in parentheses are machine-excited harmonic vibrations.

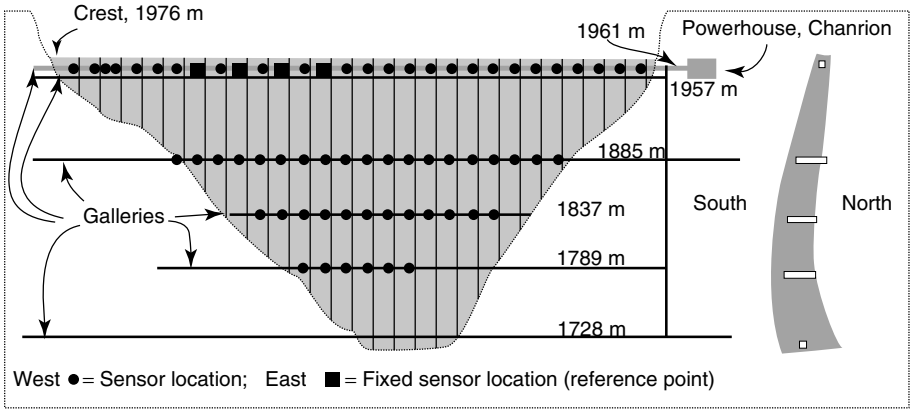


Figure 39. Mauvoisin dam measurement point grid as seen from downstream.

between the signals of a reference sensor, degree of freedom (DOF), and a roving sensor DOF is calculated for the frequency line manually picked from the spectrum. The advantage of this procedure is that peaks representing machine-induced harmonic vibrations can easily be kept out of the signal processing (Figure 38). This is not the case for traditional modal analysis procedures used for FVT tests where signal processing is automated. The same would also apply to modern stochastic subspace identification (SSI) techniques [11], used to process ambient tests in the time domain. In the last two years, the problem of automatically removing "disturbing" peaks related to machine-induced, purely harmonic vibrations from automated signal-processing algorithms, has been successfully dealt with in [12]. These algorithms consider the fact that the kurtosis of a natural structural vibration and of machine-excited purely harmonic vibration is not the same. Hence, the algorithm checks the kurtosis for every peak in the frequency spectrum (transferred back to time domain) and takes out the "bad" ones.

Finally, Felber's software packages did not deal with damping. Therefore, damping is not a topic here. Helmut Wenzel later on extended the Felber software packages with a routine to estimate damping.

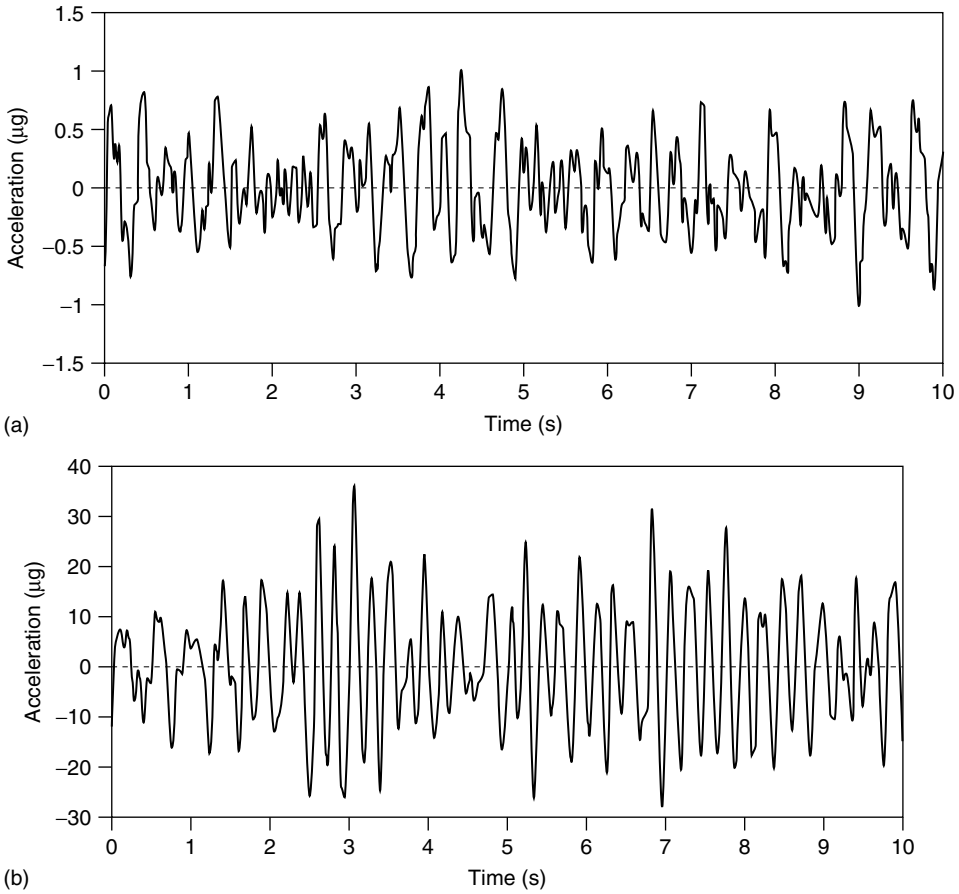


Figure 40. Typical acceleration time signal as acquired at block 7 of Mauvoisin dam for phase 1, water level 1849 m (a) and for phase 2, water level 1924 m (b).

Estimation of damping coefficients from ambient tests is included in modern commercial signal-processing software packages.

4.7 Modal parameters

Figure 41 shows the Mauvoisin dam crest shape for the first eight modes determined for a water level of 1.849 m. As for Norsjö dam, there is no node-free mode shape for horizontal bending.

4.8 Influence of the reservoir water level

Figure 42 shows the natural frequencies of the first 16 modes of Mauvoisin dam as a function of the

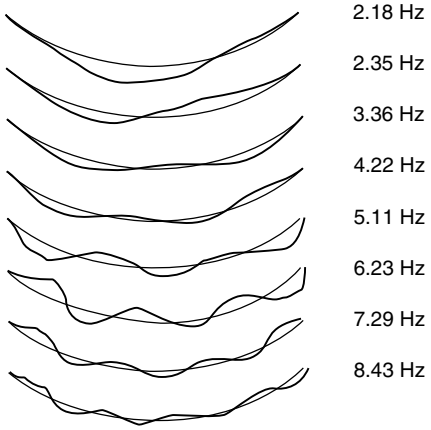


Figure 41. Mode shapes of Mauvoisin dam crest as determined for phase 1, water level 1849 m.

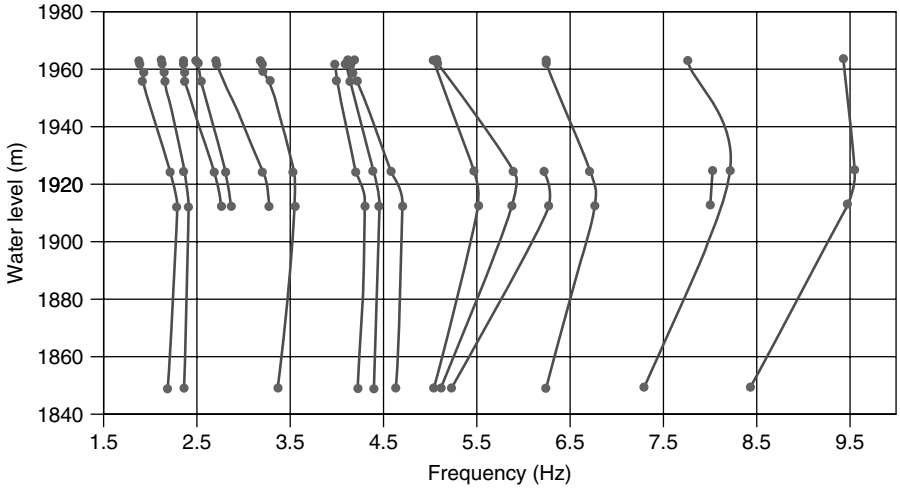


Figure 42. Frequencies of modes 1–16 of Mauvoisin dam as identified for different reservoir water levels.

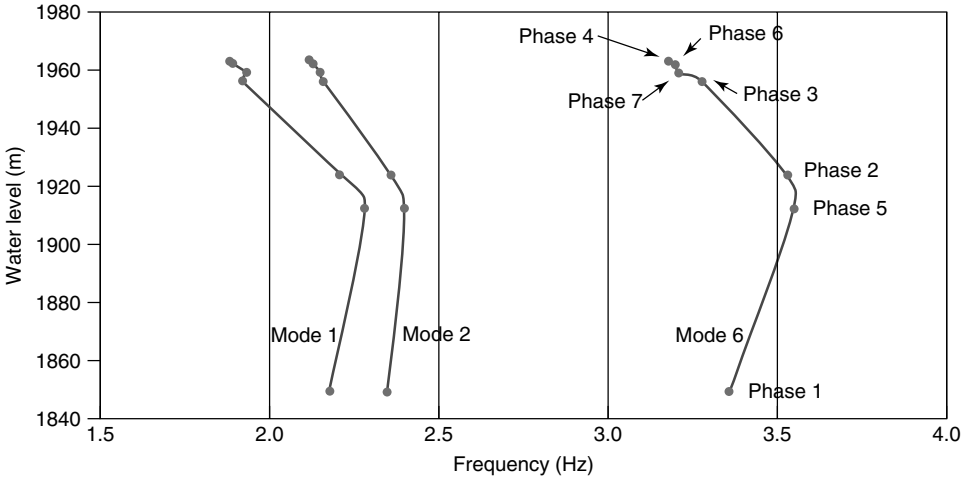


Figure 43. Frequencies of modes 1, 2, and 6 of Mauvoisin dam as identified for different reservoir water levels.

reservoir water level. Missing points indicate that it was not possible to identify the respective mode. This was mainly due to a bad signal-to-noise ratio for the respective frequency.

The results for the "best identified" modes 1, 2, and 6 are shown in Figure 43 and the phases 1–7 are identified.

Citing from the conclusions given in [25]: "The stiffening of the dam due to increasing hydrostatic pressure is more important than the added hydrodynamic masses for lower water levels. This trend is reversed for higher water levels."

REFERENCES

[1] Darbre GR. Strong-motion instrumentation of dams. *Earthquake Engineering and Structural Dynamics* 1995 **24**:1101–1111.

[2] Wyss A. Swiss National Strong Motion Network. *Strong Motion Bulletin January 2004—December 2004*, *Publication Series of the Swiss Seismological Service 117*. Swiss Federal Institute of Technology: Zurich, 1995.

[3] Duron ZH, Hall JF, Fink K, Strasser E. Measuring hydrodynamic pressures during forced vibration testing of dams. *Dam Engineering* 1991 **2**:337–355.

[4] Proulx J, Paultre P. *Etude Expérimentale et Numérique du Comportement Dynamique du Barrage-Poids Outardes 3*, Rapport de Recherche SMS-94/03. Université de Sherbrooke, 1994.

[5] Fanelli M, Giuseppetti G, Bettinali F, Galimberti C, Castoldi A, Casirati M, Pizzigalli E, Lozza S, Ruggeri G. Seismic monitoring of dams, A new active surveillance system: basic criteria, operating methods and results obtained. *Proceedings of the Ninth World Conference on Earthquake Engineering*. Tokyo-Kyoto, August 2–9, 1988; pp. VI-409–VI-414.

[6] Fanelli M, Giuseppetti G, Castoldi A, Bonaldi P. Dynamic characterisation of Talvacchia Dam: experimental activities, numerical modelling, monitoring. *Earthquake Engineering. Tenth World Conference*. Balkema Rotterdam. ISBN 90 54 10 060 5, 1992; 2689–2694.

[7] Mendes P, Oliveira Costa C, Almeida Garrett J, Oliveira S. Development of a monitoring system to Cabril Dam with operational modal analysis. *Proceedings of the 2nd International Conference on Experimental Vibration Analysis for Civil Engineering Structures, EVACES'07*. Porto, October 24–26, 2007; pp. 1015–1023.

[8] Cooley JW, Tukey JW. An algorithm for machine calculation of complex Fourier series. *Mathematics of Computation* 1965 **19**:297–301.

[9] Felber A. *Development of A Hybrid Bridge Evaluation System*, Ph.D. Thesis. University of British Columbia: Vancouver, BC, 1993.

[10] Van Overschee P, De Moor B. *Subspace Identification for Linear Systems: Theory Implementation—Applications*. Kluwer Academic Publishers: Dordrecht, 1996.

[11] Brincker R, Andersen P. Ambient response analysis—modal analysis for large structures. *Proceedings of the 6th International Congress on Sound and Vibration*. Copenhagen, 1999.

[12] Brincker R, Andersen P, Jacobsen NJ. Automated frequency domain decomposition for operational modal analysis. *Proceedings of the 25th International Modal Analysis Conference*. Orlando, FL, 2007.

[13] Tilly GP (ed). *Dynamic Behaviour of Concrete Structures*, Report of the RILEM 65 MDB Committee, Vol. 13. Developments in Civil Engineering, Elsevier Science Publishers: Amsterdam, ISBN 0-444-426 24-8, 1986.

[14] Calciati *et al. Experience Gained During in situ Artificial and Natural Dynamic Excitation of Large Concrete Dams in Italy*, ICOLD 13, Q51, R32.

[15] Severn RT, Jeary AP, Ellis BR. Forced vibration tests and theoretical studies on dams. *Proceedings of the Institute of Civil Engineers*. Vol. 69, Part 2, 1981; pp. 605–634.

[16] Deinum PJ, Dungar R, Ellis BR, Jeary AP, Severn RT, Read GAL. Vibration tests on Emosson arch dam, Switzerland. *Earthquake Engineering and Structural Dynamics* 1982 **10**:447–470.

[17] Flesch R, Eiselmayer M. Dynamic in-situ tests on Kölnbrein arch dam. *Proceedings of the 7th World Conference on Earthquake Engineering*. Istanbul, 1980.

[18] Flesch R, Eiselmayer M. Dynamic behavior of arch dams. *Proceedings 7th European Conference on Earthquake Engineering*. Athens, 1982.

[19] Cantieni R, Deger Y, Pietrzko S. Modal analysis of a concrete gravity dam: experiment, finite element analysis and link. *Proceedings 12th International Modal Analysis Conference*. Honolulu, HI, 1994; pp. 441–448.

[20] Pietrzko S, Cantieni R. Modal testing of a gravity dam—Influence of the exciter placement on the quality of the identified modal parameters. *Proceedings 12th International Modal Analysis Conference*. Honolulu, HI, 1994; pp. 1342–1348.

[21] Deger Y. Modal analysis of a concrete gravity dam—linking FE analysis and test results. Paper presented at *The MSC European User's Conference*. Vienna, 1993.

[22] Cantieni R, Wiberg U, Pietrzko S, Deger Y. Modal investigation of a dam. *Proceedings of the 16th International Modal Analysis Conference*. Santa Barbara, CA, 1998; 1151–1157.

[23] Cantieni R. Assessing a dam's structural properties using forced vibration testing. *Proceedings IABSE International Conference on Safety, Risk and Reliability—Trends in Engineering*. Malta, 2001; pp. 1001–1006.

[24] Strömberg A, Wiberg A. *En Valvdamms Dynamiska Egenskaper—Numerisk Medellering Med FEM*, TRITA-BKN. Examensarbete 48, Byggnadsstatik 1995, Kungl Tekniska Högskolan, Institutuonen för Byggkonstruktion, S-100 44 Stockholm; ISSN 1103-4297, ISRN KTH/BKN/EX—48-SE.

[25] Darbre GR, Proulx J. Continuous ambient-vibration monitoring of the arch dam of Mauvoisin. *Earthquake Engineering and Structural Dynamics* 2002 **31**:475–480.

[26] de Smet CAM, Krämer C, Darbre GR. Ambient tests at the dam of Mauvoisin. *Proceedings of the 16th International Modal Analysis Conference*. Santa Barbara, CA, 1998; pp. 1144–1150.

[27] Darbre GR, de Smet CAM, Kraemer C. Natural frequencies measured from ambient response of the arch dam of Mauvoisin. *Earthquake Engineering and Structural Dynamics* 2000 **29**:577–586.

Chapter 137
Condamine Floating Dock, Monaco

Luis M. Ortega and Manuel A. Floriano
GEOCISA, Madrid, Spain

1 INTRODUCTION—DESCRIPTION OF THE DOCK

The surface area of the Condamine Marina in Monaco was enlarged by 60 000 m^2 by building a semifloating dock. This floating dock is a double hull structure 352.72-m long, 28-m wide (44 m at the bottom slab level) and 19-m high (24.5 m with building superstructures included). This 167 000-t caisson was built in a dry dock prepared for this purpose in Algeciras Bay (Cádiz) (Figure 1) and was towed to Monaco in August 2002 (Figure 2). Detailed description of the work and its transport process can be found in [1–7].

A four-level car park with a capacity of 380 vehicles occupies 192 m of the dock, and another 136 m is used for storing cargo and small boats in two levels, 6-m high each. There are 12 cells, distributed on both sides of the structure (zones X and Z), which contain the liquid ballast (water) needed to control the floating level of the caisson (Figure 3).

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 MONITORING OBJECTIVES

The idea of monitoring the dock enables to meet the owner's requirement to check that the structure had not suffered any damage during its transportation to Monaco and its installation over there [1, 4].

The main objective of the monitoring was, therefore, to provide the necessary information for assessing the flexural forces produced during these phases and to verify that they did not exceed those established for the project.

Given this basic objective, the design of the instrumentation developed by GEOCISA was carried out in permanent collaboration with INTECSA-INARSA, the engineering consultant, commissioned with the design and supervision of floating dock and the control of the bending forces during the launching and transport phases to the Port of Monaco.

The installation of the instrumentation was carried out when the caisson was dry in the Crinavis dock. Preliminary tests for checking the monitoring system were performed during the floating phases of the

Figure 1. The caisson during the controlled flooding of the basin.

Figure 2. The caisson during the transport to Monaco.

dock and when changing the ballasts from the floating phase to the transport phase.

3 DESCRIPTION OF THE MONITORING SYSTEM

In order to meet the above-mentioned basic objective, an automated monitoring system was installed and designed to accomplish the following:

- recording the movements produced in the dock by measuring the longitudinal and transverse rotations using six servoclinometers;
- assessing the flexural forces in the dock as a box beam, from the longitudinal strains measured using 39 fiber-optic sensors;
- checking that no problems arose in the ballast cells by permanently controlling the water level in those cells using 12 pressure sensors;

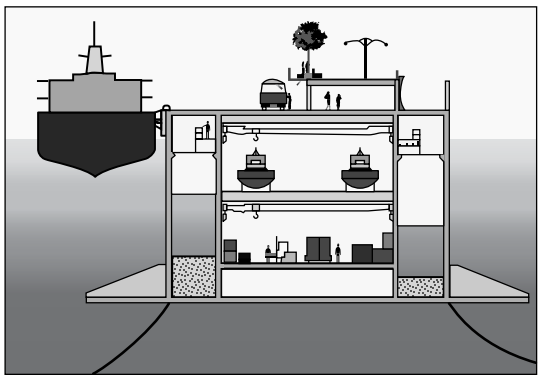


Figure 3. Typical cross section of the caisson.

- registering the overpressures, which could be produced on the upper slab or the port and starboard side walls of the caisson, in the areas next to forward and astern due to the swell. To do so, 18 sensors were used for the control of these potential hydrodynamic overpressures and 6 sensors for the control of the draught level.

All the above was completed with 16 sensors for temperature control on the external and internal concrete surface areas. There were a total of 97 sensors, 61 of which (those corresponding to rotations, strains, and temperatures) were concentrated in the three sections indicated as A, B, and C in Figure 4.

3.1 Strains

The deformations were measured using long-base (2 m) fiber-optic extensometers, with a silicon cover by *OSMOS* (Figure 5). Thirteen sensors were installed in each section in the positions indicated in Figure 6.

Each extensometer was connected using a multimode fiber-optic cable to an optoelectronic device called an *Opto-Box* (Figure 7), which generates an infrared light beam that is propagated through the extensometer. The light intensity of the optic signal from the extensometer, related to the unit deformation between its ends, returns to the Opto-Box and is converted into an electric signal, which is subsequently sampled by the data-acquisition system (DAS).

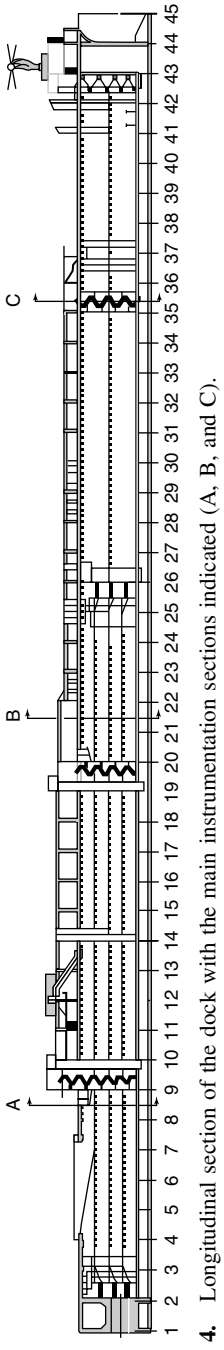


Figure 4. Longitudinal section of the dock with the main instrumentation sections indicated (A, B, and C).

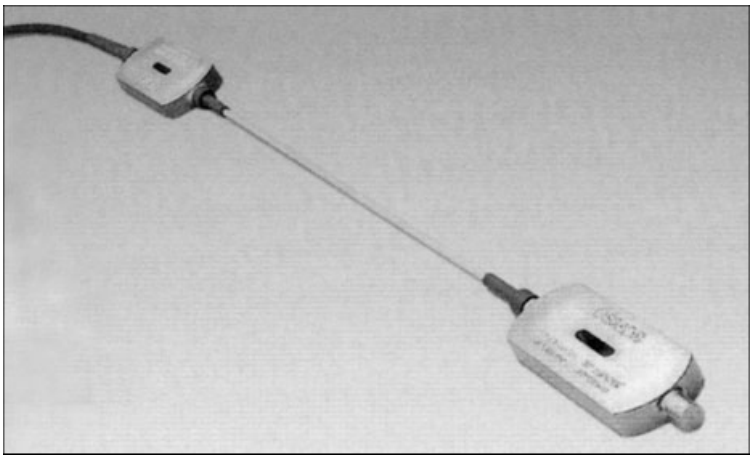


Figure 5. Long-base fiber-optic extensometers.

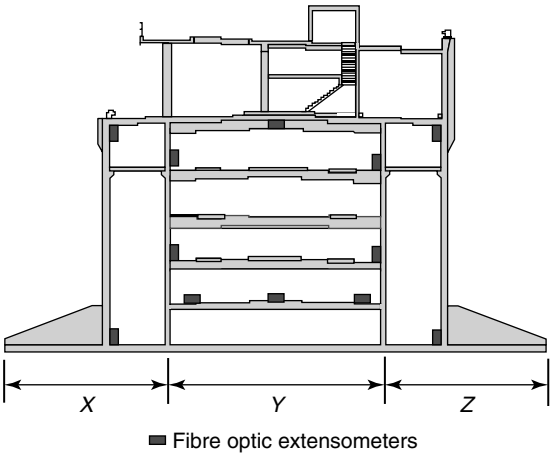


Figure 6. Cross section of the dock, showing the typical location of the fiber-optic extensometers in each instrumented section.

The sensors were installed by fixing their ends to the brass plates anchored to the concrete face. As a lot of work continued to be performed on the dock while installing the monitoring equipment, the extensometers were protected from possible mechanical impacts by using a polyvinyl chloride (PVC), semi-tube (Figure 8).

3.2 Rotations

For knowing the swell-induced movements produced during transport, both pitching (rotation in the vertical plane according to the longitudinal axis of the dock) and rocking (rotation in the vertical plane normal to the longitudinal axis of the dock) were measured in various points on the structure.

To do so, six inertial servoclinometers by *Jewell Instruments* were installed on metal supports with leveling screws. Their output signal was proportional to the pitch or rock angle with regard to the vertical line, and thus a resolution of 0.1 s was obtained.

Two sensors were installed in each of the three main measurement sections: one for measuring the pitching and another for the rocking. As the expected rotations in the event of a storm during transport could be considerable, it was decided that four of them should have a range of $\pm 14.5^{\circ}$ in order to avoid saturation, even in extreme conditions. The range of the other two was considerably lower ($\pm 3^{\circ}$) in order to try to accurately measure the rotations if, as expected, the movements were far less than such maxima. The installation of the clinometers was carried out in watertight boxes fitted to vertical faces on the deck (Figure 9).

3.3 Ballast and draught levels

One of the situations that had to be controlled during the transport was that of a possible break of the ballast cell walls. If such a break occurs, for any accidental circumstances, it could cause an alteration in the liquid ballast (outlet/inlet of water or even its flow

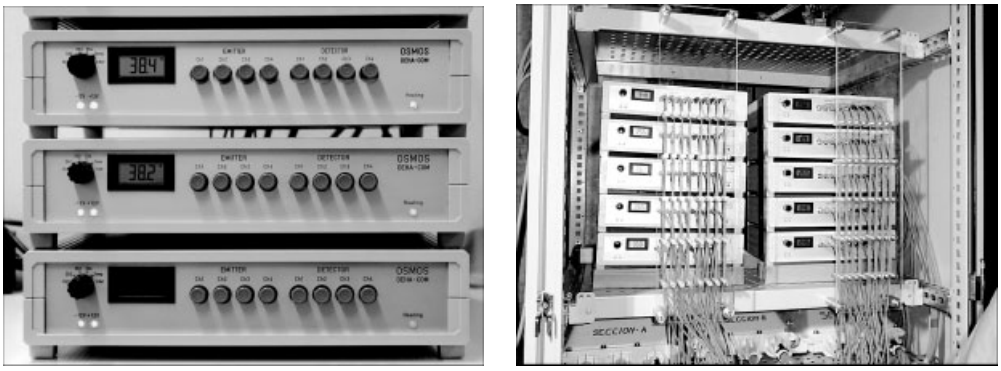
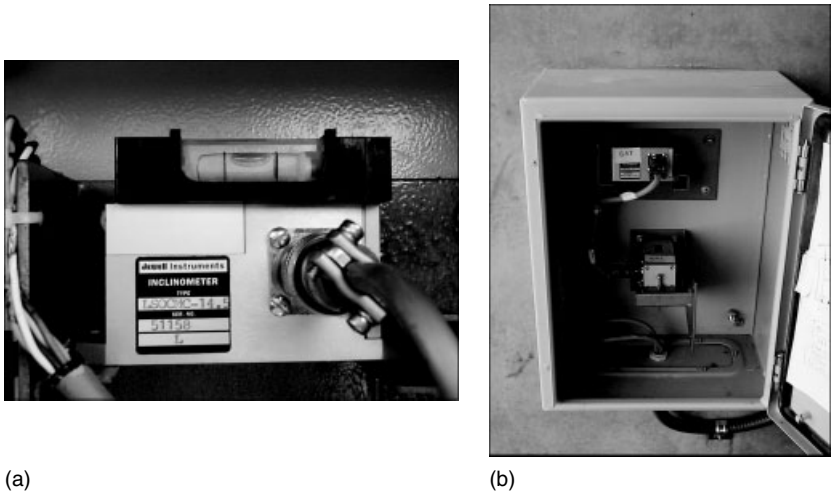


Figure 7. Opto-Box *units*. On the right are the 10 units used, arranged in the measurement centralization cabinet and with the sensors connected.

(a) (b)

Figure 8. Fiber-optic sensor installed (a) and once the PVC protection is fitted (b).



(a) (b)

Figure 9. Servoclinometer Jewell LSOC (a) and cabinet with the two clinometers installed in section C (b).

from some cells to others, as not all of them would have the same water level), which could compromise the dock's stability. As such, it was decided to control the water level in the 12 watertight cells fitted to both sides of the structure (Figure 10, zones *X* and *Z*) by installing a relative pressure transmitter in each of them, *DRUCK* PTX-1730, with a measurement range equivalent to a water column of 20 m.

Identical transducers were installed outside on the lower part of the dock, both in each of its four corners and on both sides of its central section B, to measure the dock draught at different points at any time. This was a complementary measurement to that of the rotations, which would allow to detect and characterize strong swell periods.

3.4 Hydrodynamic pressures

In order to determine the hydrodynamic pressures produced by swell during the transport, 18 hydrodynamic pressure transmitters *DRUCK* PTX-530 with a range of 0.15 Mpa were installed: 10 were distributed on the deck slab along the 50 m nearest to forward and astern; and the other 8 were placed on the external face of the caisson side walls at a level corresponding to the theoretical floating level during transport (4 in each of the port and starboard faces, in the zones near to forward and astern) (Figure 11).

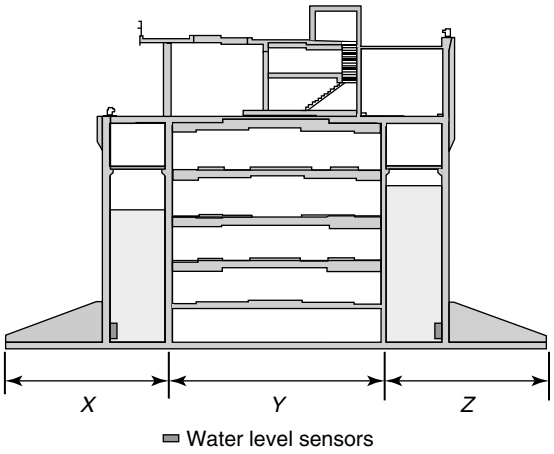


Figure 10. Location of level transmitters in the liquid ballast cells.

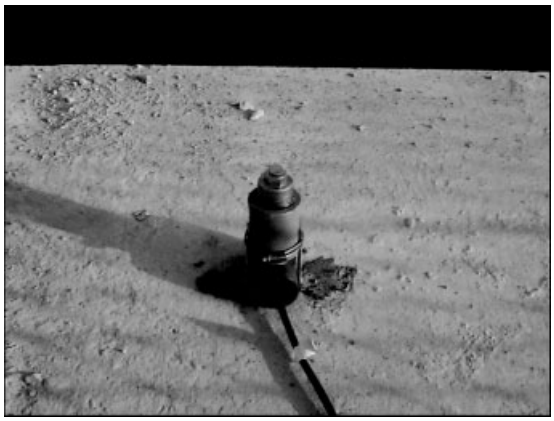


Figure 11. Hydrodynamic pressure sensors arranged on deck.

3.5 Temperatures

In order to make the interpretation of the strain measurements easier (fundamentally during the floating and ballasting operations), it was also decided to register the temperatures on the surface of the concrete using 16 PT-100-type sensors with galvanically isolated outputs.

3.6 Data-acquisition system

A data acquisition system (DAS) captured the information from the 97 sensors. It was located in an air-conditioned cabinet inside the dock (the central process unit (CPU), the Opto-Box units, the power sources, the galvanic isolation converters and connection boxes were also installed in such cabinet) (Figure 12).

The controlled magnitudes were of different nature; some of them (strain, hydrodynamic pressure, and rotations) could vary dynamically, while the others (temperature, ballast, and draught level) would only show slow variations with time. Accordingly, it was decided to record each of the measurement channels corresponding to the dynamic ones at a rate of 10 samples per second, while the other sensors were measured at a rate of 1 sample per second.

The system continuously took recordings lasting for 10 min of the signals from all sensors at the sampling speeds indicated. The real-time status of each sensor could be displayed both in numerical and graphical forms.

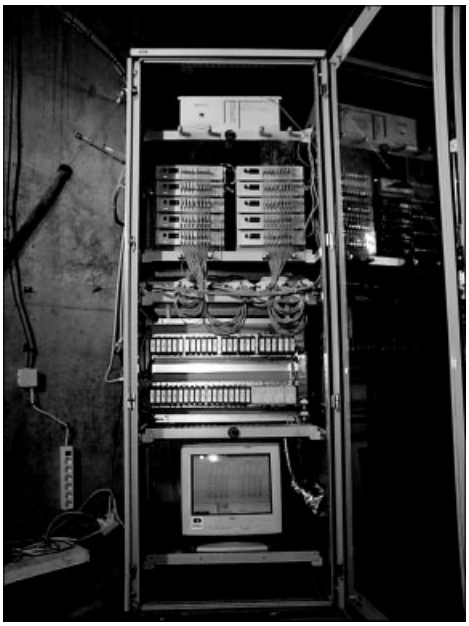


Figure 12. System centralization cabinet.

The owner demands to verify that the bending forces caused during the crossing were below the maximums in the project, called for saving all the 10-min recordings obtained throughout the complete monitoring period, regardless of nothing significant having been detected initially. By saving all records, an analysis could be made at a later date if needed. On the other hand, the number of recordings to be performed over various months (if the floating process was included) would be very large and the calculation of bending forces from the measured strain deformations required a certain amount of time. Taking into account all these circumstances, it was not logical to carry out a detailed analysis, systematically, of each and every one of the recordings, whether in real time or at a later date.

There was a secondary objective of the monitoring: the control of the transport process in real time. This would allow detecting any anomalous situation that could require urgent decisions to be taken (within the limited possibilities for action, which existed in the event of an emergency). This secondary objective was required to perform a rapid analysis of the large volume of data provided by the system and to summarize it into basic information, which could be transmitted to land for the monitoring of the transport.

To do so, it was decided to establish a preliminary processing of the data to help review and transmit them to land and to rapidly take decisions if necessary. This preliminary processing also made it easier to identify the periods during which more significant phenomena had been detected and to concentrate on the detailed analysis of the recordings obtained, and this led to the calculation of the flexural forces from the strain measurements only in these significant periods.

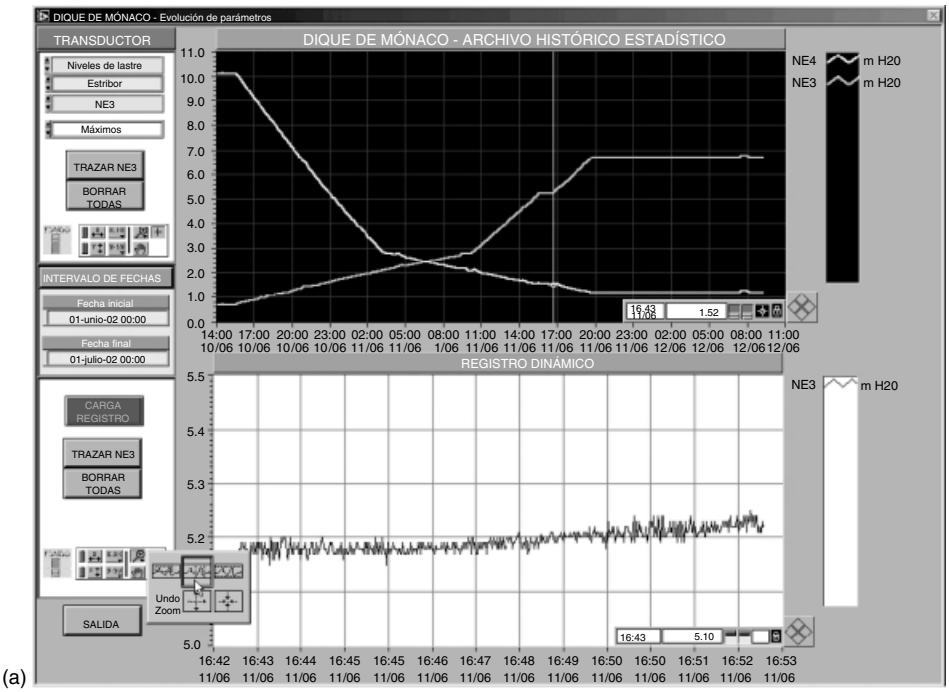
This preliminary processing determined the maximum, minimum, and mean values, and the corresponding deviations of each of the 10-min dynamic recordings for each of the channels were measured. These values characterized each 10-min dynamic recording and they were stored, along with the corresponding start time, on a historical log of the statistical values, which allowed following the evolution of all the channels easily.

A special purpose software developed by GEOCISA facilitated the consultation, on various synoptic panels, of either the time evolution of the statistical values of any sensor and of any of the dynamic recordings carried out, or the visualization of the statistical values of various sensors at the same time, corresponding to a specific period of 10 min (Figure 13).

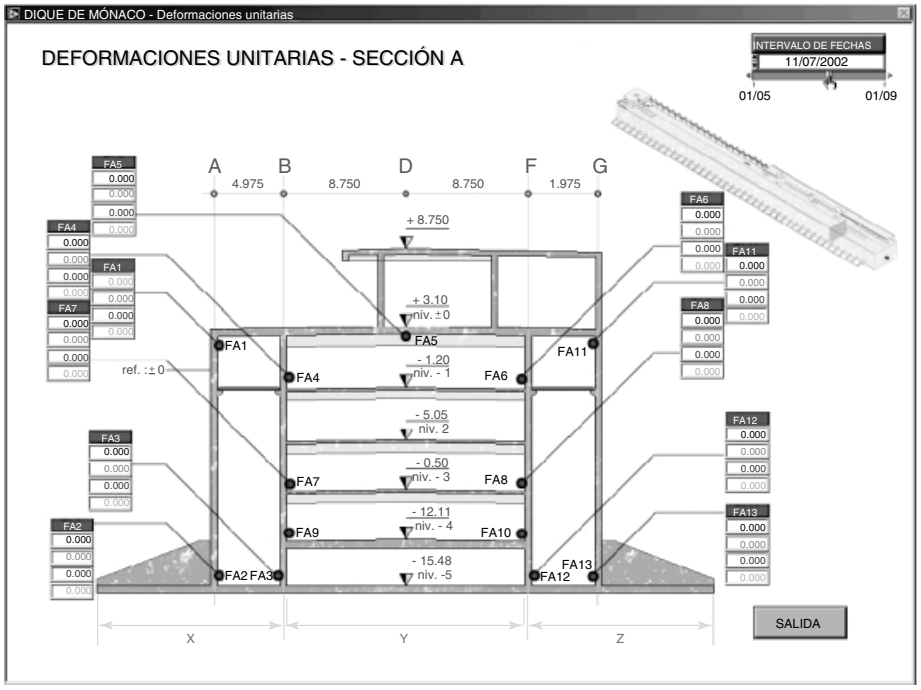
Additionally, upper and lower warning and alarm limits were established for each of the sensors to highlight their status if they exceeded the limits and to facilitate the task of supervising the monitoring by making the operator aware of the occurrence of the most significant situation.

For safety reasons, the physical presence of persons on the dock during the crossing was prohibited except in specific situations. This is the reason a radio link was established between the main computer of the monitoring system, located on the dock, and a second computer for the follow-up and analysis of data, located on one of the tugs. Through this *RadioLink* at 2.4 GHz, which provided a data transmission speed of approximately 4–5 mbps and the use of *PCAnywhere 10.0*, by *SYMANTEC*, both the transfer of files from the main computer, fitted in the caisson, and its remote operation from the remote computer was possible.

The application for displaying the historic files of the aforementioned statistical data, as well as the one for calculating bending forces from the measured



(a)



(b)

Figure 13. (a) Historic evolution of the statistical value of two channels and visualization of a 10-min dynamic register of one of them. (b) Synoptic panel of statistical values of various channels for a period of 10 min.

strains (which is commented on below), was installed on this remote computer. The use of these software applications made it necessary to keep also an up-to-date copy of the historic files of statistical parameters as well as—at least—of those dynamic registers, which could be most significant or of greatest interest on this remote computer. This task was carried out through occasional communication between both computers in order to transfer different files from their initial location in the main computer to the remote computer, installed on board the support tug "*Typhoon*", which accompanied the dock during its crossing to Monaco (Figure 14) [4, 5].

The information was permanently examined and reviewed in the remote computer by two operators taking it in turns. They transmitted it once a day, by e-mail, according to a protocol previously defined, to the follow-up team on land, in order to cover the event of any critical situation that would require extraordinary decisions to be taken.

3.7 Estimating bending forces

Two different phases were considered for estimating the bending forces from the strain measurements. The first corresponded to the floating procedure and the subsequent operations (ballasting with water, tests for the water tightness of the ballast cells, caisson exit maneuvering from the basin to the bay and changing the ballast to the transport situation) until the dock was ready for crossing to Monaco. The second corresponded to the transport itself until completion of the connection operation of the ball-and-socket joint.

3.7.1 Floating phase

During this first phase, the possible bending forces caused in the dock corresponded to very slow processes and the strains had to be calculated by the difference between the measurement at any time and the initial "zero" reading. As the period involved was quite long (various weeks), thermal variations could occur, which would affect both the dock itself and the sensors, as well as the measurement system. As such, it was recommendable to control the temperatures, which could help in interpreting the measurements.

Furthermore, as these quasi-static loads varied very slowly, the measurements over a measurement period of 10 min would have to be noticeably constant. Accordingly, the bending forces during this phase were calculated on the basis of the mean values of the effective strains for each 10-min period. These effective strains were obtained as the difference between the mean readings of each sensor during this period and the period that was considered as origin or zero situation (which was normally the initial point of measurement before starting to flood the dock to float it).

For each measurement section, the program calculated the deformation plane. From this and the geometric and mechanical characteristics of the transversal caisson section, it also estimated the bending forces. The calculation program allowed obtaining the bending moments in two different cases: assuming that the axial strain is nil (forcing the deformation plane to give strain zero in the center of gravity of the section) or not adopting such an assumption (and calculating also the axial

(a) (b)

Figure 14. Remote computer (a) installed on the support tug (b).

strain). Depending on the hypothesis, two or three values were obtained (the bending moments M_y, M_z according to two perpendicular planes and axial force N, in the event of not assuming it as zero) for each 10-min period studied, in each of the sections A, B, and C.

3.7.2 Crossing phase

During the second phase, interest centered exclusively on the bending forces that could be caused by swell. As this was a dynamic phenomenon with a nil average, it was decided to set the average of the readings obtained during each 10-min period to zero and thus only the dynamic variations produced by the swell remained. Any slow variation process was thus eliminated (sensor drift, if it were to exist, influence of the temperature changes on the sensors or on the structure, creep), which would have given a spurious result for the purposes pursued (estimation of the bending moments caused exclusively by swell).

As during the floating phase, the deformation planes were calculated from the effective strains and, based on each of these, the bending moments M_y and M_z were obtained in the three instrumented sections A, B, and C. Similar to a previous hypothesis, it was assumed that the swell did not cause significant axial forces. Therefore, it was always ensured that the deformation plane corresponded to zero strain in the section's center of gravity. As the strain did vary dynamically in this phase, the calculation of the moments was carried out for all the instants measured (at a rate of 10 times per second), providing 6000 values of each bending moment (M_y and M_z) for each 10-min period studied in each of the sections A, B, and C.

Both during the floating phase and the crossing, the program allowed the user to eliminate the values of one (or various) sensor, in order not to take them into account, if it was suspected that, for any reason, this sensor or sensors were not operating correctly.

4 RESULTS OBTAINED

Discussion of results obtained by the monitoring instrumentation is made separately for the two phases mentioned above:

- **Phase 1**

Work performed in Algeciras Bay (flooding of the dock, floating of the caisson, exit of the dock from the basin, and change of ballast in order to have the dock ready for transport) from June 10 to August 13, 2002.

- **Phase 2**

Check during the crossing from Algeciras to Monaco (from August 14 to 26, 2002) and the subsequent operations for connection of the ball-and-socket joint (until September 6, 2003).

4.1 Phase 1—Algeciras

Two curious sets of circumstances are worth highlighting with regard to the analysis of the results of this phase, both related with the actual analysis of the data.

Figure 15 shows the evolution graphs for two strain sensors in section B, located on the lower face of the caisson's upper slab (Figure 2) throughout the whole of August 13. On the one hand, it shows a slow variation over the day (13–15 $\mu\varepsilon$ peak to peak), which responds to the daily cyclical temperature change. A ripple with a much smaller amplitude (2–3 $\mu\varepsilon$) is superimposed onto this daily variation wave. After a careful analysis, it was concluded that this ripple corresponded to the warm-up/cooling cycles of the Opto-Boxes, caused by the operation inputs and shutdowns of the actual cabinet air-conditioning system. This emphasized the importance of maintaining the operating temperature of these Opto-Boxes under a specific threshold, as recommended by the manufacturer. At the same time, the result highlighted the magnificent stability and sensitivity of these fiber-optic sensors.

Figure 16 shows the evolution of two strain sensors (FB-4/FB-6) in section A throughout August 11. A sudden jump in the measurement ($\approx$100 $\mu\varepsilon$) can be observed, which did not correspond to similar situations in the other sensors of the section nor with possible real strain caused in the structure. As both sensors were placed in the bottom area of one of the liquid ballast cells, the graph depicting the evolution of the water level in this cell (NB-2) was examined. It was noted that the jump coincided with the moment in which the water level in the

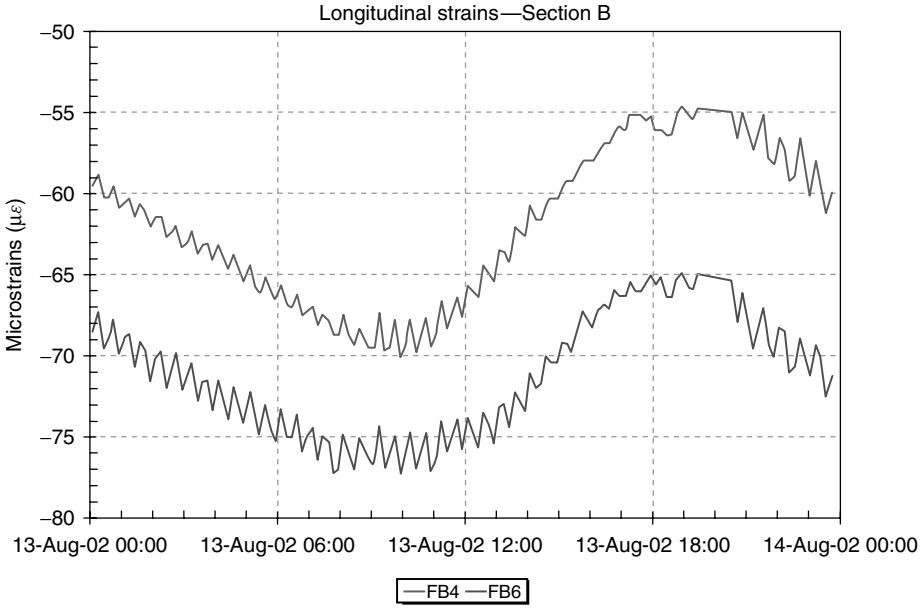


Figure 15. Evolution of unit deformations over one day.

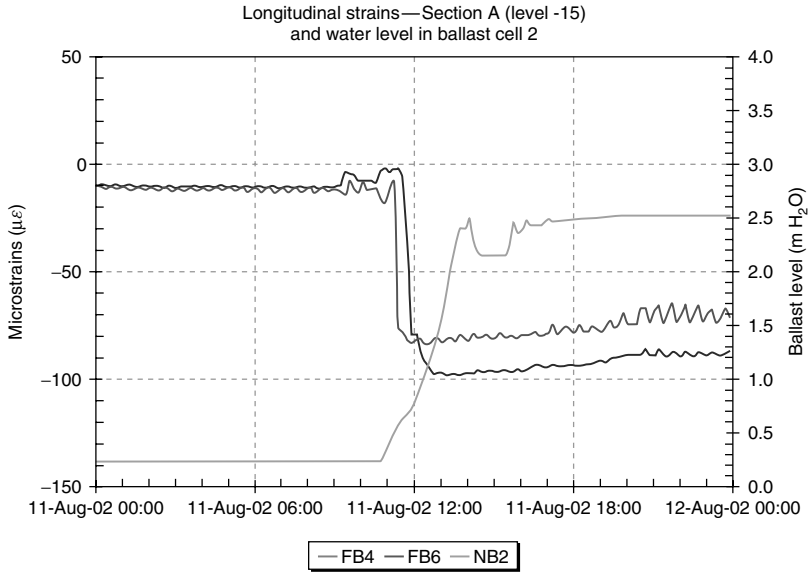


Figure 16. Effect of the hydrostatic thrust on the deformation sensor.

cell passed the level at which these sensors were positioned.

It was checked that the effect was produced when the water level was rising (and, accordingly, the sensor changes from being dry to being submerged) and that the opposite process was produced when the water level was going down.

These apparent strains were mainly caused by the effect of the hydrostatic thrust acting on the actual fiber optic anchored at its ends and, to a far lesser

extent, by the effect of the sudden change of the sensor temperature (and of that of the concrete) when coming into contact with the cold water. This effect was produced in all of the sensors located in the lower section of ballast cells when, as a result of the filling and emptying tests on the cells, there was a change from being submerged to dry or vice versa. Fortunately, once detected, this was easy to correct.

Concerning the bending moment calculation, the main objective of its evaluation during this phase was the possible contrast or calibration of the entire process. In this sense, the fact that various tasks were being carried out on the structure during almost the whole of this phase (water tightness tests, ballast transfer from one cell to another, etc., and even the continuation of works on the dock itself) made it practically impossible to perform a reference theoretical assessment. This difficulty in assessing the actions, actually applied to the caisson, prevented from obtaining a reference to compare the experimental measurements. Nevertheless, an attempt was made to experimentally contrast the flexural forces produced by the change of ballast level performed just before transporting to Monaco and after the exit of the dock from the basin. In that process, there was a relatively good control of the actions (the ballast levels in the cells and, therefore, draught of the dock) on the caisson, except for thermal effects, and a reasonable estimation of the theoretical bending moments caused by them was possible.

The effective strains were obtained as a difference between measurements taken just before the exit of the dock from the basin (August 6 at 9 p.m.—floating ballast situation) and on August 13 at 9 p.m. (transport ballast situation). The bending moments estimated experimentally, both under the assumption of zero and nonzero axial forces, were quite similar in both hypotheses (differences less than 10%) and clearly less than those estimated theoretically (differences of approximately 30%).

An analysis of the sensitivity of the estimated bending moments with regard to small variations in the strains measured by the sensors furthest away from the center of gravity was performed. It showed that relatively small variations of these measurements ($15-20\,\mu\varepsilon$), due to temperature changes or any other effect (impossible to prevent in a process lasting for various days), could lead to variations in the estimated bending moments of the same order of magnitude as the differences detected with regard to the theoretical values.

4.2 Phase 2—crossing Algeciras—Monaco

During the 12-day crossing, the weather was pleasant and no episodes of significant swell occurred. The liquid ballast levels, a fundamental parameter for control as already commented, maintained good stability throughout the crossing.

The evolution graphs corresponding to the maximum or minimum rotations every 10 min, both transversally and longitudinally, enable a quick detection of episodes of greatest caisson movements due to swell. These most significant movements started at noon on August 18 and were increasing until reaching their greatest value at noon on August 19. The greatest maximum movements were detected at approximately 11 a.m., with maximum transversal and longitudinal tilts of 0.66° and 0.14°, respectively (Figure 17). The hydrodynamic pressures recorded on the caisson side walls are in good agreement with the measurements of the rotations, with the greatest pressure also being detected on August 19. In any case, their values were very small, not exceeding 0.019 Mpa.

The maximum bending moments (in kilonewton meters) produced by the swell, corresponding to those moments of greatest movements during the transport are shown in Table 1. Even in those moments of greatest movements, such bending moments were below 5% of the design theoretically estimated maximums in the event of a storm. The evolution graphs of the experimentally estimated bending moments show very small values, which would have to be considered as practically nil.

The follow-up of the instrumentation was continued until finalizing the ball-and-socket joint connection process [1, 5]. The evolution graphs of the rotations show, perfectly well, the final setting of the first phase of this connection (dock approach until inserting the outside cone of the ball-and-socket joint into its seating in the abutment [1]) (Figure 18). This operation was performed on September 3. Between

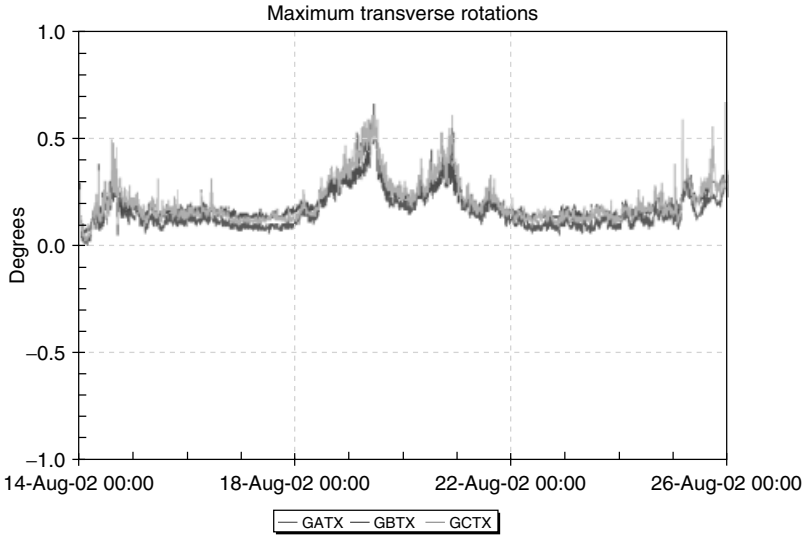


Figure 17. Maximum transversal rotations every 10 min throughout the crossing.

Table 1. Maximum moments produced due to swell during the transport

FECHA	MyA	MzA	MyB	MzB	MyC	MzC
19/08/02 08 : 33	26.431	64.084	53.481	121.492	24.889	85.413
20/08/02 21 : 10	14.759	59.753	58.487	151.119	23.981	87.907

Figure 18. Views of the approach process for the final installation of the dock in Monaco.

10:30 and 11:00 a.m., a rotation of approximately 0.1° was observed corresponding to the lifting of 40 cm, using jacks, required for aligning the ball-and-socket joint as referred to in [1]. Later, at approximately 1:00 p.m., an isolated peak of 0.2° was recorded. With a detailed analysis of the dynamic records of the 10-min period between 1:07 and 1:17 p.m., it can be seen that it corresponds to the precise moment when the ball-and-socket joint was connected (Figure 19).

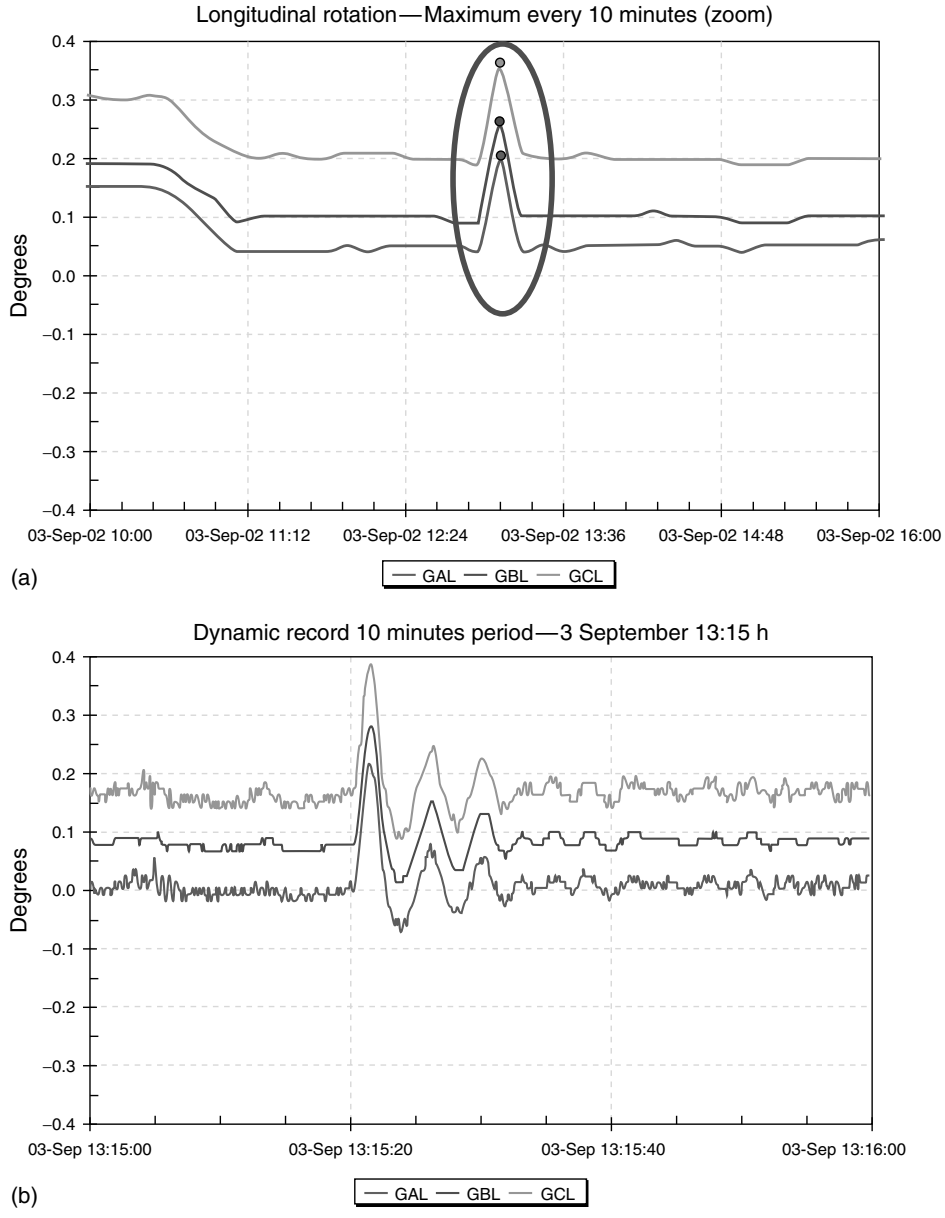


Figure 19. Longitudinal rotations during the connection operations of the knuckle joint (September 3, 2002). (a) Registered maxima every 10 min. (b) Extract of the dynamic register of the rotations corresponding to the period of 10 min between 1:07 p.m. and 1:17.

5 CONCLUSIONS

The monitoring of the floating dock for the enlargement of Port Condamine, carried out during the sea crossing from Algeciras to its definitive location in Monaco, allowed to experimentally check that the flexural forces induced on it during this transport operation was, at all times, under the maximum limits outlined in the project.

The control of the water levels in the liquid ballast cells confirmed its perfect water tightness and the good behavior of their walls and slabs.

The movements (rotations) showed values that were far below the pre-established warning levels. This is in accordance with the favorable weather conditions that prevailed throughout the crossing. In line with this almost calm-sea situation, no episode of swell over the deck was recorded, and the pressure cells installed in port and starboard in the areas close to forward and astern recorded reduced pressure values, far lower than those that could have been withstood by the outer walls.

The bending moments caused by the swell in the caisson, estimated from the strain data experimentally measured, gave values well below those estimated in the various design calculations.

REFERENCES

[1] Hue F. Puerto de La Condamine (Mónaco). El mayor dique rompeolas flotante del mundo. The Largest Floating Breakwater in the World, Ingeniería Civil No. 127, ISSN 0213-8468, Port Condamine, Monaco, July-August-September 2002; pp. 11–24.

[2] Hue F. Construcción e instalación del mayor dique-muelle semiflotante del mundo. The Building and Commissioning of the Largest Semifloating Dock in the World, Marina Civil nº 67, ISSN 0214-7238, 2002; pp. 19–27.

[3] Peset L, Barceló J and Troya L. Introducción y descripción del proyecto. Introduction and Description of the Project, Hormigón y Acero nº 223–226, Special Issue Dedicated to Monaco Floating Dock, ISSN 0439-5689, 2002; pp. 7–17.

[4] Peset L, Barceló J, López D, Hué F, Vázquez A and Ortega L. Elementos singulares en el dique de Mónaco. Special Features of the Monaco Dock, Hormigón y Acero nº 223–226, Special Issue Dedicated to Monaco Floating Dock, ISSN 0439-5689, 2002; pp. 67–117.

[5] Hué F, López D, Peset L and Troya L. Ejecución de las fases marítimas. Execution of the Maritime Stages, Hormigón y Acero nº 223–226, Special Issue Dedicated to Monaco Floating Dock, ISSN 0439-5689, 2002; pp. 139–165.

[6] Hue F and Peset L. El dique-muelle flotante de Mónaco. Cemento Hormigón nº 851, ISSN 0008-8919, 2003; pp. 58–74.

[7] Barceló J, Hue F and Peset L. Dique flotante de abrigo realizado en Algeciras para la ampliación del Puerto de Mónaco (España). Floating Breakwater Built in Algeciras for the Port of Monaco, Revista de Obras Públicas Vol. 150 nº 3433, ISSN 0034-8619, May 2003; pp. 81–110.

Chapter 138

Soil–Structure Interaction and Seismic Effects

Günther Achs
VCE – Vienna Consulting Engineers, Vienna, Austria

1 INTRODUCTION

In the last few centuries, the seismic risk of many countries was redefined due to several new findings of seismologists and geologists. This has brought considerable changes in civil engineering, concerning the development of new structures as well as the assessment of the seismic bearing capacity of existing structures. As there are various structures that have not been constructed against horizontal loading in previous times, the big challenge of today's engineers is the reassessment of those existing structures.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

Among others (material aging, damages, etc.) there are two main aspects, soil–structure interaction and the realistic system identification, which should be considered within this context. In Section 2 of this article, the effect of soil–structure interaction is described and a brief overview of the existing methods is given. The problem of a realistic identification of the dynamic behavior of existing structures and the benefit of experimental methods are described in Section 3. The main attention is attached to the measurement techniques and signal processing. Finally, in the last section of this article, an overview of seismic effects is given in general.

2 SOIL–STRUCTURE INTERACTION

2.1 Basic information

In seismic hazard analysis of building structures, two main foundation types have to be distinguished. In case of a structure founded on rock, the ground motion on the base can be considered as identical to the ground motion on the same point before the erection of the building. However, if the subsurface consists of soft soil sediments, then there will be

change in the dynamic system. In this case, the structure interacts with the surrounding subsoil, leading to a change in the seismic motion at the base.

In general, the effect of soil–structure interaction due to earthquake loading cannot be neglected and thus appropriate analysis techniques have to be used.

2.2 Theoretical background

For the seismic analysis of structures, it is necessary to apply the earthquake design loads to the foundation. The design loads arise from the inertia forces developed in the superstructure and from the subsoil deformations, which are generated by the passage of seismic waves. As mentioned in the technical literature [1], these two effects are referred to as *inertial* and *kinematic loading*. The influence of both effects depends on the characteristics of the investigated structure, the geometric parameters of the foundations, and the nature of incoming waves. Soil–structure interaction therefore covers both effects.

Most of the design engineers refer to inertial loading as soil–structure interaction, ignoring the kinematic component [1], which can be interpreted from the fact that

- in some situations kinematic interaction has negligible influences;
- kinematic interaction is often not mentioned in seismic building codes; and
- the evaluation of kinematic interaction effects is more difficult than inertial interaction effects.

In Figure 1, the key features of the soil–structure interaction problem are represented [2]. The general situation of an embedded foundation supported on

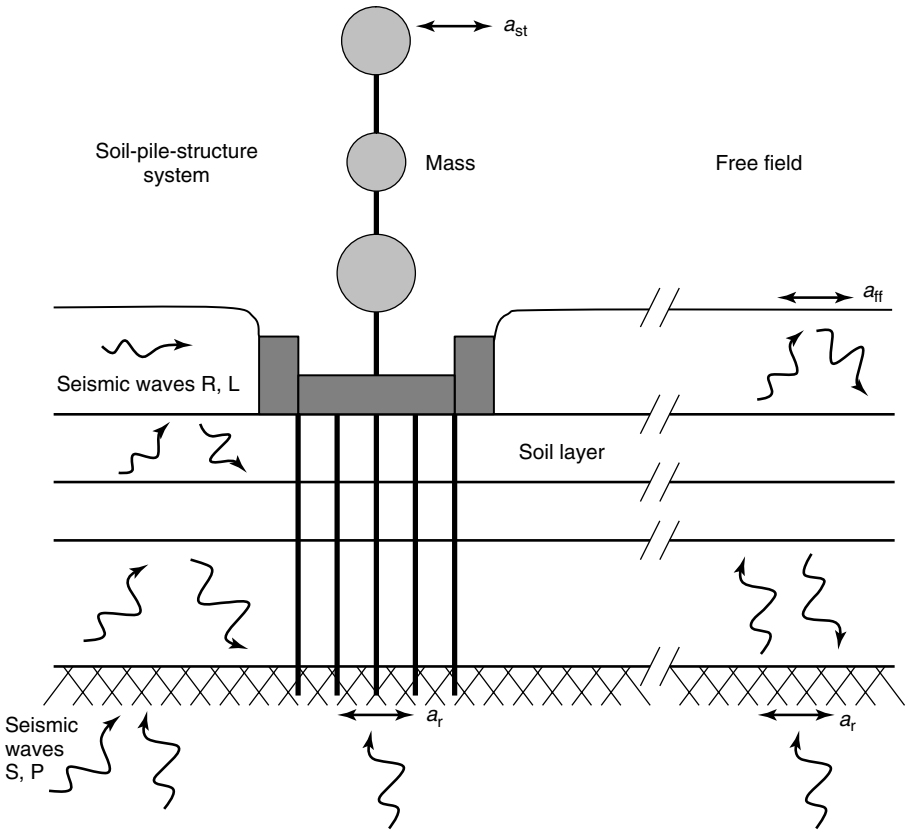


Figure 1. Illustration of soil–structure interaction on the structure response; structural acceleration a_{st}, free-field acceleration a_{ff}, and rock acceleration a_r.

piles is presented. The conclusions given in Figure 1 are also valid for any other foundation type. The subsoil layers around the structure are subjected to seismic excitation, consisting of numerous incident waves such as shear waves (S waves), dilatational waves (P waves), and surface waves (R or L waves).

In national building codes, the nature of the incoming waves is more or less given based on the global seismological conditions of the area. Nevertheless, the geometry, stiffness, and damping of the local soil deposits modify the characteristics of the free-field motion. The modified free-field motion can then be considered as the seismic input on the foundation of the structure. The determination of the free-field motion itself can be very difficult [3]. As mentioned in the technical literature [1], the design motion is usually specified at only one location (ground surface) and the complete wave field cannot be calculated back from this incomplete information. Therefore, assumptions have to be made regarding the exact composition of the free-field motion and it can be stated that no satisfactory solution is available so far.

Owing to the seismic motion around the structure and its foundation, the subsoil forces the piles and the embedded foundation to move. Even without the existence of any superstructure, the motion of the foundation will be different from the free-field motion. This can be explained by the differences in rigidity between the soil on one hand and the piles and foundations on the other hand. The incident waves are reflected and scattered by the foundation and piles. Thus, displacements and forces are imposed on the structure. This phenomenon is called *kinematic interaction*. Owing to the displacements induced at the foundation level, oscillations in the superstructure are generated, which develop inertia forces and overturning moments at its base. Hence, the foundation and the surrounding soil are exposed to additional dynamic forces and displacement; this phenomenon is generally called *inertial interaction*. It is therefore necessary to analyze the foundation of a structure for combined inertial and kinematic soil–structure interaction.

To evaluate the effects of soil–structure interaction of linear systems, it is most appropriate to use linear elastic constitutive or equivalent viscoelastic linear models for the soil. In many cases, soil nonlinearities can often be considered using an approximation by choosing appropriate values for the soil parameters.

2.3 Methods

For practical applications of soil–structure analysis, it is beneficial to use either so-called direct methods or substructure methods.

2.3.1 Direct methods

In general, the easiest way to analyze the soil–structure interaction for seismic excitation is to model a meaningful geometric section of the subsoil around the embedded building structure and to apply the free-field motion to the boundary layer of the subsoil [4]. With this approach, nonlinear behavior of the soil can also be considered. As the number of dynamic degrees of freedom can be very high, according to the geometric extension of the model and the complexity of the parameters, large computational resources are necessary.

2.3.2 Substructure methods

For big and complex models, it is more beneficial to use the substructure method, which implicitly assumes that the method of superposition has to be valid in a soil–structure interaction analysis.

The substructure method is roughly explained in Figure 2. Initially, the free-field vibration needs to be evaluated in the foundation nodes (white nodes). Afterward the interaction part is investigated in two steps, whereas the unbounded soil is analyzed as a dynamic subsystem.

To evaluate the seismic vulnerability of an existing structure based on structural health monitoring, it is indispensable to have information on the local soil parameters. The local soil parameters can be identified using laboratory tests or *in situ* measurements. As laboratory testing always points out only the parameters of a single borehole, it is generally more informative to use *in situ* experiments.

3 EXPERIMENTALLY BASED SEISMIC ASSESSMENT OF STRUCTURES

3.1 Basic information

Seismic analyses of building structures are based on the dynamic parameters of the considered object

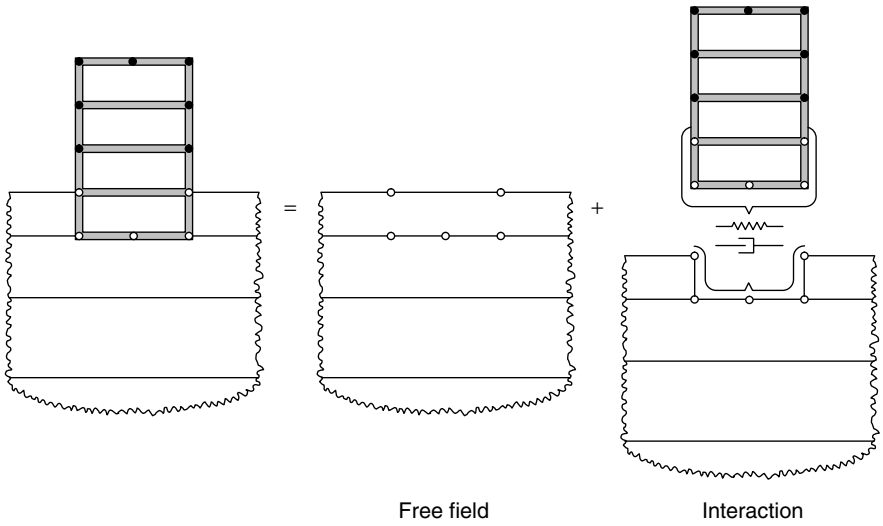


Figure 2. Illustration of the substructure method.

such as natural frequencies and the corresponding mode shapes. For existing buildings, these parameters are determined beneficially from data recorded by *in situ* measurements, in particular, when the structural system and utilized materials are unknown.

3.2 *In situ* measurements

To identify the dynamic parameters of a building structure, it is proposed to utilize accelerometers, which are distributed on the story levels on top of each other. As an example, Figure 3 shows the position of accelerometers in an investigated residential building. The sensors are placed next to the staircase because the flats are occupied, and therefore cannot be entered without additional logistic effort. It would be best to perform the measurements at night when pedestrian traffic is low and excitation from engines such as washing machines is very unlikely.

In tall buildings, measurements cannot be conducted simultaneously because the number of available sensors is customarily limited. For such objects a reference sensor is located in a story, where it remains during the entire measurement period. After each measurement, a second accelerometer is moved to a further measuring point at a different story level until the entire building is surveyed. The response of the sensor to be moved is related to the response measured with the reference sensor.

Depending on the type of building and measurement equipment, two different analysis techniques may be utilized to evaluate the dynamic parameters:

- the ambient vibrations response (ambient excitation); and
- the free vibration response after an impulsive-like load applied with a hammer (transient excitation).

It is noted that the location of excitation must remain the same when measurements are performed with a reference sensor.

3.3 Signal processing

The dynamic time–history response is recorded in three directions. However, if reasonable, the evaluation of the data is confined to one building axis, where the structure is most vulnerable, i.e., effects such as torsion or coupled-bending torsional vibrations are negligible.

The natural frequencies of the building structure are determined by transforming the time–history signals of each sensor into the frequency domain by fast Fourier transformation (FFT). Depending on the recorded data, the mode shapes are evaluated in the time domain (transient excitation) or in the frequency domain (ambient excitation).

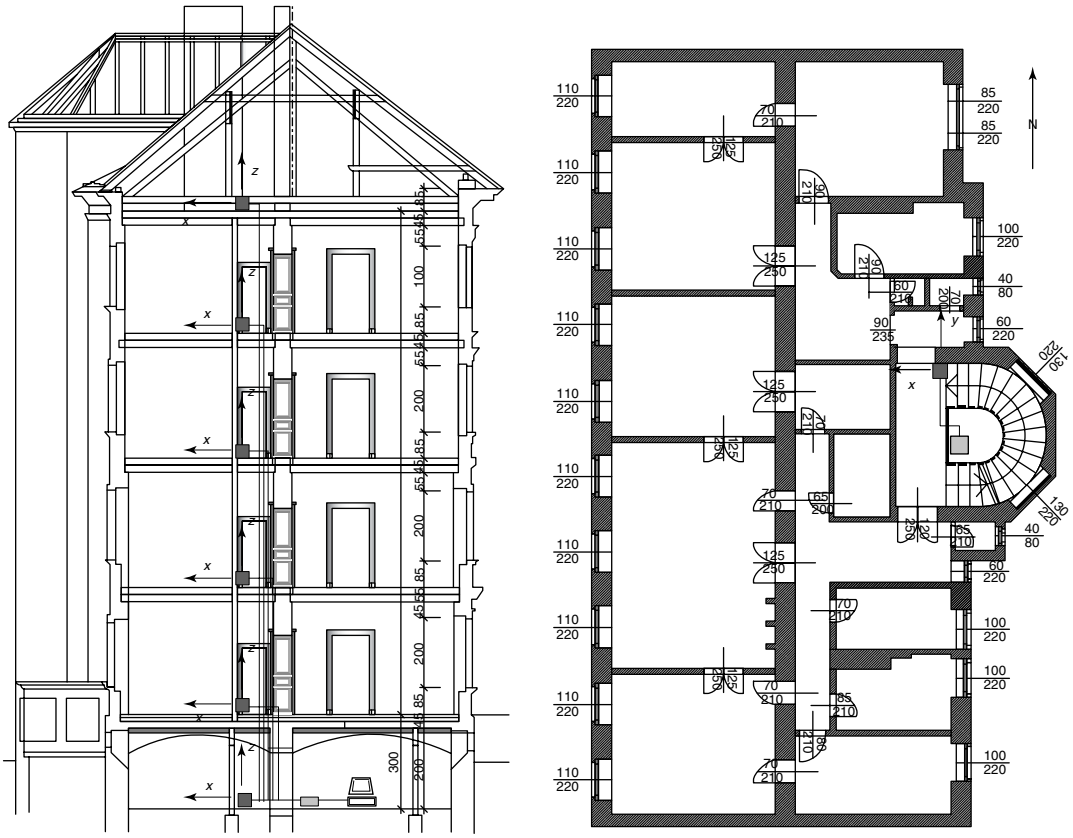


Figure 3. Sensor instrumentation of a residential building.

3.4 Transient excitation—time domain

If the building can be excited transiently by means of a hammer, the induced free vibration response is utilized to identify the dynamic parameters. The procedure of data processing in the time domain is shown in Figure 4.

The nth modal component of the undamped free displacement and acceleration response $\vec{w}_n$ and $\ddot{\vec{w}}_n$, respectively, of a multidegree-of-freedom (MDOF) system vary by the square of the nth natural circular frequency ${\omega_n}^2$ [5]:

$$\vec{w}_n = \vec{\phi}_n \cos \omega_n t, \quad \ddot{\vec{w}}_n = -{\omega_n}^2 \, \vec{\phi}_n \cos \omega_n t \quad (1)$$

In equation (1), $\vec{\phi}_n$ denotes the nth mode shape. Since the maximum amplitude of a mode shape is arbitrary, the mode shape can be extracted directly from the acceleration response.

The fractions of the response, which vibrate in the natural frequencies of interest, are extracted by filtering the raw signal. Therefore, the raw acceleration response is edited by a band-pass filter of the fourth order. The frequency band around the considered natural frequency f_n is very narrow, i.e., the upper and lower boundaries of the frequency $f_\mathrm{u,l}$ are selected according to

$$f_\mathrm{u,l} = f_\mathrm{n} \pm \frac{f_\mathrm{n}}{100} \quad (2)$$

This procedure is performed for each identified natural frequency f_n. Subsequently, the extracted time–history responses of each story is plotted on top of each other, and at certain time instants the amplitudes are read and combined to the vector $\vec{\phi}_{n,i}$. Note that the floor response must be related to the response recorded with the reference sensor

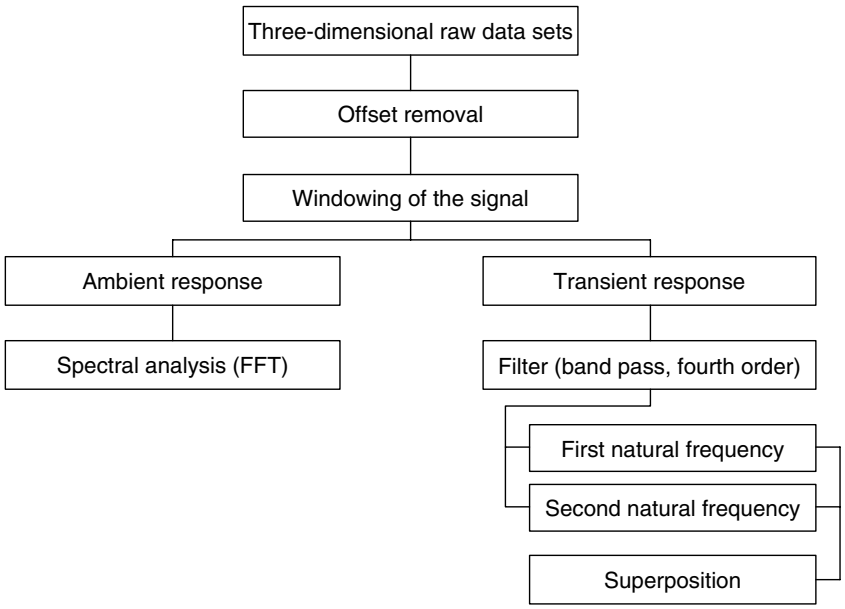


Figure 4. Evaluation of natural frequencies and mode shapes based on transient excitation. [Reproduced with permission from Ref. 5. © Prentice Hall, 2006.]

when the measurements of the floor accelerations are conducted at different time instants. Finally, the estimate of the nth mode shape $\vec{\phi}_n$ is determined by averaging

$$\vec{\phi}_n = \frac{1}{m} \sum_{i=1}^{m} \vec{\phi}_{n,i} \tag{3}$$

In Figure 5 the evaluation of the first bending mode $\vec{\phi}_1$ of an investigated building structure according to the described procedure is shown.

3.5 Ambient excitation—frequency domain

If the frequency domain is used to analyze the mode shapes, it is necessary to evaluate the phase of the acceleration response. Therefore, the phasing of the response signal in each floor of the building needs to be investigated. To increase the reliability of the results, this should be done simultaneously for each obtained record. The mode shapes are finally averaged for the different recorded signals [6]. The evaluation of the fundamental natural frequency of a structure is shown in Figure 6.

3.6 System identification—results of some investigated structures

Several existing objects were tested in order to verify the applicability and reliability of the proposed method. The *in situ* measurements were preferably performed using transient excitation. In Table 1, the investigated structures, their dimension, and employed building material are listed. Furthermore, the first and second natural frequencies f_1 and f_2, respectively, in two directions (E–W, east–west and N–S, north–south) of the objects are given in the same table.

In the following object 1 of Table 1, (Riglergasse 10, Vienna) is examined. Both ambiently and transiently excited building responses are utilized for the evaluation of this building. Measurements were performed during a long period with a sampling rate of 500 Hz. This large sampling rate was selected to window the signal with small time steps. A part of a record and the windowing of the signal are shown in Figure 7.

The structural mass was estimated consulting design documents, and lumped to the story levels. The first and second mode shapes, the story masses, and the story stiffnesses are specified in Figure 8. The

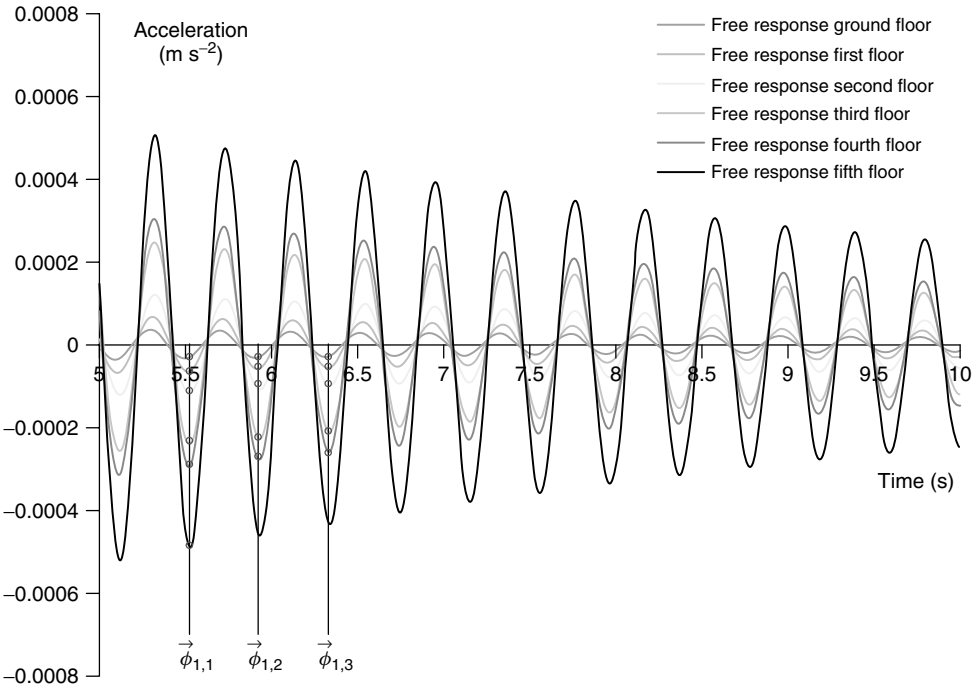


Figure 5. Evaluation of the fundamental mode of a building structure (time domain).

stiffness of the mechanical model was determined by application of the mentioned indirect finite element update procedure [7].

These results constitute the structural input values for the capacity design method. Therefore, the structural model shown in Figure 8 is subjected to a static pushover analysis [8].

To consider the global nonlinear behavior of the building structures, it is advantageous to perform large-scale *in situ* pushover tests on real structures. The outcomes of these tests can be utilized to calibrate the numerical pushover analysis. The nonlinear material parameters can be verified using several laboratory tests on specimens.

3.7 Finite element model update

An estimate of the distributed stiffness of the investigated building is calculated utilizing the equations of motion for a simplified mechanical model. Therefore, the structure is considered as an MDOF system with lumped masses at each story level. The model of the MDOF system is thus created as an approximation with several assumptions and can lead to differences in the real structure. To improve the accuracy of the system, the dynamic parameters are calibrated with the outcomes from *in situ* measurements. Therefore, a numerical updating procedure is utilized [7].

In general, finite element model updating can be performed using direct or iterative methods [9]. The former has advantages because no iteration is required and measured data are exactly reproduced. However, if the measured data are inaccurate or correspond to a highly nonlinear system, a model with no physical meaning may be obtained. In contrast, iterative methods are based on a nonlinear penalty function, which is minimized through subsequent linear steps, and more computational time is required.

4 SEISMIC EFFECTS

Earthquakes can cause tremendous damages on nature and structures. Owing to numerous different parameters (fault mechanism, epicentral depth, source and site parameters, soil conditions, etc.), the effects on the surface can have large discrepancies for every

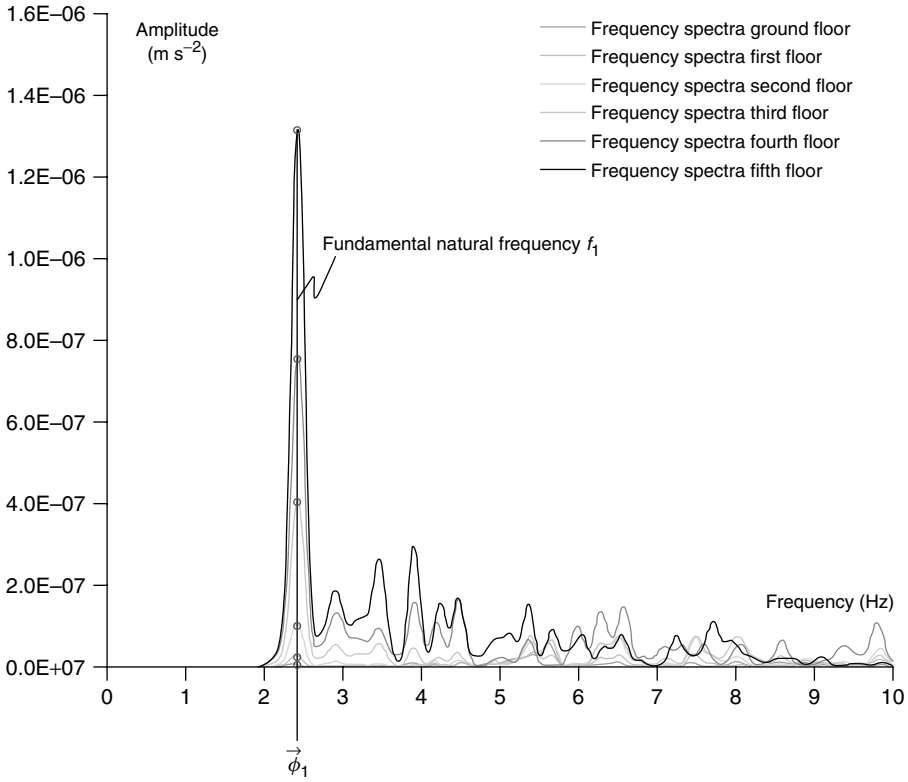


Figure 6. Evaluation of the fundamental mode of a building structure (frequency domain).

Table 1. Investigated structures—general information and natural frequencies determined from measured data

	Location	Dimensions (m)	Height (m)	Building material	$f_{1,\mathrm{E-W}}$ (Hz)	$f_{2,\mathrm{E-W}}$ (Hz)	$f_{1,\mathrm{N-S}}$ (Hz)	$f_{2,\mathrm{N-S}}$ (Hz)
1	Riglergasse 10, Vienna	18 × 16	23.2	Masonry	2.42	6.75	—	—
2	Diesterwegg. 4, Vienna	12 × 24	16.7	Masonry	3.61	11.30	—	—
3	Istanbul I—Fenerbaçe	18 × 31	14.5	RC	5.10	10.20	5.43	19.39
4	Istanbul II—Fenerbaçe	19 × 21	17.0	RC	4.85	—	5.27	10.33
5	Istanbul III—Galatasaray	4 × 14	12.0	Masonry	3.02	7.16	—	—
6	Istanbul IV—Fenerbaçe	10 × 20	24.0	RC	1.75	7.15	2.53	10.07
7	Istanbul V—Fenerbaçe	18 × 18	37.0	RC	1.61	7.58	1.63	6.06
8	Bucharest—Hotel	50 × 17	25.0	RC	—	—	3.03	23.29

earthquake event. The mechanism of the environment during an earthquake event is illustrated in Figure 9.

4.1 Local amplification—site effects

Local soil amplification has become a major topic in seismic hazard analysis in the last few decades. The influence of the local subsoil conditions on the intensity of earthquakes was well investigated. Especially in areas with high population density, the effects of different conditions can have major effects on the damage distribution of building structures.

The effects of local amplification can have an influence on all important characteristics of an earthquake,

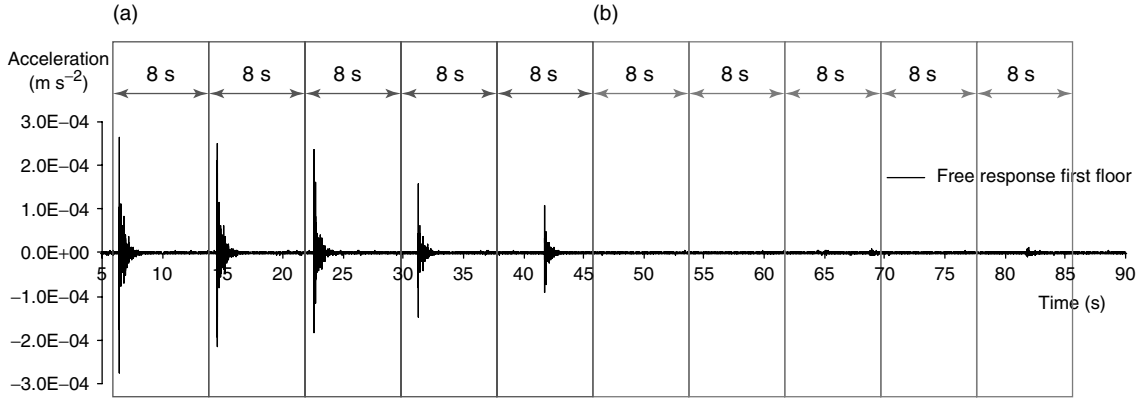


Figure 7. Transient (a) and ambient (b) windows with a length of 8 s for the acceleration record at the first floor of an investigated residential building.

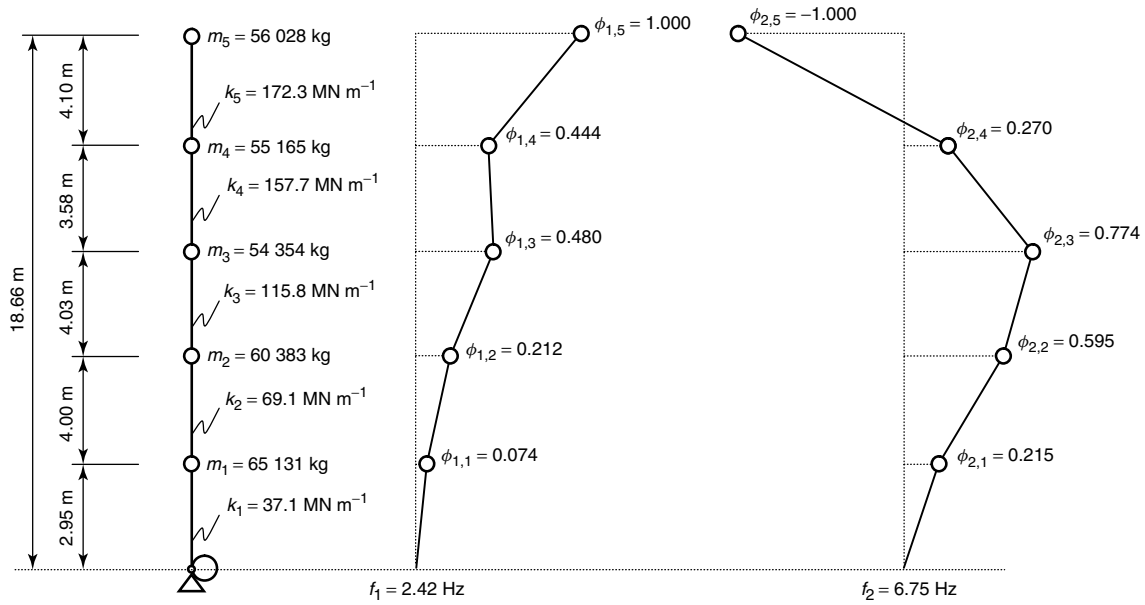


Figure 8. Final model of the investigated residential building—mode shapes (east–west direction) for the corresponding frequencies.

such as the amplitude, frequency content, and duration. Depending on the geometric extension and the material parameter of the subsoil layers, this influence is quite different.

4.1.1 Basic information

Local soil conditions have a large influence on the intensity of ground motion and earthquake damage [10]. The effects of the local soil conditions on ground motions have been demonstrated in many earthquakes that have occurred around the world. Since the use of strong motion instruments has increased over recent years, it was also possible to measure those effects. In the design of building structures, local site effects play a very important role and have to be considered as the cases arise. This can be done using site-specific design

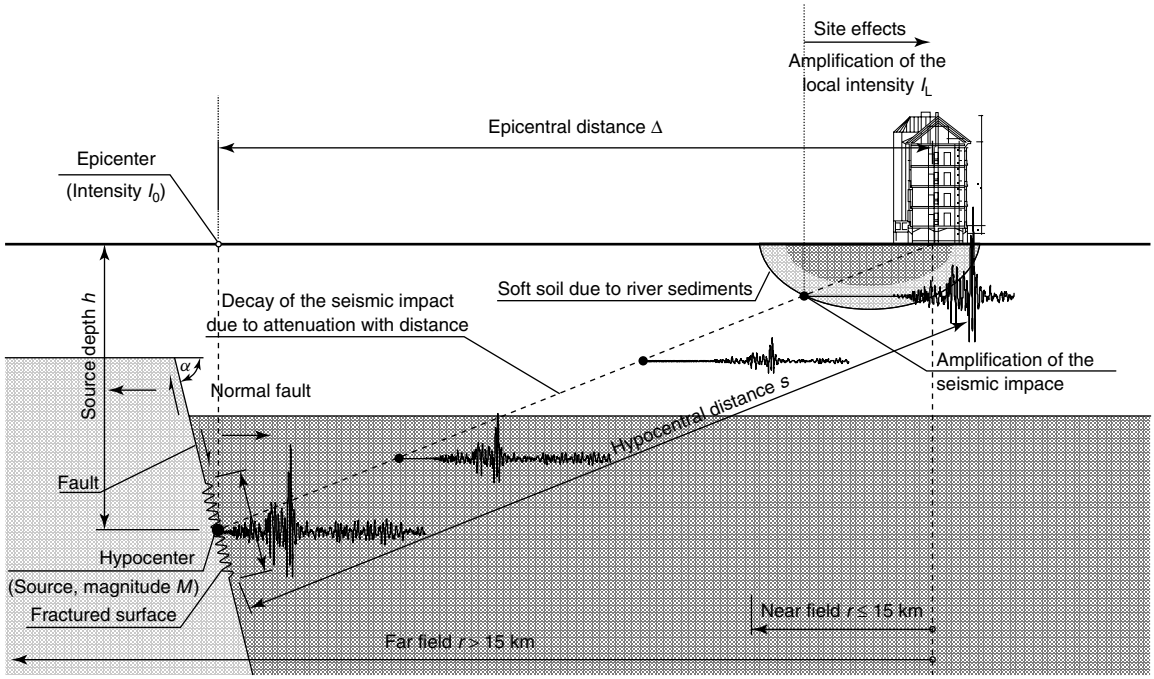


Figure 9. Schematic illustration of the seismic wave attenuation from the earthquake source to the investigated site: fault geometry, normal fault; hypocenter (source); epicenter (epicentral intensity I_0); epicentral distance Δ; hypocentral distance s; dip angle α; width of the fractured surface B.

ground motions. Nevertheless, in many cases local soil effects have been neglected during design previously. Therefore, the vulnerability of the existing buildings due to site effects needs to be evaluated subsequently.

4.1.2 Theoretical background

All important characteristics and parameters of the ground motion such as duration, frequency content, and amplitude are significantly influenced by local soil conditions. Therefore, the incoming ground motion on equal building structures founded on different subsoil can have rigorous differences. The elongation of local site effects is also dependent on the soil geometry (mainly the layer thickness), site topography, the material properties of the subsurface materials, and on the characteristics of the input motion.

In most of the cases, the S-wave velocities of materials near the surface are smaller than at greater depths [10]. The conservation of elastic wave energy under negligence of the scattering and material damping requires that the energy flow $\rho v_s \dot{u}^2$ from the depth is constant. As the density ρ and the S-wave velocity v_s decreases with ascending waves, the particle velocity $\dot{u}^2$ has to increase and leads to a higher ground motion amplitude on the surface.

The parameters of local soil effects can be evaluated theoretically using existing information about the subsoil conditions or experimentally by means of surface measurements.

4.1.3 Evaluation of site effects using theoretical investigations

Amplifications of soil site responses were simulated using several computer programs that assume simplified soil deposit conditions such as horizontal soil layers of infinite extent. One of the first computer programs developed for this purpose is SHAKE [11]. More than 25 years after its release, it is still commonly used in geotechnical earthquake engineering.

4.1.4 Measurement-based investigations of local site effects

Ambient vibration testing is a very attractive method for measurements in urban areas with moderate seismicity, because on one hand the low ambient vibration amplitudes indicate better correlation with moderate seismic events (approximately linear elastic constitutive material behavior) and on the other hand seismic cross-hole tests are complicated in urban areas because of the disturbance of building occupants and induced damages on the building structure.

A very common method to determine local site effects is the evaluation of the amplification factor using the H/V ratio according to Nakamura [12]. The method was successfully accomplished in countries with moderate seismicity where no strong motion data was available. It is therefore convenient for almost every region. In general, the input for the H/V computation can also consist of transiently or even ambiently excited signals.

In case of the evaluation of the amplification factor in urban areas, it is more beneficial to use ambient excitation because the large number of samples, which are needed for statistical evaluation would, in case of transient excitation, lead to unacceptable disturbance of the occupants.

4.1.5 Implementation in seismic hazard maps

The results of the evaluation of amplification factors can be used to generate seismic hazard maps. According to the high resolution of the measurement grid, this procedure is often called *microzonation*. The results of those microzonation studies should be implemented in existing catastrophe management plans to identify the most vulnerable parts of an urban area. This is primarily useful to organize the rescue teams in case of a tremendous earthquake event. In addition, the identification of the most vulnerable seismic zones within the city can have a direct influence on urban planning. In case of existing public utilities and infrastructure in very vulnerable areas, it may have an influence on the ongoing safety precautions. Therefore, the main utilizations of a seismic hazard map can be listed as follows:

- additional tool for the catastrophe management;
- implementation on urban and regional planning;
- attachment for national building regulations; and
- decision support for insurance companies.

4.2 Liquefaction

Nonlinear material behavior of the subsoil may have an important influence on the damage potential of earthquakes. In the last few years, numerous examples of this influence occurred, e.g., the Loma Prieta earthquake in California 1989, the Kobe earthquake in Japan 1995, and the Ismit earthquake in Turkey 1999. In the following, the nonlinear effects (mainly related to liquefaction) of the soil during an earthquake are explained.

According to technical literature [10], liquefaction is one of the most important, interesting, complex, and controversial topics in geotechnical earthquake engineering. The destructive effects of liquefaction were first recognized in 1964 during the Good Friday earthquake in Alaska and the Niigata earthquake in Japan. During both earthquakes, the phenomenon of liquefaction caused some tremendous damages, including slope failures, foundation failures, and flotation of buried structures. As liquefaction describes different phenomena, some of the most important are described, following the specification in the literature [10].

Liquefaction was first described [13] in conjunction with some phenomena that involve soil deformations caused by monotonic, transient, or repeated disturbance of saturated cohesionless soils under undrained conditions. The generation of excess pore pressure under drained loading conditions is a hallmark of all liquefaction phenomena. The tendency for dry cohesionless soils to densify under both static and cyclic loading is well known. When cohesionless soils are saturated, however, rapid loading occurs under undrained conditions, so the tendency for densification causes excess pore pressures to increase and the effective stresses to decrease. The liquefaction phenomena resulting from this process can be divided into flow liquefaction and cyclic mobility.

4.3 Earthquake-induced landslides

Landslides induced by earthquakes can be a common phenomena in mountainous regions [14]. The number

of landslides is dependent on the intensity of the earthquake and the geometric conditions of the surface (slope angles) as well as on the properties of the soil.

Multiple landslides occur almost simultaneously when slopes are shaken by an earthquake or over a period of hours or days when failures are triggered by intense rainfall or snow melting. The 1964 Great Alaska Earthquake caused widespread landsliding and other ground failure, which caused most of the monetary loss due to the earthquake. Other areas of the United States, such as California and the Puget Sound region in Washington, have experienced slides, lateral spreading, and other types of ground failure due to moderate to large earthquakes. Another phenomenon of earthquakes is widespread rockfalls caused by loosening of rocks.

Some significant examples of earthquake-triggered landslides in the past are as follows [15]:

- In May 1960, one of the world's strongest earthquakes ($M_w = 9.2$) [16] struck the coast of south-central Chile causing numerous major landslides and hundreds of surficial slides [17, 18]. The largest individual mass movements were three contiguous landslides with a total volume of 40 million m^3.
- The $M = 9.2$ Alaska earthquake in 1964 dislodged landslides from slopes over an area of about 260 000 km^2 [19].
- The 1987 Reventador earthquakes ($M = 6.1$ and 6.9) in northeastern Ecuador occurred after about 1 month of heavy rain, causing thousands of small landslides, which began as small slips on steep slopes [15, 20]. These thin slides liquefied and turned into major debris flows in the region's tributaries and main streams. Hundreds of square kilometers of the Earth's surface were modified by the landslides, which had a total volume estimated at 110–120 million m^3 [21].
- An event similar to that in Ecuador occurred in southwestern Colombia in 1994. The $M = 6.4$ Paez earthquake caused thousands of thin residual slides on steep slopes; these thin slides liquefied and turned into damaging debris avalanches and debris flows [22]. A total of 250 km^2 of the ground surface of the area was affected.

Geomorphological inventory maps can be used to study the relationships between the lithological and structural settings and the landslide types and pattern. Despite the fact that this information can prove extremely valuable for landslide hazards assessment, review of the literature shows that such studies are rare, mostly because geographical databases containing landslides, lithological, and structural information with the required accuracy are not readily available. Landslide inventory maps should be prepared after each landslide-triggering event (e.g., a rainstorm, a snowmelt event, or an earthquake), thereby covering the entire territory affected by the event. Such maps allow determining the full extent of landslide events on the structures and the infrastructure. They can also provide valuable information for evaluating the types, extent, and severity of damage caused by slope failures [23].

4.4 Surface faults and cracks

Surface faults and cracks can occur if the fracture path of the earthquake reaches the surface. This phenomenon is therefore dependent on the magnitude of earthquakes, the hypocentral depth, and the orientation of the fracture mechanism and more or less independent of the type of the subsoil. For the identification of zones with high vulnerability for surface faults and cracks, it is therefore necessary to have knowledge about active faults. These data are available in most of the cases; nevertheless, zones with active faulting are rather broad. Therefore, the influence of surface faulting and cracks can hardly be included into microzonation studies.

5 CONCLUSION

Monitoring of soil–structure interaction effects is something that needs to be intensified in the future. The topics mentioned in this article should indicate a perspective view on possible applications.

The seismic assessment of existing structures has become much more important during the last decades, as the seismic risk of many countries is redefined and increased. Within this article, the importance of the influence of the subsoil, the realistic assessment of the dynamic parameters, and local seismic effects are described. This should give the engineers a

brief overview of the complexity of a comprehensive seismic investigation of building structures.

REFERENCES

[1] Pecker A. *Dynamique des Sols*. Presses de l'Ecole Nationale des Ponts et Chaussées, 1984.

[2] Gazetas G, Mylonakis G. Seismic soil structure interaction: new evidence and emerging issues. *Geotechnical Earthquake Engineering and Soil Dynamics*. ASCE, 1998; Vol. 2, pp. 1119–1174.

[3] Lysmer J. Analytical procedures in soil dynamics. *State of the Art. ASCE Conference on Soil Dynamics and Earthquake Engineering*. Pasadena, CA, 1978.

[4] Wolf JP. *Dynamic Soil Structure Interaction*. Prentice Hall: Englewood Cliffs, NJ, 1985.

[5] Chopra A. *Dynamics of Structures*, *Third Edition*. Prentice Hall, 2006.

[6] Wenzel H, Pichler D. *Ambient Vibration Monitoring*. John Wiley & Sons, 2005.

[7] Mordini A, Savov K, Wenzel H. Damage Detection on Stay Cables Using an Open Source-Based Framework for Finite Element Model Updating, *Structural Health Monitoring*, 2008 **7**(2):91–102.

[8] Pinho R. *Using Pushover Analysis for Assessment of Buildings and Bridges*. Course Material for Advanced Earthquake Engineering Analysis—CISM: Udine, 2006.

[9] Friswell MI, Mottershead JE. *Finite Element Model Updating in Structural Dynamics*. Kluwer Academic Publishers, 1995.

[10] Kramer SL. *Geotechnical Earthquake Engineering*. Pratice Hall, 1996, ISBN 0133749436.

[11] Schnabel PB, Lysmer J, Seed HB. *SHAKE: A Computer Program for Earthquake Response Analysis of Horizontally-Layered Sites*, Report No. EERC-72/12. Earthquake Engineering Research Center, University of California at Berkeley, Berkeley, CA, 1972.

[12] Nakamura Y. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly Report of RTRI* 1989 **30**(1):25–33.

[13] Mogami T, Kubo K. The behavior of soil during vibration. *Proceedings, 3rd International Conference on Soil Mechanics and Foundation Engineering*, Zürich, Switzerland, 1953 Vol. 1; pp. 152–153.

[14] U.S. Department of the Interior. *Landslide Types and Processes*. U.S. Geological Survey, Fact Sheet 2004–3072, 2004.

[15] Schuster RL, Highland LM. *Geologic Hazards Team, Impact of Landslides and Innovative Landslide-Mitigation Measures on the Natural Environment*. U.S. Geological Survey: Denver, CO, 2003.

[16] Kanamori H. The energy release in great earthquakes. *Journal of Geophysical Research* 1977 **82**(B20):2981–2988.

[17] Davis SN, Karzulovic JK. Landslides at Lago Riñihue, Chile. *Bulletin of the Seismological Society of America* 1963 **53**(6):1403–1414.

[18] Weischet W, Von Huene R. Further observations of geologic and geomorphic changes resulting from the catastrophic earthquake of May 1960, in Chile. *Bulletin of the Seismological Society of America* 1963 **53**(6):1237–1257.

[19] Plafker G, Kachadoorian R, Eckel EB, Mayo LP. *Effects of the Earthquake of March 27, 1964*. U.S. Geological Survey Professional Paper 542-G, 1969.

[20] Schuster RL, Alberto SN, O'Rourke TD, Crespo E, Plaza-Nieto G. Mass wasting triggered by the 5 March 1987 Ecuador earthquakes. *Engineering Geology* 1996 **42**(1):1–23.

[21] Nieto AS, Schuster RL. Mass wasting and flooding. In *The March 5, 1987, Ecuador Earthquakes: Mass Wasting and Socioeconomic Effects*, *Natural Disaster Studies*, Schuster RL (ed). National. Res. Council: Wash, DC, 1991 Vol. 5 p. 51–82.

[22] Martinez J, Avila G, Agudelo A, Schuster RL, Casadevall TJ, Scott KM. Landslides and debris flows triggered by the 6 June 1994 Paez earthquake, southwestern Colombia. In *Landslides of the World*, Sassa K (ed). Japan Landslide Society, Kyoto University Press, 1999; pp. 227–230.

[23] Guzzetti F. *Landslide Cartography, Hazard Assessment and Risk Evaluation: Overview, Limits and Prospective*, Perugia, 2006.

Chapter 139

System Identification for Soil–structure Interaction

Erdal Safak

Department of Earthquake Engineering, Kandilli Observatory and Earthquake Research Institute, Bogazici University, Istanbul, Turkey

1 INTRODUCTION

The term *soil–structure interaction* (*SSI*) refers to the influence of the deformations of the ground on the dynamic response of a structure. It is an important phenomenon of the seismic response of structures. If a structure is founded on rock, it is reasonable to assume that the ground does not deform and the structure is rigidly fixed at the base. If it is founded on soil, however, the flexibility of soil will cause the foundation of the structure to have translational and rotational motions when it is vibrating. SSI can have a major influence on the seismic response of structures founded on soft soils. It can significantly alter the vibration characteristics and, consequently, the characteristics of recorded motions. If it is not accounted for in the analysis, SSI can cause erroneous interpretation of recordings from structures.

Although SSI is always present to some degree, it is generally assumed that when SSI is small, the recorded motions at the foundation level are not influenced by the motions of the superstructure. Therefore, the structure can be assumed to be fixed-based and the foundation level recordings can be taken as the base excitation. When SSI is significant, this assumption is not valid because of the feedback from the structure to the foundation and the surrounding soil medium. The feedback makes the structure a closed-loop dynamic system, where the input (i.e., the foundation motion) and the output (i.e., the motion of the superstructure) are coupled.

The identification of SSI is relatively easy if, in addition to the records from the building, there are also records available from downhole or nearby free-field stations that are not influenced by the structure's vibrations. In most cases, such records are not available.

The common approach to identify structures that are subjected to SSI has been to develop finite element models of the structure with soil springs, and then determine the characteristics of the springs

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

by trial and error until the recorded superstructure motions are matched. This approach, however, does not lead to a unique solution, since a large number of soil and structural system combinations can match the recorded motions.

There are very few studies on direct identification of SSI from recorded motions, the most notable being [1, 2], Safak [3] and [4]. This article reviews the effects of SSI on the response of structures to seismic excitations by using simple models. It presents two criteria to detect the presence of SSI in a structure, and introduces a methodology to identify the fundamental frequencies of the fixed-base building, and the characteristics of the translational and rocking motions of the structure due to SSI.

2 EFFECTS OF SSI ON RECORDED MOTIONS

We can investigate the effects of SSI on the frequency content of the vibrations of a structure by employing the simplest model available for SSI, the 2-DOF (two degrees-of-freedom) model. The model accounts for the relative translational motions of the foundation with respect to soil. The rigid-body vibrations with respect to foundation, known as *rocking motions*, will be considered later. There are more complicated and realistic models of SSI available in a number of references in the literature (e.g., [5, 6]). In spite of its simplicity, the 2-DOF model is sufficient to expose the fundamental characteristics of SSI and is widely used in practice.

A schematic of the 2-DOF model is shown in Figure 1. The lower and upper mass-spring-dashpots represent the foundation and the structure, respectively. The model is based on the assumptions that both the superstructure and the soil are linear, their motions are dominated by fundamental modes, the soil stiffness is frequency independent, and the rocking motions are negligible.

The model given in Figure 1 can be considered to represent a structure whose vibrations are dominated by its first-mode response, such as a multistory building. Let x_f, x_b, and x_0 denote the absolute displacements of the foundation, building, and the free field, respectively, and $y_f = x_f - x_0$ and $y_b = x_b - x_f$ the relative displacements of the foundation with respect to the free field and of the building with respect to the foundation. With these, we can write the following equations for the motion of the 2-DOF system subjected to base acceleration $\ddot{x}_0$

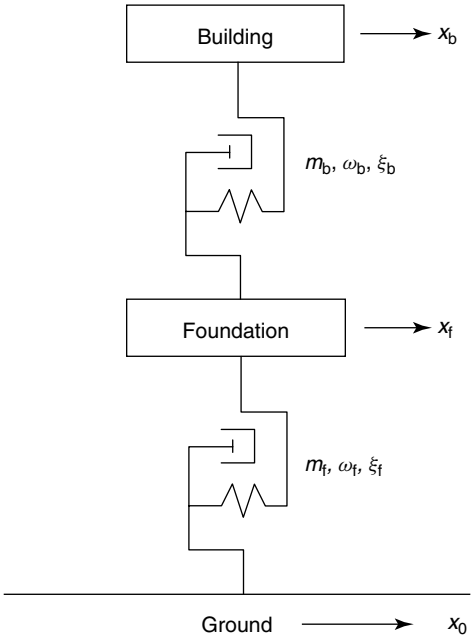


Figure 1. Two-degrees-of-freedom system representing soil–structure interaction.

$$m_f\ddot{y}_f + c_f\dot{y}_f + k_f y_f - c_b\dot{y}_b - k_b y_b = -m_f\ddot{x}_0$$
$$m_b(\ddot{y}_f + \ddot{y}_b) + c_b\dot{y}_b + k_b y_b = -m_b\ddot{x}_0 \quad (1)$$

where a dot over a variable denotes the derivative with respect to time. The standard notation used in the equations represent the mass (m), stiffness (k), and damping (c) of the fixed-base building (subscript b), and of the foundation when no building is present (subscript f), respectively. The two natural frequencies of this coupled system are

$$\omega_{1,2}^2 = \frac{1}{2}\cdot\left[\omega_f^2 + (1+\mu)\omega_b^2 \pm \sqrt{[\omega_f^2 + (1+\mu)\omega_b^2]^2 - 4\omega_f^2\omega_b^2}\right] \quad (2)$$

where $\mu = m_b/m_f$, the mass ratio, and $\omega_f = \sqrt{k_f/m_f}$ and $\omega_f = \sqrt{k_f/m_f}$ are the natural frequencies of the foundation when no building is present, and of the building when it is fixed-based, respectively.

The ratios of the natural frequencies of the coupled system to the frequency of the fixed-base building, and the frequency of the foundation when no building is present, can be expressed by the following equations:

$$\left(\frac{\omega_{1,2}}{\omega_b}\right)^2 = \frac{1}{2\eta^2}\left[1 + (1+\mu)\eta^2 \pm \sqrt{[1+(1+\mu)\eta^2]^2 - 4\eta^2}\right]$$

$$\left(\frac{\omega_{1,2}}{\omega_f}\right)^2 = \eta^2 \cdot \left(\frac{\omega_{1,2}}{\omega_b}\right)^2 \tag{3}$$

where $\eta = \omega_b/\omega_f$.

Note that the ratios depend only on μ and η, the mass, and the frequency ratios. The variation of ω_1/ω_b for η varying from 0.1 to 10, and three values of μ are plotted in Figure 2. The figure shows that for the lowest natural frequency, ω_1, the ratios are always less than one. In other words, the fundamental frequency with SSI is always lower than the frequencies of the fixed-base building and the foundation with no building. It can be shown similarly that for the second natural frequency, ω_2, the ratios are always greater than one.

By studying the limits of equation (3), we find that

$$\text{If } \eta \ll 1 : \omega_1 \approx \omega_b \text{ and } \omega_2 \approx \omega_f$$
$$\text{If } \eta \gg 1 : \omega_1 \approx \frac{\omega_f}{\sqrt{1+\mu}} \text{ and } \omega_2 \approx \sqrt{1+\mu}\cdot\omega_b \tag{4}$$

The first expression in the above equations corresponds to flexible buildings and stiff soil conditions, where SSI is negligible, and shows that the frequencies of the coupled system converge to those of the fixed-base building and the foundation, respectively, for all values of μ. The second expression corresponds to rigid buildings and soft soil conditions, where SSI is significant, and shows that the frequencies of the coupled system converge to two values, one smaller than the frequency of the foundation and the other larger than the frequency of the fixed-base building.

In general, the recorded responses in buildings are accelerations. The effects of SSI on the frequency content of recorded accelerations can be investigated by studying the frequency response functions, $H_f(\omega)$ and $H_b(\omega)$, of $\ddot{x}_f$ and $\ddot{x}_b$. To calculate frequency response functions, we take $\ddot{x}_0 = e^{i\omega t}$ in equation (1) and let $\ddot{x}_f = H_f(\omega)\cdot e^{i\omega t}$ and $\ddot{x}_b = H_b(\omega)\cdot e^{i\omega t}$. Solving

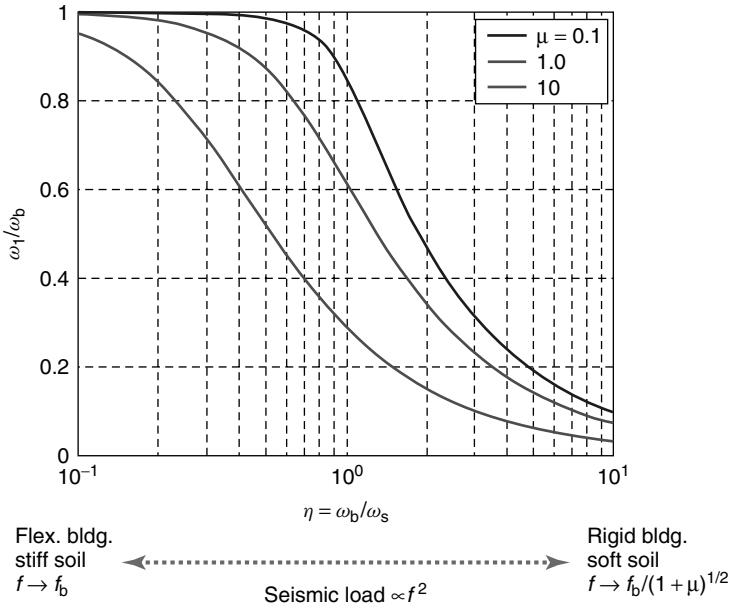


Figure 2. Effects of SSI on the natural frequency of the structure.

for $H_f(\omega)$ and $H_b(\omega)$ gives [7]

$$H_f = \frac{1}{\Delta}\cdot\left[-i\omega^3 2\xi_f\omega_f - \omega^2(\omega_f^2 + 4\xi_f\xi_b\omega_f\omega_b) + i\omega(2\xi_f\omega_f\omega_b^2 + 2\xi_b\omega_b\omega_f^2) + \omega_f^2\omega_b^2\right]$$

$$H_b = \frac{1}{\Delta}\cdot\left[-\omega^2 4\xi_f\xi_b\omega_f\omega_b + i\omega\left(2\xi_f\omega_f\omega_b^2 + 2\xi_b\omega_b\omega_f^2\right) + \omega_f^2\omega_b^2\right]$$

where

$$\Delta = \omega^4 - i\omega^3\cdot\left[2\xi_f\omega_f + 2(1+\mu)\xi_b\omega_b\right] - \omega^2\cdot\left[\omega_f^2 + (1+\mu)\omega_b^2 + 4\xi_f\xi_b\omega_f\omega_b\right] + \cdots + i\omega\cdot\left(2\xi_f\omega_f\omega_b^2 + 2\xi_b\omega_b\omega_f^2\right) + \omega_f^2\omega_b^2 \qquad (5)$$

where $\xi_b = c_b/(2m_b\omega_b)$ and $\xi_f = c_f/(2m_f\omega_f)$, the viscous damping ratios for the building and the foundation, respectively, and $i = \sqrt{-1}$.

The investigation of the changes in the transfer functions with various soil and structural parameters is given in [2]. The results clearly confirm that SSI has significant influence on the frequency content of the motions recorded at the foundation and the superstructure. Depending on the magnitude of SSI, the dominant frequencies of the recorded motions can be completely different from those of the fixed-base building, or the foundation when no building is present. The observed narrowband characteristics of the foundation transfer function for soft soils and large buildings suggest that recordings from the basements of such buildings cannot be taken as the free-field ground motions.

3 DETECTION AND IDENTIFICATION OF SSI

Since SSI alters the frequency characteristics of the recorded motions, it is important to determine if the building is subjected to SSI, prior to any identification. Normally, the motion of the building is recorded at the foundation level, top story, and several intermediate stories. If there is no SSI, we can identify the building by taking the recordings at the foundation level as the input and the recordings at upper stories as the output. A building with no SSI represents a causal system, because the input at time t_i can only influence the output at times $t > t_i$. Since there is no coupling (i.e., feedback) between the input and the output, such systems are termed *open-loop systems*. When the building is subjected to SSI, the motions of upper stories influence the motion of the foundation, i.e., the input and the output of the system are coupled. Such systems are termed *closed-loop systems*. It can be shown that the impulse response function of closed-loop systems are noncausal [8].

The causality condition provides a convenient tool to detect the SSI. For causal systems, the impulse response of the system is zero for $t < 0$, whereas for noncausal systems it is not. Therefore, we investigate the impulse response of the building from the records of the foundation and the upper stories. If the impulse response shows significant amplitudes at negative times in comparison to those at positive times, we conclude that there is SSI. Impulse response functions can be calculated by taking the foundation accelerations as the input, and the top-story accelerations as the output, and using the method given in [9]. In this method, a prewhitening filter is applied to the input so that it becomes as close to a white noise as possible. The same filter is also applied to the output. The impulse response function is obtained by scaling the cross-correlation function of the filtered input and output. A more straightforward approachto calculate the impulse response function is to find the transfer function by taking the ratio of the Fourier transforms and then to take the inverse Fourier transform of the transfer function. However, this method does not give a correct and stable impulse response because of the noise in the records. The method suggested by Ljung is more robust and mathematically more accurate for noisy signals.

An alternative to the above procedure is to first identify the building as if there were no SSI and then investigate the cross correlation of the input with the residuals (i.e., the difference between the recorded output and the output of the identified model) of the identification. If there is SSI, the cross correlation would show large amplitudes at negative time lags. A more detailed discussion of this method can be found in [10].

A less sophisticated, but a much simpler, way to detect SSI is to compare the dominant frequency of the vibration records from the roof (or, one of the upper stories) of the building to the dominant

frequency of the foundation-to-roof transfer function. The former is the frequency with SSI, whereas the latter is the frequency if the building were fixed-based (i.e., no SSI). If these two frequencies are equal or close, there is no SSI, otherwise there is SSI. We prove this by using the simple 2-DOF model of the SSI given in Figure 1. Let us first calculate the ratio $R = H_b/H_f$ of the transfer functions for accelerations $\ddot{x}_b$ and $\ddot{x}_f$. From equation (5), and by defining the nondimensional frequency ratios $r_f = \omega_f/\omega$ and $r_b = \omega_b/\omega$, we can write

$$R(\omega) = \frac{H_b(\omega)}{H_h(\omega)} = \frac{4\xi_f\xi_b r_f r_b - i(2\xi_f r_f r_b^2 + 2\xi_b r_b r_f^2) - r_f^2 r_b^2}{r_f^2 + i2\xi_f r_f + 4\xi_f\xi_b r_f r_b - i(2\xi_f r_f r_b^2 + 2\xi_b r_b r_f^2) - r_f^2 r_b^2} \tag{6}$$

Note that R is independent of $\mu = m_b/m_f$. From equation (6), we can further derive the following equations for the squared amplitude R^2 of R:

$$|R(\omega)|^2 = \frac{r_b^2(r_b^2 + 4\xi_b^2)}{1 - 2r_b^2 + r_b^2(r_b^2 + 4\xi_b^2)}$$

or

$$|R(\omega)|^2 = \frac{1 + \left(2\xi_b \frac{\omega}{\omega_b}\right)^2}{\left[1 - \left(\frac{\omega}{\omega_b}\right)^2\right]^2 + \left(2\xi_b \frac{\omega}{\omega_b}\right)^2} \tag{7}$$

By making the derivative of $R(\omega)$ with respect to ω equal to zero, we can determine the frequency ω_{max} where $R(\omega)$ has its peak as

$$\omega_{max} = \frac{\left(-1 + \sqrt{1 + 8\xi_b^2}\right)^{1/2}}{2\xi_b} \cdot \omega_b \tag{8}$$

There are two important observations that can be made from equations (7) and (8). First, equation (7) shows that $|R(\omega)|$ is independent of μ, ω_f, and ξ_f. In other words, $|R(\omega)|$ is not influenced by the dynamic characteristics of the foundation. Second, equation (8), when the coefficient of ω_b on the right-hand side is evaluated numerically, shows that $|R(\omega)|$ has its peak at or near $r_b = 1$, or $\omega = \omega_b$. For example, for $\xi_b = 0.02$, 0.05, 0.10, 0.20, and 0.50 the corresponding values of the coefficient of ω_b are 1.00, 1.00, 0.99, 0.97, and 0.85, respectively. Note that, even at $\xi_b = 0.50$, the shift of the peak frequency from the frequency of the fixed-base building is only 15%. From these two observations, we conclude that $|R(\omega)|$ describes the characteristics of the fixed-base building, and always peaks at or near the fixed-base building frequency for buildings with small damping.

The statements we have just made regarding $|R(\omega)|$ being independent of the foundation characteristics may seem contradictory to our earlier statements, where we claim that SSI causes coupling of building and foundation motions. In fact, this was used to detect the existence of SSI by showing that the coupling results in noncausal impulse response function. There is no contradiction. As clearly seen in equation (6), the real transfer function $R(\omega)$ is a complex-valued quantity, and incorporates the coupling, because the parameters of both the building and the foundation are in the equation. Since the impulse response function is merely the inverse Fourier transform of $R(\omega)$, it also incorporates the coupling, and therefore is appropriate to be used for detection. It is the amplitude of the transfer function, $|R(\omega)|$, which is independent of the foundation characteristics. In other words, the coupling between the foundation and the building motions alters the phase of the transfer function, but not the amplitude. More details on this can be found in the discussion by Zhao and closure by Safak in [3].

Considering the definitions of $H_f(\omega)$ and $H_b(\omega)$, $R(\omega)$ can be calculated as the ratio of the Fourier amplitude spectrum (FAS) of the top-story accelerations to that of the foundation accelerations. The dominant peak of this ratio gives the fundamental frequency of the building for the fixed-base case. The fundamental frequency of the foundation can be estimated from equation (3). For this, we first determine ω_1, the dominant frequency of the top-story records, and make an assumption for the value of μ based on the design specifications for the building. We then put the values of ω_1, ω_b, and μ in equation (3), and solve it for ω_b.

The damping ratios ξ_b and ξ_f corresponding to ω_b and ω_f can be estimated by first filtering the accelerations by narrowband filters centered at ω_b and

ω_f, and then calculating the displacements and the rate of decay of the amplitudes.

If, in addition to recordings from the building, there were recordings from a nearby free-field site (i.e., a site whose motion is not influenced by the vibrations of the building), the identification of SSI would be fairly straightforward. We take the free-field recordings as the input, and foundation and upper-story recordings as the output, and identify the soil–foundation–structure system as a whole using open-loop identification techniques.

4 ROCKING MOTIONS

Rocking motions are the rotational rigid-body vibrations of structures with respect to their foundations. Rocking motions are frequently encountered in tall buildings founded on soft soils and mat foundations. It is less common in structures with pile foundations.

The 2-DOF model used above can be expanded to include rocking motions by allowing the foundation mass to rotate as well as translate. This requires adding a rotational spring (to represent the rocking stiffness) and a dashpot (to represent the rocking damping) in Figure 1, which results in a 3-DOF model for SSI. The equations of vibration for this model can be found in [4]. By writing the transfer functions for the superstructure and foundation, it can be shown that the apparent frequency, f_a, and the damping, ξ_a, of the structure, with translational and rocking SSI, are

$$f_a^2 = \frac{1}{1/f_b^2 + 1/f_f^2 + 1/f_\theta^2}$$

$$\xi_a = \left(\frac{f_a}{f_b}\right)^3 \xi_b + \left(\frac{f_a}{f_f}\right)^3 \xi_f + \left(\frac{f_a}{f_\theta}\right)^3 \xi_\theta \quad (9)$$

where f_θ and ξ_θ denote the frequency and the damping ratio for the rocking.

Rotational vibrations are commonly identified from the measurements of the vertical motions of the foundation mat. For a building with a rectangular cross section and a rigid foundation mat, we need at least three vertical sensors located near the corners of the foundation. If the foundation mat is not rigid (i.e., it can deform in shear or bending), we would need more vertical sensors to separate the rigid-body rocking motions from the mat deformations. Time history, and the frequency and the damping, of the rocking motions are determined by taking the differences of the vertical motions recorded by the sensors. Typically, for building-type structures, the recorded amplitudes of vertical vibrations at the foundation level are much smaller than those of horizontal vibrations. The amplitudes of the calculated rocking motions are even smaller. Therefore, the signal-to-noise ratios in the recorded vertical motions are usually very low. Taking the difference of two noisy signals gives a signal with a noise level much worse than the original signals [11]. It is strongly recommended that first the rocking frequency is identified from the spectral analysis of vertical records before taking the difference. Next, to improve signal-to-noise ratios, the records should be band-pass filtered around the rocking frequency, and then the rocking signal should be calculated by taking the difference of the filtered records. This process would improve the accuracy of the identified rocking characteristics. To further confirm that the calculated values are those of the rocking but not the deformations of the foundation mat, the amplitude and the phase of the rocking displacements at the foundation level should be compared with those of horizontal displacements (calculated after the records are band-pass filtered around rocking frequency) at upper stories.

5 CONCLUSIONS

SSI can significantly alter the characteristics of seismic records from structures, particularly when they are founded on soft soil. It can be shown that the dominant frequency of a structure with SSI is always smaller than that of the fixed-base structure and of the foundation with no superstructure.

If there are recordings from a nearby free-field site (i.e., a site whose motions are not influenced by the vibrations of the structure), a structure with SSI can be identified by using the standard open-loop system identification techniques, i.e., by taking the free-field records as the input and the foundation and upper-story records as the output. However, if the foundation records are taken as the input, the standard identification techniques may not work because the coupling between the input and the output makes the system a closed-loop system. The impulse response function of a close-loop system is noncausal.

The causality condition provides a convenient tool to detect the existence of SSI. For this, we calculate the impulse response function from the foundation and upper-story accelerations. If the amplitudes of the impulse response function at negative times are comparable to those at positive times, we conclude that the building is affected by SSI.

The study of transfer functions of a simple 2- or 3-DOF model for SSI shows that the ratio of the FAS of an upper-story acceleration to that of the foundation acceleration gives the dominant frequency of the fixed-base building. Once this frequency is determined, the dominant frequencies of the translational and rotational SSI effects can be estimated from simple analytical relations.

REFERENCES

[1] Luco JE, Trifunac MD, Long HL. Isolation of soil–structure interaction effects by full-scale forced vibration tests. *Earthquake Engineering and Structural Dynamics* 1988 **116**(1):1–21.

[2] Safak E. Detection and identification of soil–structure interaction in buildings from vibration recordings. *Journal of Structural Engineering, ASCE* 1995 **121**(5):899–906.

[3] Zhao JX, Safak E. Discussion and closure "Detection and identification of soil–structure interaction in buildings from vibration recordings by E. Safak". *Journal of Structural Engineering, ASCE* 1997 **123**(5):690–692.

[4] Stewart JP, Fenves GL. System identification for evaluating soil–structure interaction effects in buildings from strong motion recordings. *Earthquake Engineering and Structural Dynamics* 1998 **27**: 869–885.

[5] Veletsos AS. Dynamics of structure-foundation systems. In *Structural and Geotechnical Mechanics*, Hall WJ (ed). Prentice-Hall: Englewood Cliffs, NJ, 1977.

[6] Wolf JP. *Dynamic Soil–Structure Interaction*. Prentice-Hall: Englewood Cliffs, NJ, 1985.

[7] Crandall SH, Mark WD. *Random Vibration in Mechanical Systems*. Academic Press: New York, 1963.

[8] Akaike H. Some problems in the application of the cross-spectral method. In *Spectral Analysis of Time Series*, Harris B (ed). John Wiley & Sons: New York, 1967.

[9] Ljung L. *System Identification: Theory for the User*. Prentice-Hall: Englewood Cliffs, NJ, 1987.

[10] Safak E. *Analysis of Recordings in Structural Engineering: Adaptive Filtering, Prediction, and Control*, Open-File Report, 88-647. U. S. Geological Survey: Menlo Park, CA, 1988.

[11] Safak E. On estimation of site amplification from ambient ground noise. *Proceeding of the First European Conference on Earthquake Engineering and Seismology*. Geneva, 3–8 September 2006.

Chapter 141

Environmental Factors Derived from Satellite Data of Java, Indonesia

Barbara Teilen-Willige[1], Farah Mulyasari Sule[2,3] and Helmut Wenzel[4]

[1] *Department of Hydrogeology and Bureau of Applied Geoscientific Remote Sensing (BAGF), Technical University of Berlin, Stockach, Germany*
[2] *German–Indonesian Technical Cooperation, Environmental Geological Centre, Bandung, Indonesia*
[3] *Institut Teknologi Bandung, Center for Disaster Mitigation, Bandung, Indonesia*
[4] *VCE Holding GmbH, Vienna, Austria*

1 INTRODUCTION

This contribution is concerned with natural environmental factors, especially with improving the understanding of the influence of environmental factors and of natural hazards on civil infrastructure and on damage-accumulation prognostics. It addresses problems caused by extreme geologic processes and hazards such as earthquakes, flooding, and volcanic eruptions. The use of geographic information systems (GIS) integrated remote sensing data in the scope of environmental studies has been a continuous process taking place over the last decades [1]. Data from earth-observing satellites have become a valuable supporting tool for natural hazard damage detection in Indonesia, especially during the aftermath of earthquake and tsunami disasters as in December 2004 and July 2006. Earth observation satellites as LANDSAT, SPOT, IKONOS, QUICKBIRD, ERS, or ENVISAT with increasing capabilities in terms of spatial, temporal, and spectral resolution allow a more efficient, reliable, and affordable monitoring over time. Thus, remote sensing technology has become a fundamental input for GIS, especially for natural hazard information systems. The design of a common GIS database structure—always open to new data—can greatly contribute to the homogenization of methodologies and procedures of natural hazard risk management requiring an approach by integrating remote

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

sensing data, geologic, geophysical, seismotectonic, and topographic data and catalogs of historical hazardous events. This can be demonstrated by the example of Java/Indonesia.

2 METHODS

In order to establish a cost-effective method for getting a quick overview of the determining factors influencing environment and potential damage intensity in hazard-prone areas, it is recommended to start analyzing those causal factors and their complex interactions first based on remote sensing and GIS methodologies and later, step by step, going into details. The goal is to develop a multisensor and multirisk approach in a GIS environment to assess the potential for natural hazard on a regional basis. The various data sets as LANDSAT thematic mapper (TM) data, topographic, geological, and geophysical data from the investigation areas are integrated as layers into GIS using the software ArcGIS 9.2 of ESRI. The GIS-integrated evaluation of the georeferenced satellite imageries allows the storage of the results in a standard form such as vector formats (point-, line-, or polygon shapefiles). Various digital image processing tools delivered by ENVI Software/CREASO were tested, as for finding the best-suited LANDSAT enhanced thematic mapper (ETM) band combinations or contrast stretching parameters. The imageries were merged with the panchromatic band 8 of LANDSAT ETM to get the spatial resolution of 15 m. Standard approaches of digital image processing with regard to the extraction of natural hazard relevant information used for this study are methods like classification for land use and vegetation information, and processing of the thermal band 6 for deriving surface-temperature information. For the investigation of vegetation anomalies related

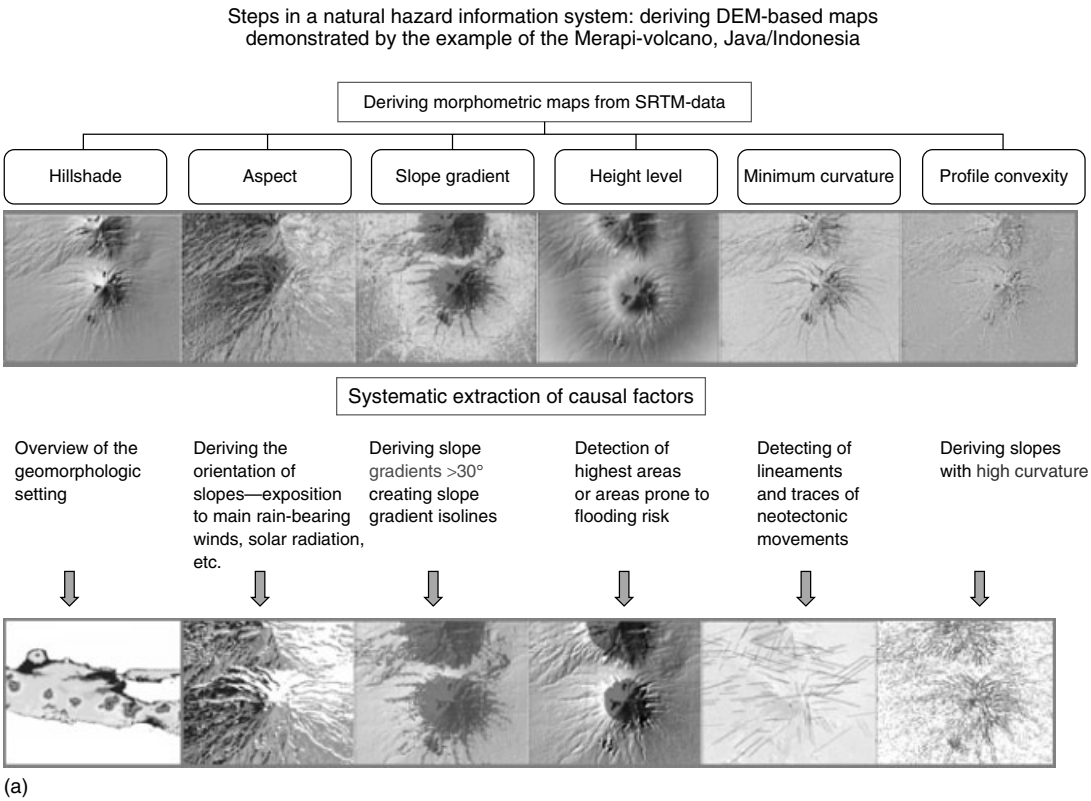


Figure 1. (a,b) Extraction of critical geomorphometric and hydromorphologic parameters based on SRTM and LANDSAT data demonstrated by the example of the Merapi volcano in Java/Indonesia.

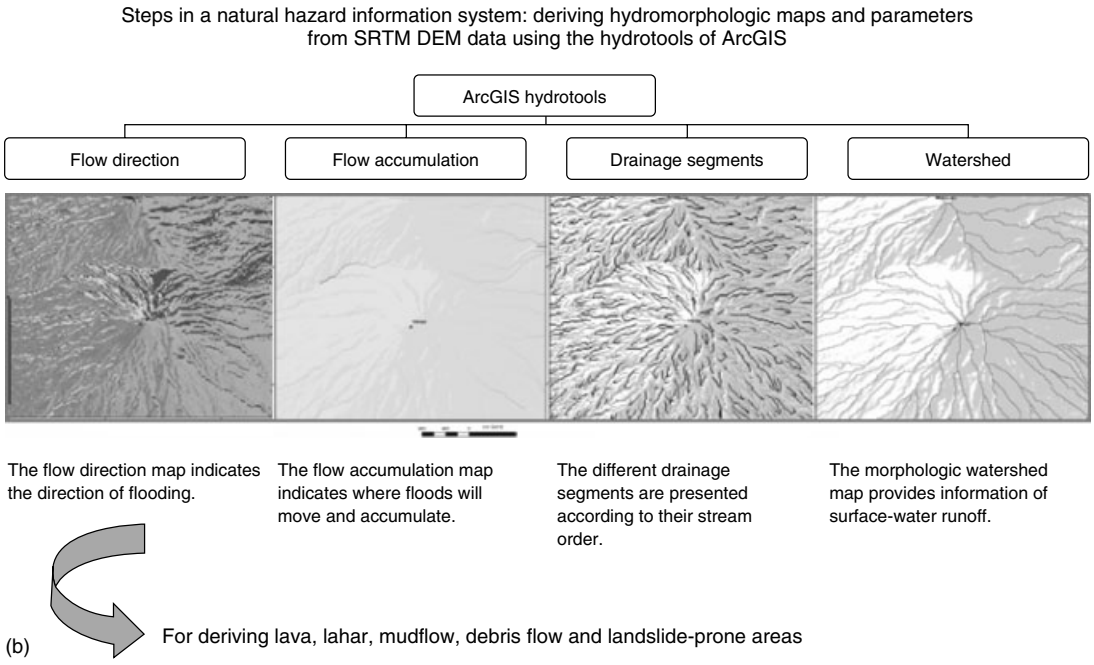


Figure 1. (*Continued*)

to subsurface structures and landslides, the NDVI (normalized difference vegetation index) was calculated on the basis of the available LANDSAT ETM bands 3 and 4. The evaluation of digital topographic data is of great importance as it contributes to the detection of the specific geomorphologic/topographic settings of hazard-prone areas. Data of the shuttle radar topography mission (SRTM, February 2000) are used to provide accurate digital elevation information. A systematic GIS approach is recommended extracting geomorphometric parameters based on digital elevation model (DEM) data as part of a multihazard information system. Figure 1(a) and (b) shows how the causal factors for natural hazards in Java are extracted systematically. From slope gradient maps, the areas with the steepest slopes are extracted, from curvature maps the areas with the highest curvature that are susceptible to landslides are extracted, from height maps the low-lying areas susceptible to flooding are extracted, and from flow accumulations maps the areas with the highest flow accumulations are extracted. Height maps help to search for topographic depressions, which are often linked with water accumulations and wetlands. Linear morphologic features (lineaments) visible on hillshade maps and LANDSAT imageries are often related to traces of faults and fractures in the subsurface.

3 EVALUATIONS OF LANDSAT AND SRTM DATA FOR DETECTING AREAS PRONE TO NATURAL HAZARDS

3.1 Earthquake damage amplification due to local site conditions

The most important geodynamic process, which determines all the tectonic and geomorphologic features of the Java Island, is the subduction of plates. Indonesia is a country where four of the earth's main plates contribute to the seismic activities in the region. Other earthquake sources are interplate and intraplate subduction seismic source zone [2]. One important factor that must be accounted for in local hazard studies is the site response caused by the surface conditions. Earthquake damage may vary locally, being a function of the type of structures in

the subsurface and/or soil mechanical ground conditions, for example, of faults and fractures, lithology or groundwater table [1] (*see* **Soil–Structure Interaction and Seismic Effects**; **System Identification for Soil–structure Interaction**). Evaluations of remote sensing data can help considerably to identify those vulnerable areas, to enhance mapping, and to improve evacuation planning. Remote sensing data can be used to map factors that are related to the occurrence of higher earthquake shocks and/or earthquake-induced secondary effects such as liquefaction or landslides (*see* **The Influence of Environmental Factors**). Previous earthquakes have indicated that the damage and loss of life are mostly concentrated in areas underlain by deposits of soft soil and high groundwater tables, for example, the Mexico City earthquake in 1985 [3]. Soft soils amplify shear waves and, thus, amplify ground shaking. The growth of Jakarta has been documented by earth observation satellites since 1972. Figures 2 and 3 illustrate the development of the Indonesian capital from 1976 to 2001 based on LANDSAT scenes. The urban areas show an enlargement of more than 2/3 in this time period. This monitoring of the urban development is of great importance for disaster preparedness as buildings constructed on former lakes and wetlands have a higher damage potential during earthquakes due to longer and higher vibrations. Most of the city of Jakarta is built on loose sedimentary covers [4] with potential risk of soil amplification, compaction, and liquefaction in case of stronger earthquakes. Therefore, it is necessary to carefully map the hydrologic and geomorphologic situation before the large spread of urban settlement. The change of urban coverage is visualized by Figure 3 showing the difference map calculated on the basis of the LANDSAT imageries from 1976 to 2001 of Jakarta. It is clearly visible where buildings were constructed on areas of former lakes and wetlands.

Another approach to detect the influence of local site conditions on earthquake damage intensity is the lineament analysis based on SRTM-derived morphometric maps and on LANDSAT data. Lineament analysis based on satellite imageries can help to delineate local fracture systems and faults that might influence seismic wave propagation and influence the intensity of seismic shock (as by causing constructive interference of multiple reflections of seismic waves at the boundaries between fault zones and surrounding rocks). Seismic waves traveling in the subsurface might be refracted at sharply outlined discontinuities as faults, and, thus, arrive at a summation effect that influences the damage intensity. Fault segments, their bends, and intersection are more apt to concentrate stress. The highest risk must be anticipated in junctions of differently oriented ruptures, especially where one intersects the other. Therefore, special attention is focused on precise mapping of traces of faults on remote sensing data, predominantly on areas with distinct expressed lineaments, as well as on areas with intersecting/overlapping lineaments. As ground movements such as liquefaction, lateral spreading, soil amplification, and compaction are important with regard to extended lifeline systems of Jakarta, a more detailed study of subsurface structures is necessary. South of Jakarta SW–NE-oriented lineaments can be detected on SRTM-derived morphometric maps and on LANDSAT imageries by the analysis of the drainage pattern, of linear valleys and hills, and of linear, tonal anomalies (Figure 4).

In the case of a stronger earthquake in this area, the highest shock can be assumed in areas with intersecting larger fault zones. The lineament density map provides an impression of such case.

3.2 Flooding susceptibility of Jakarta

The lowland area of Jakarta has always been naturally subject to regular flooding by the waterways cutting through the plain, such as the Cisadane, Ciliwung, Bekasi, and Citarum Rivers. During the wet season, when these rivers carry down the largest part of their sediment, which in the course of time has grown in quantity because of upstream human-induced eroding developments, the risk of flooding becomes paramount and has large impact [5]. Prolonged periods of rainfall can lead to plain floods that build up over days and can affect large areas of Jakarta, whereas short-lasting, very intensive rainfalls often cause flash-floods. The probability, frequency, intensity, and duration of extreme weather events will increase in Indonesia due to global climate change and due to seawater level rise and will become more common in future. This will require an efficient water management. Flood reservoir management, water steering, and retention systems for the area are necessary. One of the

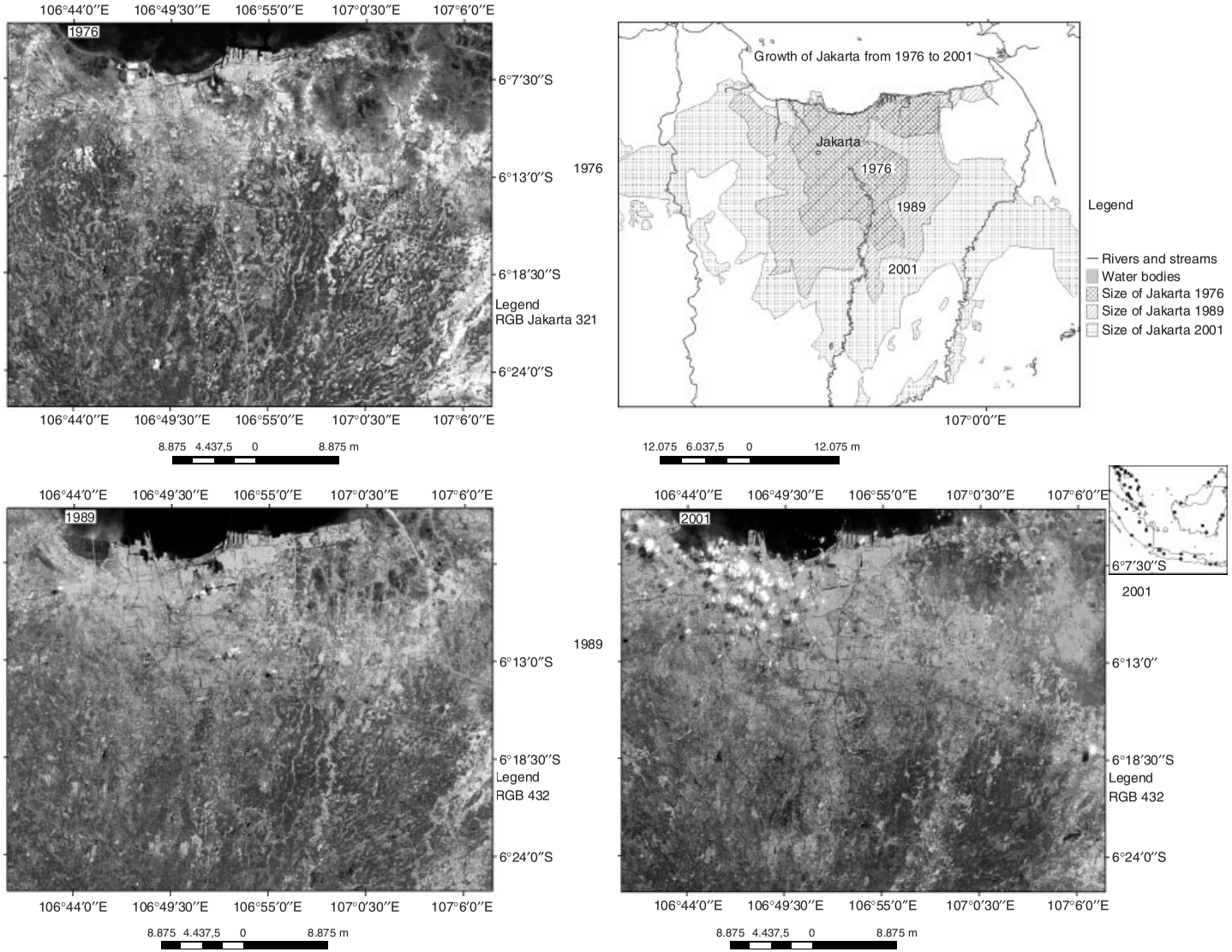


Figure 2. Development of Jakarta from 1976 to 2001 as visible on LANDSAT multispectral scanner (MSS), LANDSAT thematic mapper (TM) and LANDSAT enhanced thematic mapper (ETM) imageries. [Reproduced from LANDSAT, NASA, 1976–2001.]

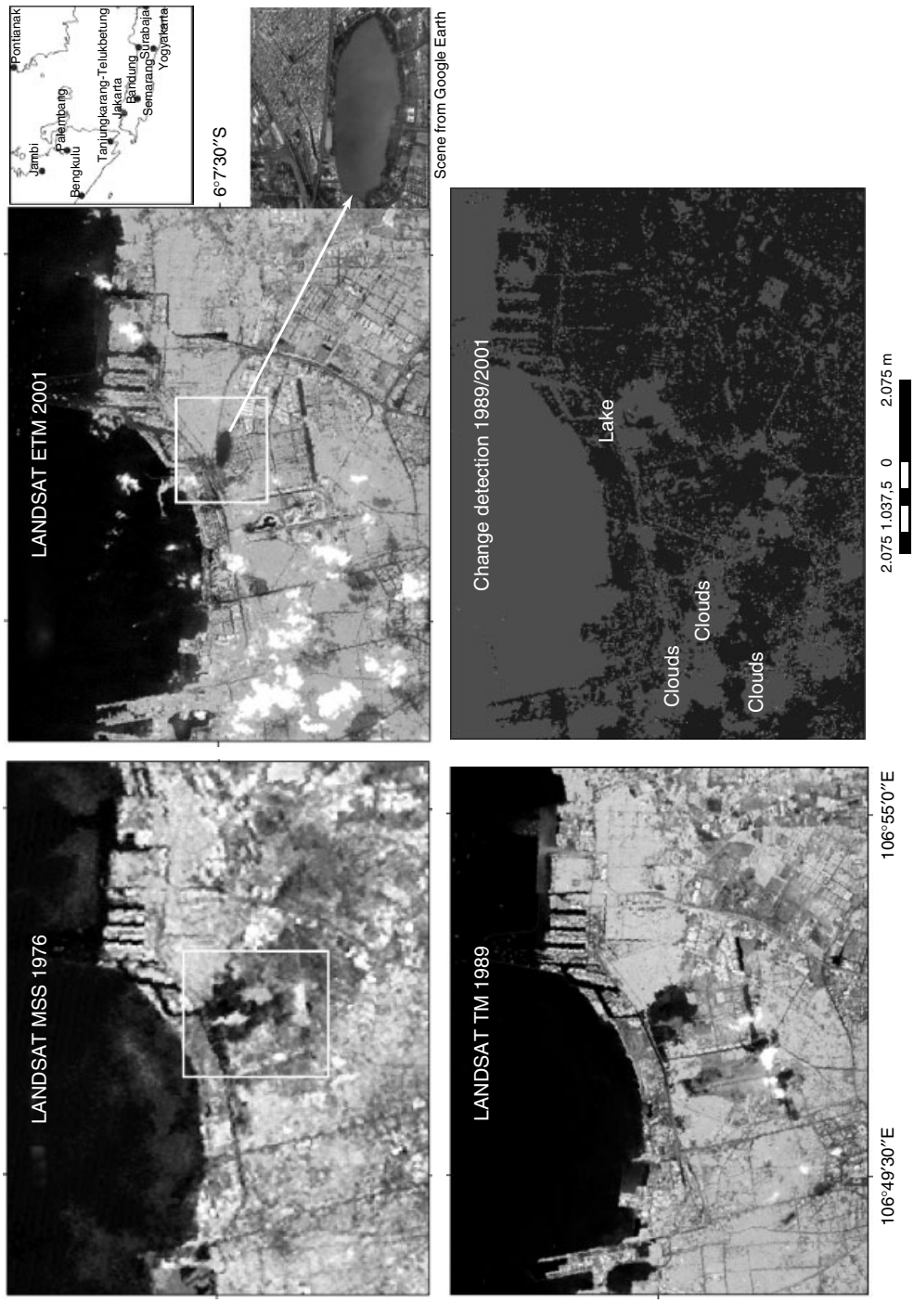


Figure 3. Change detection in Jakarta between 1976 and 2001.

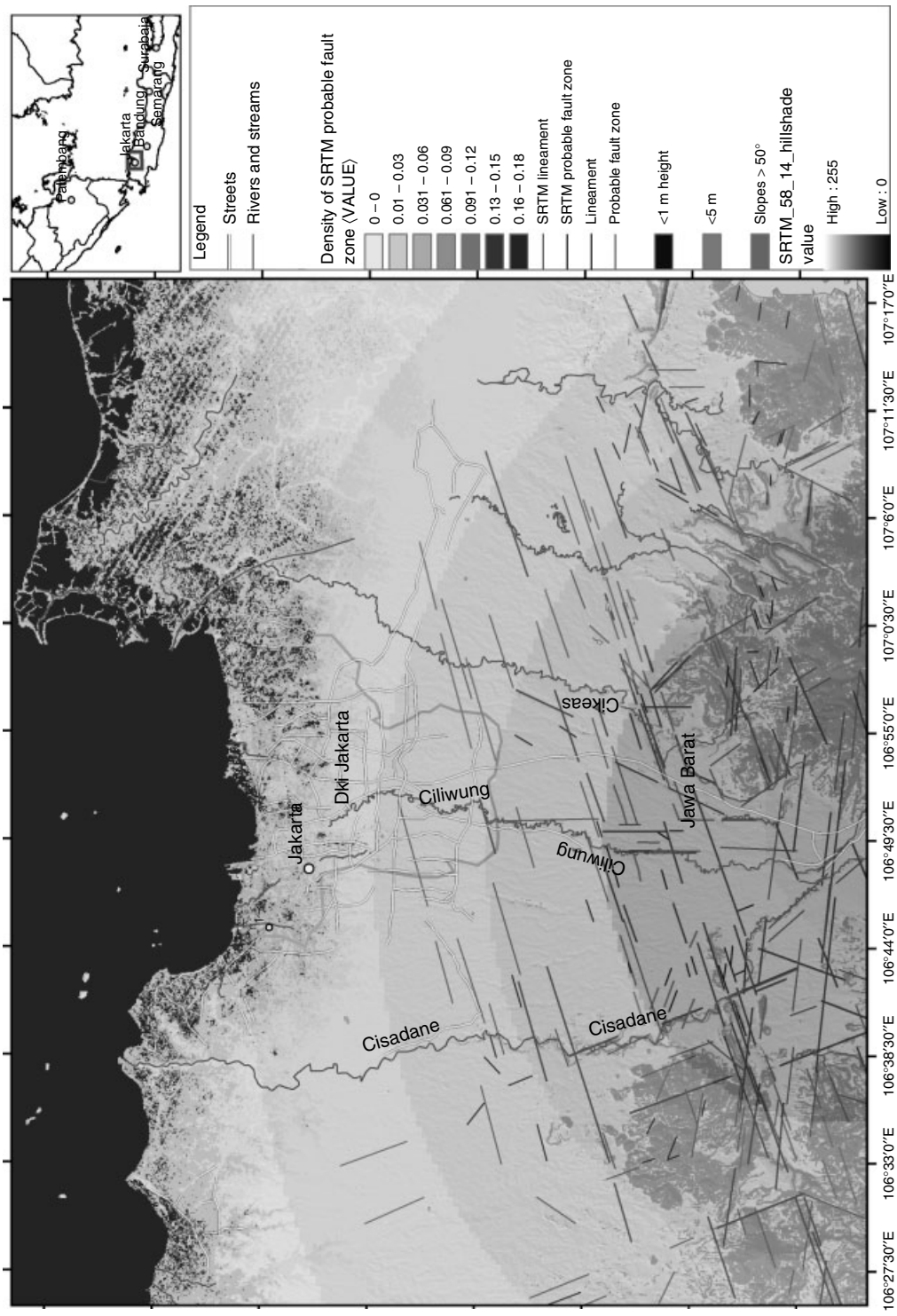


Figure 4. Lineament analysis and lineament density map—detection of linear features in the surface morphology and the drainage pattern, detection of linear tonal anomalies.

problems during these emergencies is to obtain an overall view of the phenomenon, with a clear idea of the extent of the flooded area, and, to predict the likely developments. Height maps derived from the digital elevation data of the SRTM mission indicate the lowest areas. Flow-accumulation maps calculated on the basis of SRTM data help to detect those areas, which are most susceptible to the effects of these events. The flow-accumulation map indicates, by high flow-accumulation values, that Jakarta has a very high susceptibility of being prone to flood in case of intense precipitations (Figure 5).

The GIS-implemented flow-accumulation function is applied to calculate surface runoff and lateral flow toward streams using a two-compartment distributed delay function. The flow-accumulation function routes the flow of water finding a flow direction for every cell of the elevation grid following the steepest paths. Flows from cell to cell on the same flowpath result in a flow-accumulation map in which the value for each cell represents the accumulated flow along a particular path.

The lowlands can be considered to have higher groundwater tables and to be more susceptible to flooding. It is obvious that the city of Jakarta is situated unfavorably, as it gets most of the surface runoff and groundwater flow from the southern hills. Flood-control reservoirs and flood ways retaining and focusing surface runoff are proposed for the area at the foothills and hill forelands as shown in Figure 5. The areas recommended for water-retention measurements as flood reservoirs were chosen on the basis of LANDSAT ETM imageries. However, Jakarta suffers not only seasonal flooding but is also prone to storm surge and tsunami waves. The lowlands prone to storm surge, tsunami flooding, and the consequences of rising seawater level due to global climate warming (Figure 5) need further protection measurements.

3.3 Volcanic activity in Java

One of the many natural disasters that has caused suffering and damage to human lives and nature in Indonesia is the eruption of volcanoes. Indonesia has the largest number of historically active volcanoes (76), and a total of 1171 dated eruptions. Indonesian authorities often need data of volcanic areas, not only to monitor eruptions but also to produce maps and thematic diagrams predicting the potential risk to the surrounding areas. In this field, remote sensing data is used to detect lithological differences, vegetation changes, landuse classification, variations after volcanic events, and the extent and growth of urban areas into endangered areas. One of the most dangerous, well-known Indonesian volcanoes is Merapi in Central Java. It has a height of 2911 m above mean sea level, lying on the boundary of special district of Yogyakarta Province and Central Java Province. Its distance from the capital city Yogyakarta is about 35 km. Merapi volcano is characterized by periodic, dangerous, big eruptions in the range of three to seven years. The Merapi eruptions, i.e., the explosions of the lava dome, consist of hot volcano debris and gases of 900–1200 °C temperature. More than 1.5 million people are working and living in the near surroundings of this volcano, where the danger is even extended to people living along the rivers draining the lower lands. This specific danger is caused by mudflow containing debris and rocks of volcanic origin and transported by rainwater of the often-heavy monsoon rainfall from November through April, which can amount to 40 mm within 2 h. Another danger, which could also threaten the people and nature in the relevant area, is pyroclastic flow that exists in two types. The first type results from the collapse of lava domes in the crater of the volcano. Occasionally, a dome might grow so large that it becomes unstable and collapses into several drainage catchments. The distance traveled by the flow at its extent strongly depends upon the volume of the destroyed lava, gas, pressure, and the slope angle of the flank. The second type results from the collapse of debris that is erupted vertically. Most of the eruptions have low explosivity and the pyroclastic flow usually reaches up to 6 or 7 km from the summit. The velocity of pyroclastic flow can reach up to 110 km h^{-1}. Various data set on past eruptions, administrative data, landuse data, topographical data, geological data, and the satellite images acquired from LANDSAT TM 2001 for the months of April and July have been assembled and provided by GFZ, Potsdam, a German partner institution in a joint Indonesian–German research project, not to mention the meteorological data, i.e., rainfall intensity and topographic map of the study area that is provided from Indonesian authorities. The task is

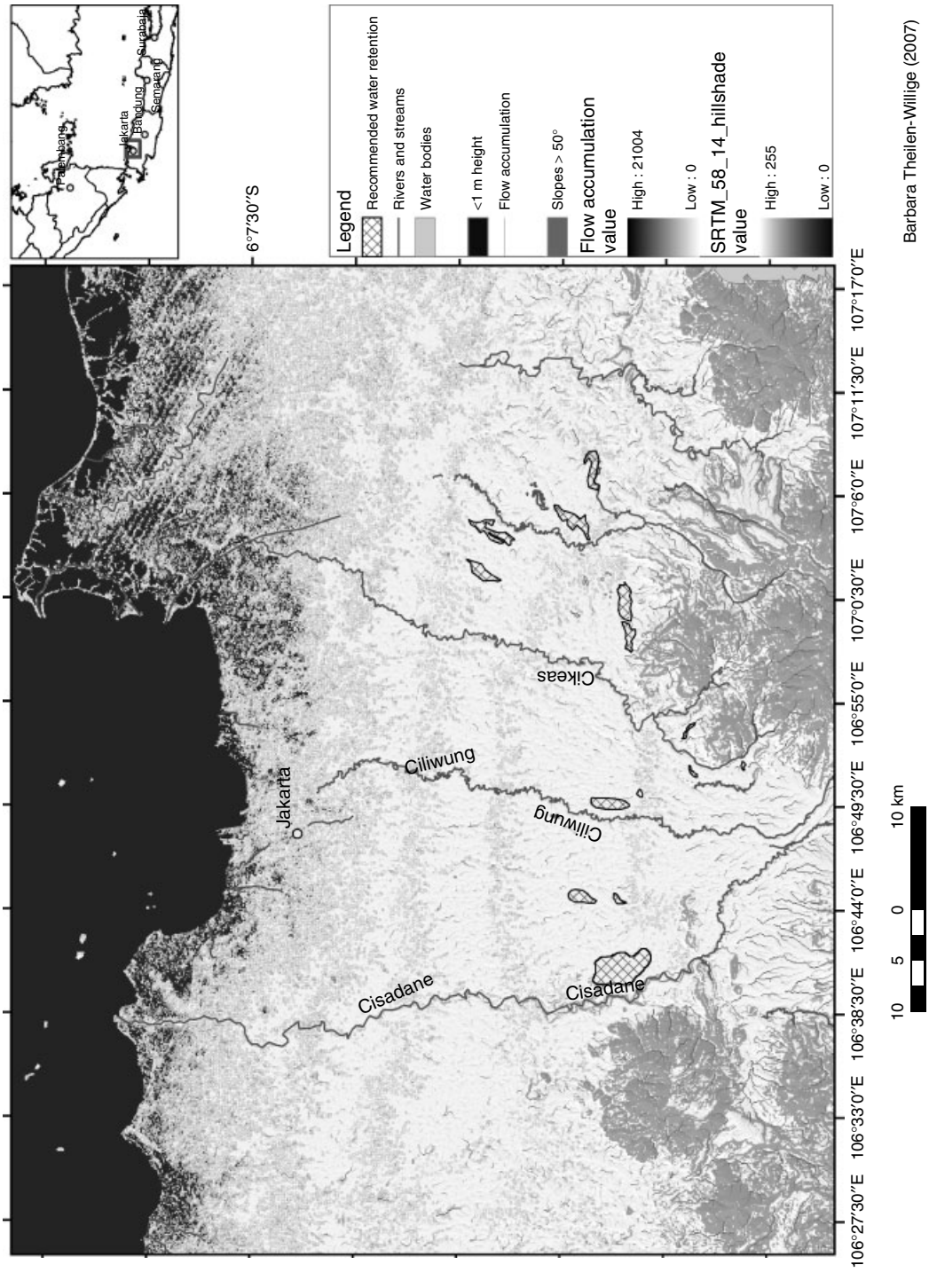


Figure 5. Flow-accumulation map indicating, in Darker grey tones, those areas with the highest flow accumulation.

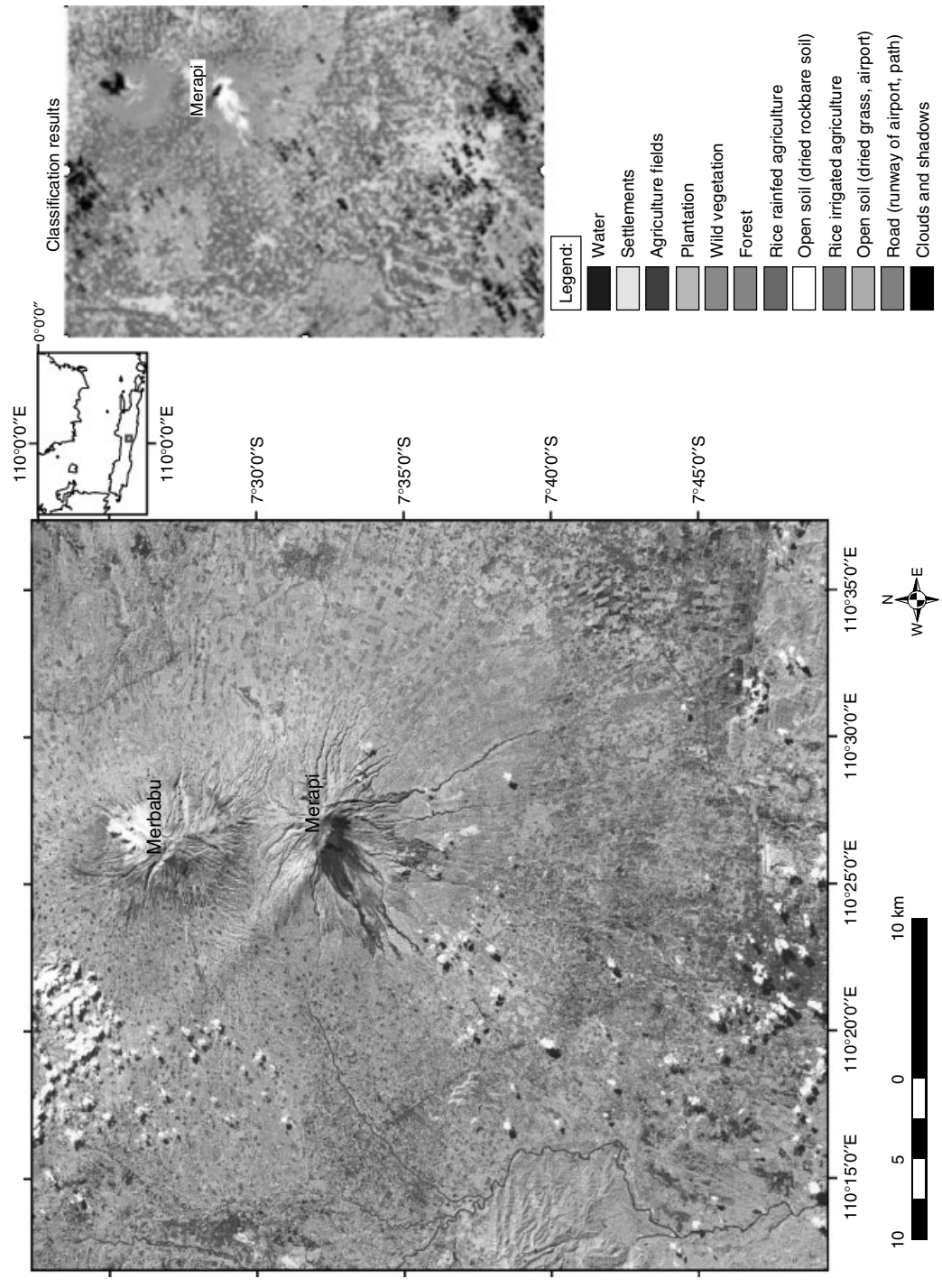


Figure 6. LANDSAT TM 2001 (TM band combination: 5, 4, 3) and classification result [6]. [Reproduced from LANDSAT, NASA, 2001.]

divided into three stages: in the first stage, remote sensing techniques are used in order to process and analyze the satellite images, classify them into several main classes as landuse classification. In the second stage, numerical model of a pyroclastic flow (typical of Merapi volcano eruption) is used, where several parameters are set in the frame of case studies, to stimulate the possible damages that affect the types of landuse within the hazard area. Finally, in the third stage, Arc View GIS technique is used for further analyzing and processing of types of landuse that are affected from the eruption, where hazard damage calculation could be produced (Figure 6). Hence, the estimated degree of damage due to the (assumed) eruption can be derived [6]. The accuracy of the total hazard damage calculation gives an overall impression as to how huge the loss (human lives and properties) will be, if the eruption occurs.

4 CONCLUSIONS

All the above-mentioned issues are important for the understanding of natural hazards, especially when located in very densely populated areas, where the welfare of human lives should be taken into consideration in the first place. The presented methods contribute to the improvement of the risk mitigation efforts at the local level, by applying theoretical knowledge, together with the relevant technology in the reduction of natural disasters that mostly occur in developing nations [6]. The design of a common GIS database structure—always open to new data—can greatly contribute to the homogenization of methodologies and procedures of natural hazard risk management. Meanwhile, the so-called Free GIS software fulfilling the basic GIS requirements, for example, DIVA-GIS, Map Window GIS, SAGA GIS, etc., can be used without costs. Additional Free GIS software is also available for the spatial analysis of DEM data. Basic LANDSAT ETM and SRTM data are provided free of charge for scientific research purposes, for example, by the University of Maryland, USA. Therefore, the use of the remote sensing and GIS technology for natural hazard site assessment and for the elaboration of hazard maps, according to the presented approach, can be recommended as low-cost approach that could be achieved by local communities in every country as a contribution to a GIS database for disaster preparedness [7, 8].

ACKNOWLEDGMENTS

Satellite data:
LANDSAT data: Global Land Cover Facility, University of Maryland, USA http://glcfapp.umiacs.umd.edu :8080/esdi/index.jsp

SRTM data: Consortium for Spatial Information (CGIAR-CSI) http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp

REFERENCES

[1] Gupta RP. *Remote Sensing Geology*. Springer: Berlin, NY, 2003.

[2] Sengara IW, Kertapati EK, Susila IGM. Seismic Hazard Assessment in Denpasar—Bali. Report. Institute of Technology, Bandung (ITB): Indonesia, The Regional Workshop on Best Practices in Disaster Mitigation, Proc., General Paper, 320–328 http://www.adpc.net/AUDMP/rllw/themes/gen-id.pdf.

[3] Steinwachs M. Das erdbeben am 19 September 1985 in Mexiko—ingeneieurseismo-logische aspekte eines multiplen subduktionsbebens. In *Ausbreitungen von Erschütterungen im Boden und Bauwerk*, Steinwachs M, (ed). 3rd annual conference, DGEB, Trans Tech. Publications: Clausthal, 1988.

[4] Geological Survey of Indonesia. *Geologic Map of Western Part of Java, 1: 500.000*, *Second Edition*. Geological Research and Development Centre: Bandung, 1998.

[5] Caljouw M, Nas PJM. Flooding in Jakarta. *The 1st International Conference on Urban History*. Surabaya, 23–25 August 2004, Pratiwo Jakarta/Leiden.

[6] Mulyasari F. *Assessment of Danger to People and Nature caused by Volcanic Eruptions Case Studies at the Merapi Volcano, Indonesia*, Master Thesis. University Karlsruhe: Germany, Resources Engineering Master Program, 2002; p. 89.

[7] Theilen-Willige, B. Remote sensing and GIS contribution to Tsunami risk sites detection in southern Italy. *Deutsche Gesellschaft für Photogrammetrie*. Fernerkundung und Geoinformation—PFG, 2006; Vol. 2, pp. 103–114.

[8] Theilen-Willige B. Emergency planning in northern Algeria based on remote sensing data in respect of tsunami hazard preparedness. *Science of Tsunami Hazards* 2006 **25**(1):3–17, http://www.sthjournal.org/251/willige1.pdf. http://www.sthjournal.org/251/willige2.pdf.

Chapter 149
Gas Turbine Engines

Michael J. Roemer
Impact Technologies, Rochester, NY, USA

1 INTRODUCTION

The history of structural health monitoring for gas turbine engines by most accounts began with the engine original equipment manufacturers (OEM's) in the 1970s and their utilization of engine component design models for tracking critical life-limited components of the engine. Models for components such as turbine inlet vanes/blades based on thermomechanical fatigue were used to track and trend life usage based on calculated stress/strain histories formed using onboard sensor readings that could be correlated (using gas path models) with turbine inlet temperature (*see* **Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance**). On the basis of the estimated turbine inlet temperatures and appropriate heat transfer models, estimates of actual metal temperatures and associated stress/strain achieved during operation could be determined. This was used as input to associated fatigue calculations, and eventually damage accumulation trending using Minor's rule was performed off-line and used as an initial process of structural health management for life-limited components.

During a similar time frame, aerothermal performance models were also being used with the gas path measurements being monitored on the engine to perform estimates of its efficiency. In many cases, only the performance margins (e.g., exhaust gas temperature—EGT) were used to infer a performance-based health indicator and determine when an engine was due for an overhaul. Hence, tracking and trending of an individual engine's performance margins became part of the off-line procedures used to manage the health of many engines. Hence, from both a mechanical and performance perspective, the current state of this technology was structured as isolated, independent health management approaches based on trending the damage and performance margins of the engine separately, relying upon physics-based models, simple

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

trending, historical information, or inspection and maintenance results in a fragmented approach.

Later in the 1980s, as information and data coming from the engine became more available based on regular downloads from engine controllers, it became more plausible to perform automated trending and analysis that could enable more consistent results for their respective applications. With the wide-scale application of digital computers and commercial products that allow for data monitoring and analysis of various types, advanced fault detection and prediction technologies became more easily implemented and cost-effectively linked to the necessary data for providing real-time assessments of engine health. Technologies such as advanced signal processing, probabilistic component life analysis, neural network classifiers, fuzzy logic decision support analysis, and Bayesian networks were just a few of the algorithmic approaches being implemented to provide for better automated equipment health state awareness and prediction (*see* **Aerospace Applications of SMART Layer Technology**).

As such technologies are currently being implemented today, we now have the data and information necessary to eliminate many of the engine faults and failures that contribute most significantly to downtime and prolonged maintenance activities. Hence, these assets are now becoming more economically viable over their life cycle and providing operators with increased profits and/or output. Therefore, increased applications of enhanced diagnostic and prognostic algorithms that can detect and predict, within a specified confidence bound, time-to-failure of critical engine faults and components have the potential to provide many benefits including.

- improved safety associated with operating and maintaining gas turbine engines;
- reduced overall life cycle costs (LCC) of engines from installation to retirement;
- ability to optimize maintenance intervals for specific engines or fleets of engines and prioritization of tasks to be performed during the planned maintenance events;
- increased uptime/availability of all engines within a fleet;
- provides engineering justification for scheduling maintenance actions with corresponding economic benefits clearly identifiable.

2 INTEGRATED ENGINE HEALTH MONITORING OVERVIEW

Figure 1 provides an illustration describing how fault detection, isolation, and prognostic technologies are

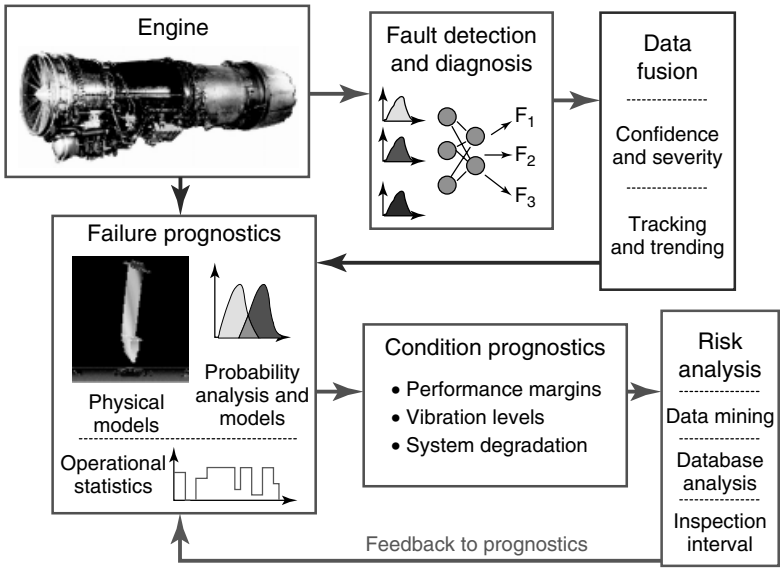


Figure 1. Integrated engine health monitoring.

currently being implemented within many military engine health monitoring programs. The integration of such technologies typically begins with validated sensor information currently measured on the engine being fed directly into the diagnostic algorithms for fault detection/isolation and classification. The ability of a diagnostic system to fuse information from multiple diagnostic sources together to provide a more confident fault assessment is emphasized along with a system's ability to estimate confidence and severity levels associated with a particular diagnosis. In a parallel mode, the validated sensor data and real-time current/past diagnostic information are utilized by prognostic modules to predict future time-to-failure, failure rates and/or degraded engine condition (i.e., vibration alarm limits, performance margins, etc.). The prognostic modules typically utilize physics of failure, stochastic models taking into account randomness in operation profiles, extreme operating events, and component forcing. In addition, the diagnostic results can be combined with past history information to train real-time algorithms (such as neural networks or real-time probabilistic models) to continuously update the projections on remaining life. A few selected and relevant approaches and algorithms for performing component prognostics are described in this article.

Once predictions of time-to-failure or degraded condition are determined with associated confidence bounds, the prognostic failure distribution projections can be used in a risk-based analysis to optimize the time for performing specific maintenance tasks. A process that examines the expected value between performing maintenance on an engine or component at the next opportunity (therefore reducing risk but at a cost of doing the maintenance) versus delaying maintenance action (potential continued increased risk but delaying maintenance cost) can be used for this purpose. The difference in risk between the two maintenance or operating scenarios and associated consequential and fixed costs can then be used to optimize the maintenance intervals or alter operational plans.

3 SENSOR VALIDATION

Sensor malfunctions have traditionally plagued many engine health monitoring systems with false alarms and an inability to distinguish between real engine-based faults and sensor faults. Hence, a necessary "front end" of any engine health monitoring system should have the capability to detect and classify specific sensor failure modes and distinguish them from actual engine faults. One particular approach that has been implemented within several gas turbine engine test cells is based on utilizing generic signal-processing techniques such as digital filtering, cross-correlation, and coherence coupled with an intelligent classifiers including a fuzzy logic rulebase to diagnose malfunctioning sensors.

The generic, signal-processing-based approaches are capable of detecting signal anomalies such as spiking, intermittent signal loss, cross talk, and clipping using modified signal correlation and coherence algorithms. Digital filtering is used to aid in the detection of spikes, noise, intermittent loss of the signal, and other anomalies, which manifest themselves by a rapid change in signal magnitude. The high degree of overlap between the two methods in detecting the most common signal faults provides a higher degree of confidence when an anomaly is flagged and will therefore minimize false alarms.

The digital filter algorithm utilizesa high-pass Butterworth or similar filter with a 3-Hz minimum cutoff frequency that is selected based on the physical response of the engine sensors themselves. Figure 2 shows how the standard deviation of a filtered signal is used to detect an anomalous signal. The top left window shows a noisy, time-domain thermocouple signal and the top right shows a "normal" thermocouple signal. The filtered version of both these signals is shown below with the variance of the filtered signal used as the feature, which will flag a signal anomaly. Although this technique is a simple one, its effectiveness and generic implementation capability makes it a practical choice for instrumentation validation.

The cross-correlation and coherence algorithms attempt to find statistically significant shifts in the quantitative relationships between signals in both the time and frequency domains. These techniques are implemented utilizing a moving time-domain window that continuously computes the correlation and coherence functions of each of the measured signal pairs. Therefore, any signal whose previous $(t - n)$ window of time-domain data becomes consistently less correlated with the current (t) windows can be identified

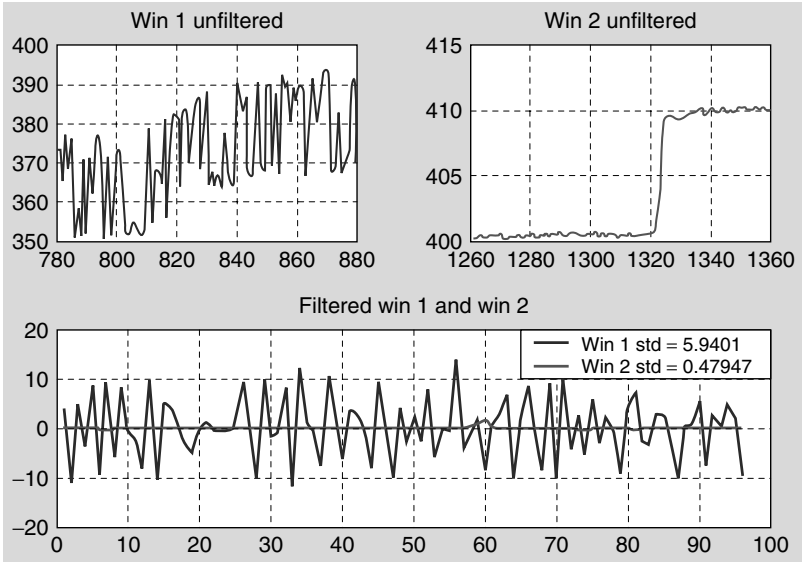


Figure 2. Digital filter detecting thermocouple noise.

as a sensor-related issue and not engine performance driven. The cross-correlation and coherence functions are as calculated as follows:

$$R_{x_1x_2}(0) = \frac{\sum x_1(n)x_2(n)}{\sqrt{\sum x_1^2(n)\sum x_2^2(n)}}$$

normalized zero-lag correlation (1)

$$\gamma^2_{X_1X_2}(\omega) = \frac{\left|\sum R_{x_1x_2}(\tau)\mathrm{e}^{-i\omega\tau}\right|^2}{\left(\sum R_{x_1x_1}(\tau)\mathrm{e}^{-i\omega\tau}\right)\left(\sum R_{x_2x_2}(\tau)\mathrm{e}^{-i\omega\tau}\right)}$$

coherence function (2)

Though only two signals are represented in equations (1) and (2), the correlation and coherence functions can be applied to any number of signals at once. The data samples utilized within these generic techniques are typically small and dependent upon the sampling frequency and the amount of time over which data is gathered. This may require the use of the "small sampling theory". In gas turbine engine monitoring applications, statistics are typically obtained from the coherence and correlation functions a priori and then the level of significance is determined by utilizing a t-test on the statistic of interest, as shown in equations (3) and (4). In many cases, the mean value of the samples is used to determine if a significant shift in the statistic has occurred.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \tag{3}$$

where

$$\sigma = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{\nu}} \tag{4}$$

$\bar{X}_1$ the mean value of the first sample

$\bar{X}_2$ the mean value of the second sample

N_1 the size of the first sample

N_2 the size of the second sample

s_1 the standard deviation of the first sample

s_2 the standard deviation of the second sample

$\nu = N_1 + N_2 - 2$ the degrees of freedom

Figure 3 shows some selected results from two of the signal-processing techniques that were used to detect sensor faults on data taken from an engine test cell. Note that multiple sensor spikes, signal noise, and a signal "drop out" are all present within

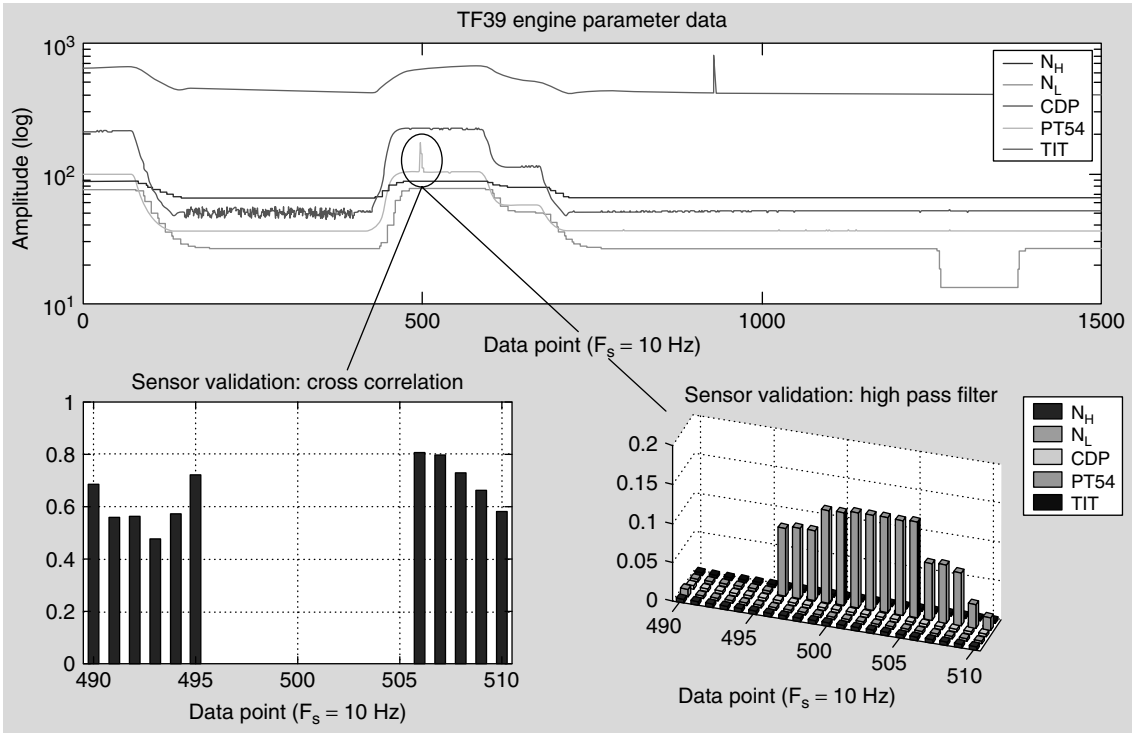


Figure 3. Results of cross-correlation and digital filtering.

this data set. As illustrated, the cross-correlation approach, shown on the bottom left, identifies an anomaly when its value drops to zero. The digital filter approach triggers an anomaly when the signal's standard deviation is high (coherence results are not shown). Both approaches indicate that the anomaly occurred and then returned to normal.

Finally, when the engine sensor data is processed by both the generic filtering and correlation techniques, the individual algorithm outputs can be fused together in a postprocessing stage to provide a more confident sensor fault detection. A Dempster–Shafer fusion algorithm can be implemented for this purpose, which accounts for uncertainty in the conditional probabilities derived by the various techniques being processed. The Dempster–Shafer methodology hinges on the construction of a statistical set, called the *frame of discernment*, in which a set size will contain every possible hypothesis. Every hypothesis (A) has a belief (lower bound) denoted by a mass probability $(m(A))$ and a plausibility (upper bound) denoted by $1 - m(A')$. Beliefs and plausibilities are combined in the following manner. A full example of the Dempster–Shafer fusion technique is provided in [1].

$$\text{Belief}(Hn) = \frac{\sum_{A \cap B = Hn} m_i(A) \cdot m_j(B)}{1 - \sum_{A \cap B = 0} m_i(A) \cdot m_j(B)} \tag{5}$$

4 ENGINE THERMODYNAMIC PERFORMANCE DIAGNOSTICS

The sensed parameters associated with the engine gas path and corresponding ambient conditions are required by performance analysis algorithms to detect and isolate performance faults. Pressure, temperature, and flow readings at different points within the gas path are required, as are any bleed flows, fuel flow, rotor speeds, and any other engine conditions that must be accounted for in performance calculations. Figure 4 shows the locations and their

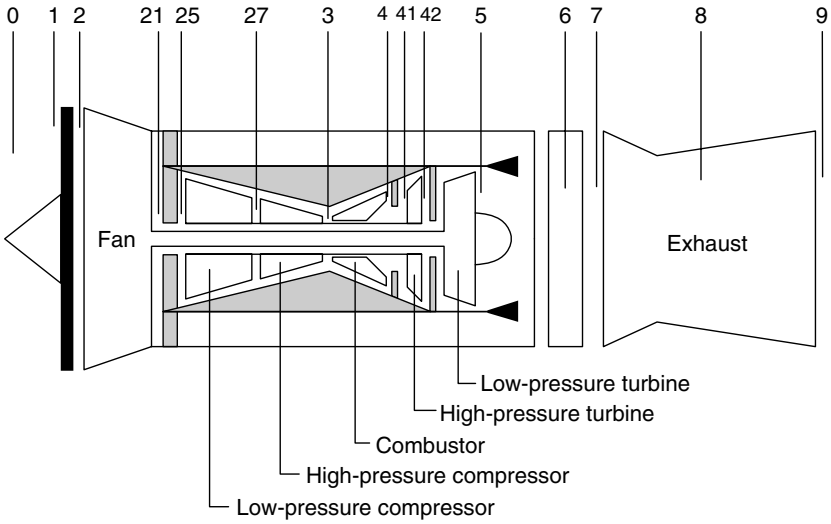


Figure 4. Engine gas path designations.

associated designations for jet engine pressure and temperature measurements. A specific engine gas path analysis model is typically run off-line under off-design conditions to develop a matrix of "diagnostic error patterns" expected under ideal engine degraded conditions. Measurement and modeling uncertainties are developed based on the variances in the modeling and measurement acquisition processes.

Using the gas path performance measurements, statistical engine parameter curves can be developed for specific engines so that comparisons can be made between current conditions and some baseline or normal condition. Measurement uncertainties can be calculated based on this recorded data. The middle plot shown in Figure 5 illustrates the corrected performance curve of the compressor discharge pressure as a function of N2 speed at pseudo-steady-state conditions. The "bands" around the curve represent the 1-sigma distribution (1 SD) levels. The left side of Figure 5 shows a 3D representation of how the percent deviation in compressor delivery pressure (CDPC) changes with speed.

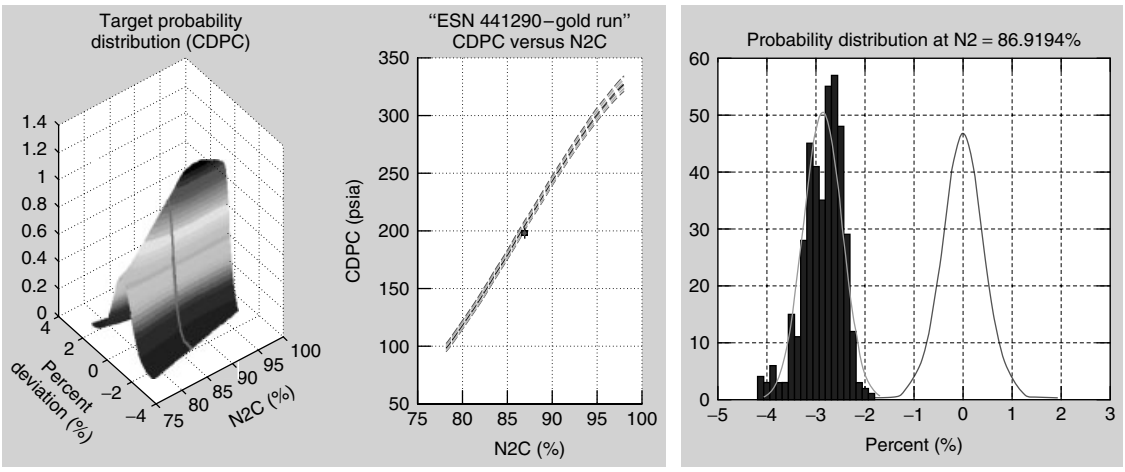


Figure 5. Statistical engine signature curves and shift detection.

Utilizing confidence intervals for the discovery of a statistically significant trend away from "normal" engine operation is a rigorous way to assess engine anomalous events. Utilizing a full set of engine signature curves described above is well suited for this type of analysis. The detection of an anomalous event will trigger the start of a more comprehensive diagnostic treatment of the detected error pattern. The key is to be capable of determining whether the mean measured parameter values have shifted with a high degree of confidence. The confidence interval approach is a well-proven statistical method for performing this calculation. The t-test is another commonly used technique to determine if a parameter has shifted appreciably from its performance curve.

In the right side of Figure 5, the distribution centered about 0% deviation is the compressor discharge pressure associated with specific engine. The histogram, with a distribution centered about -3% deviation, was acquired from a degraded engine. Clearly, this distribution is different with a high level of statistical significance, over 99% in fact. As a result, it can be confidently stated that the current engine condition has a lower compressor delivery pressure (CDP) than the baseline engine data. This type of analysis, combined with all of the other relevant parameters, is used to determine if the overall compressor efficiency of the current engine condition was lower than that of the baseline.

Next, the performance diagnostic approach relies on gauging the proximity of ALL the current system deviations to known performance faults based on the gas path analysis (GPA) model. A multiparameter, probabilistic classifier technique has been shown to be capable of identifying degraded performance in propulsion systems [2]. This approach requires that sufficient sensor information be available to assess the current condition of the system in terms of shifted trends in parameters from a baseline condition. Modeling and measurement uncertainty is accounted for with this technique utilizing the distributions on the current parameter shifts and model-based fault conditions. While a physical model, such as a gas path analysis or control system simulation, is beneficial, it is not a requirement for this approach to work. An alternative to the physical model is built-in "expert" knowledge of the fault condition.

This generic, probabilistic-based diagnostic classifier involves assigning non-normal or normal probability density functions (PDFs) to performance error patterns associated to known faults in N-dimensional space. Similarly, the current error exists as a PDF in the parameter space as well. The probability that the current condition (C, measured parameter shifts) may be attributed to a given fault (F, identified known fault conditions) is determined by the "overlap" (i.e., multidimensional integration) of their respective joint probability density functions. Figure 6 shows how this is done in two-dimensional parameter space. If C and F can be assumed to be normally distributed (not a necessary assumption, however), the probability of association (p_a) with a given fault condition F can

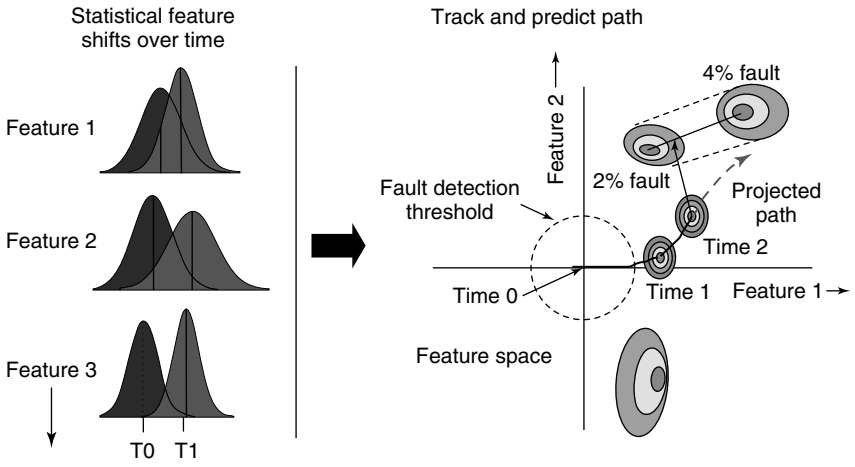


Figure 6. Multiparameter diagnostic technique.

be found using

$$p_a = 2\Phi\left(-\frac{\overline{F-C}}{\sqrt{\sigma_f^2+\sigma_c^2}}\right) = 2\Phi(-\beta) \qquad (6)$$

where:

$\bar{F}, \bar{C}$ = the mean of the distributions F and C, respectively

σ_f, σ_c = the standard deviation of the F and C distributions

The function $\Phi(\,)$ is the standard normal cumulative distribution and the β is denoted as the reliability index. The β represents the Euclidean distance between the current conditional distribution (C) and a given fault distribution (F). Hence, this approach performs diagnostics by evaluating the likelihood of the current conditions to known fault conditions and prognostics by extrapolating a fault-weighted, evolutionary path.

Figures 7 and 8 provide an example of the evolution of a performance error pattern as an engine's performance degrades over time. In the scenario shown in Figure 7, the PDF of the current error pattern initially evolves toward a 2% high-pressure compressor (HPC) efficiency fault. This is also shown in Figure 8 from the fact that from $T = 0$ to $T = 3$ the Euclidean distance between the current PDF and the HPC fault gets smaller. However, as time goes by, the current condition evolves toward, and eventually past, the HPC fault. From Figure 8, at $T = 5$ the current PDF is closest to the 2% high-pressure turbine (HPT) efficiency fault. Figure 7 illustrates that the engine's degradation has indeed evolved beyond association with the HPC fault to high association with the HPT fault. In this example, the final position rests at 9.98% association with the HPC fault and 22.8% association with the HPT fault.

The final step in the performance diagnostic process involves mapping the measured parameter deviations to the modeled engine faults. Hence, this performance model or alternatively an OEM DEC model is used to generate performance error patterns or diagnostic scalars used for fault isolation. As illustrated in Figure 9, the diagnostic scalars are simply the differences between the model estimate and the measured values of key performance parameters shifts (Wf,N2,T45,P25,T25,T3,P3) at referred conditions. When these error patterns are generated,

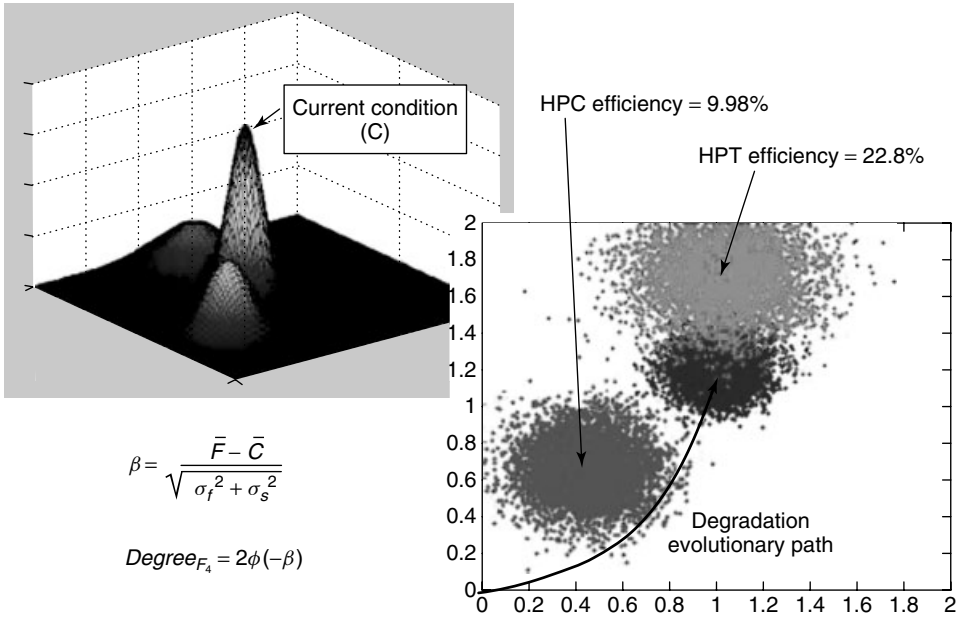


Figure 7. Degree of fault association.

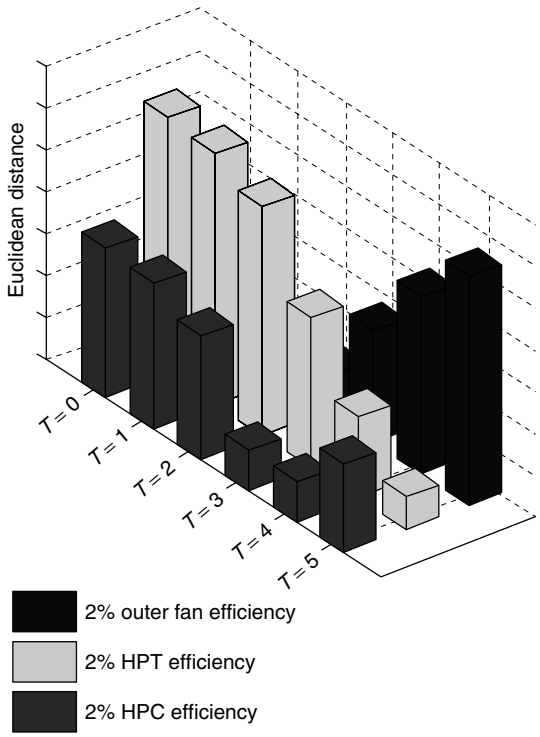


Figure 8. Performance fault evolution.

the next step is to consider the influence of root cause performance faults on the key performance parameters. A set of root cause performance faults for typical engines are fan efficiency degradation, low-pressure compressor (LPC) efficiency, HPC efficiency, HPT efficiency, low-pressure turbine (LPT) efficiency, 2.5 bleed flow, 2.9 bleed flow, discharge area, and stator vane misrigging.

Different severities of these faults will create different performance parameter error patterns that must be contained in an "engine-specific baseline characteristics" database. Although actual engine faults would be ideal for producing the fault error patterns, simulating them with the engine model a priori is typically done. With the error pattern calculated, the database of known faults must be autonomously compared with the current delta scalar pattern to enable real-time performance assessments. In reality, the data-driven error pattern will not exactly match any of the calculated (ideal) error patterns, so the most likely root causes must be determined. There are potentially many methods for performing this classification task. A few of these include least squares pattern match algorithms, Kohonen map/neural network classification and the previously described probabilistic classifier technique, shown in Figure 9. These error pattern classification techniques have already been shown to be capable of identifying degraded performance in propulsion systems [3]. Modeling and measurement uncertainty is accounted for with this technique utilizing the distributions on the current parameter shifts and model-based fault conditions.

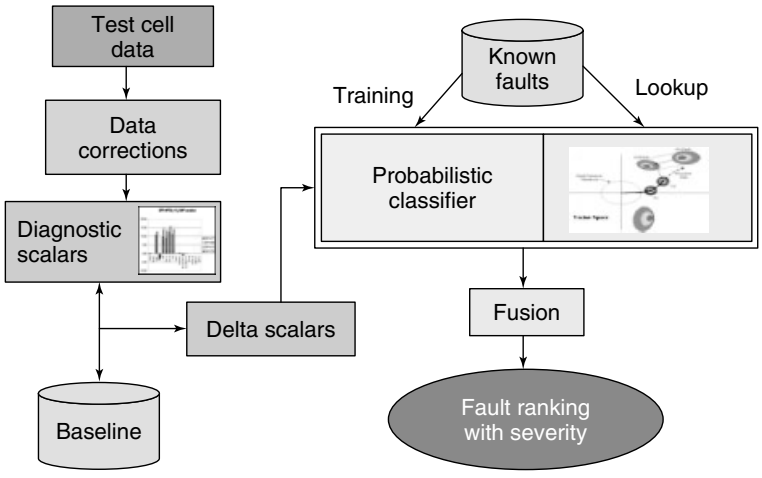


Figure 9. Performance diagnostic process.

5 ENGINE VIBRATION DIAGNOSTICS

Typical engine vibration monitoring systems utilizes piezoelectric-based accelerometers to measure the dynamic response of the engine. One of these accelerometers is often located on the compressor rear frame, while another is typically on the turbine rear frame. Vibrations from various sources, such as engine core rotor, fan rotor, hydraulic pump, etc., are detected by the transducers and converted into electrical signals. The signals are then routed to a vibration signal conditioning analyzer where the signals are amplified and antialiased filtered. If the system is functioning properly, the signal is typically modulated by primarily two components that are proportional to the fan and core speeds.

Real-time assessment of mechanical faults (i.e., bearing, rotordynamic, and structural) based on analysis of the vibration signatures at specified locations on the engine has been developed using feature-based diagnostic techniques. Domain knowledge associated with particular vibration fault frequencies, fixed frequency ranges, per-rev excitations, and structural resonances are extracted from the vibration spectrums acquired from the engine. These spectrums are used to develop a knowledge base from which fuzzy logic membership functions and an associated rulebase are developed. An example of a generic fault matrix and feature extraction process is illustrated in Figure 10 for a typical waterfall vibration plot.

The functionality associated with a feature-based vibration diagnostics capability is primarily based on the development of two particular analysis approaches. These approaches can best be described using the two plots in Figure 11, which illustrate the following: (i) waterfall plot of vibration spectrums (0–1000 Hz, typically) over several different core engine speeds and (ii) tracked-order plots of 1XRev through 2XRev amplitudes for all engine rotors plotted as a function of core speed. Examples of these two plots are shown for a large military engine.

The amplitude associated with the waterfall plot is typically in mils (Pk–Pk), but can also be given in velocity units of (cm/s Pk). Once these measurements are obtained, plots of the transient vibration waterfall are stored and analyzed for rotordynamic health. Using the data set of vibration measurements from a specific engine or group of engines, a table of spectral features and corresponding "normal" amplitude bands are developed for detecting anomalous

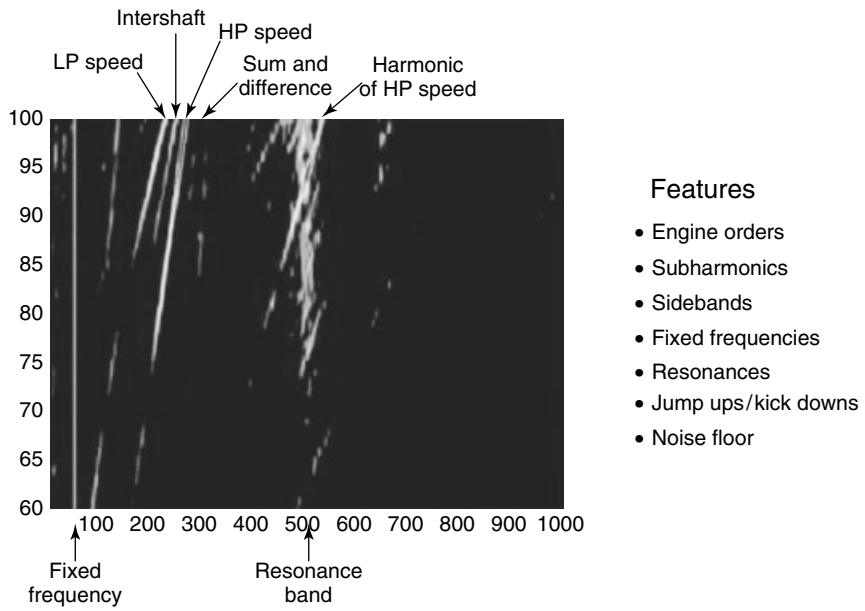


Figure 10. Feature extraction from vibration analysis.

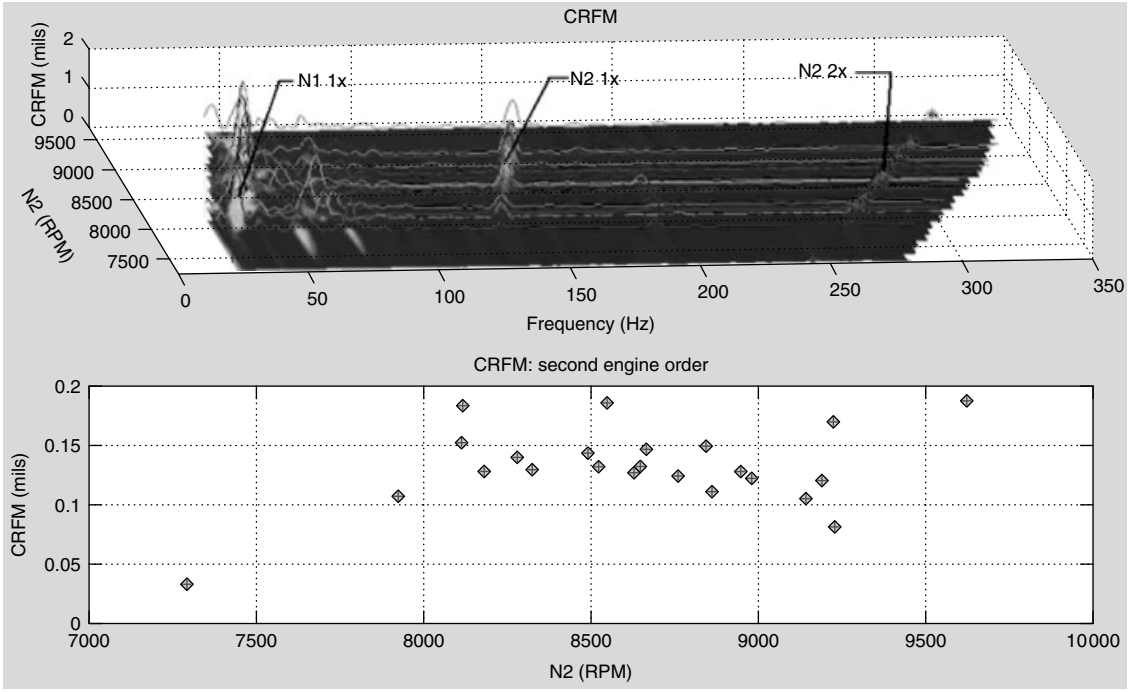


Figure 11. Waterfall and tracked-order plots for TF39 engine.

vibration signatures and associating them with particular diagnostics where applicable. In Figure 11, one can easily see that vibration peaks are associated with the core engine and fan speeds.

The vibration analysis schemes also utilize a shape-based statistical analysis of the tracked orders to detect and diagnose mechanical faults. The combination of these techniques allows for a more robust and sensitive diagnostic capability. Figure 12 shows a particular shape of a high-pressure (HP) shaft tracked-order and ± 2 SD determined from testing of multiple engines. The bold line in Figure 12 shows a simulated tracked order of an engine with a different structural resonance. In this example, which is for illustration purposes, an amplitude level band on the tracked order would not detect a problem; however, a statistical analysis of the shape of the tracked order would be able to detect a fault.

Owing to the varying amount of data and diagnostic information that exists in regard to vibration anomalies that occur on particular engines, many programs propose to use a master database to collect these vibration signature features associated with the waterfall plot and tracked-order statistics. The engine-specific vibration signatures will be used to develop a statistical description of engine shaft tracked-order amplitudes, broadband amplitudes, and nonsynchronous vibration amplitudes. The statistical description can then be utilized directly in the vibration anomaly detector algorithms using normal distribution membership functions within the fuzzy logic decision process.

6 ENGINE LIFE-LIMITED COMPONENT PROGNOSTICS

A prognostic model for an engine's life-limited component such as a hot section blade and nozzle must have ability to predict or forecast the future condition or risk of failure given the past operating environment and some prediction on how the engine will be operated in the future. To be clear, oftentimes the term prognostics is further defined by either failure or condition prognostics. Failure prognostics often refers to the continuous accumulation

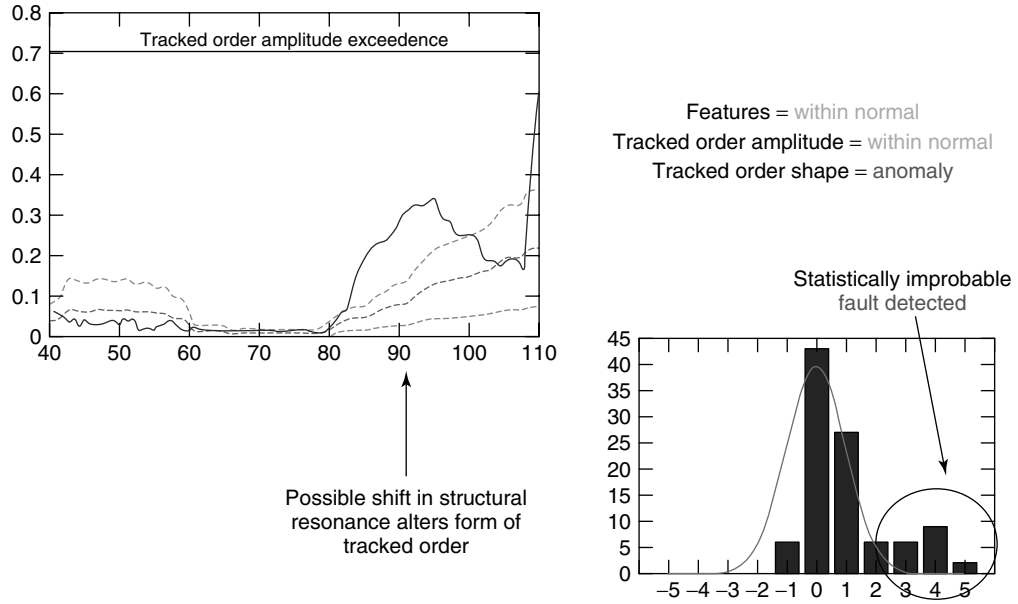


Figure 12. Tracked-order analysis.

of damage and/or life on components or systems of components, with or without the presence of any identified faults. Components governed by mechanical wear and failure often fit into this category (i.e., prediction of crack initiation without the presence of a fault detected). In contrast, condition prognostics is most often associated with a fault being diagnosed prior to a vibration or performance-related limit being exceeded. A detected fault must be isolated and assessed for severity so that the remaining useful life can be determined. This useful life is defined by the operating time between detection and an unacceptable level of degradation.

For life-limited components, failure prognostics is applied using a physics-based stochastic model that typically incorporates mechanical (finite element) or thermodynamic (through-flow model) deterministic models as their basis. The probabilistic procedure for addressing inherent modeling uncertainties must be built into these models using statistical distributions of the parameters that most directly affect the component life-limiting factors. Some of these factors include the material properties, dynamic forcing, and process variability. The distribution on the current remaining life in a component life prediction may be determined by calculating all possible combinations of these life-limiting factors in a stochastic process given past operating conditions. Operating hours can be statistically analyzed, trended and projected into the future to provide the prognosis of remaining life. More advanced stochastic models that represent failure-mode uncertainties, projected operational parameters, and rare/random events can be used to help predict failure-mode propagation. This physics-based model should be calibrated using in-service data to clearly reflect the root cause of the in-service failure-mode experiences. In the case where a finite element model can be used (gearing, blading, impellers, or rotors, for example), crack initiation regions should agree with any in-field experience and inspection data. More empirically analyzed components such as bearings should have clearly identified relationships between diagnosed fault severity and life consumption.

6.1 Example of a stochastic physics-based model

Sophisticated fracture mechanics and damage accumulation analysis have shown that accelerated crack nucleation and microcrack formation in components

can occur due to start-ups and shutdowns, transient load swings, higher than expected intermittent loads, or defective component materials. More commonly, normal wear causes configuration changes (loose fit of assembled parts, work-hardened surfaces, and reduced structural section areas) that contribute to increased or unexpected dynamic loading conditions. High cycle dynamic and transmission loads can lead to micro-crack incubation and formation of micro-cracks at the material grain boundaries in stress concentrated regions (especially between hardened surfaces and softer subsurface material interfaces, and at acute changes in component material geometry). The majority of crack growth evolves in a subcritical propagation process of crack tip blunting, unstable crack formation, and crack elongation. As supercritical loading in the cracked material region is approached, growth accelerates resulting in material dislocation and detachment. Subcritical crack evolution is highly dependent on a component's material, geometry, loading conditions, and the particulars of the unique component crack growth cycle. This kind of failure-mode knowledge is often times overlooked in determining the potential usefulness of a particular prognostic or diagnostic algorithm.

The available time to take corrective or compensatory actions during specific periods of microcrack incubation, formation, and subcritical propagation in the material of a faulted component must be considered. On the basis of this understanding, either of two beneficial actions could be taken: a corrective one to perform maintenance to repair or replace the part, or a compensatory one to reduce system operational loads to extend the life of the faulted part. The informed decision exists only if the diagnostic/prognostic system has the ability to detect that the fault exists, isolate it to the specific component, and assess its severity in a timely manner.

A stochastic physics-based model of a turbine blade is used to describe the modeling approach described above and is shown below in Figure 13. Although each component prognostic modeling procedure is different based on the failure modes being predicted, a process that utilizes the raw, database and processed diagnostic data through a physical-based model is still applicable.

The factors and associated level of uncertainty that most directly affect the remaining useful life on a component must be identified in this physical model. One of these factors specific to a turbine blade is

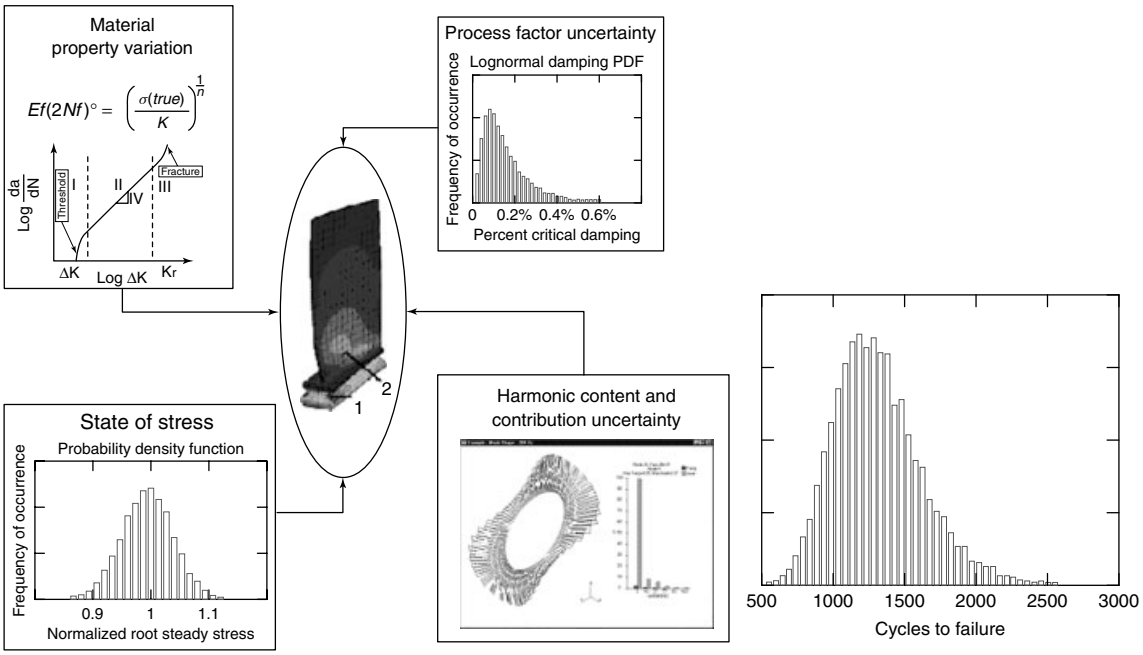


Figure 13. Physics-based stochastic model.

the steady stress at the critical locations in the root region. The uncertainty associated with this steady stress due to variations in the operating environment, temperatures, manufacturing tolerances, etc., must be accounted for with a statistical distribution computed by testing and field service data. Another factor critical in predicting turbine blade life is the dynamic stress as a function of the uncertainty in harmonic excitation and response characteristics. The strength or resistance capability of the material must also be considered as a function of the uncertainty in material properties.

In an effort to describe the process of prognostic modeling for a specific component failure mode (only one aspect of this turbine blade model), the low-cycle fatigue (LCF) fatigue life for the root location that experiences stress cycling in excess of the material's yield strength is described in equation (7):

$$Nf1_L = \frac{1}{2}\cdot\sigma_L(\text{true})^{\left[\frac{1}{(n-c)}\right]}\cdot K^{\left[\frac{1}{(n-c)}\right]}\cdot Ef^{\left(\frac{-1}{c}\right)} \quad (7)$$

All of the parameters involved in calculating LCF life have levels of uncertainty associated with them and are therefore given as probability distributions that may or may not be Gaussian. The distributions are combined using a Monte-Carlo simulation. The Monte-Carlo simulation is an automatic process that randomly selects thousands of different values from each of the life-limiting factor distributions. Over the entire simulation, the randomly chosen values are combined to generate the distribution of a parameter that may have been very difficult or impossible to calculate in a strict analytical sense. The result of the simulation is also shown in Figure 13.

The damage due to the LCF may be given by a nonlinear damage accumulation rule proposed by Halford at NASA Langley [4]:

$$\text{Damage} = \left(\frac{n_1}{Nf1_L}\right)^{r_1} \quad (8)$$

The complete turbine blade prognostic model must further account for the other failure modes at other critical locations on the blade; however, this is outside the illustrative scope of this article. The net result, however, is the path that has been taken to determine the current component life consumption as shown in Figure 14.

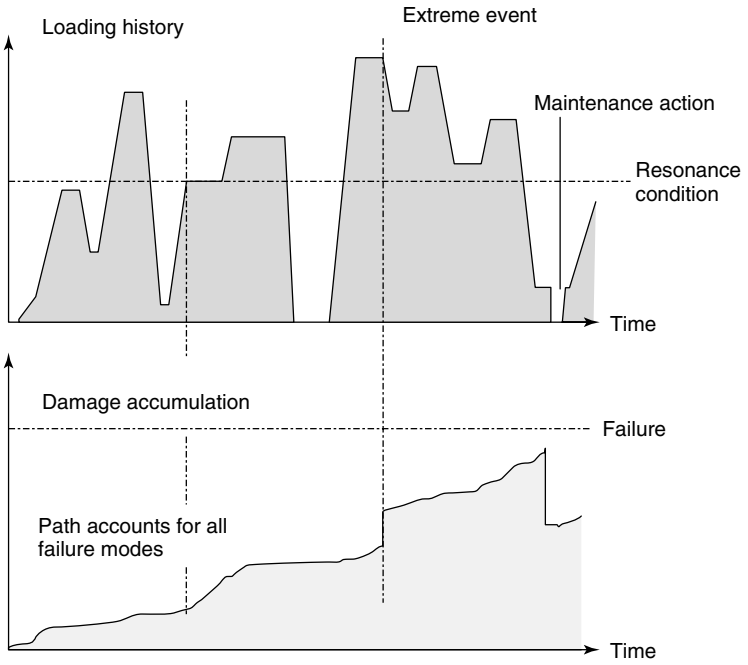


Figure 14. Damage accumulation and projected remaining useful life.

When statistics on the past operating profile of the machine are tracked, projected future operating conditions and maintenance actions can be estimated and utilized by the prognostic model in order to forecast the remaining life in the blade.

7 CONCLUSIONS

A cross section of the current state of the art in applying integrated gas turbine engine health monitoring, and diagnostic and prognostic technologies has been presented, which can offer significant potential for reducing current engine LCC. These technologies can be implemented across the entire spectrum of both military and industrial engines from land-based gas turbines to aircraft engines. Implementation of these technologies is advantageous in nearly eliminating sensor problems, improving maintenance decision effectiveness by providing early warning of incipient performance and vibration faults, and gauging remaining life and predicting future usage associated with critical components.

This article has also introduced specific methods currently being implemented for automated sensor, vibration and performance diagnostics for engine health management. Specifically, the techniques have been developed and implemented for various military engine test cells and industrial gas turbines. Initial results on implementing these technologies has been encouraging, with several sensor faults and vibration-related fault frequency being detected. The capabilities of the engine health monitoring approaches will satisfy system requirements of reduced support costs and improved engine testing and accountability. The initial findings are favorable in the quality of the signals, capabilities to store and network data, and implement diagnostics and intelligent troubleshooting.

REFERENCES

[1] Roemer MJ, Kacprzynski GJ, Orsagh RF. Assessment of data and knowledge fusion strategies for prognostics and health management. *IEEE Aerospace Conference*. Big Sky, Montana, 2001.

[2] Roemer MJ, Ghiocel DM. A probabilistic approach to the diagnosis of gas turbine engine faults. *Paper 99-GT-363, ASME and IGTI Turbo Expo 1999*. Indianapolis, Indiana, June 1999.

[3] Roemer MJ, Kacprzynski GJ, Schoeller M. Advanced test cell diagnostics for gas turbine engines. *IEEE Aerospace Conference*. Big Sky, Montana, 2001.

[4] Halford G. Cumulative fatigue damage modeling—crack nucleation and early growth. First International Conference on Fatigue Damage. Hyannis, Massachusetts, 22–27 September 1996.

FURTHER READING

Agosta JM, Weiss, JW. Active fusion for diagnosis guided by mutual information measures. *Proceeding of the 2nd International Conference on Information Fusion*. Orlando, FL, 6–8 July 1999.

Bjerager P. On computational methods for structural reliability analysis. *Proceedings of the International Workshop on Structural System Reliability*. Boulder, CO, 1988.

Brooks RR, Iyengar SS. *Multi-Sensor Fusion*. Prentice Hall: Upper Saddle River, NJ, 1998.

Dietz WE, Kiech EL, Ali M. Jet and rocket engine fault diagnosis in real time. *Journal of Neural Network Computing* 1989 Summer.

Eshleman RL. Detection, diagnosis and prognosis: an evaluation of current technology. *Proceedings of MFPG 44*. Vibration Institute, 1990.

Gladney ED. *NASA Launches an Automated Data Acquisition System*. Sensors, September 1998.

Hall D, Llinas J. An introduction to multisensor data fusion. *Proceedings of the IEEE* 1997 **85**:6–23.

Kohonen T. *Self Organizing and Associative Memory*. Springer-Verlag: New York, 1987.

Leferve E, Colot O. A classification method based on the Dempster-Shafer's theory and information criteria. *Proceeding of the 2nd International Conference on Information Fusion*. Orlando, FL, 6–8 July 1999.

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Statistical Models*. IRWIN: Chicago, 1996.

Orsagh RF, Roemer MJ, Kacprzynski GJ. Development of metrics for mechanical diagnostic technique qualification and validation. *COMADEM Conference*. Houston, TX, December 2000.

Pusey HC. An historical view of mechanical failure prevention. *Proceedings of the 11th Biennial Conference on*

Reliability Stress Analysis and Failure Prevention. ASME, 1995.

Roemer MJ, Kacprzynski GJ. Advanced diagnostics and prognostics for gas turbine engine risk assessment. *Paper 2000-GT-30, ASME and IGTI Turbo Expo 2000*. Munich, Germany, May 2000.

Roemer MJ, Atkinson B. Real-time engine health monitoring and diagnostics for gas turbine engines. *Paper 97-GT-30, ASME and IGTI Turbo Expo 1997*. Orlando, Florida, June 1997.

Zadeh L. Application of fuzzy set theory. *Fuzzy Sets, Information and Control* 1965 **8**:338–353.

Chapter 156
Integrated Sensor Durability and Reliability

James L. Blackshire and Kumar V. Jata
Air Force Research Laboratory, Wright Patterson Air Force Base, Dayton, OH, USA

1 INTRODUCTION

A significant amount of interest currently exists in the area of integrated systems health management (ISHM). Although this basic concept can take on many different functional forms, for civil, mechanical, and aerospace systems, the use of structural health monitoring (SHM) as part of a larger ISHM strategy offers the potential for monitoring the structural health of a system with far-reaching consequences and benefits [1–3]. Improved fleet management using condition-based maintenance (CBM) strategies, for example, would use ISHM and SHM to provide critical diagnostic information for assessing the health of an aircraft system. In a CBM framework, aircraft would be maintained more efficiently and effectively using critical health status indicators, which would be provided by an integrated health diagnostic system (*see* **Usage Management of Military Aircraft Structures**).

A critical aspect of the ISHM and SHM concepts involves the use of integrated sensing systems to interrogate, inspect, and diagnose the structural systems. In practice, the inspection system uses distributed sensors, which are attached to an existing structure or integrated directly within a newly fabricated structure [4–8]. In either case, the integrated sensors become a part of the structure, and are subjected to similar or identical environmental conditions (*see* **The Influence of Environmental Factors**). This places additional requirements on the performance of the sensor system with regard to long-term durability, reliability, and operational performance over extended time periods, and in a wide variety of varying environmental conditions. These environmental conditions would likely include variations in moisture, chemical attack, vibration, temperature, and mechanical loading of the structural components [1, 3–7, 9]. In addition, the sensors would need to behave in a consistent and reliable manner throughout the duration of the inspection process, which for most applications could span several years to decades.

1.1 Historical background

The durability of onboard sensing systems has been an important and long-standing problem for more

 ISBN: 978-0-470-05822-0.

than a century. Much of the early development and use of sensor technologies involved wired sensing systems in the manufacturing industries for process control in the form of temperature, humidity, pressure, displacement, force, fluid flow, acceleration, and proximity sensors [10, 11]. Significant advances in sensor system design, speed, accuracy, miniaturization, packaging, and overall performance were made in the 1970s, which enabled the development and growth of whole new manufacturing industries. At the heart of these advances were efforts to improve sensor system durability, reliability, and survivability using newly developed sensor materials, improved sensor system designs, and robust sensor packaging methods. As a result, unprecedented sensing capabilities became available with extended dynamic ranges, sensitivities, and useful sensor lives [10–14].

The development of miniature electronic circuitry and portable battery technologies in the mid-twentieth century provided an additional opportunity for developing onboard sensing systems for remote locations and mobile systems [15–20]. The transportation and civil industries, in particular, have benefited significantly from sensors for monitoring performance and load states in automobiles, railway systems, aircraft, and civil structures. In addition to increased demands on sensor system performance, durability and reliability issues became much more prevalent and demanding for these remote/mobile sensors, which were required to operate in dynamic and sometimes unpredictable operational environments. Unlike the earlier manufacturing and processing sensors, which were used in controlled environments, the sensors used in transportation and civil applications were often subjected to uncontrolled environments, which affected their performance, reliability, and durability. The use of onboard power and control circuitry also required strategies for improving battery and circuit durability [21–23]. By the mid-1990s, rugged sensor systems had become available with advanced heat sink, vibration isolation, and connector designs, which provided virtually trouble-free sensor performance capabilities in remote locations for extended times of up to 5–10 years [10, 11, 24, 25].

Within the past decade, the concept of SHM has advanced to a state where onboard sensing systems are becoming available for assessing the structural integrity of numerous civil, mechanical, and aerospace systems [1–8, 25–29]. In general, the sensing of structural integrity involves the use of sensors that are intimately coupled to the host structure. As far as the sensor system durability and reliability are concerned, this places additional demands on the sensor with regard to its ability to withstand usage and environmental conditions that would normally be subjected to the host structure alone. As a result, new challenges and opportunities exist for advanced SHM sensor systems, which have yet to be fully addressed.

1.2 State of the art

SHM represents a wide, multidisciplinary field of engineering where integrated sensing systems are used to diagnose/monitor a structure's operational status and damage state. A large number of integrated sensing methods are currently being developed for many different structural applications [30–35]. Table 1 provides an abbreviated listing of the major types of sensors used in SHM applications along with the measurement types, physical principles involved, and reliability issues for each major sensing method.

With regard to SHM system durability and reliability, a number of technologies have advanced to a level where they are being applied to realistic structures for long-term assessment and use [36–40, 43, 48]. Traditional sensing technologies utilizing commercial-off-the-shelf (COTS) sensors, for example, have been used for more than a decade in civil and aerospace applications [4, 37, 38]. Building on flight recorder data collection systems originally developed in the 1950s–1960s, the health and usage monitoring system (HUMS) represents one of the most mature onboard sensing technologies in use today [38, 49, 50] (*see* **Experience with Health and Usage Monitoring Systems in Helicopters**). Originally developed in the early 1990s for monitoring structural vibrations (and anomalies) in helicopter rotor and gearbox assemblies, the HUMS system constitutes a complete onboard SHM system, which has had a nearly 100% reliability rate during extensive trials over the past decade. This high reliability rate has been attributed largely to the use of a variety of proven COTS technologies (accelerometer sensors, electronics, and wiring, communication elements). Only limited premature sensor failures,

Table 1. Major types of structural health monitoring sensing methods

Sensor type	Measurement type	Physical principle	Reliability issues	References
1. Ceramics and oxides				
Piezoelectric	Strain, vibration, ultrasound	Electromechanical	Brittle fracture, disbond	2–12, 14, 24–38
Pyroelectric	Temperature	Thermoelectric	Brittle fracture, disbond	10, 11, 25–35
Ferroelectric	RFIDs, vibration, temperature	Dipole moment	Brittle fracture, disbond	10, 11, 13, 25–35
2. Fiber-optic				
EFPI, Bragg grating	Strain, temperature, chemical	Optical reflectance	Brittle fracture, pullout	2, 4, 39–42
3. Thin/thick film				
Strain/crack gauges	Strain, crack growth	Electrical resistance	Electrical short, disbond	2, 10, 11, 43
Thermocouples	Temperature	Electrical resistance	Oxidation, disbond	4, 10, 11, 26–35
Electrochemical	Corrosivity, chemical	Electrical resistance	Electrical short, disbond	11, 26–35, 44
4. MEMS	Strain, vibration, force	Micromechanical motions	Fracture, wear, short	14, 15, 25–35, 45
5. Electromagnetic				
MWM, Foil EC	Cracks, fatigue, corrosion	Dielectric, eddy currents	Electrical short, disbond	11, 26–35, 46
6. Comparative vacuum	Crack growth, pressure	Vacuum release	High/low temperature, disbond	47

attributed primarily to electronic faults and inadequate mounting of the sensors, have been observed. Similar sensing approaches are beginning to appear for SHM aerospace [37], automotive [4], and train/rail system [51] applications.

The use of thin-film strain gauge and corrosivity sensors for structural assessment is also noteworthy with regard to SHM sensor reliability testing on a large scale (*see* **Loads Monitoring in Aerospace Structures**; **Aircraft Structural Diagnostic and Prognostic Health Monitoring for Corrosion Prevention and Control**). An ongoing US Air Force program for C-17 cargo aircraft is currently utilizing a suite of strain gauges on a portion of the fleet in key structural locations to compare actual usage data against design assumptions [43]. In that program, the durability of the strain gauge sensors was tracked for 22 aircraft from 1992 to the present. An important result of that program was an observation of infant mortality rate of 23%, where sensors failed very close to their installation time. Of the 160 sensors installed, 58% were still operational as of March 2003. A similar large-scale flight test program involving the evaluation of thin-film galvanic corrosivity sensors was begun in 2003, where more than 600 sensors were attached to 100+ commercial and military aircraft [44]. Roughly, 10–15% of the sensors were reported as being lost due to neglect during periodic depot maintenance (PDM) activities, where physical damage to the sensor and electronics had occurred. Environmental exposure had also caused the degradation of an additional 1–2% of the sensors within 2–3 years for internally mounted sensors and ~5% for externally mounted sensors.

The development of custom and more nontraditional sensing approaches has also progressed in recent years. To a large degree, many of the leading SHM methods have completed laboratory proof-of-concept evaluations and are being considered for transition opportunities in real-world applications. Six technologies in particular (fiber-optic, piezoelectric, thin-film, electromagnetic, microelectromechanical sensors (MEMS), and vacuum sensors) have matured to a level where initial reliability testing has been completed, and real-world testing trials have begun [41, 42, 52–54]. Building on two decades of research and development, fiber-optic

sensors are being used in load and strain monitoring applications for bridges [54] (*see* **Fiber-optic Sensor Principles**; **Fiber Bragg Grating Sensors**; **Hybrid PZT/FBG Sensor System**; **Fiber-optic Sensors** and **Reliable Use of Fiber-optic Sensors**) and for damage detection in composite aerospace structural components [41]. Miniature piezoelectric sensors arrays have also been installed on commercial and military aircraft recently, where preliminary flight testing has demonstrated 1151 flight hours over 246 take-off/landing cycles with only minor problems being reported [42] (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**). Additional operational testing programs have been reported recently for comparative vacuum monitoring (CVM) crack sensors [55] (*see* **Comparative Vacuum Monitoring (CVM™)**), meandering winding magnetometer (MWM) damage sensors [47], (*see* **Electric and Electromagnetic Properties Sensing**) and MEMS sensors [46] (*see* **Microelectromechanical Systems (MEMS)**).

2 TECHNICAL BACKGROUND

At a fundamental level, the reliability of an SHM sensing system involves the durability of the materials that make it up. Both individually and collectively, the long-term effectiveness of an onboard sensing system relies on how the various component materials of an SHM system perform/degrade with time and usage. In this section, the various degradation mechanisms that can affect an SHM system are briefly described, where the influence of environmental factors is highlighted (*see* **The Influence of Environmental Factors**).

2.1 Environmental effects on SHM system performance

The environmental effects on an SHM sensor system can be loosely categorized into four basic areas: (i) mechanical effects, (ii) thermal effects, (iii) chemical effects, and (iv) electrical effects. Mechanical effects include structural loading, vibration, and acceleration, which can range from small, long-duration stresses to large dynamic 'shock' events (e.g., a quick acceleration event) (*see* **Civil Infrastructure Load Models for Structural Health Monitoring**; **Static Damage Phenomena and Models**; **Damage Evolution Phenomena and Models**; **Failure Modes of Aerospace Materials** and **Principles of Structural Degradation Monitoring**). Thermal effects include direct/indirect heating, natural environmental temperature variations, high/low temperature extremes, and thermally induced mechanical loading (*see* **Thermomechanical Models**). Chemical effects include corrosion, moisture, and fluid susceptibility, which can selectively attack the various parts of an SHM system through chemical interactions. Electrical effects include electrical conduction/insulation loss, electrical short circuits, and electromagnetic field interactions (e.g., electromagnetic interference (EMI)) (*see* **Electric and Electromagnetic Properties Sensing**).

2.1.1 Mechanical loading, vibration, and acceleration effects

With regard to SHM reliability, mechanical effects involve the ability of an SHM material or system to resist applied forces, where the strength of a material or system determines whether degradation or damage will occur (*see* **Damage Evolution Phenomena and Models**; **Fatigue Life Assessment of Structures**). Material strength is usually categorized in terms of compressive strength, tensile strength, bending strength, and shear strength, depending on the direction of the applied force relative to the material response. The strength is also affected by how the force is applied in time, that is, static (i.e., constant) or dynamic. In general, the yield strength of a material becomes important for static loading, while dynamic loading involves fatigue and wear processes.

One of the simplest expressions used in strength assessments involves uniaxial stress:

$$\sigma = \frac{F}{A} \tag{1}$$

where F is the force (in newton) acting on an area A (in square meter). Compressive stress involves loading where the material tends to compact, while tensile stress results in a stretching of a material by pulling forces [45, 56, 57]. Materials such as metals, which can tolerate nominal tensile stresses of $\sim$100 Mpa or more, are termed elastic/linear, while materials such as ceramics, which can be susceptible to failure at low nominal tensile stress levels, are considered to be inelastic/nonlinear. Bending stresses

typically involve combinations of compressive and tensile stresses, while shear stresses result in the sliding of a material along planes that are parallel to the direction of the applied force.

The durability and reliability of an SHM material or system is directly related to its strength limits. If a material is stressed beyond its limits, it will fail on the basis of its yield strength or yield point, where a fundamental transition from elastic behavior to plastic behavior occurs. Knowledge of the yield point is vital when designing a material, component, or system because it generally represents an upper limit to the load that can be applied [45, 56, 57]. In general, yield will occur when the largest principal stress exceeds the uniaxial tensile yield strength [45]:

$$\sigma_{\text{applied}} \geq \sigma_{\text{YS}} \tag{2}$$

In many materials, the relation between applied stress and the material's elastic deformation response (i.e., strain) is directly proportional, where the "modulus" of a material can be used to determine stress–strain relationships [45]. In particular, Young's modulus (E) describes the material's response to linear strain, the bulk modulus (K) describes the material's response to uniform pressure, and the shear modulus (G) describes the material's response to shearing strains.

Mathematically, Young's modulus is related to the tensile stress and tensile strain by [45]

$$E \equiv \frac{\sigma}{\varepsilon} = \frac{F/A_0}{\Delta L/L_0} = \frac{FL_0}{A_0 \Delta L} \tag{3}$$

where σ is the tensile stress, ε is the tensile strain, F is force, A_0 is the original cross-sectional area, L_0 is the original length, and ΔL is the change in length. Using equation (3) and Hooke's law, the force exerted by a material under a specific strain can be calculated [45]:

$$F = \frac{EA_0\Delta L}{L_0} = \left(\frac{EA_0}{L_0}\right)\Delta L = kx;$$
$$k = \frac{EA_0}{L_0}; \quad x = \Delta L \tag{4}$$

where again F is the force exerted by a material when it is compressed/extended by length ΔL.

The shear modulus, G, is important for material interfaces (e.g., a sensor bond interface), and is defined as the ratio of shear stress to the shear strain [45]:

$$G = \frac{F/A}{\Delta x/h} = \frac{Fh}{\Delta x A} \tag{5}$$

where F is the shearing force, A is area, F/A is the shear stress, h is the initial length, Δx is the transverse displacement, and $\Delta x/h$ is the shear strain.

2.1.2 Thermal effects

Temperature is an important aspect of most SHM applications, affecting both sensing performance and system reliability (*see* **Thermomechanical Models** and **Loads and Temperature Effects on a Bridge**). The temperature of a system is related to the average energy of microscopic motions in the materials that make up the system. In a solid material, the microscopic motions are primarily due to vibrations of the constituent atoms and molecules in the solid. A closely related principle to temperature is the concept of heat (symbolized by Q), which determines how thermal energy is transferred within a material/system [58]. With regard to SHM reliability, temperature and heat transfer cause three basic degradation or damage effects in a sensing system: (i) thermal effects on sensing performance, (ii) thermal changes in material structure/phase, and (iii) thermal expansion/contraction of materials.

Thermal effects on sensing performance are largely dependent on the sensor type and the materials used in the sensing system. Electronic circuitry, metallic wires, sensor channels, and connections suffer from reliability issues due to the generation of heat when they are used. This is primarily due to the fact that freely moving valence electrons in a metal transfer both electric current and heat energy efficiently. Most electrical components generate heat and can malfunction if they are overheated, or in extreme cases the SHM system can become permanently damaged.

With regard to sensor material properties, temperature and heat transfer can also impact performance, reliability, and durability by permanently changing

the chemical phase and physical response characteristics of a sensor material. The index of refraction in fiber-optic sensors, piezoelectric properties in elastic wave/vibration sensors, and electrical resistance properties of semiconductor strain gauges can all be altered because of high–low temperature extremes. The *Curie temperature* (T_C), for example, refers to a characteristic property of ferromagnetic and piezoelectric materials where sensing performance can be reduced or lost completely if reached [59]. At temperatures below T_C, the magnetic moments of ferroelectric materials and net dipole moments of piezoelectric materials provide the needed sensing characteristics. As the temperature is increased, however, thermal fluctuations destroy preferred alignments in the materials, resulting in the loss of net magnetization and spontaneous polarization when T_c is reached.

Perhaps, the most common temperature-related factor that can affect SHM system reliability involves the thermal expansion/contraction of a material. Thermal expansion is the tendency of matter to increase its volume/size in response to an increase in temperature. The degree of expansion divided by the change in temperature is called *the material's coefficient of thermal expansion*, α, which in one-dimensional form can be written as [58]

$$\alpha = \frac{\Delta L/L_o}{\Delta T}; \quad \Delta L = L_o \alpha \Delta T; \quad L = L_o(1 + \alpha \Delta T) \tag{6}$$

where L_o is the original material length, ΔL is the length change, and ΔT is the temperature change. Figure 1 provides a schematic diagram of the thermal expansion process for an increase in temperature in a 1D solid, and also a schematic of the resulting tension/compression stresses for a rigidly bonded PZT sensor on aluminum substrate material, which has been subjected to a temperature increase and differential material expansion.

For solid materials, the coefficient of thermal expansion is positive, which means that the solid will expand on heating, and contract on cooling [58]. For coefficient of thermal expansion values of $4 \times 10^{-6}/^{\circ}\mathrm{C}$ and $23 \times 10^{-6}/^{\circ}\mathrm{C}$ in piezoelectric and aluminum materials, respectively, a thermal expansion of 22.5 and 131 $\mu\mathrm{m\,m^{-1}}$ would be expected in each respective material for a $+100\,^{\circ}\mathrm{F}$ ($+55.5\,^{\circ}\mathrm{C}$) increase in temperature. This fivefold difference in expansion level results in compressive stresses in the substrate material and tensile stresses in the piezoelectric sensor, which can cause physical damage to the sensor material [60–62].

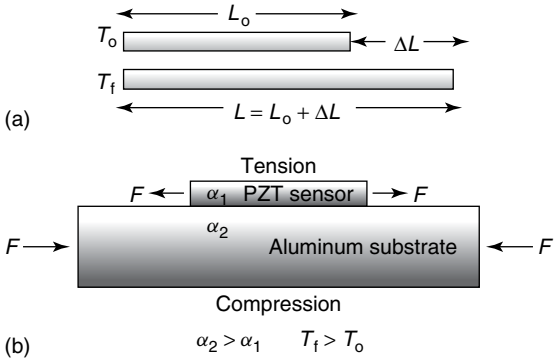


Figure 1. Schematic representation of thermal expansion due to temperature increase (a) and resulting tension/compression stresses in bonded PZT sensor on aluminum substrate (b).

2.1.3 *Moisture and fluid susceptibility effects*

There are two primary types of chemically induced damage that can impact the reliability of an SHM system—moisture and chemical attack. Moisture, acting as an electrolyte, can cause corrosion and electrical shorting problems, while chemical attack can occur from a wide variety of solvents, lubricants, cleaning solutions, fuels, and other fluids used in SHM sensing environments. In general, almost every material used in an SHM sensing system is susceptible to chemical attack, where protective coatings or encapsulation provide the best protection.

2.1.4 *Electrical, magnetic, and electromagnetic radiation effects*

The major electrical effects that can impact the reliability of an SHM system include electrical conduction/insulation loss, electrical short, and EMI. Electrical shorting and the loss of electrical connections for transmitting sensor/signal information, in particular, represent two of the most common reliability and durability issues for most SHM applications. Electrical shorts can take place because of a number of reasons including mechanical, thermal, and chemical breakdown of the wiring as described above.

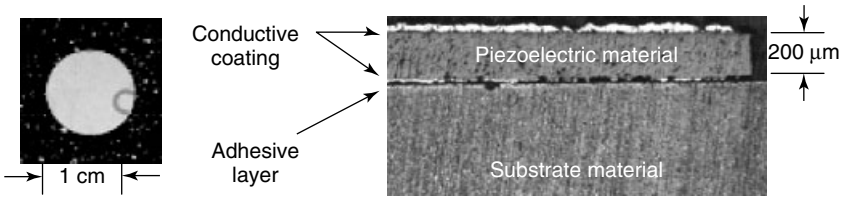


Figure 2. Digital image and cross-sectional cut through piezoelectric sensor disk.

In addition to causing reliability issues, a short circuit can also cause circuit damage, overheating, fire, and battery explosions.

3 EXAMPLES OF SHM SENSING SYSTEM RELIABILITY

Previously published work has shown that the reliability of an SHM system can be compromised by environmental stresses, resulting in system performance loss, sensor degradation, and sensor damage [9, 40, 43, 60]. Both gradual and abrupt performance losses have been reported, being attributed to a number of factors including undesired load transfer [9, 60, 61], thermal assisted fracture, and simple electrical/electronic failures. In this chapter, four examples are provided for SHM sensor reliability, including (i) a study of damage introduced in surface-bonded piezoelectric sensors, (ii) a study of the reliability of thermocouple sensor arrays, (iii) durability testing for CVM sensors, and (iv) performance testing results for layered piezoelectric actuator/sensor networks.

3.1 Reliability of surface-bonded piezoelectric sensors

Surface-bonded piezoelectric sensors are considered to be one of the most attractive integrated sensor concepts for SHM (*see* **Piezoceramic Materials—Phenomena and Modeling**; **Piezoelectricity Principles and Materials**; **Piezoelectric Wafer Active Sensors**). By converting mechanical energy into electrical energy, piezoelectric materials provide an effective means for monitoring structural vibrations and for probing structures with elastic waves. Damage is typically characterized through elastic wave scattering, where the amplitude, phase, and frequency content of the waves is used to understand the damage state of a structure. Surface-bonded piezoelectric sensors provide an inexpensive, lightweight, and minimally intrusive sensing solution for SHM.

For most applications, single-crystal or polycrystalline ceramic piezoelectric materials are used, where miniature disks are engineered for a specific application. A typical piezoelectric sensor disk consists of a layer of piezoelectric material coated on the top and bottom with a thin conductive layer (Figure 2). A very thin adhesive bond layer is then used to attach the sensor to a substrate material, which consists of an approximate 10-μm-thick layer of Vishay's M-Bond 200 cyanoacrylate adhesive and catalyst as shown in Figure 2.

Previous work has shown that static load transfer between the host structure and an attached piezoelectric sensor disk can have a significant impact on the sensor's durability and its performance characteristics [9, 60, 61]. In particular, the partitioning of load between the substrate and sensor was found to be dependent on the relative stiffness between the two materials, which can result in premature failure of the adhesive or the sensor due to high stress levels. This situation is depicted schematically in Figure 3, where a cross-sectional representation of the problem is depicted, along with key material properties: elastic modulus E_r and E_p; layer thicknesses t_r, t_a, and t_p; adhesive shear modulus G_a, and sensor dimensions y and L.

Following the analysis of Martin and Blackshire [61], the piezoelectric sensor stress, σ_r, can be expressed in terms of the stiffness ratio, $S = (E_r t_r / E_p t_p)$, as follows:

$$\sigma_r(y) = \left[\frac{FS}{((1+S)t_r)}\right]\left\{1 - \left[\frac{\cosh(\beta y)}{\cosh(\beta L)}\right]\right\} = \sigma_r(sp)\left\{1 - \left[\frac{\cosh(\beta y)}{\cosh(\beta L)}\right]\right\} \quad (7)$$

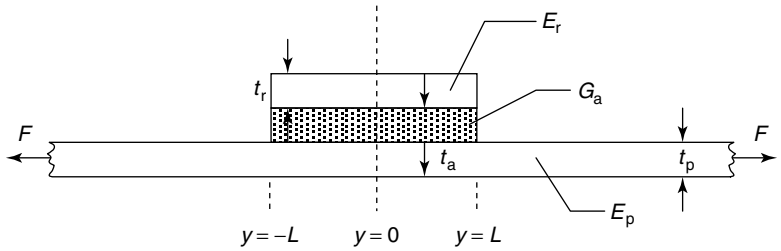


Figure 3. Schematic representation of bonded piezoelectric sensor disk (cross-sectional view).

where F is the load per unit length and β^2 is given by the expression:

$$\beta^2 = \left(\frac{G_a}{t_a}\right)\left[\left(\frac{1}{E_p t_p}\right) + \left(\frac{1}{E_r t_r}\right)\right] \tag{8}$$

The corresponding shear stress distribution in the adhesive layer is given by the expression:

$$\tau_a(y) = -t_r \sigma_r'(y) = \beta\left[\frac{FS}{((1+S)\cosh(\beta L))}\right][\sinh(\beta y)] \tag{9}$$

Using equation (7), the stress conditions in a surface-bonded sensor for compliant, rigid, and infinitely rigid bonds can be calculated (Figure 4a), with parameters of $E_p = 7.31 \times 10^{10}$ Pa, $E_r = 8.40 \times 10^{10}$ Pa, $G_a = 7.00 \times 10^6$ Pa (for compliant) and 7.00×10^8 Pa (for rigid, and infinite), $t_p = 0.001$ m, $t_r = 0.0001$ m, $t_a = 0.0001$ m, $L = 0.005$ m. To lower stress transfer between the substrate and the sensor, the size of the sensor or adhesive shear modulus must be decreased, the adhesive layer thickness must be increased, or the stiffness ratio must be decreased.

Using equation (9), the adhesive shear stress for a compliant and rigid bond can be calculated (Figure 4b). From Figure 4, it is clear that the use of a compliant bond helps with load/stress transfer to the sensor, and also with shear stress levels in the adhesive layer, both of which help enhance sensor durability and bond integrity. Thus, to lower the stress transfer to the reinforcement, the size of the reinforcement or adhesive shear modulus must be decreased or the adhesive layer thickness must be increased. Decreasing the stiffness ratio will also decrease the stress transfer.

Blackshire *et al.* experimentally studied SHM environmental usage effects and demonstrated that these

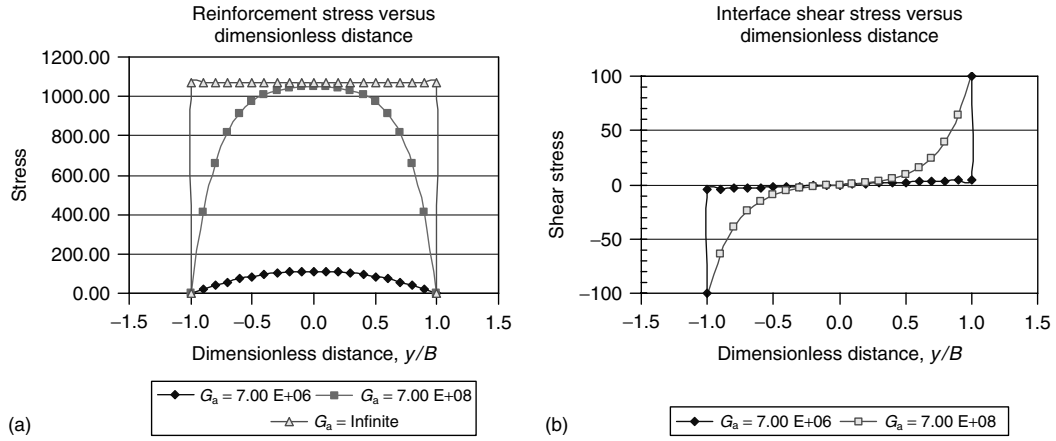


Figure 4. (a) Piezoelectric sensor disk stress distribution for compliant, rigid, and infinitely rigid bonds and (b) interface shear stress distribution for compliant and rigid bonds.

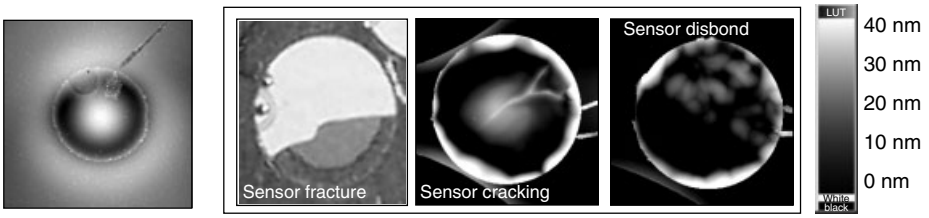


Figure 5. Examples of (a) undamaged piezoelectric sensor and (b) piezoelectric sensors damaged by mechanical and thermal loading of the substrate material.

factors can cause degradation and damage to surface-bonded piezoelectric sensor systems [9, 60–62]. Figure 5 shows typical examples of their observations, which included sensor fracture, cracking, and disbonding. The damage was attributed to induced stresses in the bonded sensors arising from biaxial bending, and thermal loading of the substrate material. The effects of static biaxial stress in the substrate, and the influence of adhesive bond, was also modeled and studied by Martin and Blackshire [61]. These studies pointed to the need for mechanically decoupling the bonded sensor from the stresses present in the structure due to static and low-frequency (i.e., vibration) loading.

In an attempt to understand the thermal expansion damage effects observed in Figure 5(b), Martin *et al.* used an equivalent spring analysis [60–62] to determine the stress levels within the piezoelectric and bond material for a 37.8 °C (+100 °F) thermal loading event. In their analysis, the initial (unloaded) spring heights corresponded to the lengths of the piezoelectric sensor and aluminum substrate after free thermal expansion had occurred in the materials. The springs were then coupled in parallel to determine the equilibrium spring height that would occur when the force of compression on the taller spring (aluminum, in the case shown) equaled the force of elongation on the shorter spring (piezoelectric sensor) so that their deformed lengths were equal. An equivalent spring stiffness per unit depth could then be calculated by

$$k_i = \frac{E_i t_i d}{l_i} \tag{10}$$

where $l_i \sim l_j$. The difference in spring heights was then found to be

$$\Delta h = \Delta l_i - \Delta l_j = \Delta x_i - \Delta x_j \tag{11}$$

where the force equation could be written as follows:

$$F_i = k_i \Delta x_i \tag{12}$$

and where, from equilibrium considerations, $F_i = -F_j$. Substituting and solving for Δx_i gives

$$\Delta x_i = \frac{\Delta h}{(1 + k_i/k_2)} = \frac{\Delta h}{\left(\dfrac{k_j + k_i}{k_j}\right)} \tag{13}$$

The relationship between force per unit length and stress is then given by

$$\sigma_i = \frac{F_i}{t_i} = \frac{k_i \Delta x_i}{t_i} \tag{14}$$

which results in a stress level of 53 MPa in the piezoelectric sensor for $E_1 = 7.31 \times 10^{10}$ Pa, $E_2 = 8.40 \times 10^{10}$ Pa, $t_1 = 0.001$ m, $t_2 = 0.0002$ m, and $L = 0.01$ m. This analysis ignores the axial symmetry of the sensor, the bending stresses, and transient thermal effects, which would tend to increase stress levels within the piezoelectric sensor material. For failure stress levels estimated at 70–80 MPa for the piezoelectric materials used, the possibility of fracture of the sensor is a real possibility for a +100 °F increase in temperature and the resulting thermal expansion mismatch conditions.

The effects of piezoelectric sensor disbond on the generation and reception of elastic wave energy were also studied. Figure 6(a) depicts the received signals from a pair of surface-bonded piezoelectric sensors spaced at 2.5″ (63.5 mm) for two different measurement cases: (i) a pair of fully bonded piezoelectric sensor disks and (ii) a 33% disbonded transducer sending a signal to a fully bonded sensor [62]. The

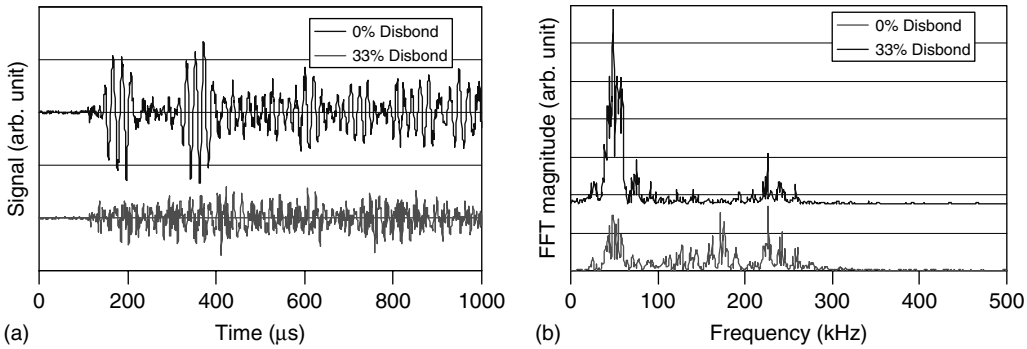


Figure 6. (a) Pitch–catch signals between two fully bonded sensors and one good and one 33% disbonded sensor and (b) frequency content of received signals.

drive signals in this case were at 50 kHz for a five-cycle toneburst at ±5 V. The received signal for the two fully bonded sensors (top) shows the distinct five-cycle waveform for the various expected Lamb modes, while the signals for the bonded and 33% disbonded pair shows a reduced signal level and loss of drive frequency fidelity in the received signal. The frequency content of the signals is depicted in Figure 6(b), which shows this basic effect, with an increase in frequency content for the 100–300 kHz band, which is likely because of the modal vibration effects in the disbonded sensor. A noticeable degradation in sensor system response was, therefore, observed for disbonded piezoelectric sensors in this case.

With regard to surface-bonded piezoelectric sensor durability, the use of advanced silicon, polymer, and epoxy adhesives has recently shown promise for improved sensor bonding compared to rigid cyanoacrylate adhesives [63]. Research involving electroactive polymer (EAP), electroactive paper (EAPap), and piezoelectric polymer materials has also progressed recently, where flexible/conformable sensor concepts are becoming available and more common [64–66], improving piezoelectric sensor system ruggedness and durability.

3.2 Reliability of direct-write thermocouple sensor array

The reliability of thermocouple sensors integrated onto the surface of a titanium plate by direct-write plasma spray methods was recently studied by Ackers *et al.* [67]. In that study, the direct-written sensors were subjected to various thermal soak conditions at elevated temperatures, where the modal frequency response characteristics of the plate–sensor combination were monitored to assess changes in SHM system reliability. Figure 7 provides a schematic depiction of the titanium plate, sensor locations, and modal impact point locations used in the study.

The modal impact testing used four reference accelerometers to measure vibration frequency response functions (FRFs) of the instrumented plate up to 4 kHz for the 109 impact locations depicted in Figure 7. Three baseline measurements were first collected before any thermal soaks were conducted, followed by two sets of data collected after each thermal soak. The Ti 6242 titanium plate was subjected to thermal soaks of increasingly higher temperature loads, at 400 °C for 24 h, 450 °C for 24 h, 475 °C for 48 h, and a final 2-h period cyclical soak ranging from 450 to 525 °C over 48 h, simulating operational conditions.

Figure 8 provides examples of the major results for the study, where an upward shift in the FRFs of the plate–sensor combination was observed owing to the elevated temperature soak conditions. The results were consistent with analytic considerations, mode-shape curvature effects, and transmissibility function analysis, each of which provides information on global parameter changes in a system based on stiffness, density, and inertial properties [67]. In particular, the upward shift in FRF frequency depicted in Figure 8 was attributed to a mass change resulting from oxidation at high-temperature exposure of the plate–sensor system [67].

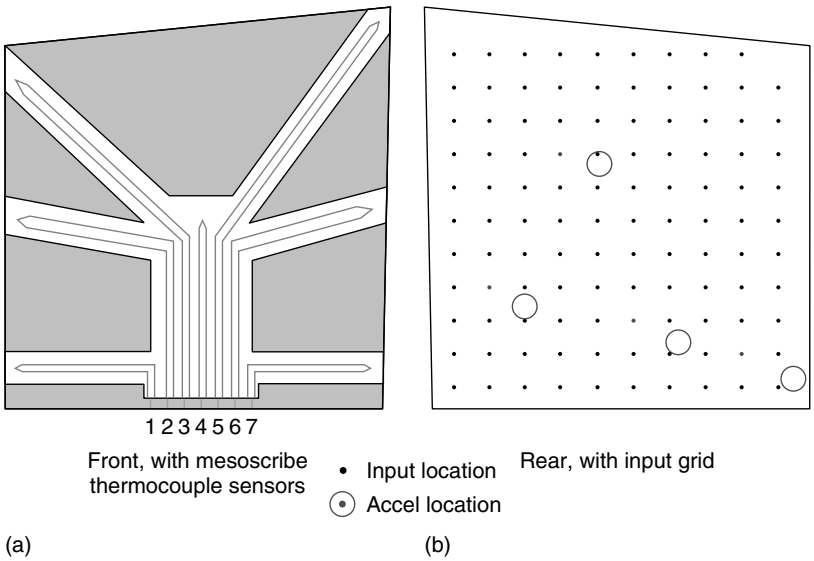


Figure 7. Schematic of titanium plate with positions of direct-write thermocouple sensors depicted (a) and modal impact grid pattern (b). [Reproduced from Ref. 67.]

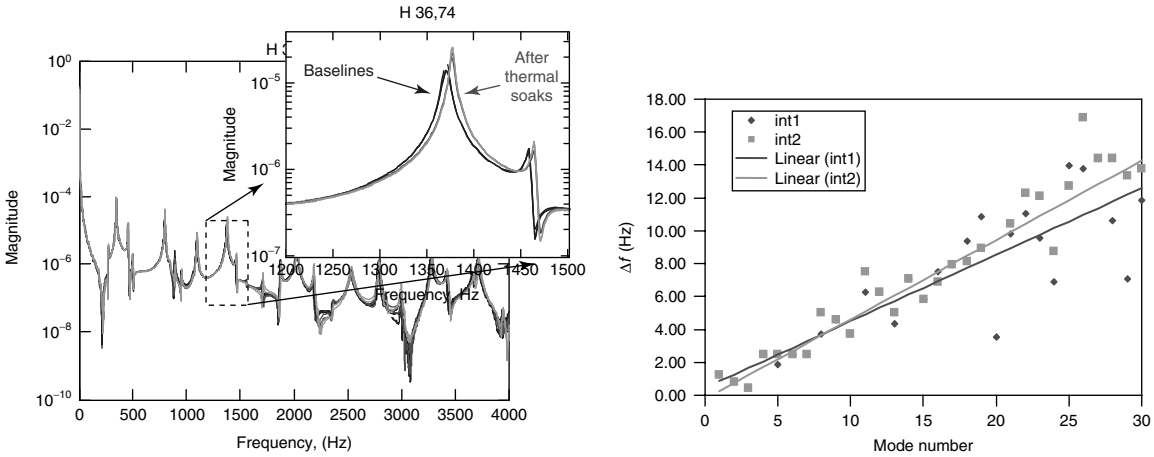


Figure 8. Frequency response function data taken before and after thermal soaks showing frequency shifts in modal frequency response after elevated temperature exposures. [Reproduced from Ref. 67.]

Figure 9 provides representative measurement results for the direct-write thermocouple output readings for the 400 °C (752 °F) and 450 °C (842 °F) thermal soak cases. In both cases, the thermal soak lasted for a 24-h period. The measurement results showed consistency between all of the channels with a nominal standard deviation of ∼1.5 °C (2.5 °F) during the 24-h period for each measurement. The measurements also showed a reliable and accurate measurement output between consecutive thermal soak cycling events.

From these results, it was concluded that the thermocouple sensors were reliable under the thermal loads examined. The results from the FRF analysis did show changes occurring in the global properties of the plate, which were attributed to minor changes in mass or stiffness. Modal curvature and transmissibility analyses, however, indicated that there was

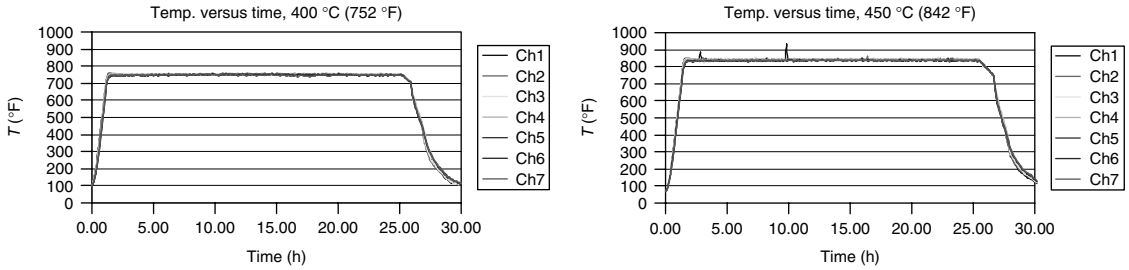


Figure 9. Direct-write thermocouple readings for 400 and 450 °C thermal soak cases. [Reproduced from Ref. 67.].

no localized damage or disbonding of the thermocouples themselves. Additionally, the thermocouples continued to output thermal data with reasonable accuracy throughout the experiments. Further studies are currently being carried out to understand the reliability of direct-write thermocouple sensors to withstand other types of loads (e.g., mechanical load and thermal shock).

3.3 Reliability of comparative vacuum monitoring (CVM) sensors

The reliability of CVM sensors was recently studied by Roach *et al.* [68, 69]. In that effort, the performance and reliability of CVM sensors for crack detection in thin and thick metallic structures was evaluated under controlled laboratory conditions and during a long-term flight testing program (*see* **Comparative Vacuum Monitoring (CVM™)**). As part of a performance validation effort, a series of 26 sensors were installed in four different DC-9, 757, and 767 aircraft in the Northwest Airlines and Delta Airlines fleet [68]. The testing was used to study the long-term operation of the sensors in actual operating environments, where complementary laboratory flaw detection testing was undertaken as part of an overall CVM certification effort. Additional CVM sensor durability testing was also conducted by the Australian Defense Science and Technology Organization (DSTO) and Airbus, where temperature, chemical, and ultraviolet (UV) exposure studies showed no loss in sensor functionality for long-term testing up to 36 months.

A typical CVM sensor bonded to an aircraft component is depicted in Figure 10 along with vacuum tube connections and SHM's PM200 portable data collection module. The sensor includes a self-adhesive, elastomeric material, which has had fine channels laser machined along the bottom surface to form alternating pressurized (atmospheric) and vacuum galleries or channels. When a surface-breaking crack occurs under the CVM sensor, a leakage path forms between the atmospheric and vacuum channels, producing a measurable change in

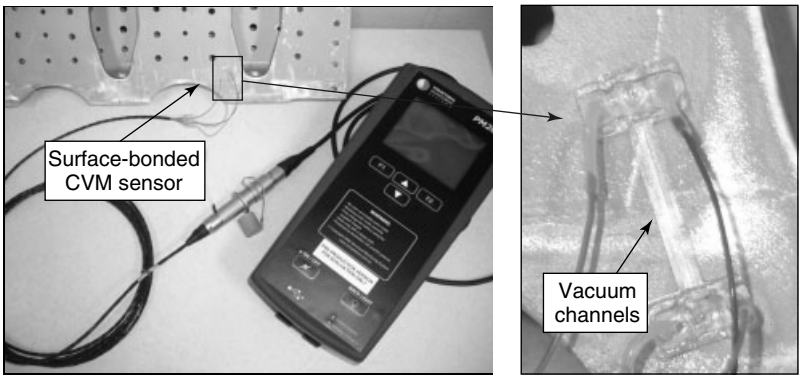


Figure 10. Comparative vacuum monitoring (CVM) sensor bonded to aircraft component.

the vacuum state, which is a positive indication that a crack exists under the sensor.

Regarding CVM sensor durability and reliability, the use of load-bearing elastomer materials and flexible/compliant adhesives enables the sensor to withstand high loading stresses and temperature changes in the aircraft environment. The use of self-test procedures also provides fail-safe conditions, which ensure that the sensor is attached to the structure and working properly. Because the sensor physics relies on pressure measurements, the inherent reliability issues related to electrical wire connections and interference signals are minimized, improving overall system reliability. The use of a protective sealant (e.g., PR1440 sealant) over the sensor and its connection points, and durable Teflon tube attachments to the sensor has also helped to ensure good reliability under operational use in aircraft environments.

3.4 Reliability of layered piezoelectric actuator/sensor networks

The performance and reliability of layered piezoelectric actuator/sensor networks was recently evaluated by Kusaka and Qing [70] and Beard *et al.* [71, 72] (*see* **Stanford Multiactuator–Receiver Transduction (SMART) Layer Technology and Its Applications**). In those efforts, the mechanical loading durability, environmental reliability, and overall system reliability of Acellent's Stanford multiactuator receiver transduction (SMART) layer technology [73] were studied, respectively. Kusaka *et al.*, for example, studied the performance and degradation of layered piezoelectric actuator/sensor networks under monotonic and fatigue loading conditions. Both surface-bonded (metallic and composite substrates) and fully embedded (within composites) sensor networks were investigated, where the piezoelectric sensor performance remained unchanged for strain levels below the static failure limit of the sensor material (lead zirconate titanate (PZT)). For strain levels above the static PZT failure strain limit, reductions in electromechanical coupling efficiency were observed at ∼20% for both monotonic and cyclic loads [70]. The measurement results were consistent with previously published work by Mall and Hsu [74], Mall and Coleman [75], and Paget *et al.* [76] for isolated piezoelectric sensors embedded within composite materials.

The work of Beard *et al.* considered practical issues in the real-world implementation of an SHM system, where sensor integration, calibration, reliability, and environmental compensation were identified as key issues that need to be considered. Testing results for the SMART Layer system were highlighted in that effort, where environmental conditions (temperature, moisture, chemical attack) were considered as well as mechanical shock/loading, and intense fatigue testing. In general, the results indicated that the layered piezoelectric actuator/sensor networks did not lose functionality for most realistic operational conditions [71]. The need for self-diagnosis, sensor redundancy, and environmental compensation was also discussed with regard to improving overall system reliability [72].

Regarding sensor durability and reliability, the use of sealed electrical connectors, sensor system encapsulation/coating, optimized adhesive bonding, and sensor networks integrated on a flexible thin-film material (polyamide) provided for a rugged SMART Layer design [71–73]. The use of multiple distributed sensors also provided the system with sensor redundancy and opportunities for implementing self-test procedures. Ongoing systematic test and evaluation of the technology is taking place, where standardized test procedures (e.g., MIL-STD-810F) are being used to maximize sensor system performance, durability, and reliability.

4 CONCLUSIONS

The long-term reliability, durability, and survivability of integrated sensor systems and their associated hardware represent a critical aspect of any SHM system. Although significant improvements in SHM system reliability and durability have been made in recent years, much more work is needed to help ensure the long-term performance of SHM systems in fielded, operational environments. In general, an SHM sensing system's design must include hardening against mechanical, thermal, chemical, and electrical environmental effects, where intelligent choices in component materials and packaging are needed to maximize system reliability. Standardized testing and evaluation, calibration procedures, and self-test diagnosis can also provide additional means for ensuring

long-term reliability for SHM systems. The interested reader is encouraged to explore the numerous references provided, and particularly the books by Rao [1], Inman [2], Adams [4], Giurgiutiu [25], Holnicki-Szulc and Rodellar [26], Staszewski *et al.* [30], and the SHM Workshop series edited by Chang [31–34].

REFERENCES

[1] Rao B. *Handbook of Condition Monitoring*. Elsevier Publishing: New York, 2005.

[2] Inman D. *Damage Prognosis: For Aerospace, Civil and Mechanical Systems*. John Wiley & Sons: Hoboken, NJ, 2005.

[3] Giurgiutiu V, Zagrai A, Bao J. Piezoelectric wafer embedded active sensors for aging aircraft structural health monitoring. *International Journal of Structural Health Monitoring* 2002 **1**:41–61.

[4] Adams D. *Health Monitoring of Structural Material and Components*. John Wiley & Sons: Hoboken, NJ, 2007.

[5] Dalton R, Cawley P, Lowe M. The potential of guided waves for monitoring large areas of metallic aircraft fuselage structures. *Journal of Nondestructive Evaluation* 2001 **20**:29–45.

[6] Giurgiutiu V, Zagrai A, Bao J, Redmond J, Roach D, Rackow K. Active sensors for health monitoring of aging aerospace structures. *International Journal of the Condition Monitoring and Diagnostic Engineering Management* 2003 **6**(1):1–46.

[7] Wu F, Chang F-K. Built-in active sensing diagnostic system for civil infrastructure systems. *Proceedings of the SPIE*. Newport Beach, CA, 2001; Vol. 4330, pp. 27–35.

[8] Basheer M, Derriso M, Rao V. Self organizing wireless sensor networks for structural health monitoring. In *Structural Health Monitoring—From Diagnostics and Prognostics to Structural Health Management*, Chang FK (ed). DEStech Publications: Lancaster, PA, 2003, pp. 1193–1207.

[9] Blackshire J, Giurgiutiu V, Cooney A, Doane J. Characterization of sensor performance and durability for structural health monitoring systems. *Proceedings of the SPIE*. San Diego, CA, 2005; Vol. 5770, pp. 66–75.

[10] Rhodes T, Carroll G. *Industrial Instruments for Measurement and Control*. McGraw-Hill: New York, 1972.

[11] De Sliva C. *Sensors and Actuators—Control System Engineering*. CRC Press: Boca Raton, FL, 2007.

[12] Janocha H. *Actuators—Basics and Applications*. Springer: New York, 2004.

[13] Brauer J. *Magnetic Actuators and Sensors*. John Wiley & Sons: Hoboken, NJ, 2006.

[14] Tzou H, Fukuda T. *Precision Sensors, Actuators and Systems*. Springer: New York, 1992.

[15] Mead C. Schottky barrier gate field effect transistor. *Proceedings of the IEEE* 1966 **54**(2):307–308.

[16] Amos S, James M. *Principles of Transistor Circuits*. Butterworth-Heinemann: Burlington, MA, 1999.

[17] Streetman B. *Solid State Electronic Devices*. Prentice Hall: Englewood Cliffs, NJ, 1992.

[18] Carson R. *Principles of Applied Electronics*. McGraw-Hill: New York, 1961.

[19] Horowitz P, Hill W. *The Art of Electronics*. Cambridge University Press: Cambridge, MA, 1989.

[20] Warnes L. *Analogue and Digital Electronics*. Macmillan Press: New York, 1998.

[21] Coombs C. *Printed Circuits Handbook*, *Sixth Edition*. McGraw-Hill: New York, 2007.

[22] Fink D, Beaty H. *Standard Handbook for Electrical Engineers*, *Eleventh Edition*. McGraw-Hill: New York, 1978.

[23] Linden D, Reddy T. *Handbook of Batteries*. McGraw-Hill: New York, 2001.

[24] Preumont A. *Vibration Control of Active Structures*. Springer: New York, 2002.

[25] Giurgiutiu V. *Micromechatronics*. CRC Press: Boca Raton, FL, 2004.

[26] Holnicki-Szulc J, Rodellar J. *Smart Structures—Requirements and Potential Applications in Mechanical and Civil Engineering*. Springer: New York, 1999.

[27] Holnicki-Szulc J, Soares C. *Advances in Smart Technologies in Structural Engineering*. Springer: New York, 2004.

[28] Bullough W, Worden K, Haywood J. *Smart Technologies*. World Scientific Publishing: Hackensack, NJ, 2003.

[29] Janocha H. *Adaptronics and Smart Structures Second Edition*. Springer: New York, 2007.

[30] Staszewski W, Boller C, Tomlinson G. *Health Monitoring of Aerospace Structures—Smart Sensor Technologies and Signal Processing*. John Wiley & Sons: Hoboken, NJ, 2003.

[31] Chang FK (ed). *Structural Health Monitoring—Current Status and Perspectives*. CRC Press: Boca Raton, FL, 1997.

[32] Chang FK (ed). *Structural Health Monitoring—From Diagnostics and Prognostics to Structural Health Management*. CRC Press: Boca Raton, FL, 2003.

[33] Chang FK (ed). *Structural Health Monitoring—Advances and Challenges for Implementation*. CRC Press: Boca Raton, FL, 2005.

[34] Chang FK (ed). *Structural Health Monitoring—Current Status and Perspectives*. CRC Press: Boca Raton, FL, 2007.

[35] Ansari F (ed). *Sensing Issues in Civil SHM*. Springer: New York, 2005.

[36] Giurgiutiu V. *Structural Health Monitoring with Piezo Wafer Active Sensors*. Academic Press: Burlington, MA, 2007.

[37] Renouf M. Ultra electronics airborne acoustic integrity monitoring system (AAIMS). *Aircraft Structural Integrity Program*. San Antonio, TX, 28–30 November 2006.

[38] Newman J, Irving P, Lin J, Le D. Crack growth predictions in a complex helicopter component under spectrum loading. *Fatigue and Fracture of Engineering Materials and Structures* 2006 **9**(11):949–958.

[39] Inaudi D, Glisic B. *Fibre Optic Structural Health Monitoring*. John Wiley & Sons: Hoboken, NJ, 2007.

[40] Hegner H, Whitesel H. Study of fiber optic sensor reliability, durability, and failure modes for shipboard machinery. *Proceedings of the SPIE*. Bostan, MA, 1994; Vol. 2072, pp. 12–21.

[41] Fernandez-Lopez A, Wagner W, Guemes A. Embedded sensors at the root of a helicopter blade. In *Structural Health Monitoring—Quantification, Validation, and Implementation*, Chang FK (ed). DEStech Publications: Lancaster, PA, 2007, pp. 256–263.

[42] Kearns J, Pena-Macias J, Criado-Abad A, Southward T, Evans D, Malkin M. Development and flight demonstration of a piezoelectric phased array damage detection system. In *Structural Health Monitoring—Quantification, Validation, and Implementation*, Chang FK (ed). DEStech Publications: Lancaster, PA, 2007, pp. 93–100.

[43] Ware R, Reams R, Woods A, Selder R. Sensor reliability in fielded C-17 aircraft strain gauges. In *Structural Health Monitoring—Advances and Challenges for Implementation*, Chang FK (ed). National Technical Information Service: Springfield, VA, 2005, pp. 478–486.

[44] Abbott W, Kinzie R. Aircraft corrosion sensing and monitoring program. *The 9th Joint FAA/DOD/NASA Conference on Aging Aircraft*. Atlanta, GA, 6–9 March 2006.

[45] Mott R. *Applied Strength of Materials*, *Fourth Edition*. Prentice Hall: Englewood Cliffs, NJ, 2002.

[46] Beeby S. *MEMS Mechanical Sensors*. Artech House: Boston, MA, 2004.

[47] Goldfine N, Grundy D, Washbaugh A, Schlicker D, Sheiretov Y, Huguenin C, Lovett T. Corrosion and fatigue monitoring sensor networks. In *Structural Health Monitoring—Advances and Challenges for Implementation*, Chang FK (ed). Artech House: Norwood, MA, 2005, pp. 1217–1224.

[48] Qing P, Beard S, Kumar A, Yu P, Chan H, Zhang D, Ooi T, Marotta S. Practical requirements for implementation and usage of SHM systems on aerospace structures. In *Structural Health Monitoring—Advances and Challenges for Implementation*, Chang FK (ed). DEStech Publications: Lancaster, PA, 2005, pp. 1502–1509.

[49] Cronkhite J, Dickson B, Martin W, Collingwood G. *Operational Evaluation of a Health and Usage Monitoring System (HUMS)*, DOT/FAA/AR-97/64 Report, April 1998.

[50] Dickson B, Cronkhite J, Bielefeld S, Killian L, Hayden R. *Feasibility of a Rotorcraft Health and Usage and Monitoring System (HUMS): Usage and Structural Life Monitoring Evaluation*, DOT/FAA/AR-95/9 Report, February 1996.

[51] Frankenstein B, Hentschel D, Pridoehl E, Schubert F. Hollow shaft integrated health monitoring system for railroad wheels. *Proceedings of the SPIE*. San Diego, CA, 2005; Vol. 5770, pp. 46–55.

[52] Culshaw B, Dakin J (eds). *Optical Fiber Sensors—System and Applications*. Artech House: Norwood, MA, 1989; Vol. II.

[53] Chambers J, Wardle B, Kessler S. Lessons learned from a broad durability study of an aerospace SHM system. In *Structural Health Monitoring—Quantification, Validation, and Implementation*, Chang FK (ed). DEStech Publications: Lancaster, PA, 2007, pp. 247–255.

[54] Gebremichael Y, *et al.* Integration and assessment of fibre Bragg grating sensors in an all-fibre reinforced polymer composite road bridge. *Sensors and Actuators, A: Physical* 2005 **118**(1):78–85.

[55] Barton D. Comparative vacuum monitoring: a new method of in-situ real time crack detection and monitoring. *10th Asia-Pacific Conference on Non-Destructive Testing*. Brisbane, 17–21 September 2001.

[56] Timoshenko S. *Strength of Materials*, *Third Edition*. Krieger Publishing: Malabar, FL, 1976.

[57] Popov E. *Engineering Mechanics of Solids*. Prentice Hall: Englewood Cliffs, NJ, 1990.

[58] Incropera F, DeWitt D. *Fundamentals of Heat and Mass Transfer*, *Fifth Edition*. John Wiley & Sons: Hoboken, NJ, 2001.

[59] Kittel C. *Introduction to Solid State Physics*. John Wiley & Sons: Hoboken, NJ, 1996.

[60] Blackshire J, Cooney A. Evaluation and improvement in sensor performance and durability for structural health monitoring systems. *Proceedings of the SPIE*. San Diego, CA, 2006; Vol. 6179, pp. 137–146.

[61] Martin SA, Blackshire JL. Effect of adhesive material properties on induced stresses in bonded sensors. In *Review of Progress in Quantitative Nondestructive Evaluation*, *AIP Conference Proceedings—Volume 894*, Thompson DO, Chimenti DE (eds). American Institute of Physics: Melville, NY, 2006; Vol. 26, pp. 1524–1532.

[62] Blackshire J, Martin S, Na J. Disbonding effects on elastic wave generation and reception by bonded piezoelectric sensor systems. *Proceedings of the SPIE*. San Diego, CA, Vol. 6530, pp. 65300L-1–65300L-8.

[63] Blackshire J, Martin S, Na J. The influence of bond material type and quality on damage detection for surface-bonded piezoelectric sensors. In *Structural Health Monitoring—Current Status and Perspectives*, Chang FK (ed). CRC Press: Boca Raton, FL, 2007, pp. 203–211.

[64] Bar-Cohen Y (ed). *Electroactive Polymer (EAP) Actuators as Artificial Muscles: Reality, Potential, and Challenges*. SPIE: Bellingham, WA, 2001.

[65] Yun S, Kim J, Song C. Performance of electro-active paper actuators with thickness variation. *Sensors and Actuators, A* 2007 **133**(1):225–230.

[66] Zhang Q, Bharti V, Zhao X. Giant electrostriction and relaxor ferroelectric behavior in electron-irradiated poly(vinylidene fluoride -trifluoroethylene) copolymer. *Science* 1998 **280**:2101–2104.

[67] Ackers S, Adams D, Sathish S, Jata K. Reliability study of thermocouple array instrumented on a titanium plate using modal impacts and piezoelectric actuation. *Proceedings of the 3rd European Workshop on Structural Health Monitoring*. Granada, 5–7 July 2006; pp. 783–790.

[68] Roach D, Rackow K, DeLong W, Yepez S, Reedy D, White S. *Use of Composite Materials, Health Monitoring and Self-Healing Concepts to Refurbish Our Civil and Military Infrastructure*, Sandia Report SAND2007-5547, September 2007, pp. 223–265.

[69] Roach D, Kollgaard J, Emery S. Application and certification of comparative vacuum monitoring sensors for in-situ crack detection. *Air Transport Association Nondestructive Testing Forum*. Fort Worth, TX, October 2006.

[70] Kusaka T, Qing P. Characterization of loading effects on the performance of SMART layer embedded or surface-mounted on structures. In *Structural Health Monitoring—From Diagno stics and Prognostics to Structural Health Management*, Chang FK (ed). DEStech Publications: Lancaster, PA, 2003, pp. 1539–1546.

[71] Beard S, Kumar A, Qing X, Chan H, Zhang C, Ooi T. Practical issues in real-world implementation of structural health monitoring systems. *Proceedings of the SPIE*. San Diego, CA, 2005; Vol. 5762, pp. 196–203.

[72] Beard S, Liu B, Qing P, Zhang D. Challenges in implementation of SHM. In *Structural Health Monitoring—Current Status and Perspectives*, Chang FK (ed). CRC Press: Boca Raton, FL, 2007, pp. 65–81.

[73] Lin M, Qing X, Kumar A, Shawn J. SMART layer and SMART suitcase for structural health monitoring application. *Proceedings of the SPIE*. San Diego, CA, 2001; Vol. 4332, pp. 98–106.

[74] Mall S, Hsu T. Electromechanical fatigue behavior of graphite/epoxy laminate embedded with piezoelectric actuator. *Smart Materials and Structures* 2000 **9**:78–84.

[75] Mall S, Coleman J. Monotonic and fatigue loading behavior of quasi-isotropic graphite/epoxy laminate embedded with piezoelectric sensor. *Smart Materials and Structures* 1998 **7**:822–832.

[76] Paget C, Levin K, Delebarre C. Actuation performance of embedded piezoelectric transducer in mechanically loaded composites. *Smart Materials and Structures* 2002 **11**:886–891.

Chapter 35
Novelty Detection

Lionel Tarassenko[1], David A. Clifton[1,2], Peter R. Bannister[1], Steve King[3] and Dennis King[3]

[1] *Department of Engineering Science, University of Oxford, Oxford, UK*
[2] *Oxford BioSignals Ltd, Abingdon, UK*
[3] *Rolls-Royce Civil Aero-Engines, Derby, UK*

1 INTRODUCTION

1.1 The need for novelty detection

The complexity of modern high-integrity systems is such that only a limited understanding of the relationships between the various system components can be obtained. An inevitable consequence of this high degree of system complexity is the large number of possible failure modes, the effects of which on observable (sensor) data are often poorly defined. To compound this, examples of abnormal behavior in high-integrity systems are few and far between; usually, there are insufficient examples of failure to construct accurate fault-detection systems. As a result, conventional fault-specific failure-detection schemes are usually limited to identifying a small subset of known, well-understood modes of failure.

An alternative to identifying rare and unexpected modes of failure is the *novelty detection* paradigm [1–5], in which a model of normality is constructed from normal system data. Departures from normal behavior are classified as novel events. Novelty detection is alternatively known as *one-class classification* [6] or *outlier detection* [7].

1.2 Overview of novelty detection methods

The concept of "novelty" can be related to the probability of observing data that do not belong to the distribution characterizing normal data. Fundamental to this approach is the assumption that normal data are generated from an underlying data distribution, which may be estimated from example data. The classical approach to this estimation problem is based on the

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

use of density-estimation techniques to determine the underlying data distribution [8]. The resultant distribution may then be thresholded to define the boundaries of "normal" areas of data space [3].

With *parametric* techniques, assumptions are made about the form of the underlying data distribution, and the parameters of the distribution are estimated from observed data. The most commonly used form of distribution for continuous variables is the Gaussian distribution, which is defined by its mean and variance parameters. These parameters are estimated from the data using the *maximum likelihood* method, which has a closed-form analytical solution for a Gaussian distribution. More complex forms of data distribution may be modeled by using Gaussian mixture models [9, 10] or other mixtures of different types of distributions such as the gamma distribution [11, 12]. Model parameters are again estimated from data by maximizing the likelihood, but this now requires the use of numerical techniques such as the expectation-maximization algorithm [13]. Mixture models, however, can suffer from the requirement of large numbers of training examples to estimate model parameters accurately [5].

A further limitation of parametric techniques is that the chosen functional form for the data distribution may not be a good model of the distribution that generates the data. *Nonparametric* approaches, which make as few assumptions as possible about the form of the distribution, include kernel density estimators and neural networks. With kernel estimators, the probability density is estimated using large numbers of kernels distributed over the data space. The estimate of the probability density at each location in data space relies on the data points that lie within a *localized* neighborhood of the kernel [10]. The kernel density estimator used in this article, Parzen windows, places a Gaussian kernel on *each* data point and then sums the local contributions from each kernel over the entire dataset. Hence there is no computation required while "training" the model, other than for storage of the dataset. This implies that the cost of estimating the data density grows with the size of the dataset, but this shortcoming is addressed in Section 4 below.

Neural networks are an alternative nonparametric method, although their main use is not in density estimation but in *multiclass* classification problems [14]. However, neural networks can also be used for novelty detection by generating artificial data around the normal data to simulate the patterns from an "abnormal" class [15]. This approach is fraught with danger, however, as it requires the use of strong assumptions about the distribution of abnormal data beyond the boundaries of normal data. Furthermore, any approach that depends on the generation of artificial data is plagued by the curse of dimensionality: a very large number of artificial patterns are required to populate high-dimensional spaces.

The self-organizing map (SOM), initially proposed for the clustering and visualization of high-dimensional data [16], provides an unsupervised representation of training data using a neural network. Various applications of the SOM to novelty detection have been proposed [17, 18]. In this article, we present results from our use of the Sammon map and its parameterized version, NeuroScale, to visualize high-dimensional feature vectors, but this is purely for data understanding prior to the design of an appropriate density estimator for novelty detection.

Another type of classifier that has been adapted for novelty detection is the support vector machine (SVM), in which a number of hyperplanes are found that best separate data from different classes, after their transformation by a kernel function [19]. In the application of an SVM to one-class classification or novelty detection, two main approaches have been taken. The first finds (in the transformed space) a hypersphere of minimum radius that best surrounds most of the normal data [20]. The second approach separates the normal data from the origin with maximum margin [21]. In work related to that described in this article, we have extended this latter approach to novelty detection in jet-engine vibration data [22].

Hidden Markov models (HMMs), which include temporal dependence through the use of a state-based representation updated at every time step, are ideal for novelty detection in a sampled *time series*. While the features are directly observable, the underlying system states are not, and hence they are "hidden". The transitions between the hidden states of the model are governed by a stochastic process [23]. Each state is associated with a set of probability distributions describing the likelihood of generating observable "emission" events. These distributions may be thresholded to perform novelty detection [24]. Novelty detection with HMMs may also be achieved by constructing an "abnormal" state, a transition into

which implies abnormal system behavior [25]. A related state-based approach to novelty detection in time-series data relies on factorial switching Kalman filters [26]. This is a dynamic extension of the switched Kalman filter [27], which models time-series data by assuming that a continuous, hidden state is responsible for data generation, the effects of which are observed through a modeled noise process. Again, an explicit "abnormal" mode of behavior is included within the model, which is used to identify departures from normality. The weakness of this approach is that it requires assumptions to be made about the distribution of features for the "abnormal" state.

In this article, rather than using dynamic models such as HMMs or Kalman filters, we focus instead on *static* novelty detection, in which each pattern is treated independently. Our overall approach is outlined in Section 1.3.

1.3 A framework for novelty detection

The first step in the process of novelty detection, as with any other pattern-recognition technique, is *feature extraction* (Section 2). The aim here is to derive features that characterize normality but which are also likely to be affected by the occurrence of abnormal events. This usually requires some prior knowledge of the system under study.

We next make use of *data visualization* techniques (Section 3) for increasing our understanding of the data. In particular, we are interested in how the D-dimensional feature vectors (where D is the number of features) are distributed over the space of normal data, especially near the boundaries of normality.

Normality is then characterized by learning the probability density function (pdf) $p(x)$ of normal feature vectors using a nonparametric method, as it is best to make as few assumptions as possible about the distribution of normal data. In Section 4, we describe how we use Parzen windows as a nonparametric density estimator, and show how we deal with the problem of having very large numbers of feature vectors in the training dataset.

To classify new examples of vibration spectra as either "normal" or "abnormal", a *novelty threshold* has to be placed on the value of $p(\mathbf{x})$ for the new feature vector. We address the question of how to set the novelty threshold in principled fashion by introducing extreme value statistics in Section 5. These are methods that allow us to explicitly model the abnormal areas at the boundaries of normal feature space given the normal data.

1.4 Application of novelty detection to jet-engine vibration data

Throughout this article, we illustrate our approach to novelty detection in the context of jet-engine health monitoring, using vibration data from the development program for a three-shaft Rolls-Royce aerospace gas-turbine engine. The dataset covers 135 test flights, spanning a period of six months. The data were acquired over a wide range of operating conditions, and the behavior of the engine is characterized by three distinct periods:

- period A: normal, consistent operation of the engine from flights 1–127;
- period B: a change in engine condition, as a component from one of the three engine shafts works its way loose and moves freely inside the engine between flights 128 and 133;
- period C: a final engine event during flight 134, as the component is ejected through another one of the engine shaft stages and the engine surges. A subsequent engine test ("flight 135") is undertaken on the ground for experimental purposes.

We show that, by modeling normality using vibration parameters acquired at the start of period A, it is possible to detect both the initial shift in engine behavior due to the loose component (period B) and the transient event associated with the final component ejection during period C.

2 PREPROCESSING AND FEATURE EXTRACTION

2.1 Introduction

To prepare a real-world dataset for analysis, *preprocessing* is invariably performed. This is necessary because the analysis of sensor signals in their original form usually yields poor results. The main issues to consider are as follows:

- **Data quality**

The acquisition of data for novelty detection is susceptible to noise, both from the environment in which the system is operating and from the characteristics of the sensor. To facilitate analysis, noise-reduction techniques are applied. Typically, this involves filtering of the acquired data according to prior knowledge (or estimation) of the noise process.

In the example of the gas-turbine engine introduced in Section 1, analog measurements of the amplitude of engine vibration are made, using sensors mounted on the engine casing. These signals are initially low-pass filtered to remove high-frequency noise. Prior to quantizing these analog signals into digital signals for computerized analysis, antialiasing filtering is applied to the time-domain waveform to prevent errors in digitization.

- **Suitability of the data domain**

The domain in which the data are acquired may not be best suited for characterizing differences between "normal" and "abnormal" behavior of the monitored system. Transforming the data into a new domain can often allow more effective novelty detection.

For example, a gas-turbine engine contains rotating shafts and clearly exhibits peaks in vibration amplitude at the fundamental rotational frequency of those shafts when observed in the frequency domain. These vibration amplitudes can be used (as described later) to characterize the condition of the engine. However, as data are acquired in the time domain, these characteristics are not easy to identify. In this case, acquired data are transformed into the frequency domain using short-time fast Fourier transforms (FFTs).

- **Tractability**

Novelty detection problems can involve large datasets that are intractable to analyze in terms of both computing time and memory storage requirements. This is typical for datasets in which samples are obtained at a high sampling rate, and for datasets that consist of channels of data obtained from many sensors. To make such analysis tractable, some form of data compression must be performed to reduce the size of the dataset. In most cases, data compression involves the loss of information. The challenge is to ensure that sufficient information is retained such that "normal" and "abnormal" behavior may be characterized.

Techniques such as down-sampling or data averaging can be applied to reduce the size of datasets for which the sampling rate is particularly high, though it is critical to retain sufficient information to allow novelty detection.

The process of *feature extraction*, described in the next subsection, provides a means of reducing data size, while still retaining sufficient information to separate "normal" and "abnormal" behavior, provided that appropriate features are chosen.

2.2 Feature extraction

Feature extraction is the representation of signals using a smaller set of quantities, termed *features*. Prior knowledge of the system for which novelty detection is to be performed can often be used for *feature selection*. For example, if an engine shaft rotates at 50 Hz, a peak in vibration energy occurs at 50 Hz, with corresponding harmonics occurring at multiples of 50 Hz. Engine manufacturers use the values of vibration amplitude at the fundamental frequency and its harmonics (termed *tracked orders*) as features deemed suitable for characterizing the condition of the engine.

The gas-turbine engine used as an exemplar within this article contains three shafts that rotate concentrically, termed the *low-pressure* (LP), *intermediate-pressure* (IP), and *high-pressure* (HP) shafts. Air at atmospheric pressure enters the engine at the LP shaft, is then forced into the smaller IP shaft, and then into the yet smaller HP shaft, increasing in pressure at each stage. The HP shaft forces the air into the combustion chamber at many times atmospheric pressure, where it is mixed with fuel and ignited to generate thrust.

The tracked order corresponding to the fundamental rotational frequency of the LP shaft is termed $1LP$, and its harmonics are termed $2LP, 3LP, 4LP$, etc. Harmonics also occur at $0.5 \times$ fundamental frequency ($0.5LP$), and at $1.5 \times$ fundamental frequency ($1.5LP$). In practice, harmonics above $4LP$ are not usually considered owing to their low amplitude. Similarly, tracked orders for the IP and HP shafts are termed $1IP, 1HP, 2IP, 2HP$, etc.

2.3 Features for in-flight novelty detection

For the example dataset considered in this article, analog vibration measurements recorded by sensors mounted on the engine casing are used to compute 1024-point FFTs every 0.2 s. Using measurements of the shaft speeds recorded by tachometers mounted on the shafts, the various tracked orders are located as peaks in the 1024-point FFT. Each vibration spectrum is reduced to a small number of extracted features (the vibration amplitudes at the fundamental frequency and its harmonics), for each of the three shafts, thereby removing the need to store the original spectra.

Thus, feature extraction here results in a time series of tracked order amplitudes at 0.2 s intervals. To remove spurious transients in the data, these time series of features are smoothed using a three- or five-point median filter.

2.3.1 Constructing a feature vector

After feature extraction, a single *feature vector,* $\mathbf{x}_i$, is constructed for each shaft. For example, the feature vector for the IP shaft is

$$\mathbf{x}_i = [0.5IP, 1IP, 1.5IP, 2IP, 3IP, 4IP, RE] \tag{1}$$

which is a 7-D feature vector formed from six of the IP-shaft tracked orders and the *residual energy RE*, defined to be

$$RE = E - T_{\mathrm{LP}} - T_{\mathrm{IP}} - T_{\mathrm{HP}} \tag{2}$$

where E is the total signal energy in the vibration spectrum (the sum of all the energy values in the first 512 values of the 1024-point spectrum), and T_{S} is the total energy in all the tracked orders related to shaft S. T_{S} is calculated by summing the energy values around the spectral peaks corresponding to the six tracked orders. Typically, this is performed over a very small number of frequency bins (three or five) around each peak. The residual energy is included as a seventh feature to detect gross changes in vibration energy not captured by the tracked orders.

The feature vector $\mathbf{x}_i$ for the IP shaft is used for the remainder of this article to illustrate in-flight novelty detection. (The LP or HP shaft feature vectors could equally well have been chosen.) A feature vector is computed every 0.2 s, allowing novelty detection to occur in real time, the aim being to provide early warning of abnormal engine operation during a flight.

2.3.2 Utilizing system expertise

We have considered the use of prior system knowledge for selecting features to extract from original data, in this case, the vibration tracked order values for a shaft concatenated into a single feature vector. It is known, however, that engine condition varies throughout the phases of a typical flight, from takeoff, through climbing, to cruising, and then landing, and the levels of vibration for each shaft are different for each flight regime or phase. It is not a trivial exercise to partition the time series of feature vectors according to flight regime automatically because of the speed overlap between the different regimes. Instead, the total speed range for each shaft is divided into four speed bands of equal width, $\{\mathrm{S1}, \ldots, \mathrm{S4}\}$, with a separate model constructed for each speed band.

2.3.3 Acquiring a balanced dataset

When building a dataset for training a data-driven model, it is important that data are acquired as uniformly as possible over the system's entire *operating range*. A jet engine spends much of its time operating at "cruise" conditions, with much less time spent operating under takeoff or landing conditions. If training data were drawn randomly from the entire *flight*, then this would result in a strong bias toward cruise conditions, simply because there would be many more feature vectors derived from periods of cruise operation.

A balanced dataset is constructed by rejecting feature vectors during steady-state conditions if the change in engine speed with respect to that associated with the previous feature vector is below a given threshold. The effect of this is shown in Figure 1, in which a histogram of feature vector speeds from the original dataset shows a very dominant peak in the range from 80 to 85% of maximum shaft speed, corresponding to cruise conditions. After rejection of consecutive feature vectors with similar speeds, the distribution is much more uniform across the whole speed range. This type of approach is used

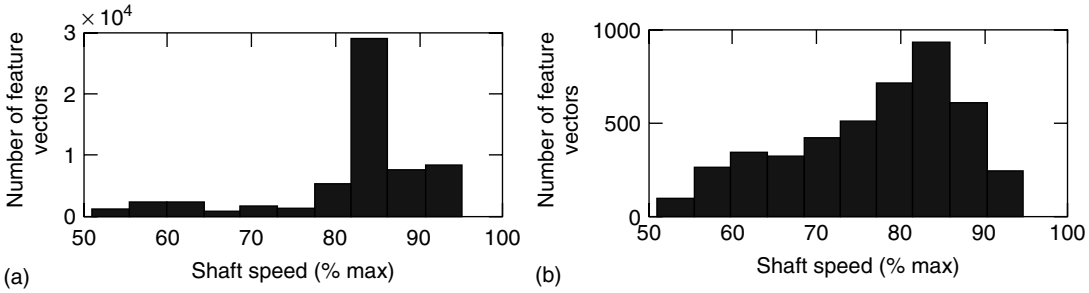


Figure 1. Histograms showing the effect of introducing a criterion for rejecting feature vectors that are too similar to each other during steady-state conditions. The histogram corresponding to the original dataset is shown in (a), with a predominance of speeds between 80 and 95% of maximum shaft speed. Speed-based rejection gives the more uniformly distributed histogram (b). Note the change in scale on the y axis from 30 000 (a) to 1000 (b).

to construct a balanced dataset of 7-D feature vectors acquired during a number of consecutive flights. The use of this balanced dataset for training in-flight models of normality is described in Sections 4 and 5 of this article.

2.4 Features for flight-by-flight novelty detection

As a complement to the real-time on-the-engine monitoring described in Section 2.3, there is also a need to perform novelty detection on a flight-by-flight basis. The latter approach is adopted for ground-based analysis of summaries of engine condition throughout an entire flight. A *vibration signature* for the entire flight can be constructed for each tracked order, using a speed-based representation in which the complete range of shaft speeds is divided into subintervals (*speed bins*) of equal width. Within each speed bin, the vibration amplitude of the corresponding tracked order is averaged across the entire flight.

An example of this is shown in Figure 2, in which a flight summary of 1*IP* vibration amplitude is shown,

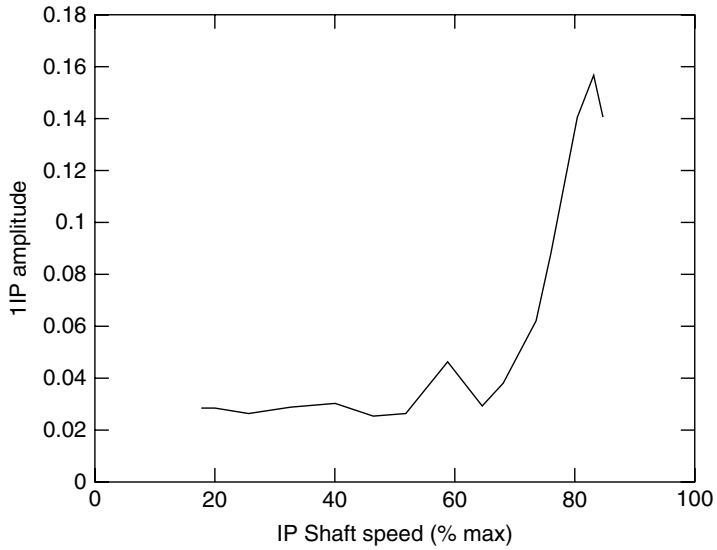


Figure 2. The 1*IP* vibration signature, summarizing 1*IP* vibration amplitude throughout a flight, using a 20-D feature vector.

displayed against IP-shaft speed. The range of shaft speeds has been divided into 20 speed bins, resulting in a 20-D feature vector.

2.5 Normalization

In general, we cannot assume that each feature in the vector has the same dynamic range. It is therefore desirable to normalize feature vectors such that their individual elements may be compared, regardless of their absolute values, while preserving information necessary for novelty detection. Though many normalization schemes exist for normalizing a D-dimensional feature vector $\mathbf{x}_n = [x_n^1, \ldots, x_n^D]$ from a set of N feature vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ [28], the most appropriate, in practice, is *component-wise normalization*, defined to be

$$\mathbf{x}'_n = \left[\frac{x_n^1 - \mu_1}{\sigma_1}, \frac{x_n^2 - \mu_2}{\sigma_2}, \ldots, \frac{x_n^D - \mu_D}{\sigma_D}\right] \tag{3}$$

where μ_d and σ_d are the mean and standard deviation of element d across the whole dataset:

$$\mu_d = \frac{1}{N}\sum_{i=1}^{N} x_i^d \quad \sigma_d = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i^d - \mu_d)^2} \tag{4}$$

3 VISUALIZATION

3.1 Introduction

The process of feature extraction leads to feature vectors $\mathbf{x}$ that combine D features extracted from original data. Typically, D is large; for example, feature vectors from the in-flight engine dataset have $D = 7$ dimensions, and feature vectors from the flight summary dataset have $D = 20$ dimensions. The initial analysis of datasets containing high-dimensional feature vectors is performed using *visualization* techniques that allow us to explore the structure of the datasets.

In visualization, feature vectors are mapped from their original D dimensions into a 2- (or 3-) D space such that each vector is mapped onto a corresponding single point in this visualization space. Feature vectors similar to one another in D-dimensional space map to points that are close to one another in the 2D space, while feature vectors that are significantly far apart in D-dimensional space should map to points that are far apart in the 2D space. As a result, groups of similar feature vectors form clusters in the visualization space.

In datasets suitable for the application of novelty detection techniques, feature vectors that correspond to "normal" data (which are often similar) form clusters. Feature vectors that correspond to "abnormal" data (which are dissimilar to "normal" feature vectors) appear as outlying points, separated from the normal clusters. Thus, we can determine the effect of selecting various features for inclusion in the feature vectors by examining the separation of "normal" and "abnormal" data in the visualization, with the goal of selecting features that result in maximum separation.

Similarly, the effects of various methods of normalizing the feature vectors can be determined using visualization, again with the goal of choosing a normalization method that maximizes separation between "normal" and "abnormal" data, such that abnormal data may be clearly identified.

3.2 Mapping with NeuroScale

To map N feature vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ from their original D-dimensional space (the *data space*) into 2-D space (the *visualization* or *latent space*), we consider the distance $d_{i,j}$ between two feature vectors $(\mathbf{x}_i, \mathbf{x}_j)$ in data space and the distance $d^*_{i,j}$ between the corresponding pair of mapped points (y_i, y_j) in latent space. The *Sammon stress metric* is defined to be the difference between the Euclidean distance $d_{i,j}$ in the high-dimensional space and the Euclidean distance $d^*_{i,j}$ in the low-dimensional space, summed over all N feature vectors:

$$E_{\text{sam}} = \sum_{i=1}^{N}\sum_{j>i}^{N}(d_{i,j} - d^*_{i,j})^2 \tag{5}$$

The NeuroScale training algorithm seeks to minimize the error function E_{sam} to best preserve the original interpoint distances $d_{i,j}$ after mapping the points into latent space [29]. The mapping is learnt,

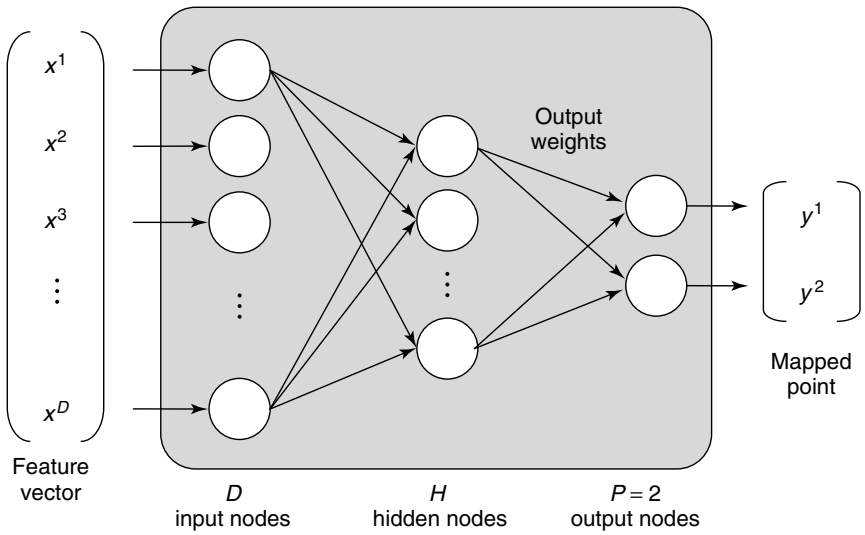


Figure 3. Mapping a D-dimensional feature vector into two dimensions using a NeuroScale neural network.

using a radial basis function (RBF) neural network with D inputs, $P = 2$ output nodes, and a single hidden layer of H hidden nodes, as shown in Figure 3.

Each of the H nodes in the hidden layer corresponds to a basis function in D-dimensional data space. H is typically selected to be an order of magnitude greater than the dimensionality, D, of the feature vectors, and an order of magnitude smaller than the number, N, of feature vectors. For example, if $D = 10$ and $N = 1000$, an appropriate number of hidden nodes would be $H = 100$.

To construct the NeuroScale mapping, the parameters of the H basis functions and the values of the output weights (i.e., those from the hidden layer to the output layer) must be determined. To achieve this, the neural network is first initialized, and then trained using the *training set* of N feature vectors as described below:

1. Initialization

The initial centres of the H radial basis functions are set to be the D-dimensional co-ordinates of H feature vectors randomly selected from the training set. There are a number of algorithms which may be used for setting the initial width of the H radial basis functions. Typically, the width is set to be the distance between the two centers that are furthest apart to achieve good coverage of input space.

The output weights from the H hidden nodes to the P output nodes are initialized using principal component analysis (PCA). The eigenvectors of the training set covariance matrix with the largest eigenvalues are found. These *principal components* are the P vectors in data space that account for maximum variance in the data, and they are used as the *initial* values for the target outputs from which the initial values of the output weights are computed [14].

2. Training

Training the RBF network is a two-stage process. In the first stage, the parameters of the basis functions are set so that they approximately model the unconditional data density of the training set. In the second stage, the output weights are learnt by optimizing the Sammon stress metric using methods from linear algebra [30].

Once training is complete, feature vectors may be visualized by presenting them at the input layer of the NeuroScale network, which then gives the corresponding 2-D coordinates of the mapped feature vector at its output.

3.3 Application to in-flight engine data

3.3.1 Selecting a "normal" training set of feature vectors

To examine how the 7-D feature vectors extracted from in-flight vibration data vary from flight to flight, a NeuroScale network was trained for each of the speed bands $\{S1, \ldots, S4\}$ using the process described above. The training set in each speed band consisted of feature vectors acquired during a set of "normal" flights, as explained below.

Flights 1–30 corresponded to recordings during which the engine did not exceed takeoff speed. Thus, these flights could not be used for training NeuroScale networks for each of the speed bands. Flights 31–79 were the first "normal" flights to cover the full range of shaft speeds, of which 14 flights contained full 7-D data. These 14 flights were therefore selected to provide the feature vectors used to train the NeuroScale network for each speed band.

3.3.2 Training the NeuroScale network

The optimization of the NeuroScale network requires the minimization of E_{sam}. There are $N(N-1)/2$ distances to consider for a training set of N feature vectors, and so the computational demands of the training process increase as $O(N^2)$. The set of 14 flights selected in the training set contains $N1 = 8770$, $N2 = 14\,295$, $N3 = 28\,232$, and $N4 = 4603$ feature vectors in speed bands S1 to S4, respectively. For the training time and computational requirements to be manageable, each of the four training sets is first reduced to a smaller representative set of feature vectors, using k-means clustering.

The k-means clustering algorithm [23] is an iterative method for producing a set of cluster centers μ_j (for $j = 1, \ldots, k$) that represents the distribution of a (typically much larger) set of feature vectors $\mathbf{x}_i$. For the purposes of training NeuroScale networks for the speed bands S1 to S4, we reduce the training set of N1 to N4 feature vectors to $k = 500$ cluster centers in each case, as shown in Figure 4.

After application of the k-means algorithm, each original training set with $\{N1, \ldots, N4\}$ feature vectors is reduced to a new training set consisting of the $k = 500$ cluster centers. Using the rule of thumb that the number of nodes H in the hidden layer of the NeuroScale network should be an order of magnitude smaller than the size of the training set, we set $H = 50$. The network is then trained as described above.

To confirm the appropriateness of the $k = 500$ cluster centers, we can compare the visualization of the cluster centers with that of the original feature vectors.

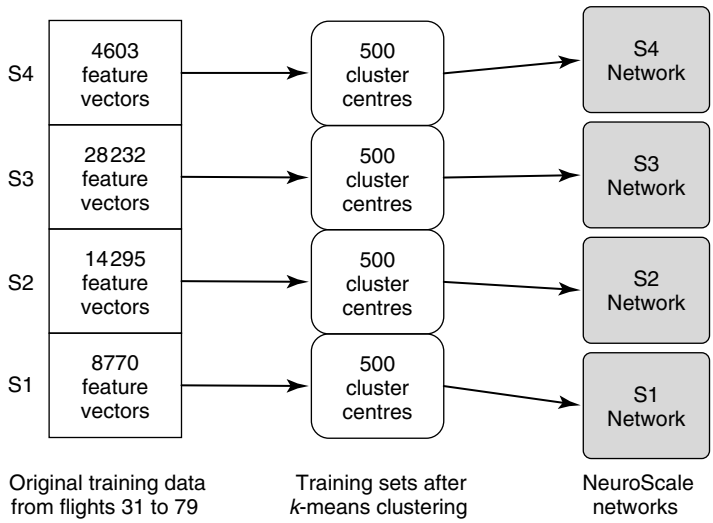


Figure 4. Training NeuroScale visualization networks for each of the four speed bands $\{S1, \ldots, S4\}$, using training sets of $k = 500$ cluster centers derived from the $\{N1, \ldots, N4\}$ feature vectors.

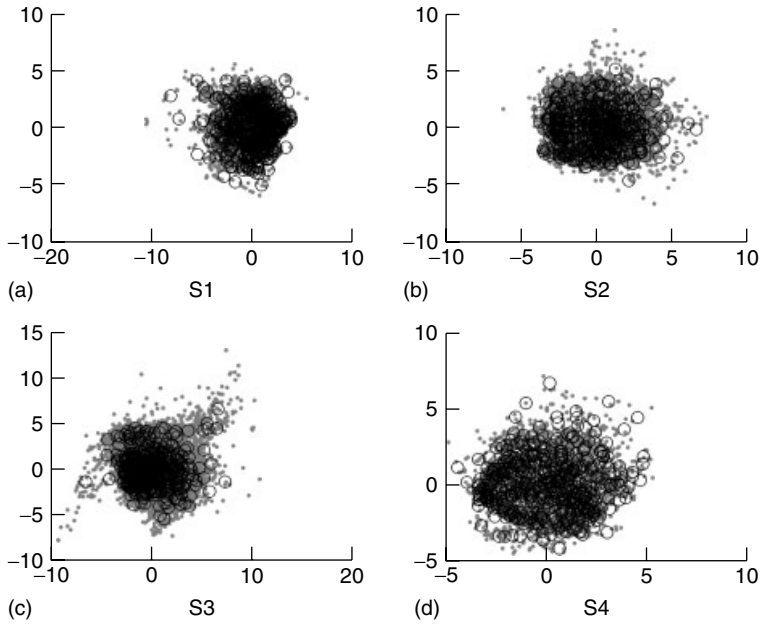


Figure 5. Visualization of 7D in-flight feature vectors in 2D, using NeuroScale. Data for the four speed subranges $\{S1, \ldots, S4\}$ are shown in (a–d), respectively. 7D cluster centers are shown as circles in the visualization. The projections of the original 7-D feature vectors from which the cluster centers were generated are shown as gray dots. In each case, the cluster centers are seen to be a valid representation of the original feature vectors.

Figure 5 shows visualizations of the $k = 500$ cluster centers, and the original data, for each of the four speed bands $\{S1, \ldots, S4\}$. In each plot, it can be seen that the 7-D feature vectors generally overlay the 7-D cluster centers (shown by circles) when visualized in two dimensions.

In any dimensionality-reduction mapping, the $N(N-1)/2$ intervector distances are not perfectly preserved. A small number of the original feature vectors lie slightly outside the main distribution of cluster centers, as may be seen in the bottom left of Figure 5(c), for example. As seen later, however, these distances from "normal" feature vectors to the nearest cluster centers are small compared to the distances from mapped "abnormal" feature vectors to the nearest cluster centers.

3.3.3 Visualizing test data

Figure 6 shows a visualization of the 7-D feature vectors from flight 131 within period B, in which a previously undetected change in engine condition occurred. The NeuroScale networks trained previously using data from "normal" flights are used to map the feature vectors from flight 131 into two dimensions.

Figure 6(a) shows feature vectors from speed band S2, with the cluster centers from the "normal" flights shown for comparison. It can be seen that some of the feature vectors within this speed band are not closely aligned with the cluster centers. A few are well separated, appearing in the lower left of the plot.

Figure 6(b) shows test data from speed band S3. Here, a large proportion of the test data is well separated from the "normal" cluster centers, indicating that majority of the feature vectors in this speed band appear to be considerably "abnormal". Speed band S3 includes that part of the flight in which the airplane is climbing, which engine experts describe as being a period during which evidence of any abnormal engine behavior is likely to become apparent.

Figure 7 shows the visualization of the 7-D feature vectors from flight 134 from period C, during which an engine event occurred. The same NeuroScale networks are used to map the feature vectors from the test flight into two dimensions.

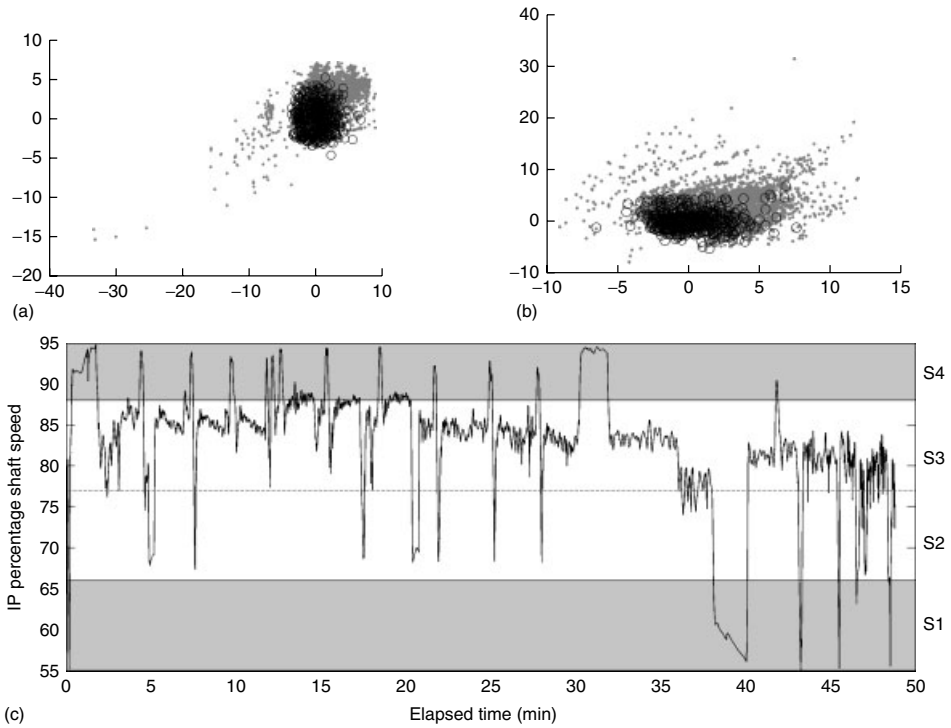


Figure 6. Visualization of in-flight feature vectors from flight 131 (shown as gray dots). 7-D cluster centers from "normal" flights are shown in the visualization as circles. (a) Feature vectors from speed band S2 do not lie close to the "normal" cluster centers, with a few extending out to the lower left of the plot. (b) Feature vectors from speed band S3 lie mostly away from the "normal" cluster centers. (c) Engine speeds throughout flight 131, showing speed bands $\{S1, \ldots, S4\}$.

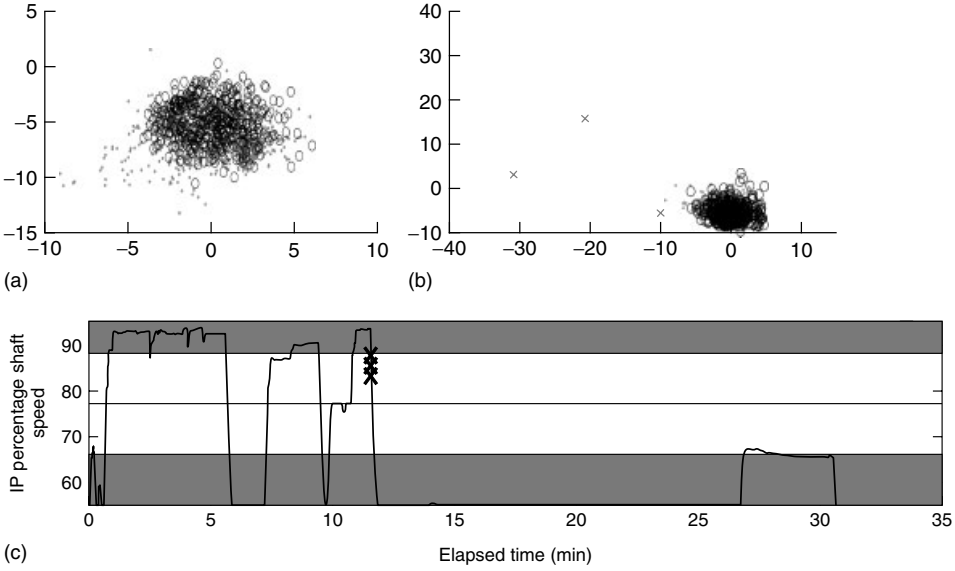


Figure 7. Visualization of in-flight feature vectors from flight 134 (shown as gray dots). 7-D cluster centers from "normal" flights are shown in the visualization as circles. (a) Feature vectors from speed band S2 extend significantly from the main cluster of "normal" centers. (b) Three of the feature vectors from speed band S3 (marked by gray crosses) lie even further from the "normal" cluster centers. (c) Engine speeds throughout flight 134, showing speed bands $\{S1, \ldots, S4\}$.

Figure 7(a) shows the visualization of the feature vectors from speed band S2 with respect to the "normal" cluster centers. Again, feature vectors within this speed band are separated from the normal cluster centers, extending into the lower left of the plot.

Figure 7(b) shows the visualization of the feature vectors from speed band S3. Retrospective analysis of the original IP-shaft vibration data by engine experts showed that the damage caused by the loose component led to an engine surge early in the flight (after about 11 min). The engine deceleration that follows shortly afterward—marked by three crosses on (c)—gives rise to a highly abnormal vibration pattern—also marked by three crosses on (b).

3.3.4 Visualizing flight-by-flight summaries using 7-D vectors

We have so far described the visualization of feature vectors from individual flights, comparing them with the set of cluster centers previously derived from "normal" flights. We describe here how the same visualization method can also be used for the direct comparison of flights.

Given N 7-D feature vectors $\mathbf{x}_n = [x_n^1, \ldots, x_n^7]$ from a flight, we can obtain a summary representation for that flight, a single vector $\hat{\boldsymbol{\mu}}$, which we define as follows:

1. For the N feature vectors in the flight, compute the component-wise mean vector $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \frac{1}{N}\left[\sum_{n=1}^{N} x_n^1, \sum_{n=1}^{N} x_n^2, \ldots, \sum_{n=1}^{N} x_n^7\right] \quad (6)$$

2. The mean $\boldsymbol{\mu}$ is a "synthetic" feature vector. The best equivalent representation for the flight data is the feature vector $\hat{\boldsymbol{\mu}}$ from $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ that is closest to $\boldsymbol{\mu}$, using Euclidean distance inversely weighted by the variance in each dimension as the distance metric.

Figure 8 shows a visualization of the $\hat{\boldsymbol{\mu}}$ vectors from speed band S3, for all flights. The $\hat{\boldsymbol{\mu}}$ vectors from the "normal" period A form a cluster to the left of the plot, while the $\hat{\boldsymbol{\mu}}$ vectors from periods B and C lie to the right of this cluster, indicative of the abnormal vibration behavior of the IP shaft in these latter flights.

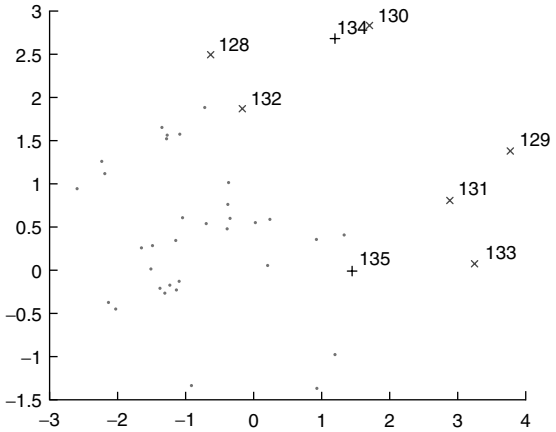


Figure 8. Visualization of $\hat{\boldsymbol{\mu}}$ vectors for each flight. Vectors from periods A, B, and C are shown as dots, crosses, and plus symbols, respectively. It can be seen that the $\hat{\boldsymbol{\mu}}$ vectors from period A form a cluster to the left, while vectors from periods B and C lie to the right of this cluster, corresponding to the abnormal IP-shaft vibration behavior observed during these flights.

The NeuroScale visualization shown here is only an approximate summary of the "average" behavior of IP-shaft vibration in one speed band. The optimal summary representation of a flight is the 20-D vibration signature for each shaft across the entire speed range, and the NeuroScale visualization techniques are applied to this representation in the next subsection.

3.4 Visualization of 20-D flight summary signatures

A 20-H-2 NeuroScale network was trained to map the 20-D vibration summary signatures for the IP shaft into a 2-D visualization space. As with the in-flight data in Section 3.3, the network was trained using "normal" flights $\{31, \ldots, 79\}$ from period A, and thus the number of examples in the training set is 49. The number of nodes in the hidden layer of the network was set to $H = 30$.

Note that the rule of thumb given in Section 3.2, in which the value of H is set to be an order of magnitude greater than the number of input nodes and an order of magnitude less than the number of

examples in the training set, cannot be met in this case. The best practice is to train several networks using different values of H, and compare the resultant visualizations to ensure that an appropriate value of H is selected for use. In practice, NeuroScale visualizations are robust to changes in H, and it is often possible to construct useful visualization maps despite the number of available training examples being not much greater than the number of inputs.

Figure 9 shows the result of using the trained NeuroScale network to visualize the 20-D vibration signatures for the IP shaft from all flights. It can be seen that signatures from "normal" flights (in period A, plotted with dots) form a cluster in the upper right quadrant of the visualization map, close to the (0, 0) point in the map. This indicates that the signatures are similar in 20-D space, and thus the behavior of the IP shaft during these "normal" flights is consistent.

Signatures in which a change of engine condition took place (in period B, plotted with crosses) are seen to be well separated from the cluster of signatures from "normal" flights, highlighting departure from normality in the IP-shaft vibration signature.

The signature corresponding to the engine event (flight 134 in period C) is the furthest away from the cluster of "normal" signatures, confirming that the highly abnormal vibration behavior of the IP shaft during this flight was captured in this summary signature.

4 CONSTRUCTING MODELS FOR NOVELTY DETECTION

The objective of a novelty detection system is the generation of reliable and robust alerts if the condition of the system being monitored is deemed to have deteriorated. The techniques described in the previous section allow us to visualize the data to be used to construct the model of normality.

4.1 Parzen window model of normality

The Parzen window kernel density estimator method [31] is the model adopted here to estimate the pdf, $p(\mathbf{x})$, for the training (normal) data. With this method [10], $p(\mathbf{x})$ is estimated using the following steps:

1. Locate a hyperspherical Gaussian window, or kernel, with width σ, on *each* of the D-dimensional feature vectors in the training

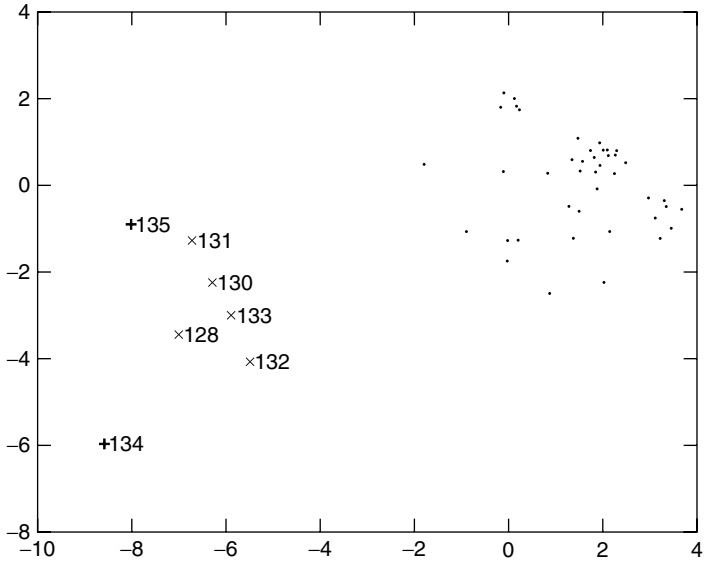


Figure 9. Visualization of 20-D flight summary signatures. Signatures from periods A, B, and C are shown by dots, crosses, and plus symbols, respectively. Signatures from periods B and C are clearly separated from the signatures from "normal" period A.

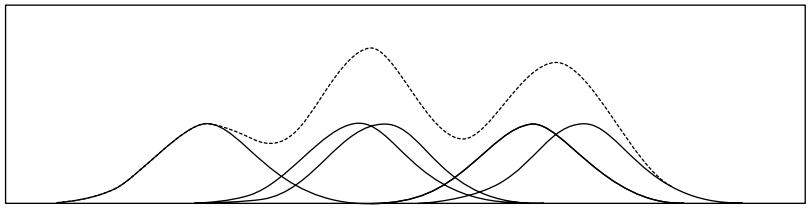


Figure 10. A simple 1-D example of the Parzen windows method of estimating a probability density function. Five Gaussians (solid lines) of uniform width, centered on five observations of a random variable x, are shown along with their sum (dashed line). Wherever the data density is highest, the probability density function (pdf) estimate takes a higher value.

dataset, $\mathbf{x}_i$, where $i = 1, \ldots, N$; (a one-step "training phase").

2. Evaluate the sum of the Gaussian distributions using the squared Euclidean distances between the test feature vector $\mathbf{x}$ and the training vectors $\mathbf{x}_i$, normalized by a factor that ensures $p(x)$ integrates to 1.

This gives the following formula for the estimate of $p(\mathbf{x})$:

$$p(\mathbf{x}) = \frac{1}{N(2\pi)^{D/2}\sigma^D} \sum_{i=1}^{N} \exp\left\{\frac{-\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right\} \tag{7}$$

By placing a Gaussian kernel over each feature vector $\mathbf{x}_i$ in our training dataset, we construct a probability density estimate of $p(\mathbf{x})$ that will have a higher value of p where the concentration of training data is greatest. This is illustrated in the simple 1-D example shown in Figure 10.

There are two parameters that must be carefully chosen for an accurate estimation of the pdf of the training (normal) data according to the above equation: the number N of training vectors and the width parameter σ. N must not be too large to avoid undue computational complexity. Hence the number of training patterns is reduced to 500 by using the k-means clustering algorithm as described in Section 3.3 and illustrated in Figure 4. Thus the 500 cluster centers used for training the NeuroScale network for speed band S3 (as shown in Figure 4) are also the N feature vectors $\mathbf{x}_i$, where $i = 1, \ldots, N$, used in the Parzen windows model of normality.

If desired, some of the outliers in the visualization map can be removed from the set of N Gaussian kernels used in the Parzen windows model of normality, so that only the "most normal" feature vectors are used to define normality. This strategy was not adopted here and all 500 cluster centers were retained for defining normality.

The width parameter σ acts as a smoothing parameter for the estimator. If its value is too large, local variation in the D-dimensional space is not captured. If the value is too small, the estimate of the pdf is too noisy and follows the data too closely (overfitting). There are a number of methods for setting the value of σ; for example, by using a validation set. Here, we use a simple heuristic method whereby σ is set to be the average distance of the 10 nearest neighbors from each vector in the (normal) training dataset, averaged across all 500 feature vectors [2].

Novelty increases as $p(\mathbf{x})$ decreases; we define the *novelty score* as $\log_e 1/p(\mathbf{x})$ which is equivalent to $-\log_e p(\mathbf{x})$. Thus feature vectors a long distance away from the Gaussian kernels centered on the normal training vectors will have a low value of $p(\mathbf{x})$ and hence a high novelty score.

4.2 Application to in-flight vibration data (7-D feature vectors)

A Parzen window model of normality is first constructed for the 7-D vibration feature vectors introduced in Section 2.3 (IP-shaft tracked orders in speed band S3, plus residual energy, to which component-wise normalization is applied). The 500 cluster centers are selected, using the k-means clustering algorithm, from the 28 232 feature vectors available in speed band S3 in the data from flights 31–79 (Figure 4) and visualized in 2-D in Figure 5(c). The width σ is set using the 10 nearest-neighbor heuristic described above.

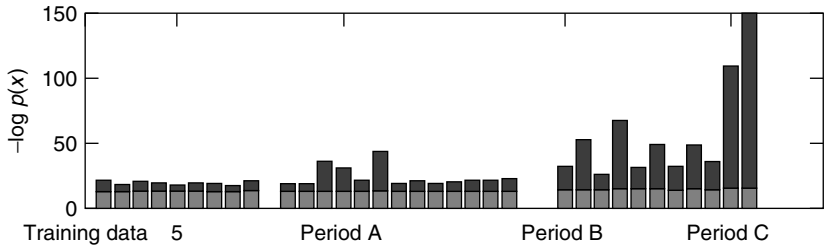


Figure 11. Novelty scores (mean shown in light gray and max in dark gray) for 7-D vibration feature vectors, computed using a Parzen windows model of normality for the IP shaft based on 500 cluster centers. The model was constructed using the same 500 prototype vectors as in the visualization studies of Section 3.3. These vectors were extracted from the IP-shaft vibration data in speed band S3 between flights 31 and 79 (shown as training data in the above figure).

The mean and maximum novelty scores for each flight are calculated using the Parzen window estimate of $p(x)$. These scores are shown in Figure 11 for the training data (flights 31–79), the rest of period A (up to flight 127), the eight flights from period B (flights 128–133), and finally the two "flights" from period C (flights 134 and 135). The deterioration at the start of period B is evident from the increase in maximum novelty score. The final event, which causes the engine to surge in flight 134 (penultimate flight on Figure 11) has an even greater effect on this parameter.

Figure 12 shows the novelty scores computed every 0.2 s during two flights (flight 80 from period A and flight 134 from period C), using the same 500-center Parzen window model of normality for the 7-D vibration feature vectors for the IP shaft.

The consistently low novelty scores throughout flight 80 (apart from one noise transient) are evidence of the normal condition of the engine at this point in the test flight program. The novelty scores recorded for a short period of time during flight 134, after which an engine shutdown was carried out, are far in excess of the scores seen in earlier flights.

4.3 Application to flight summary data (20-D vibration signatures)

A Parzen window model of normality can also be constructed for the 20-D vibration signatures that provide a flight summary for each shaft. The fundamental tracked order data for the IP shaft from flights 30–79, previously used to train the NeuroScale network of Section 3.4, are now used to construct a Parzen window model of normality for IP-shaft vibration, using the 49 available training signatures (i.e., $N = 49$, for this Parzen window model) and the same heuristic for setting the width σ for each of the 49 Gaussian kernels.

The resulting novelty scores for the 20-D vibration signatures shown in Figure 13 are significantly higher for the flights in periods B and C (flight 128 onward) than during period A when the engine was known to be operating at normal condition.

Thus the 20-D vibration signature model confirms the results obtained with the in-flight 7-D vibration model, with both novelty detection models detecting increased abnormality in the engine's vibration patterns for the IP shaft. The 20-D vibration model, which is based on a vibration signature averaged across the entire flight, provides a clearer distinction between normality (period A) and the onset of deterioration at the start of period B. Methods for setting the novelty threshold for the automatic detection of abnormality are considered in Section 5 of this article.

5 SETTING NOVELTY THRESHOLDS

5.1 Introduction

To decide whether a feature vector $\mathbf{x}$ is "normal" or "abnormal", we require a decision boundary, termed the *novelty threshold*. This threshold p_H can be defined using the unconditional probability distribution $p(\mathbf{x}) < p_H$. However, because $p(\mathbf{x})$ is a *distribution* function, it is necessary to integrate to give the cumulative probability $P(\mathbf{x})$. This is then used to

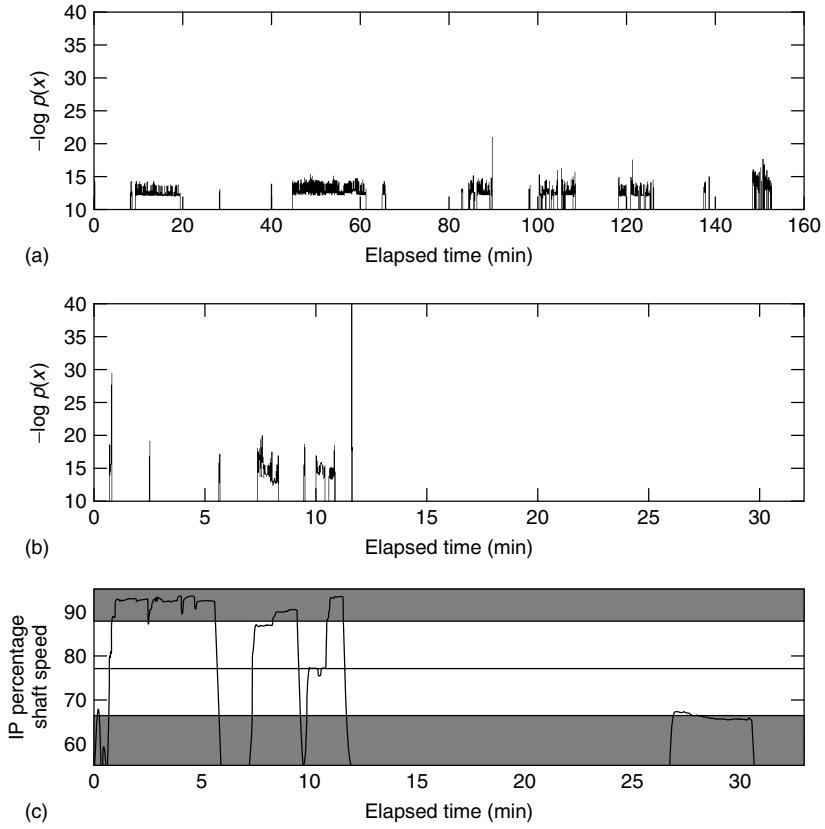


Figure 12. Novelty scores for the 7-D vibration feature vectors, computed using a Parzen window model of normality based on 500 cluster centers. Flight data from speed band S3 of the IP shaft acquired from flight 80 at the beginning of period A (a) produces low (albeit noisy) novelty scores. The acceleration during the climb at the beginning of flight 134 gives rise to a high novelty score for a short time at the beginning of plot (b), and an even higher score at the end (deceleration after surge). Plot (c) shows the engine speeds throughout flight 134, showing speed bands {S1, . . . , S4}.

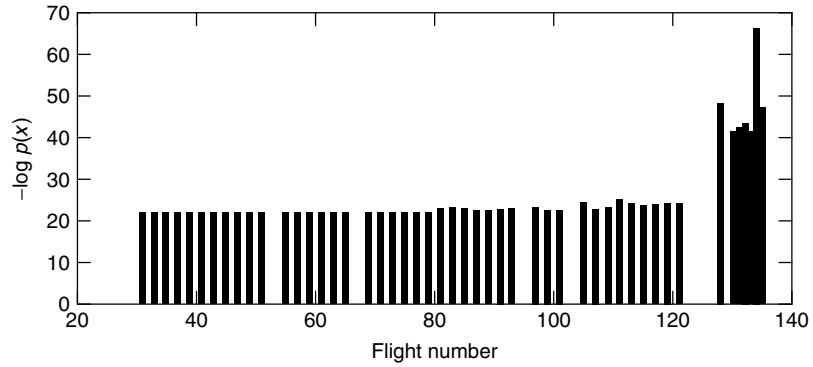


Figure 13. Novelty scores for the 20-D vibration signatures from each flight, computed using a Parzen window model of normality with $N = 49$. The model was trained using signatures from flights 31–79. Only those flights that had a sufficiently broad coverage of the speed range to create a valid 20-point summary signature are included.

set thresholds in relation to the actual probability of observing an abnormal event (e.g., $P(\mathbf{x}) = 10^{-6}$).

This section discusses the disadvantages of using conventional methods for setting novelty thresholds, and then describes a principled probabilistic approach suitable for novelty detection, illustrated using the models described in Section 4.

5.2 Disadvantages of conventional methods

Conventional methods of creating a probabilistic model of normality, as described in Section 4, are well suited for estimating the distribution of *normal* data. However, vectors close to the boundary of normality, which is defined by the novelty threshold, lie in regions of input space with very low data density (the tails of the $p(\mathbf{x})$ distribution). Conventional density-estimation methods may not accurately model the boundary of normality in input space.

5.3 Extreme value statistics

Extreme value theory (EVT) is a branch of statistics that provides a probabilistic method of directly estimating the boundaries of normality in input space. Given a set of normal training data $\mathbf{X} = \{\mathbf{x}_1 \ldots \mathbf{x}_m\}$, EVT estimates the probability distribution of the *maximum* of that set, $\max(\mathbf{X})$. The threshold is set according to where we believe the maximum of the normal data will occur, and thus provides a principled method of setting novelty thresholds.

The case studies described in this article focus on the detection of novelty in multidimensional data: the analysis of 7-D feature vectors of in-flight data and that of 20-D vibration signatures summarizing an entire flight. For the purposes of explanation, this section first considers the setting of thresholds with EVT using 1-D data.

Extreme value statistics originated in the field of civil engineering, used for modeling the likelihood of observing extreme loads in structures [32]. Here, we focus on "classical" EVT as previously used in novelty detection [33, 34] for biomedical applications.

According to the Fisher and Tippett theorem [35] upon which classical EVT is based, the distribution H describing the location of $\max(\mathbf{X})$ must belong to one of the following three families of extreme value distributions:

Type I (Gumbel):

$$H(y_m) = \exp\left(-\exp(-y_m)\right) \tag{8}$$

Type II (Fréchet):

$$H(y_m) = \begin{cases} 0 & \text{if} \quad y_m \leq 0 \\ \exp\left(-y_m^{-\alpha}\right) & \text{if} \quad y_m > 0 \end{cases} \tag{9}$$

Type III (Weibull):

$$H(y_m) = \begin{cases} \exp\left(-(-y_m)^{\alpha}\right) & \text{if} \quad y_m \leq 0 \\ 1 & \text{if} \quad y_m > 0 \end{cases} \tag{10}$$

where the Fréchet and Weibull distributions have a shape parameter α, and y_m is termed the *reduced variate*, being a linear transformation of the normal data x.

For novelty detection purposes, we have described the modeling of normal data using Parzen window density estimators in Section 4, which consist of a mixture of Gaussian kernels. Maxima from these Gaussian components are modeled using a Gumbel distribution [32, 33].

5.4 Setting novelty thresholds in univariate models

We first illustrate threshold setting for novelty detection using EVT with a univariate (1-D) example. Figure 14(a) shows a histogram of the maximum vibration levels for each flight observed within one of the speed bins in the 20-D vibration signature. A conventional density-estimation process is used to construct a model of normality, in which a distribution is fitted to the data. If we assume that the distribution of vibration levels in this speed bin is approximately Gaussian, the most likely distribution, given the data (the *maximum likelihood* distribution), is shown by the curved line.

A novelty threshold can then be set in the tails of this distribution. A typical threshold for this application is $P(x) = 1 \times 10^{-6}$ [36], which corresponds to $x = 0.018$. As can be observed from Figure 14(a), the gradient of the fitted distribution

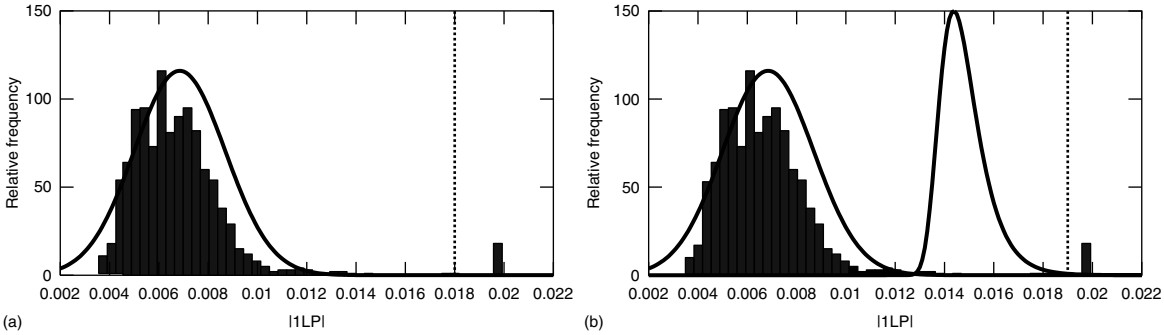


Figure 14. (a) Normal data are used to fit an estimated data distribution (curved line). A novelty threshold (dashed line) is set in the tail of this distribution. (b) An EVT distribution is estimated from the normal data (second curved line), which is then used to set a novelty threshold (dashed line).

is approximately zero when $x = 0.018$, making the position of the threshold sensitive to small changes in normal data around the mean of the distribution if the data are modeled with this single Gaussian distribution.

Assuming that the normal data are normally distributed $x \sim N(\mu, \sigma^2)$, we can set a novelty threshold using the expanded form of the Gumbel distribution, given that extreme values of the Gaussian distribution are described by the Gumbel:

$$p_E(x) = \frac{1}{\sigma d} \exp\{-y_m - \exp(-y_m)\} \qquad (11)$$

where the reduced variate is $y_m = (x' - c)/d$, with location c and scale d parameters:

$$c = \sqrt{2 \ln m} - \frac{\ln \ln m + \ln 4\pi}{2\sqrt{2 \ln m}}$$
$$d = (\sqrt{2 \ln m})^{-1} \qquad (12)$$

using normalized data $x' = (x - \mu)/\sigma$, and where m is the number of data in $\mathbf{X}$. Thus, the value of $\max(\mathbf{X})$ has distribution $p_E(x)$, shown by the second curve in Figure 14(b). A novelty threshold is then set at $P_E(x) \leq 1 \times 10^{-6}$ using the cumulative Gumbel distribution

$$P_E(x) = \exp\{-\exp(-y_m)\} \qquad (13)$$

corresponding to $x = 0.019$, shown by the vertical line in Figure 14(b). Unlike the threshold found using conventional density-estimation techniques, the novelty threshold found using EVT is robust to small changes in normal data. It is based on the EVT distribution $P_E(x)$, and is thus only dependent on m, μ, and σ, which are insensitive to small changes in normal data.

5.5 Setting novelty thresholds in multivariate models

Section 4 described the Parzen window method of constructing a model of normality for multivariate data, in which the probability density of normal data $p(\mathbf{x})$ is estimated using a mixture of Gaussian kernels, with a kernel centered on each normal training example. In multivariate data space, the novelty threshold will form a contour. Data falling within the region contained by the contour are classified "normal", while data falling outside it are classified "abnormal".

To illustrate some of the difficulties encountered when setting a novelty threshold with multivariate data, we first consider using the cumulative probability $P(\mathbf{x})$ associated with the data density $p(\mathbf{x})$ from the Parzen window model. We then show how EVT can be used to set the novelty threshold in multivariate data.

5.5.1 Multivariate novelty thresholds using P(x)

As before, we may wish to set the novelty threshold using $P(\mathbf{x})$, yet no cumulative distribution is defined for multivariate Gaussian distributions in closed form.

Instead, we define the cumulative distribution $P(\mathbf{x})$ at a certain contour on the data density $p(\mathbf{x}) = C$ to be

$$P(\mathbf{x}) = \int_C p(\mathbf{x})\, d\mathbf{x} \tag{14}$$

i.e., $P(\mathbf{x})$ is the volume contained by integrating $p(\mathbf{x})$ around contour C. Thus, we recast the problem of setting the multivariate novelty threshold into finding that contour $p(\mathbf{x}) = C$ which contains the desired probability volume; e.g., $P(\mathbf{x}) \leq 1 \times 10^{-6}$.

In order to find this contour C that contains some probability volume $P(\mathbf{x}) = H$, we can draw a large number of samples N from the Parzen window model, and compute the values of $p(\mathbf{x})$ with respect to that model. The novelty threshold $p(\mathbf{x}) = C$ can then be set such that the proportion H of the total number of samples N is contained by that contour C i.e., for $H \times N$ samples, $p(\mathbf{x}) < C$.

For example, if $H = 1 - 10^{-6}$ and we draw $N = 10^8$ samples from the Parzen window model, the novelty threshold C can be set such that $H \times N = 100$ samples are contained by the contour C i.e., 100 of the 10^8 samples have $p(\mathbf{x}) < C$.

Thus, we could use this method to set a novelty threshold as a contour of $p(\mathbf{x})$ that contains the desired probability volume $P(\mathbf{x})$. However, this method is based on a stochastic sampling process, and we can use EVT to set the multivariate novelty threshold in a more principled manner.

5.5.2 *Multivariate novelty thresholds using EVT*

As before, our goal is to find a contour $p(\mathbf{x}) = C$ such that the desired probability volume is contained within it, this time using the cumulative Gumbel distribution $P_E(\mathbf{x})$ instead of the cumulative distribution $P(\mathbf{x})$; e.g., $P_E(\mathbf{x}) \leq 1 \times 10^{-6}$.

In a Parzen window model, probabilities $p(\mathbf{x})$ along the radius of each kernel vary with the distance from the kernel center according to a single-sided Gaussian distribution $|N(\mu_n, \sigma)|$, and so tend to the Gumbel distribution for extreme values along that radius. This is illustrated in Figure 15, in which an example density $p(\mathbf{x})$ is formed from two Gaussian kernels, $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$. Along the radius shown by the solid line, the Gumbel distribution $p_E(\mathbf{x})$ quantifies our belief in where extreme values generated from the kernel should lie.

Thus, we can use the cumulative Gumbel distribution $P_E(\mathbf{x})$ to determine where extreme values should

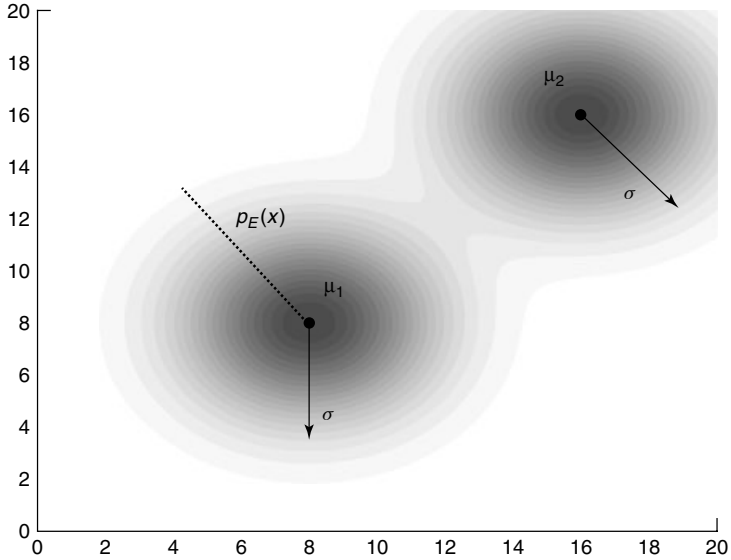


Figure 15. An example bivariate Parzen window model, where $p(\mathbf{x})$ is formed by two kernels $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$. Probabilities $p(\mathbf{x})$ along the radius of each kernel vary with distance from its center according to the single-sided distribution $|N(\mu_1, \sigma)|$, and so tend to the Gumbel distribution $p_E(\mathbf{x})$ as shown.

lie along that radius. Rearranging (13) and using the definition of the reduced variate y_m in terms of (12), we find the radius x where $P_E(x) \leq H$:

$$x = \frac{\sigma}{\sqrt{2\ln m}}\left[2\ln m - \ln(-\ln H) - \frac{\ln\ln m + \ln(4\pi)}{2}\right] \quad (15)$$

This radius x occurs at some multiple of the kernel width $x = K\sigma$, and has an associated kernel probability $p(K\sigma)$. The contour $p(\mathbf{x}) = C$ that corresponds to $P_E(\mathbf{x}) \leq 1 \times 10^{-6}$ is thus given by $p(\mathbf{x}) = p(K\sigma)$, and defines the desired EVT novelty threshold.

This is illustrated in Figure 16, in which the novelty threshold that corresponds to $P_E(\mathbf{x}) \leq 1 \times 10^{-6}$ is shown as a black contour, $p(\mathbf{x}) = p(K\sigma)$.

5.6 EVT novelty thresholds for in-flight vibration models

Figures 17 and 18 show the result of using the above method to set an EVT-based novelty threshold in the 7-D models constructed using in-flight engine data. Using equation (15) with $m = 500$ and $H = 1 \times 10^{-6}$, the novelty threshold for this model that contains probability volume $P_E(\mathbf{x}) \leq 1 \times 10^{-6}$ occurs at the contour $p(\mathbf{x}) = 1.2 \times 10^{-13}$, or $-\log_e p(\mathbf{x}) = 29.7$.

Flight 80 from period A is shown in Figure 17(a), in which $p(\mathbf{x})$ is well below the novelty threshold, as expected for "normal" data. Flight 134 from period C, in which the engine surge occurs at the beginning of the flight, is shown in Figure 17(b). Here, $p(\mathbf{x})$ exceeds the novelty threshold shortly after takeoff and again very clearly at the time of the deceleration after the engine surge.

5.7 EVT novelty thresholds for flight summary models

We can also set novelty thresholds using the same EVT methods for the 20-D model of vibration signatures summarizing entire flights. As before, we consider only those 49 signatures deemed to be valid, which covered over 50% of the total speed range.

Using equation (15) with $m = 49$ and $H = 1 \times 10^{-6}$, the novelty threshold for this model that contains the probability volume $P_E(\mathbf{x}) \leq 1 \times 10^{-6}$ occurs at the contour $p(\mathbf{x}) = 7.4 \times 10^{-18}$.

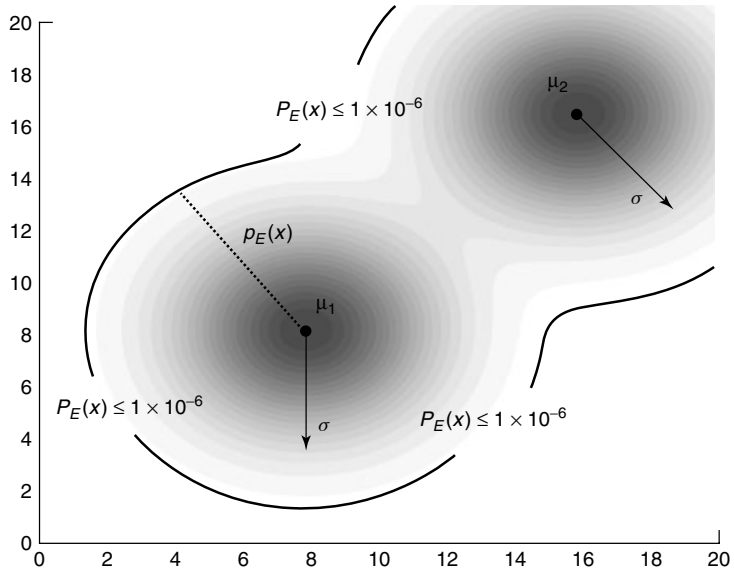


Figure 16. The EVT novelty threshold at $P_E(\mathbf{x}) \leq 1 \times 10^{-6}$ is shown as a contour line in this 2-D example. In 3-D, the novelty boundary would be described by a plane, and a hyperplane in higher dimensions.

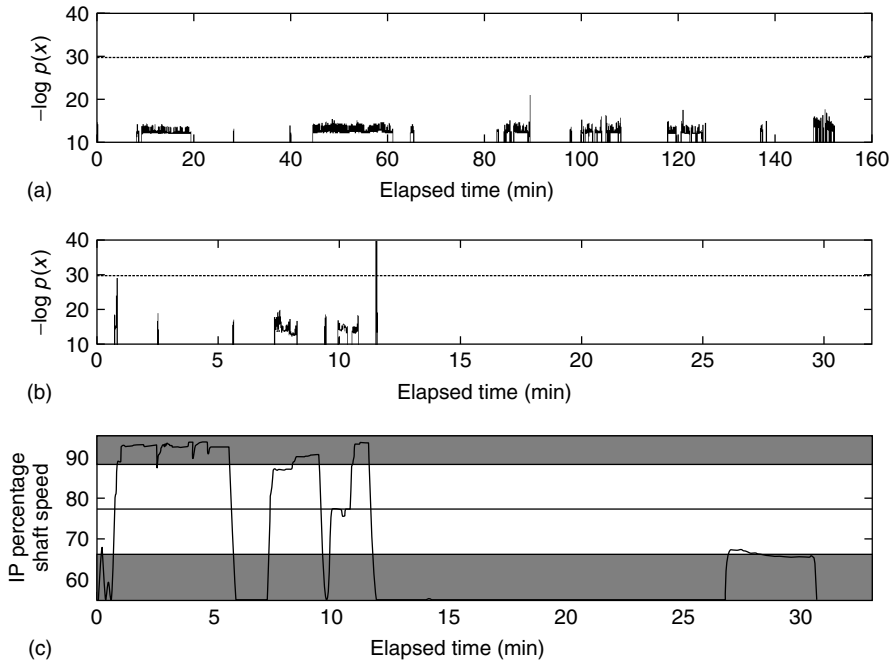


Figure 17. Plot (a) figure shows the novelty scores for flight 80, which all lie well below the novelty threshold computed using EVT, shown as the horizontal line at $-\log_e p(\mathbf{x}) = 29.7$. For flight 134, plot (b), the EVT novelty threshold is exceeded at the beginning (during climbing) and again at the end (deceleration after surge). Plot (c) shows the engine speeds throughout flight 134, showing speed bands $\{S1, \ldots, S4\}$.

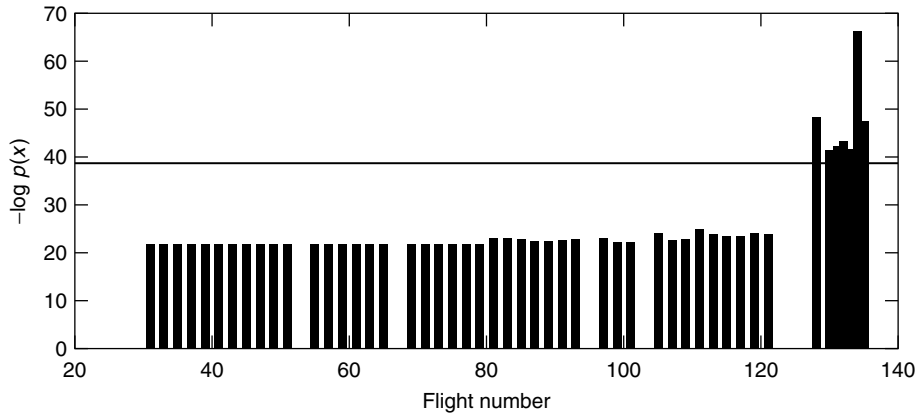


Figure 18. Novelty scores $-\log_e p(\mathbf{x})$ for flight summary vibration signatures, computed using a 20-D Parzen window model of normality. The novelty threshold (shown as a horizontal line) was determined using EVT, and takes the value $-\log_e p(\mathbf{x}) = 39.4$.

Figure 18 shows the novelty threshold compared with $p(\mathbf{x})$ for each of the 49 valid flight signatures. For all signatures in period A (the normal flights), $p(\mathbf{x})$ is significantly below the novelty threshold. It can be observed from the figure that even the "normal" flights of period A have novelty scores of $-\log_e p(\mathbf{x}) \approx 22$. This is because the model is 20-D and so even the peaks of the probability distribution

$p(\mathbf{x})$ will take small values. Assuming that the peak of $p(\mathbf{x})$ occurs at the mean of a Gaussian kernel, $\boldsymbol{\mu}$, the associated maximum probability $p(\boldsymbol{\mu})$ can be determined from the Gaussian distribution:

$$p(\mathbf{x} = \boldsymbol{\mu}) = \frac{1}{(2\pi\sigma^2)^{20/2}} \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\mu}\|^2}{2\sigma}\right) = 3.6 \times 10^{-10} \quad (16)$$

and so the lowest possible novelty score, corresponding to this peak in $p(x)$, is $-\log_e p(x) = 21.7$.

It may also be seen from the figure that the novelty scores for all flights during periods B and C exceed the novelty threshold. Flight 134 has the highest novelty score ($-\log_e p(\mathbf{x}) = 68.0$), which corresponds to the occurrence of the major engine problem.

6 CONCLUSION

In this article, we have presented results showing how abnormal behavior during an early flight test of a new three-shaft jet engine may be identified using novelty detection. Two types of vibration feature detectors have been investigated: one, which consists of the vibration levels at harmonically related frequencies and can be computed in real time during a flight; the other, which is a speed-based vibration signature, summarizing the entire flight. Models of normality have been constructed for both representations, following the application of the k-means clustering algorithm to reduce the number of training vectors. Results presented in Section 4 of this article show how the novelty score increased for both the real-time and the end-of-flight summary models, for a test flight during which the development engine surged following the ejection of a loose component. We have also demonstrated how extreme value statistics, which models the tails of the normal data distribution, can be adapted to set a robust novelty threshold for the reliable detection of unexpected events such as the one highlighted in this article.

RELATED ARTICLES

Ship and Offshore Structures

Wind Turbines

Large Rotating Machines

Gas Turbine Engines

SHM and Lifetime Management of Industrial Piping Systems

Fatigue Monitoring in Nuclear Power Plants

REFERENCES

[1] Roberts S, Tarassenko L. A probabilistic resource allocating network for novelty detection. *Neural Computation* 1994 **6**:270–284.

[2] Bishop CM. Novelty detection and neural network validation. *Proceedings of IEE Vision, Image, and Signal Processing* 1994 **141**(4):217–222.

[3] Tarassenko L, Hayton P, Cerneaz N, Brady M. Novelty Detection for the Identification of Masses in Mammograms. *Proceedings of 4th International Conference on Artificial Neural Networks*, Cambridge, 1995; pp. 442–447.

[4] Hayton P, Schölkopf B, Tarassenko L, Anuzis P. Support vector novelty detection applied to jet engine vibration spectra. In *Advances in Neural Information Processing Systems 13*, Leen TK, Dietterich TG, Tresp V (eds). MIT Press: Boston, 2001, pp. 946–952.

[5] Markou M, Singh S. Novelty detection: a review. *Signal Processing* 2003 **83**:2481–2497.

[6] Moya M, Hush D. Network constraints and multi-objective optimization for one-class classification. *Neural Networks* 1996 **9**(3):463–474.

[7] Ritter G, Gallegos M. Outliers in statistical pattern recognition and an application to automatic chromosome classification. *Pattern Recognition Letters* 1997 **18**:525–539.

[8] Silverman BW. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London, 1986.

[9] McLachlan GJ, Basford KE. *Mixture Models: Inference and Applications to Clustering*. Dekker: New York, 1988.

[10] Bishop CM. *Pattern Recognition and Machine Learning*. Springer-Verlag: Berlin, 2006.

[11] Agusta Y, Dowe DL. Unsupervised learning of gamma mixture models using minimum message length. In *Proceedings of 3rd IASTED Conference Artificial Intelligence and Applications*, Hamza MH (ed), Acta Press: Calgary, 2003, pp. 457–462.

[12] Mayrose I, Friedman N, Pupko T. A gamma mixture model better accounts for among-site rate heterogeneity. *Bioinformatics* 2005 **21**(2):3151–3158.

[13] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 1977 **39**:1–38.

[14] Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press: Oxford, 1995.

[15] Markou M, Singh S. A neural network-based novelty detector for image sequence analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2006 **28**(10):1664–1677.

[16] Kohonen T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 1982 **43**:59–69.

[17] Ypma A, Duin RPW. Novelty detection using self-organising maps. *Proceedings of Connectionist Based Information Systems*, Springer: London, 1998; Vol. 2, pp. 1322–1325.

[18] Labib K, Vemuri R. *NSOM: A Real-Time Network-Based Intrusion Detection System Using Self-Organizing Maps*. Networks Security, 2002.

[19] Vapnik V. *The Nature of Statistical Learning Theory*. Second Edition, Springer-Verlag: Berlin, 2000.

[20] Tax DMJ, Duin RPW. Data domain description using support vectors. In *Proceedings of ESAN99*, Verleysen M (ed). D.Facto: Brussels, 1999; pp. 251–256.

[21] Schölkopf B, Williamson R, Smola AJ, Shawe-Taylor J, Platt J. Support vector method for novelty detection. In *Advances in Neural Information Processing Systems 12, (NIPS'99)*, Solla KMSA, Leen TK (eds). MIT Press: Boston, 2000, pp. 582–588.

[22] Hayton P, Utete S, King D, King S, Anuzis P, Tarassenko L. Static and dynamic novelty detection methods for jet engine health monitoring. *Philosophical Transactions of the Royal Society Part A* 2007 **365**:493–514.

[23] Duda RO, Hart PE, Stork DG. *Pattern Classification*. John Wiley & Sons: New York, 2001.

[24] Yeung DY, Ding Y. Host-based intrusion detection using dynamic and static behavioral models. *Pattern Recognition* 2002 **36**:229–243.

[25] Smyth P. Markov monitoring with unknown states. *IEEE Journal on Selected Areas in Communications* 1994 **12**(9):1600–1612.

[26] Quinn J, Williams CKI. Known unknowns: novelty detection in condition monitoring. *Proceedings of 3rd Iberian Conference on Pattern Recognition and Image Analysis*, *Lecture Notes in Computer Science*. Springer-Verlag: Berlin, 2007; Vol. 4477, pp. 1–6.

[27] Ghahramani Z, Hinton GE. Variational learning for switching state-space models. *Neural Computation* 1998 **12**(4):963–996.

[28] Clifton DA, Bannister PR, Tarassenko L. Learning shape for jet engine novelty detection. In *Advances in Neural Networks III*, Wang J (ed). *Lecture Notes in Computer Science*, 3973. Springer-Verlag: Berlin, 2006, pp. 828–835.

[29] Lowe D, Tipping M. Feed-forward neural networks and topographic mappings for exploratory data analysis. *Neural Computing and Applications* 1996 **4**(2):83–95.

[30] Nabney I. *NETLAB: Algorithms for Pattern Recognition*. Springer-Verlag: Berlin, 2002.

[31] Parzen E. On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 1962 **33**(3):1065–1076.

[32] Coles S. *An Introduction to Statistical Modelling of Extreme Values*. Springer-Verlag: Berlin, 2001.

[33] Roberts SJ. Novelty detection using extreme value statistics. *IEE Proceedings* 1999 **146**(3):124–129.

[34] Roberts SJ. Extreme value statistics for novelty detection in biomedical dataprocessing. *IEE Proceedings on Science, Measurement and Technology* 2000 **147**(6):363–367.

[35] Fisher RA, Tippett LHC. Limiting forms of the frequency distributions of the largest or smallest members of a sample. *Proceedings of the Cambridge Philosophical Society* 1928 **24**:180–190.

[36] Clifton DA, Bannister PR, Tarassenko L. A framework for novelty detection in jet engine vibration data. *Key Engineering Materials* 2007 **347**:305–312.

Chapter 31
Machine Learning Techniques

Fulei Chu, Shengfa Yuan and Zhike Peng
Department of Precision Instruments and Mechanology, Tsinghua University, Beijing, China

1 INTRODUCTION

Condition monitoring and fault diagnosis (CMFD) is a process usually being implemented in aerospace, civil and mechanical structures, etc., to ensure their safe performance [1]. Many researchers in both academia and industry have been engaged in the development of these techniques. In a CMFD system, three basic units are necessary: a signal acquisition unit, a signal analysis unit for fault feature extraction, and a fault classification unit. Recent advances in both hardware and sensor technology mean that signal acquisition with high accuracy and high sampling frequencies is now possible. Attention was therefore turned to the study of how to extract fault features and classify faults from the observed signals. The feature extraction and fault classification are essentially an object-orientated knowledge-acquisition process, and the performances of the CMFD systems are crucially determined by the quantity and quality of the obtained knowledge. Owing to the complexity of modern structures and machines and the continuous emergence of new knowledge, it is hard for the conventional knowledge-acquisition methods (KAMs), where the knowledge is usually gained through the knowledge of engineers, to meet the demand of health monitoring for the modern structures. The deficiencies of the conventional KAMs have promoted the application of machine learning in condition monitoring and fault diagnosis.

Machine learning simulates the learning ability of human by using computers to gain knowledge and skills automatically, and by learning to improve performances continuously and to realize self-improvement. The objective of machine learning is to design some methods that can effectively find the inherent relations in data by learning from the known data, thus predicting the unknown data or judging their characteristics. Therefore, generalization is the most concerned issue for machine learning.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

It is sometimes stated that learning theory is designed to address three main problems [2]:

1. classification, i.e., the association of a class or set label with a set or vector of measured quantities;
2. regression, i.e., the construction of a map between a group of continuous input variables and a continuous output variable based on a set of samples;
3. density estimation, i.e., the estimation of probability density functions from the samples of measured data.

The excellent capability of machine learning in classification makes it to be of great potential in the development of condition monitoring and fault diagnosis. Since direct engagement between machine learning and the fault diagnosis system is the classification, the article is mainly focused on the introduction of implementing machine learning in the fault diagnosis studies as classifier.

2 A BRIEF INTRODUCTION OF MACHINE LEARNING

2.1 The idea of machine learning

From the view of the theory of knowledge, machine learning is very similar to human learning [3]. First, the purposes of both machine learning and human learning are to become intelligent. Secondly, both of them are a knowledge-increasing process. Thirdly, their aims are to understand the objective things and this is also the most essential common ground between machine learning and human learning.

The differences between machine learning and human learning [4] are as follows:

1. Human learning is a long-term process, while machine learning is usually fast and short.
2. Humans are forgetful and can only remember some knowledge but machines can remember all knowledge that it has learned.
3. For human learning, knowledge is untransportable, namely, one person's knowledge cannot be directly copied to another person, but machine learning can copy the learned knowledge to any other system.
4. A distinct characteristic of human learning is generating ideas in a best way, while the ideas obtained by machine learning are usually not the best.
5. The connection and inspiration of humans are hard to be simulated by a machine. Human learning can be of jumping style, while machines always follow rules docilely. This is caused by the different logics followed by human learning and machine learning, respectively.

The essential differences between machine learning and human learning make it impracticable to directly introduce the mechanism of human learning to machine learning. Designing the mechanism of machine learning must take into account the computer specialty and it is necessary to study algorithms for machine learning.

When computers are applied to solve a practical problem, the methods used to derive the desired outputs from a set of inputs usually are required to be able to be described explicitly. However, for solving complex problems, the situations in which there is no method available to compute the desired output from a set of inputs can arise, or the computation may be very expensive [5]. Examples of such situations could be found while modeling a complex chemical reaction, where the precise interactions of the different reactants are generally not known. An effective strategy to solve the problems is to learn the input/output functionality from examples in the similar way of children learning what cars are sport cars simply by being told which cars are sporty rather than by being given a precise specification of sportiness. The approach of using examples to synthesize programs is known as the *learning methodology* and, in the particular case, where the examples are input/output pairs it is called the *supervised learning* [6]. The examples of input/output functionality are referred to as the *training data*.

When an underlying function from inputs to outputs exists, it is referred to as the *target function* [6]. The estimation of the target function learnt by the learning algorithm is known as the *solution of the learning problem*. In the case of the classification, the target function is sometimes referred to as *the decision function.* The solution is chosen from a set of candidate functions that build a map from the input space to the output domain. Usually, a set of candidate functions known as *hypotheses* is chosen before

trying to learn the correct function. Hence, the choice of the set of hypotheses (hypotheses space) [7] is the first important part of the learning strategy. The next important part is the learning algorithm [8], which takes the training data as input and selects a hypothesis from the hypotheses space. When a hypothesis that is consistent with the training data is found, it may not make correct classifications for the unseen data. The ability of a hypothesis to correctly classify the data not in the training set is known as its *generalization* [9]. The learning algorithm is actually a kind of optimization algorithm that can be described in geometric ways or in algebraic ways.

2.2 The key elements of machine learning

Machine learning constitutes of four key elements [9] including the environment, the learning segment, the knowledge base, and the implement segment, as shown in Figure 1.

2.2.1 Environment

The environment is the outside source from which the machine gains knowledge. It includes working objects of the system and their external conditions, such as the diagnosed objects and fault symptoms in the intelligent fault diagnosis systems.

2.2.2 Knowledge base

The knowledge base stores or remembers various kinds of knowledge gained by learning. The knowledge includes

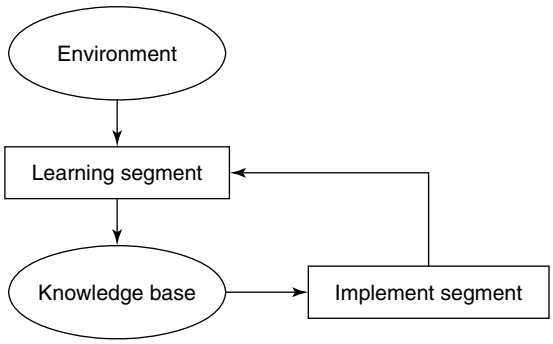


Figure 1. Key elements of machine learning.

1. the object's basic concepts, definitions, theorems, and axioms;
2. laws, rules, and special information of the concrete objects in various conditions;
3. data, information, intermediate result and so on, which can reflect the change in the environments.

2.2.3 Implement segment

The implement segment is also called *working segment* or *decision-making segment*. Using the knowledge stored in the knowledge base, it makes decision or action to complete various work such as the pattern recognition and the expert consultation, and reports the operation results to the learning segment.

2.2.4 Learning segment

The learning segment does searching, control, and logical thinking, such as abstraction, comparison, summarization, synthesizing, and reasoning, to produce, amend, and update the knowledge.

3 REVIEW OF LEARNING METHODS

Machine learning is applied in a wide range of fields and each field has its own characteristics and therefore machine learning methods vary for different fields. Learning methods can be sorted from various aspects or on the basis of different ways. A relatively reasonable way to categorize the machine learning methods is based on the systematicness [3, 9]. Using the systematicness-based way, the machine learning methods can be classified into following categories.

3.1 Inductive learning

Inductive learning [3, 9] is a learning method that derives general rules from the particular cases. Environment provides a series of positive cases and negative cases to the system, and the system conducts the generalization operation with these cases by induction learning and engenders a series of conceptual descriptions. Inductive learning algorithms fall into

two categories: learning from examples and learning from observations and discoveries.

3.2 Analytic learning

Analytic learning [3, 9] applies the domain knowledge to do analysis and learns with one or a few examples. Its main features are as follows: (i) Its inference strategy is deduction rather than induction; (ii) The experiences gained by solving past problems are used to solve new problems or to produce control rules of searching that enables the applications of the domain knowledge to be more effective. The goal of analytic learning is to improve the efficiency of the system rather than to enlarge the range of the concept descriptions. Analytic learning mainly includes the explanation-based learning, the analogy-based learning, and the case-based learning.

3.3 Genetic algorithm and classifier system

Classifier system [3, 9] is a kind of high parallel message-passing rule based systems. It learns by means of the trust-allot and the rule discovering. The genetic algorithms (GAs) [3], which were motivated by the analogy to the biological evolution, provide an effective rule discovering method. Rather than searching from the general-to-specific hypotheses or from the simple-to-complex, GAs generate successor hypotheses by repeatedly mutating and recombining parts of the best-so-far hypotheses. Hypotheses are often described using bit strings whose interpretation depends on the application and, sometimes, may also be described by symbolic expressions or even computer programs. GAs have been applied successfully to a variety of learning tasks and optimization problems.

3.4 Connectionist learning (or neural net)

A connectionist model [3, 9] (neural net) consists of some simple cells similar to the nerve cells and the weighted connections between these cells. Each cell has a state determined by the inputs of other cells to which the cell links. The connectionist model trains the net with different categories of examples and produces internal representations for the net. With the obtained internal representations, other input examples can be identified. Learning is mainly the adjustment of the connection weights. The connectionist learning is nonsymbolic and possesses the ability of the high parallel distributed processing. Neural networks have some disadvantages such as the local optimal solution, the low convergence rate, and the "over-fitting". The neural network methods are based on the empirical risk minimization principle that assumes the sample number is infinite. However, in practice, the sample number is always finite, and the insufficiency of fault samples can make the neural network poor in generalization and, therefore, is a bottleneck problem for the fault diagnosis applications.

4 SUPPORT VECTOR MACHINES

As a principled and very powerful learning method that is based on statistical learning theory, the support vector machines (SVMs) were developed by Vapnik and coworkers [6, 10]. It has already outperformed most other methods in a wide variety of applications in the few years since its introduction. SVMs are especially suitable for the small-sample cases and their aim is to find the optimal solution in the finite available information rather than in the infinite large samples. SVM can satisfactorily overcome the problems of "over-fitting", local optimal solution, and low convergence rate and, moreover, has good generalizations even when the samples are few [11–14].

The basic idea of the pattern recognition with SVM is to project the sample space into a high-dimension *eigenspace*. In the eigenspace, the optimal separating hyperplanes of the original sample set can be found, but the calculation complexity will not increase significantly. In this section, the SVM algorithm and its motivation are briefly described, and the more detailed description of the SVM can be found in [6].

For the typical binary pattern recognition, given a training set $\{(x_1, y_1), \ldots, (x_l, y_l)\}$ where $x_i \in X$ and X is an input vector of l dimension, and $y_i \in \{0, 1\}$ is called *target* or *label*. Symbol H represents the set of all possible hypotheses that may be considered as classification functions by a learning machine.

The notion of shattering is defined as a *set of instances*. Consider a subset of instances $S \subseteq X$. Each hypothesis h from H can impose dichotomy on S by partitioning S into two subsets: $\{x \in S | h(x) = 1\}$ and $\{x \in S | h(x) = 0\}$. Given an instance set S of size $|S|$, there are $2^{|S|}$ possible dichotomies but, however, H may be unable to represent all of them. It is said that H shatters S if every possible dichotomy of S can be represented by the hypothesis from H. Therefore, a set of instances S is shattered by hypothesis space H if and only if for every dichotomy of S there exists some hypothesis in H consistent with it.

The Vapnik–Chervonenkis (VC) dimension d of hypothesis space H defined over instance space X is the size of the largest finite subset of X shattered by H. If arbitrarily large finite sets of X can be shattered by H, then $d \equiv \infty$. The larger the subset of X that can be shattered, the more expressive H is. The VC dimension of H is precisely this measure [3]. VC dimension of the class H measures the richness or flexibility of the function class, which is also often referred to as its *capacity*. Controlling the capacity of a learning system is one way of improving its generalization accuracy [10].

Theorem: Let H be a hypothesis space of VC dimension d. For any probability distribution $\Im$ on $X \times \{-1, 1\}$, with probability $1 - \delta$ over random examples S, any hypothesis $h \in H$ that makes k errors on the training set S has error no more than

$$\underset{\Im}{\text{err}}(h) \leq \varepsilon(l, H, \delta) = \frac{2k}{l} + \frac{4}{l}\left(d \log \frac{2el}{d} + \log \frac{4}{\delta}\right) \tag{1}$$

provided $d \leq l$.

The theorem suggests that a learning algorithm for hypothesis class H should seek to minimize the number of training errors, since everything else in the bound has been fixed by the choice of H. In inductive principle this is known as *empirical risk minimization*, since it seeks to minimize the empirically measured value of the risk functional. The theorem can also be applied to a nested sequence of hypothesis classes

$$H_1 \subset H_2 \subset \cdots \subset H_i \subset \cdots \subset H_M \tag{2}$$

By using δ/M, the probability of any one of the bounds failing to hold would be less than δ. If a hypothesis h_i with minimum training error is sought in each class H_i, then the number of errors k_i that it makes on the fixed training set S will satisfy

$$k_1 \geq k_2 \geq \cdots \geq k_i \geq \cdots \geq k_M \tag{3}$$

while the VC dimension $d_i = \text{VC dim}(H_i)$ is from a nondecreasing sequence. The bound of the above theorem can be used to choose the hypothesis h_i for which the bound is minimal, that is, the reduction in the number of errors (first term) outweighs the increase in the capacity (second term). This induction strategy is known as *structural risk minimization* [10].

The introduction of SVM can start with a simple case of two classes that is linearly separable. Assume a data set

$$D = \{(\mathbf{x}_i, y_i)\}, \quad (i = 1, \ldots, l), \quad y_i \in \{-1, +1\} \tag{4}$$

where $\mathbf{x}_i$ is the input sample and y_i is the output class, y has two values (-1 or $+1$) that stand for two classes, l is the sample number. It is required to determine, among all linear separating planes that can separate the input samples into two classes, which one will have the smallest generalization error. In Figure 2, rings and diamonds stand for the two classes of sample points respectively and H is a separating plane. H_1 and H_2 are the planes that are parallel to H and pass through the sample points closest to H in the two classes. Define the distance between H_1 and H_2 as *margin*. The optimal separating plane that has the smallest generalization error is the one that not only correctly classifies all sample points into the

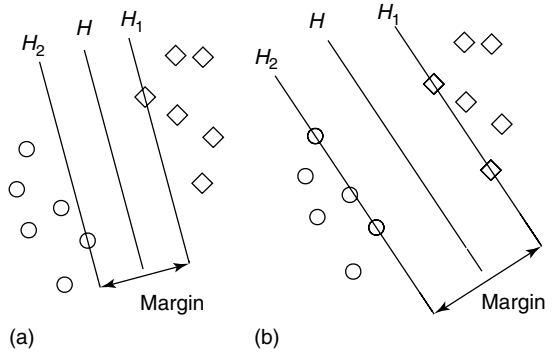


Figure 2. Separating plane with different *margin* (a) and (b). [Reproduced from Ref. 21. © Elsevier, 2006.]

two classes but also has the maximal *margin* between H_1 and H_2.

As a starting point for the analysis and construction of more sophisticated SVMs, the maximal margin classifier is an important concept but, however, it cannot be used in many real-world problems. For example, in the case where the data are noisy, there will in general be no linear separation in the feature space unless very powerful kernels are used. In order to overcome the defect, Vapnik proposed a *soft margin* algorithm to allow some data points misclassified by using the margin slack variables. In the *soft margin* algorithm, the optimal separating plane is a linear classifier that has the classification function as follows [15]:

$$f(\mathbf{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{l} \lambda_i y_i \mathbf{x}^{\mathrm{T}} \mathbf{x}_i + b\right\}, \quad (i = 1, \ldots, l) \tag{5}$$

where sgn( ) is the sign function. If $u > 0$, then $\operatorname{sgn}(u) = 1$; if $u \leq 0$, then $\operatorname{sgn}(u) = -1$. The Lagrange coefficient λ_i is the solution of the following quadratic programming (QP) problem:

$$\begin{aligned} \underset{\mathbf{\Lambda}}{\text{Maximize}}\ W\ (\mathbf{\Lambda}) &= -\mathbf{\Lambda}^{\mathrm{T}}\mathbf{1} + \frac{1}{2}\mathbf{\Lambda}^{\mathrm{T}} D \mathbf{\Lambda} \\ \text{subject to} \quad \mathbf{\Lambda}^{\mathrm{T}} y &= 0 \\ \mathbf{\Lambda} - C &\leq 0 \\ -\mathbf{\Lambda} &\leq 0 \end{aligned} \tag{6}$$

where $(\mathbf{\Lambda})_i = \lambda_i$, $\mathbf{1}_i = 1$ and $D_{ij} = y_i y_j \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j$, C is a penalty constant for the sample points misclassified by the optimal separating plane. Its role is to strike a proper balance between the calculation complexity and the classifying error. When $C \to +\infty$, it is the ideal case in which all samples are theoretically correctly separated by the optimal separating plane and there is no classifying error at all, but the calculation complexity could be the biggest. It is found that only a few coefficients λ_i are not zero, and since every coefficient λ_i corresponds to a particular sample point, this means that only the sample points with nonzero λ_i determine the optimal separating plane, and only these few sample points called *support vectors* can affect the classification result while other sample points could be removed from the sample set and the optimal separating plane would be almost unaltered. The *support vectors* are usually few in the sample set, and Vapnik has shown that the sample number is proportional to the generalization error of the classifier.

Since there are certain problems in practice that are not linearly separable, SVM has to be developed for the classification of nonlinear problems. By projecting the original sample space into a high-dimension *eigenspace* with a kernel function $K(\mathbf{x}, \mathbf{x}_i)$, the nonlinear separable problem becomes linearly separable in the *eigenspace*. The classification function for the SVM classifier in the eigenspace is shown as follows:

$$f(\mathbf{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{l} \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + b\right\} \tag{7}$$

In Table 1 some kernel functions $K(\mathbf{x}, \mathbf{x}_i)$ proposed by Vapnik are listed together with the associated classifiers, which are known to have good properties.

The SVM initially proposed for the two-class pattern recognition is often called as *two-class SVM classifiers*, which can be generalized to the multiclass pattern recognition by some methods. These methods can be summarized into two kinds.

The first one is the multioutput SVM algorithm, which has only one SVM classifier but many outputs. When constructing the classification function, all classes are considered, and the classification function is constructed by revising the optimization functions and constraint conditions used in two-class SVM classifiers. This algorithm has only one SVM classifier but its classification function is very complex and, accordingly, the calculation is very complex as well and the training and recognition is very time-consuming, and therefore the classifying error is big when the number of samples is large [16].

Another kind is to combine some two-class SVM classifiers for the purpose of multiclass pattern

Table 1. Some kernel functions and the type of classifiers defined by them

Kernel function	Type of classifier
$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^{\mathrm{T}}\mathbf{x}_i + 1)^d$	Polynomial of degree d
$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\lVert\mathbf{x} - \mathbf{x}\rVert^2/2\sigma^2)$	Gaussian radial basis function (RBF)
$K(\mathbf{x}, \mathbf{x}_i) = \tanh(\mathbf{x}^{\mathrm{T}}\mathbf{x}_i + \theta)$	Multilayer perceptron

recognition. This method mainly consists of two submethods as follows.

One is the "one to one" algorithm. This algorithm constructs all two-class SVM classifiers between any two classes; therefore, $k(k-1)/2$ two-class SVM classifiers can be constructed in all for the case of k classes. At the recognition stage, new coming sample $\mathbf{x}$ is input to the established classifier [17]:

$$f^{mn}(\mathbf{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{l} \lambda_i^{mn} y^{mn}{}_i K(\mathbf{x}, \mathbf{x}_i) + b^{mn}\right\} \quad (8)$$

If the classifier shows that $\mathbf{x}$ belongs to the m class, then a ballot is cast for class m. After recognized by all these $k(k-1)/2$ classifiers, $\mathbf{x}$ is then judged to belong to the class that has the most ballots.

Some disadvantages of this algorithm are as follows: (i) the number $k(k-1)/2$ of two-class SVM classifiers increases greatly with the class number k and, consequently, the calculation increases greatly and, moreover, the rates of training and recognition are very slow; (ii) when two or more classes have the same ballots, it is hard to judge which class the new sample $\mathbf{x}$ belongs to; (iii) there is at least a class that has the most ballots; therefore, a new sample $\mathbf{x}$ that does not belong to any of the k classes would be misjudged to belong to them and therefore the wrong classification rises.

The other submethod is the "one to rest" algorithm. This algorithm takes each class, e.g., class m of these k classes as a separate category and the rest $k-1$ classes as another category and then constructs a two-class SVM classifier and names it as SVM m. In this way, k two-class SVM classifiers can be constructed in all. The classification function of SVM m is as follows:

$$f^{m}(\mathbf{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{l} \lambda_i^{m} y^{m}{}_i K(\mathbf{x}, \mathbf{x}_i) + b^{m}\right\} \quad (9)$$

At the recognition stage, the new sample $\mathbf{x}$ is input to all the obtained k classifiers SVM m ($m = 1, \ldots, k$). There are k outputs in all. The new sample $\mathbf{x}$ is judged to belong to the class whose corresponding classifier has the largest output [18].

The disadvantages of this algorithm are as follows: (i) all samples have to be taken into the training of the k classifiers and, in the recognition stage, the classification result can be obtained only after the new sample $\mathbf{x}$ has been recognized by all these k classifiers. Therefore the calculation is very time consuming and the rate of training and recognition is slow as well; (ii) it also shares the disadvantages (ii) and (iii) of the above algorithm [11, 14, 19, 20].

5 THE MACHINE-LEARNING-BASED FAULT DIAGNOSIS SYSTEM

For many kinds of complicated machines like turbopump rotors [21], it is difficult to use the mathematical model to describe their faults. Artificial intelligence methods [3–5], which do not involve mathematical models, have been widely used in the fault diagnosis in the recent years, especially in the fault diagnosis for dynamic procedures such as the starting, the stopping, and the changing of working mode. Knowledge is the foundation of artificial intelligence. Generally, there are two ways to obtain knowledge. One is that knowledge is induced and sorted by human, and is input to the computer in acceptable and processible forms. Expert systems usually adopt this way. The other is that computers are endowed with the ability of self-learning, called *machine learning*. For many complicated machines, knowledge sources are not rich and, therefore, knowledge acquisition is very difficult but extremely necessary. Machine learning can acquire knowledge beyond human ability, learn new knowledge with the changing of environment, and improve itself automatically with experience.

5.1 The framework of the machine-learning-based fault diagnosis system

The framework [21] of the machine-learning-based fault diagnosis system is shown in Figure 3. Four modules are used: the signal preprocessing module, the feature extraction module, the training module, and the recognition module.

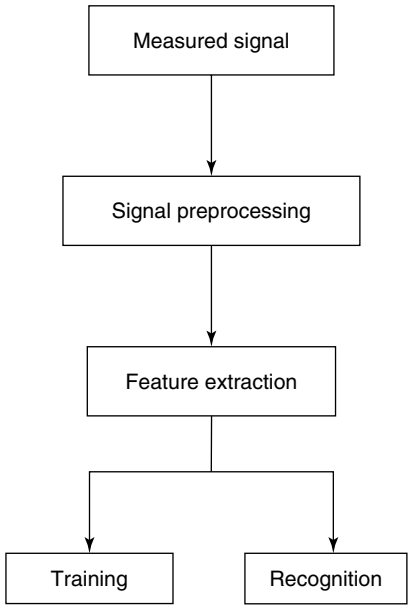


Figure 3. Pattern-recognition-based fault diagnosis system. [Reproduced from Ref. 21. © Elsevier, 2006.]

5.2 Signal preprocessing

Signal processing is important for condition monitoring and fault diagnostics, which aims at finding a simple and effective transform to the original signals. Therefore, the important information contained in the signals can be seen and then the dominant features of signals can be extracted for fault diagnostics. Various signal processing methods are available including the most popular method—Fourier transform [4], and the wavelet transform [5] and so on.

5.3 Feature extraction

Fault feature vectors often are of multiple dimensions and consist of a number of variables, between which there exist correlations and redundancies. To eliminate the redundancies and to be convenient for the classification, feature selection methods are often employed, such as principal component analysis (PCA) [4]. By these methods, the fault feature vectors with high dimension can be transformed to the low dimension vectors that preserve enough information in the raw fault feature vectors and, therefore they should be adequate to present the system.

5.4 Training

After obtained the fault features, the next step is the pattern classification (or training). For many complicated machines like the turbo-pump rotor, it is difficult to use the mathematical model to explicitly describe the faults. On the other hand, the machine learning does not involve the mathematical models and it can learn the input (fault features)/output (fault types) functionality from examples.

5.5 Recognition

When a machine learning system has been trained, it obtains the knowledge related to the faults. When a new coming fault feature is inputted to this machine learning system, the machine learning system can recognize the fault type of this fault feature. It is called *diagnosis process*. In addition, to avoid the situation that a normal state is recognized as a fault by the machine learning system, normal features are required to be used to train the machine learning system together with the fault features; therefore, the machine learning system can judge whether the object is normal or not.

6 APPLICATION OF SVMs IN TURBO-PUMP FAULT DIAGNOSTICS

Turbo-pump rotors are complicated machines whose working condition is special. Faults frequently occur on the turbo-pump rotors, and some of them are quite dangerous. Fault diagnosis technique for the turbo-pump rotor is therefore very important. Table 2 shows 14 familiar fault patterns of the turbo-pump rotors, based on which the "one to others" SVM algorithm is used to construct 13 two-class SVM classifiers. These 13 two-class SVM classifiers are ranged as a binary tree in terms of the fault priority where more common or more dangerous faults are placed with precedence, and then they form a multiclass fault diagnosis system as shown in Figure 4. The Gaussian function $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2/2\sigma^2)$ is chosen as the kernel function. Vapnik has shown that different kernel functions used make only slight differences to the classification results.

Table 2. Fault patterns of turbo-pump rotor and vibration energy distribution

	Fault pattern	Frequency band (f_r is the frequency of rotor rotation)								
		$<0.4f_r$	$0.4-0.49f_r$	$0.5f_r$	$0.51-0.99f_r$	$1.0f_r$	$2.0f_r$	$3.0-5.0f_r$	$>5.0f_r$	Mesh frequency
1	Gear damage	—	—	—	—	0.22	—	—	—	0.82
2	Structure resonance	0.21	0.27	—	—	0.51	0.11	—	—	—
3	Rotor radial touch friction	0.13	0.04	0.04	0.12	0.31	0.11	0.11	0.11	—
4	Rotor axial touch friction	0.04	0.04	0.04	0.11	0.32	0.22	0.15	0.16	—
5	Shaft crack	—	—	—	—	0.41	0.21	0.23	0.21	—
6	Bearing damage	0.12	0.13	—	—	0.44	0.21	0.21	—	—
7	Body joint looseness	0.24	0.26	—	—	0.31	0.11	0.13	—	—
8	Bearing looseness	0.73	—	—	—	0.21	—	0.04	—	—
9	Rotor parts looseness	0.43	0.41	—	—	0.12	—	0.12	—	—
10	Pressure pulse	0.21	0.25	—	—	0.11	0.11	0.31	0.11	—
11	Cavitation	0.55	0.32	—	—	0.04	0.04	0.04	—	—
12	Vane rupture	—	—	—	—	0.10	0.06	0.06	—	—
13	Rotor eccentricity	—	—	—	—	0.90	0.06	0.06	—	—
14	Shaft bend	—	—	—	—	0.92	0.04	0.04	—	—

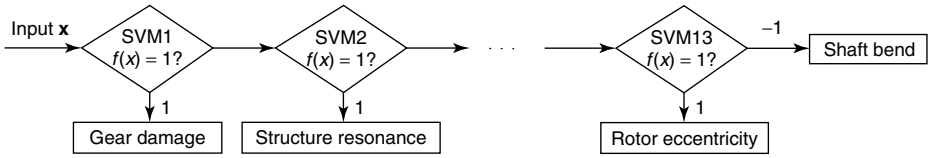


Figure 4. "One to others" multiclass fault diagnosis system. [Reproduced from Ref. 21. © Elsevier, 2006.]

By PCA, nine-dimensional fault feature vectors are transformed to two-dimensional fault feature vectors. Some of the obtained two-dimensional fault feature vectors are shown in Figure 5, where rings, triangles, diamonds and several other graphics denote the feature data for different faults. It is evident that they are classifiable. The two-dimensional vectors possess more than 80% information contained by the original nine-dimensional fault feature vectors. Therefore, they are adequate to present the system.

After the training procedure, the obtained 13 two-class SVM classifiers produce 13 optimal separating planes. The area I, II, III, ..., XIV are associated to the 14 kinds of faults respectively. The classification results in Figure 6 show that "one to others" SVM algorithm has classified 240 data into 14 kinds of faults accurately, and there are no reject regions.

When a new coming fault sample **x** is to be recognized by the "one to others" multiclass fault diagnosis system, its two-dimension feature data is first input to SVM1 in Figure 4; if the output of SVM1 is 1, then the new sample **x** is recognized as the fault "gear damage" and the recognition is finished; or if the output of SVM1 is −1, then

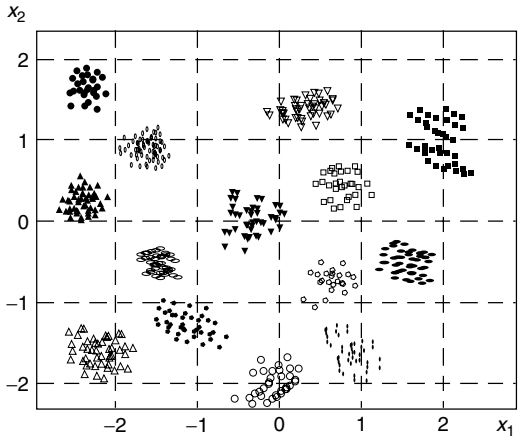


Figure 5. Two-dimensional fault feature vectors. [Reproduced from Ref. 21. © Elsevier, 2006.]

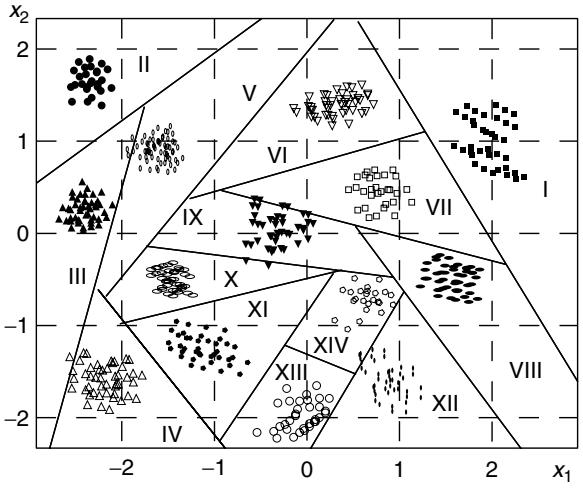


Figure 6. Optimal separating hyperplane of SVM. [Reproduced from Ref. 21. © Elsevier, 2006.]

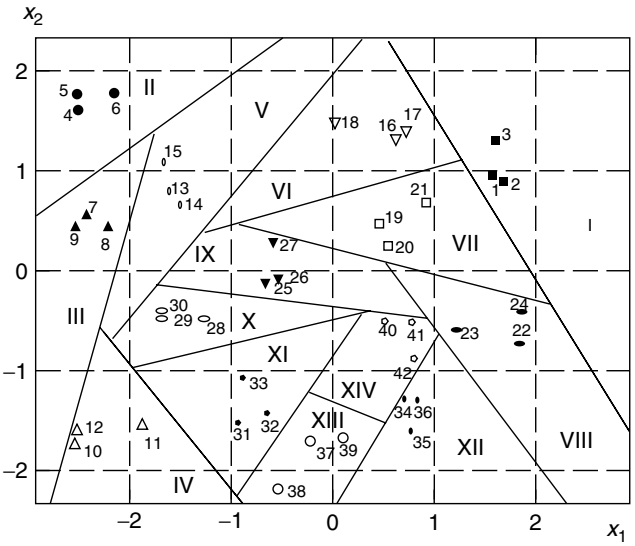


Figure 7. Recognition results. [Reproduced from Ref. 21. © Elsevier, 2006.]

the feature data of **x** is automatically transferred to SVM2; and so on.

To test the generalization ability, 42 samples that have not been used in the training stage have been input to the "one to others" multiclass fault diagnosis system. The nine-dimension data of the 42 samples have been transformed to two-dimension data by PCA method. Since they are testing samples, their fault types are known as *prior*. Figure 7 shows the recognition results of the "one to others" multiclass fault diagnosis system.

The positions of the 42 testing samples in Figure 7 clearly tell which type of faults they belong to, and the recognition results are consistent with the actual

fault types. This verifies that the "one to others" SVM algorithm can correctly recognize the fault samples.

To compare the "one to others" SVM algorithm with other algorithms, five different algorithms including one to rest SVM, one to one SVM, one to others SVM, RBF neural net, and back propagation (BP) neural net are used to diagnose 14 kinds of faults in Table 2 using the exact same fault samples. By averaging 10 best results for each algorithm, average diagnosis accuracies are 93, 96, 88, 83%, respectively, for the methods of one to rest SVM, one to one SVM, RBF neural net and BP neural net, and the average relative diagnosis time is 82, 100, 70, 65%, respectively, for the above four different methods, while the average diagnosis accuracy is 97% and the average relative diagnosis time is 54% for the method of "one to others" SVM. Therefore, the "one to others" SVM algorithm turns out to be the most effective one as it has the shortest diagnosis time and the highest accuracy.

A problem in the practical application of SVM is how to select some SVM parameters so that the performance of SVM can be the best. These SVM parameters mainly include the penalty constant C, the relaxation factor ξ, and the parameters in the kernel function, for instance, the width parameter σ in RBF kernel function. These parameters affect the SVM performance more or less. Up to now, there is no effective method to determine these SVM parameters. Generally, the cross verification trial or the gradient step-down operation is used to select them, but these methods depend on human experiences too much or require the kernel function to be continuously differentiable and, therefore, the resulted SVM classifier is likely to fall into the local minimum. These problems present an obstacle to the more effective application of the SVM.

Moreover, in the real fault diagnosis application, the use of dimensionality-reducing methods such as PCA to extract the fault features might lead to the loss of useful information while discarding some redundant variables. In addition, when there are many fault-correlated variables, the number of principal components gained by PCA could be still large if enough fault messages are to be retained. The great number of principal components would equally bring in many irrelevant messages and consequently reduce the efficiency of the fault diagnosis system. Therefore, more efficient feature selection methods must be developed. In order to overcome the shortcomings mentioned above, an attempt is made to jointly optimize the feature selection and the SVM parameters with GAs to improve the SVM performance that is taken as the target function in the optimization.

The performance of SVM is mainly referred to its generalization capability, namely the capability of recognizing new data that mainly consists of three indexes: error ratio $E_{\xi\alpha}$, recall ratio $R_{\xi\alpha}$, and accuracy ratio $P_{\xi\alpha}$. Here, a correct ratio $E = 1 - E_{\xi\alpha}$ is suggested to estimate the performance of the SVM [22, 23].

In the training of SVM, the proper choice of the kernel function is very important, which is closely related to the complexity of the classification functions set and can affect the performance of SVM. Two most commonly used kernel function in SVM is the Gaussian function $K_{\mathrm{RBF}}(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2/\sigma^2\right)$ and the polynomial function $K_{\mathrm{poly}}(x_i, x_j) = (x_i{}^{\mathrm{T}} x_j + 1)^d$, where x_i and x_j are vectors in the input space, σ and d are the parameters of the Gauss function and the polynomial function, respectively. The ability of a single kernel function to improve the performance of SVM is limited, and a parameter λ is introduced to construct a mixture kernel function:

$$K_{\mathrm{mix}}(x_i, x_j) = \lambda K_{\mathrm{RBF}}(x_i, x_j) + (1 - \lambda) K_{\mathrm{poly}}(x_i, x_j), \quad 0 \le \lambda \le 1 \tag{10}$$

The penalty constant C in SVM makes a considerable influence on the performance of SVM, and it decides the complexity of the SVM model and the penalty degree to those sample points misclassified by the optimal separating plane. When C is too big or too small, the generalization capability of SVM can be weakened [24, 25]. For unevenly distributed data, using different penalty constants for each class could reduce the real risk. For the binary classification, the penalty constants for each class are used as C_+ and C_-.

The relaxation factor ξ in SVM indicates the error expectation in the classification process. The value of ξ affects the number of *support vectors* generated by the SVM classifier. When ξ is too big, the number is small, but the classification error is high, and vice versa [26–28].

Combining the above parameters, the training model of SVM can be established as

$$\begin{aligned} \boldsymbol{M} &= \{\lambda, \sigma, d, C_+, C_-, \xi\} \\ & 0 \le \lambda \le 1, \sigma, d, C_+, C_- \ge 0 \end{aligned} \tag{11}$$

The SVM performance E is taken as the target function in the optimization of the training model of SVM, which can be described as

$$\begin{aligned} &\max_{M} E(\boldsymbol{M}) \\ \text{s.t.} \quad &\boldsymbol{M} = \{\lambda, \sigma, d, C_+, C_-, \xi\} \\ & 0 \le \lambda \ge 1, \sigma, d, C_+, C_-, \xi \ge 0 \end{aligned} \tag{12}$$

For a feature set $F = \{f_1, f_2, \ldots, f_i \ldots, f_N\}$, where N is the size of the feature set and f_i is a feature. The following binary vector is introduced to denote the feature selection: $\boldsymbol{S} = \{s_1, s_2, \ldots, s_i, \ldots, s_N\}$, $s_i \in \{0, 1\}, i = 1, \ldots, N$. The values 0 and 1 for s_i stand for whether the corresponding feature f_i in $\boldsymbol{F}$ is selected or not. Considering the SVM performance E as the target function, the optimization problem of the feature selection can be expressed as

$$\begin{aligned} &\max_{S} E(\boldsymbol{S}) \\ \text{s.t.} \quad &\boldsymbol{S} = \{s_1, s_2, \ldots, s_i \ldots, s_N\} \\ & s_i \in \{0, 1\}, i = 1, \ldots, N \end{aligned} \tag{13}$$

The training model $\boldsymbol{M}$ of SVM must be given in advance before optimizing the feature selection $\boldsymbol{S}$, and the feature selection $\boldsymbol{S}$ must be done before optimizing the training model $\boldsymbol{M}$. Since the optimization targets in formulas (11) and (13) are all to improve the SVM performance E, the feature selection $\boldsymbol{S}$ and the training model $\boldsymbol{M}$ can be jointly optimized at the same time, which can be expressed as follows:

$$\begin{aligned} &\max_{M} E(\boldsymbol{S}, \boldsymbol{M}) \\ \text{s.t.} \quad &\boldsymbol{S} = \{s_1, s_2, \ldots, s_N\}, s_i \in \{0, 1\}, \ i = 1, \ldots, N \\ &\boldsymbol{M} = \{\lambda, \sigma, d, C_+, C_-, \xi\} \\ & 0 \le \lambda \ge 1, \sigma, d, C_+, C_-, \xi \ge 0 \end{aligned} \tag{14}$$

where $(\boldsymbol{S}, \boldsymbol{M})$ is a hybrid vector that jointly describes the feature selection and the training model of SVM. This joint optimization problem can be solved by GAs. The optimization of $\boldsymbol{S}$ and $\boldsymbol{M}$ are mutually correlated.

With the above experimental data, the feature selection and SVM are first optimized separately and then optimized jointly. The comparison results of the diagnosis of fault 1 (gear damage) are shown in Figure 8. Cases 1–4 are the results obtained using the optimized SVM and the nonoptimized feature selection, the optimized feature selection and the nonoptimized SVM, the jointly optimized feature selection and SVM and no further optimization, respectively. It can be seen that both the separate optimization and the joint optimization on the feature selection and the SVM can improve the SVM performance E. The joint optimization improves the E most and has the fastest convergence and, in addition, solely optimizing the feature selection can improve the E better than solely optimizing the SVM [29, 30].

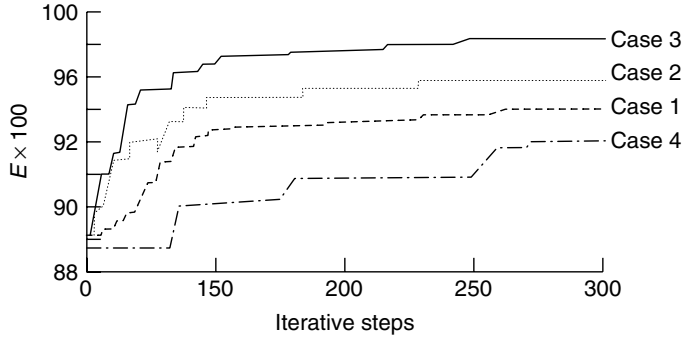


Figure 8. Optimization results with GAs.

7 CONCLUSIONS AND REMARKS

Owing to the complexity of modern structures and machines, the conventional knowledge engineer based knowledge-acquisition methods are no longer able to meet the demands of the condition monitoring and fault diagnosis for modern structures and machines. More sophisticated knowledge-acquisition methods are therefore expected. Machine learning, which is a significant development in the artificial intelligence in recent years, has already shown excellent capability in discovering knowledge from observed data through constructing computational relationships between quantities on the basis of observed data and rules. Especially, it is expected that the machine learning can greatly promote the development of the condition monitoring and fault diagnosis as it has the excellent capability of classification, and the performances of classifier play a crucial role in the condition monitoring and fault diagnosis system. This article has given an outline to machine learning including the background and some key algorithms and theories that form the core of machine learning. More efforts have been made to the introduction of SVM and an application example of using SVM to successfully conduct fault diagnosis on a turbo-pump rotor, with which the great potential of the machine learning in fault diagnosis has been validated.

RELATED ARTICLES

Statistical Pattern Recognition

Artificial Neural Networks

REFERENCES

[1] Worden K, Farrar CR. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A* 2007 **365**:303–315.

[2] Worden K, Manson G, Surace C. Aspects of novelty detection. *Key Engineering Materials* 2007 **347**: 3–16.

[3] Mitchell TM. *Machine Learning*. McGraw-Hill, 2000.

[4] Hunt EB. *Artificial Intelligence*. Academic Press: New York, 1975.

[5] Reed S. Artificial neural networks. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, 2008.

[6] Vapnik VN. *The Nature of Statistical Learning Theory*. Springer-Verlag: New York, 1999.

[7] Bishop CM. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

[8] Sohn H. Pattern recognition. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, 2008.

[9] Vidyasager M. *A Theory of Learning and Generalization*. Springer, 1997.

[10] Cristiannini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[11] Guo G, Li SZ, Chan KL. Support vector machine for face recognition. *Image and Vision Computing* 2001 **19**:631–638.

[12] Steve RG. *Support Vector Machines for Classification and Regression*, Technical Report, http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf or http://www.citeulike.org/user/lboussou/article/2815861. University of Southampton, 1998, pp. 1–28.

[13] Chapelle O, Haffner P, Vapnik VN. Support vector machine for histogram-based image classification. *IEEE Transactions on Neural Networks* 1999 **10**:1055–1064.

[14] Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation* 2001 **13**:637–649.

[15] Vapnik VN. *Statistical Learning Theory*. John Wiley & Sons, 1998.

[16] Bredensteiner EJ, Bennett KP. Multicategory classification by support vector machines. *Computational Optimization and Applications* 1999 **12**:53–79.

[17] Daniel JS, James AB. Support vector machines and the multiple hypothesis test problem. *IEEE Transactions on Signal Processing* 2001 **49**:2865–2872.

[18] Ma XX, Huang XY, Chai Y. 2PTMC Classification algorithm based on support vector machines and its application to fault diagnosis. *Control and Decision* 2003 **18**:272–284 (in Chinese).

[19] Schölkopf B. Statistical learning and kernel methods. In *CISM Courses and Lectures, International Centre for Mechanical Sciences*, Della Riccia G, Lenz HJ, Kruse R (eds). Springer: Vienna, 2000; Vol. 431, pp. 3–24.

[20] Scholkopf B, Smola A. *Training with Kernels-support Vector Machines, Regularization, Optimization and Beyond*. MIT Press: Cambridge, MA, 2002, pp. 165–189.

[21] Yuan SF, Chu FL. Support vector machines based fault diagnosis for turbo-pump rotor. *Mechanical Systems and Signal Processing* 2006 **20**(4): 939–952.

[22] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach. *Neural Computation* 2000 **12**:2385–2404.

[23] Kim KI, Jung K, Kim HJ. Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters* 2002 **9**:40–42.

[24] Moghaddam B. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002 **24**:780–788.

[25] Yang MH. Kernel eigenfaces vs kernel fisherfaces:face recognition using kernel methods. *Proceedings of 5th IEEE International Conference on Automatic Face and Gesture Recognition*. Washington, DC, 2002; pp. 215–223.

[26] Scholkopf B, Smola A, Williamson R, Bartlett PL. New support vector algorithms. *Neural Computation* 2000 **12**:1207–1245.

[27] Platt JC, Cristianini N, Shawe-Taylor J. Large margin DAG's for multiclass classification. In *Advances in Neural Information Processing System*. MIT Press, 2000, pp. 547–553.

[28] Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks* 2002 **13**:415–425.

[29] Yuan SF, Chu FL. Fault diagnosis based on support vector machines with parameter optimisation by artificial immunisation algorithm. *Mechanical Systems and Signal Processing* 2007 **21**(3): 1318–1330.

[30] Yuan SF, Chu FL. Fault diagnosis based on particle swarm optimization and support vector machines. *Mechanical Systems and Signal Processing* 2007 **21**(4):1787–1798.

Chapter 86

Risk Monitoring of Aircraft Fatigue Damage Evolution at Critical Locations

Michael Shiao

Airport and Aircraft Safety Group, Aviation Research & Development Office, Federal Aviation Administration, Atlantic City International Airport, NJ, USA

1 INTRODUCTION

Traditionally, structural health of aircraft is monitored by scheduled or unscheduled inspections using nondestructive inspection (NDI) methods. However, human error, cost of inspections, reliability of crack detection, damage inspectability, etc., motivate the development of automated inspection techniques. Recent development in structural health monitoring (SHM) via on-board surface or embedded sensor systems allows continual NDIs for aircraft airworthiness. One key element for risk monitoring of aircraft fatigue damage is to characterize the damage evolution due to fatigue. In the Encylopedia, there are several articles also discuss the damage evolutionary process in a structure (*see* **Damage Evolution Phenomena and Models** and **Fatigue Life Assessment of Structures**). For metallic structures, damage tolerance (DT) analysis has been developed. The DT approach recognizes the existence of initial anomalies or flaws and incorporates inspection and repair/replacement as an important method to sustain structural reliability and safety.

The DT approach complements SHM and provides updated damage information for proper maintenance actions. It also provides SHM with the ability to forecast and update the risk using the information from damage monitoring.

For fatigue damage risk management, probabilistic DT analysis is one of the methods to predict the risk and to characterize the uncertain damage state of structures associated with damage initiation, accumulation, inspection, detection, and other maintenance

ISBN: 978-0-470-05822-0.

effects. However, there are several difficult issues in the DT approach that need to be addressed.

- **Assumptions of initial flaw sizes and other uncertainties in the DT models**

Assuming a deterministic flaw size is not realistic and the reliability of the design cannot be quantified. On the other hand, the initial flaw sizes are often too small to be detected by NDI tools or on-board SHM systems, and it is practically impossible to develop initial flaw-size probability distributions with confidence. To address the issue, a common approach is to use measurable defects at a later time, such as from a teardown inspection, and apply fracture mechanics crack growth models to back-extrapolate the defect sizes to develop the equivalent initial flaw size (EIFS). However, it is well known that the EIFS derived under a specific condition is not suitable for different geometries or loading conditions.

- **Uncertainties and sensitivities of NDI methods and advanced on-board SHM systems**

Characterizing NDI capability and sensitivity using probability of detection (PoD) is a well-accepted approach for conventional off-board NDI methods. PoD obtained for conventional off-board NDI methods represents the inspection reliability due to device sensitivity and variability; inspectors' physical and mental conditions; detection variability due to component geometry, defect shape, size, location, orientation, effect of operating environments, etc. To improve the inspection reliability, one can use on-board sensors for automated inspection on the fatigue hot spot and to eliminate human error.

The concept of a PoD curve is still viable for on-board sensors. The PoD curve for a particular on-board sensor of a specific manufacturing and application process needs to be developed. A PoD curve for a sensor indicates that not all sensors are manufactured and operated identically; there is variation in the detectability between sensors. For example, the PoD for an on-board sensor would be developed on the basis of the manufacturing process of the sensor, the corresponding electronics, and the *in situ* environment of application. Development of the PoD would involve a test article with multiple sensors and a series of cracks of various sizes. Multiple test articles would be used to determine the PoD as a function of crack sizes. Potentially, in the future, computational methods incorporating the physics of the sensor mechanics, structural mechanics, and probabilistic methods will be used to develop computational PoD curves [1].

- **Inspection schedule and frequency**

The simplest approach to select inspection intervals is to divide the service time by an integer and use equal intervals for conventional NDI methods. Clearly, this would not be optimal, as the best inspection intervals should reflect the fact that the defect growth rate is nonlinear, and it is also more effective to detect defects when they are easily detectable, but before they are near critical sizes. For on-board sensors, owing to their continual inspection feature, the question will be when to assess the data obtained from the continued monitoring. When an on-board sensor is installed, inspection frequency can be drastically increased, since cost incurred for manual inspection is of no concern. However, inspection dependency needs to be carefully investigated. Using conventional NDI methods, inspections are assumed to be independent of one another since the inspection process is a result of many independent probabilistic events. However, the continual inspections become highly correlated for advanced SHM systems since a specific SHM system, with known detection sensitivity, is used repeatedly throughout the aircraft life.

- **Maintenance-induced damage and quality of repairs and replacements**

When damage is detected, repair should follow immediately. It has been recognized that poor repair quality may cause future aircraft failure. In addition, there have been reports regarding maintenance-induced damage. Although in many risk analyses repair quality is assumed to be perfect for computational simplicity, repair quality may play an important role in probabilistic risk assessment when inspection and repair frequency increase.

In this article, a probabilistic framework for the risk forecasting and updating and associated methodologies are described. The framework includes risk forecasting probabilistic algorithms considering the uncertainties in inspection sensitivity, repair quality, fatigue crack growth characterization, and risk updating using updated uncertainties by Bayesian updating approach. Risk forecasting is based on the integration

of damage accumulation methodology, simulation-based probabilistic methods for uninspected structures, and special probabilistic algorithms to account for uncertainties associated with materials, structures, and loadings/environments, as well as uncertainties related to maintenance practices, such as inspection scheduling and techniques, replacement and retirement decisions, and repair quality (including maintenance-induced damage). The measured damage information through scheduled, nonscheduled, or continual inspection is used to update prior knowledge of uncertainties and hence improve the risk prediction.

2 PROBABILISTIC FRAMEWORK FOR RISK FORECASTING

The framework, illustrated in Figure 1, includes a wide range of uncertainties including:

- random or uncertain parameters in material (e.g., crack growth threshold, modulus of elasticity, fracture toughness);
- defect or flaw (including size, shape, and location, and the frequency of occurrence);
- loading, type of usage (with frequency of occurrence);
- finite-element model (including modeling error);
- crack growth model (including modeling error);
- maintenance (including inspection schedules, frequency of inspections, PoD curves, repair/replacement methods, and effects).

In the following, we discuss the maintenance event tree and repair quality modeling first, since they are the main reason risk forecasting using SHM is computationally challenging. It is followed by the discussion of probabilistic methods for risk prediction.

2.1 Maintenance event tree

In the case of damage accumulation processes, it may be uneconomical to design a structure in such a way that the reliability is sufficient for an entire design life. A more economical solution can be obtained by establishing a maintenance scheme. In

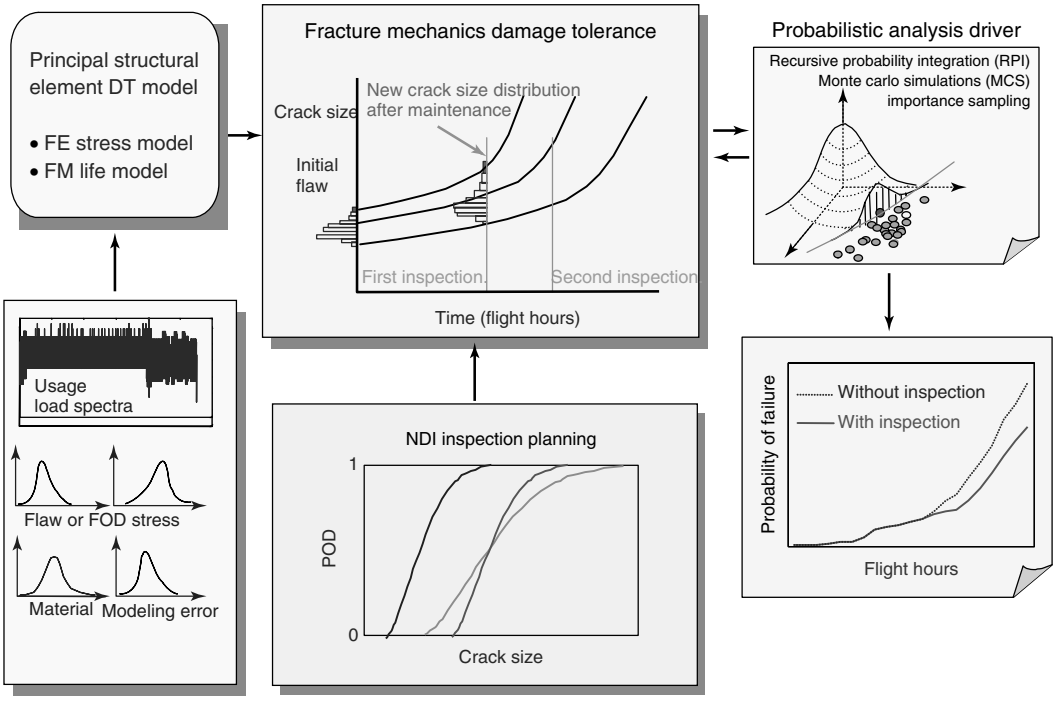


Figure 1. Probabilistic framework for risk forecasting.

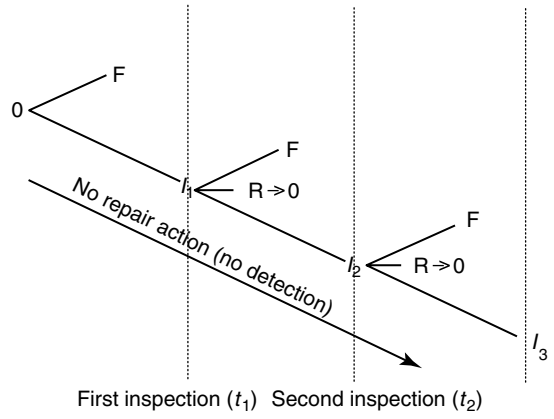


Figure 2. Maintenance event tree.

these cases, failures will not occur if the inspection detects damages and repair action is taken immediately. The sequence of inspection and repair events can be represented in an event tree, as shown in Figure 2.

Let the first inspection I_1 be planned at time t_1. At this node, there exists a branch with three possible events: (i) event of failure F—a failure occurs before t_2, (ii) event of repair R—the inspection detects a serious defect and repair is conducted, and (iii) event of no-finding I_2—no serious defect is detected and the next inspection at $t = t_2$ is planned. When a detected defect has been repaired, a new sequence is started and a new branch is generated at the next inspection.

2.2 Repair quality modeling

Perfect repair is unlikely. Often, field repair may produce inferior quality owing to insufficient equipment, inability to acquire high-quality materials on time, poor repair skills, etc. In addition, repair most often removes the crack tip of the propagating crack, yet it often leaves behind potential initiation sites for cracks to grow during future operation. For these reasons, assuming a perfectly repaired structure may be unconservative.

In general, a postrepair crack size distribution that is worse than that of the original parts has to be estimated. Berens *et al.* [2] used the equivalent repair quality to quantify the possible repair flaws inherited in the structure after repair. Although the effect of repair quality on the failure probability is not immediate, it can have a major effect on aging structures. In addition, the effect of the potential maintenance-induced damage should be modeled by further adjusting the postrepair crack size distribution.

2.3 Probabilistic methods for SHM using conventional NDI methods

In the last decade, several probabilistic methods and associated software have been developed for DT applications, considering uncertainties and inspection planning. While each of the methods works well for the damage scenario considered and the uncertainty model adopted, these methods are limited in their ability to solve general problems. Therefore, general and accurate methods are needed for DT analysis with maintenance planning under various uncertainties, including EIFS and location, material properties, loads and environmental effects, inspection scheduling, PoD, and repair quality and frequency.

Monte Carlo simulation (MCS) offers the most robust and reliable solution framework for general problems. The major issue is that MCS is usually time consuming. For maintenance planning and risk monitoring, the computational issue is further amplified because the conventional approach requires an MCS for each different maintenance plan. As a result, many sets of MCSs are required to search for the optimal solution by exploring the design space that consists of many possible combinations of inspection scheduling, techniques, and repair/replacement/retirement strategies.

To relieve the computational burden of traditional MCS methods and to further reduce the computational time for generating many sets of crack growth histories for maintenance planning, this article describes an efficient method that combines the generality of MCS with the efficiency of analytical probabilistic methods. The core of the method is a recursive probability integration (RPI) method that allows repeated use of baseline MCS-based crack growth histories for various maintenance plans. The fundamental concept of RPI is based on branching out the probable events after each maintenance action following an inspection where the PoD is applied. The probability of occurrence of each branched event is then determined on the basis of the probability of crack detection. In addition to allowing the reuse of

crack growth histories, RPI has an additional benefit of improving the MCS sampling efficiency, especially when a maintenance plan significantly reduces the probability of failure. When the RPI algorithm is used, inspections are assumed to be independent. This assumption is appropriate for the traditional NDIs, i.e., different operators, equipment, and locations, and widely separated by the time of inspections, but may not be so for on-board sensors.

The MCS-based RPI method randomizes non-inspection-related random variables only and has been demonstrated to be several orders of magnitude faster than traditional MCS methods, which randomize all uncertain random variables [3, 4]. The computational efficiency can be further improved by a combination of the RPI method and the conditional expectation method (CEM) [5, 6]. Both methods, as well as the proposed analysis procedure, are described in detail in the following sections.

2.3.1 MCS-based recursive probability integration method

Figure 3 shows the possible fatigue crack growth paths from an MCS without inspections. Failure of each fatigue path is indicated by the cross symbol for any failure requirement, such as net section yield, reaching critical crack size, or fracture. To illustrate the RPI concept using a graphical presentation for clarity, Figure 4 shows fatigue life simulations without inspections. In the figure, the horizontal line indicates the beginning and end of fatigue life of a simulation without inspection. This line is called the *fatigue path*, since it implicitly represents the fatigue crack growth path from the beginning to the end.

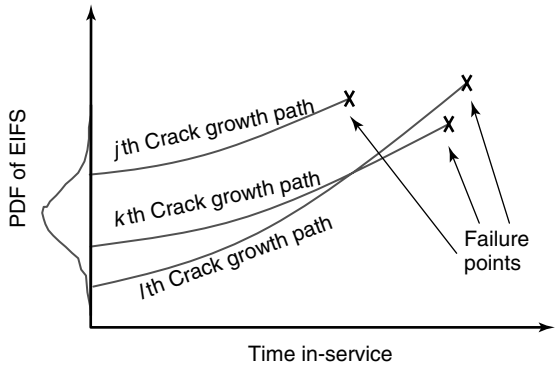


Figure 3. Fatigue life and fatigue crack growth paths without inspection.

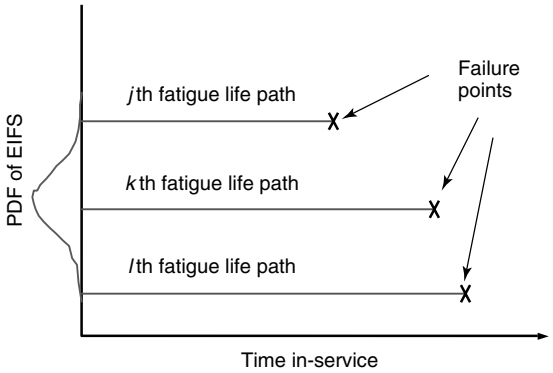


Figure 4. Fatigue life paths without inspection.

Now let us investigate the probability calculation. When n simulations are conducted in an MCS, the probability of occurrence of each path (simulation) is $1/n$. The conditional probability of failure of jth fatigue (simulation) path, given the occurrence of this path, ${P_f}^c(j)$, is defined by equation (1), where c stands for conditional.

$$\begin{aligned} {P_f}^c(j) &= 0 \quad j\text{th fatigue life} > \text{design life} \\ {P_f}^c(j) &= 1 \quad j\text{th fatigue life} < \text{design life} \end{aligned} \tag{1}$$

The probability of failure for an MCS without inspection can be obtained by equation (2).

$$P_f = \frac{1}{n}\sum_{j=1}^{n} {P_f}^c(j) \tag{2}$$

Now let us add one inspection to the analysis. Figure 5 shows an inspection at time t that is conducted on the jth fatigue path.

The crack size at the time of inspection is denoted as $a(j,t)$. Owing to the inspection, the jth fatigue path will branch out to two possible paths. If the crack is not detected, the jth fatigue path will continue its original fatigue path as indicated by path 1 in Figure 5. When a crack is detected and repaired immediately, the following path can be either one of the multiple repair paths, indicated by path 2 in Figure 5. As a result of probability branching, ${P_f}^c(j)$ in equation (2) needs to be updated accordingly by equation (3).

$$\begin{aligned} P_f^c(j) = {} & PoD(a(j,t)) \times P_f(\text{path 1}) \\ & + (1 - PoD(a(j,t))) \times P_f(\text{path 2}) \end{aligned} \tag{3}$$

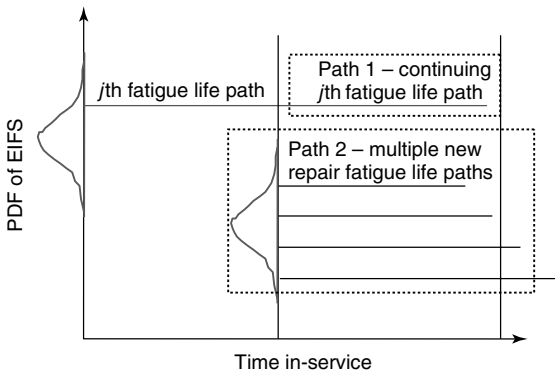


Figure 5. Probabilistic branching due to inspection and repair.

$PoD(a(j,t))$ is the PoD of crack size $a(j,t)$. P_f(path 1) is the conditional probability of failure of the original (jth) fatigue path, as defined in equation (1). P_f(path 2) is the probability of failure of repair paths, which can be determined by an MCS without inspection starting at the time of inspection using equations (1) and (2). Once $P_f{}^c(j)$ is updated by equation (3), the probability of failure with one inspection can be computed by equation (2). Using the same analogy, probability of failure for multiple inspections can be derived. Several key parameters used for RPI are defined below and illustrated in Figure 6.

t_s: service life (or design life);
m: number of inspections;
$MCS(k)$: an MCS without inspections starting at the kth inspection, where $k = 0$ to m; $k = 0$, representing simulations with original parts; and $k > 0$, representing simulations with repair parts;
n_k: number of simulations for $MCS(k)$;
$t_f(k, j)$: fatigue life of the jth simulation of $MCS(k)$;
$a(k, j, i)$: the crack size at the ith inspection of the jth fatigue path of $MCS(k)$; $k = 0$ to m; $j = 1$ to n_k; for each k, $i = k + 1$ to m;
Full Path i: a probability event consists of complete fatigue paths of an MCS considering subsequent inspections starting at the ith inspection, $i = 0$ to m;
$Br(k, j, i)$: a branched probability event consists of all possible fatigue paths of an MCS considering subsequent inspections starting at the ith inspection with the following conditions.

The initial condition of the fatigue branch $Br(k, j, i)$ is inherent from the condition at the ith inspection of the jth fatigue path of $MCS(k)$ where $k = 0$ to m; for each $k, i = k$ to $m - 1$; for a pair of k and i, $Br(k, j, i)$ consists of fatigue branch $Br(k, j, i + 1)$ and *Full Path*$(i + 1)$

Figure 6 shows a full repair path starting at the $(m - 2)$th inspection. The dashed line represents the jth fatigue path of $MCS(m - 2)$, and the circle indicates that an inspection has been conducted along this path. A diamond indicates that a repair action has been taken after detecting a crack. The solid line with a diamond at the beginning and an arrow at the end represents an MCS considering subsequent inspections. It also represents a Full Path starting at the time indicated by the diamond.

Figure 6 also shows that $Br(m - 2, j, m - 2)$ consists of a single (jth) realization of $MCS(m - 2)$ and two possible branched probability events, *Full Path*$(m - 1)$ and *Full Path*(m), after subsequent inspections along this realization. From the definition of Full Path, *Full Path*$(m - 2)$ is also represented

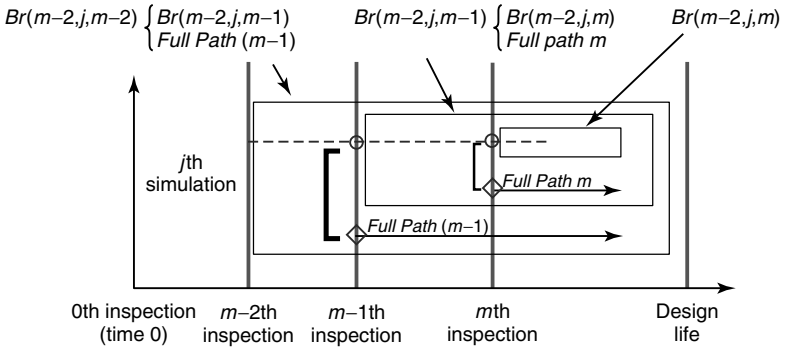


Figure 6. *Full Path*$(m - 2)$ with two subsequent inspections.

by the probability events $Br(m-2, j, m-2)$, where $j = 1, n_{m-2}$. The following are also shown in Figure 6: $Br(m-2, j, m-2)$ consists of $Br(m-2, j, m-1)$ and the *Full Path*$(m-1)$, $Br(m-2, j, m-1)$ consists of $Br(m-2, j, m)$ and the *Full Path* m, and $Br(m-2, j, m)$ represents the last segment of the jth fatigue path of $MCS(m-2)$.

Now let us follow the jth simulation path to investigate the probability of occurrence of each probability event. At the $(m-1)$th inspection, if a crack is detected and a repair action is taken, the probability of occurrence of a repair action (or *Full Path*$(m-1)$) is equal to the PoD for the crack size $a(m-2, j, m-1)$. If no crack is detected, the probability of occurrence of a fatigue branch $Br(m-2, j, m-1)$ is equal to $1 - PoD$. As defined before, *Full Path*(i), where $i > 0$, represents the probability event after a repair action is taken at the ith inspection. It is noted that *Full Path*(i) depends upon index i but is independent of indices k and j (same indices for $a(k, j, i)$ and $Br(k, j, i)$), which means that this event can be used repeatedly for any k and j index. In other words, the probability of failure of this probability event just needs to be calculated once. However, the probability of occurrence of this event will be the PoD of crack size $a(k, j, i)$, which is a function of indices k and j.

The basic concept described above leads to the development of the RPI algorithm, as shown in equations (4) and (5).

$$P_f^{\,Full\,Path\ k} = \frac{1}{n_k}\sum_{j=1}^{n_k} P_f^{\,Br(k,j,k)} \tag{4}$$

where $k = m$ to 0

In equation (4), $P_f^{\,Br(k,j,k)}$ for $k < m$ is calculated recursively by equation (5).

$$\begin{aligned} P_f^{\,Br(k,j,i)} &= (1 - PoD(a(k, j, i+1))) \\ &\quad \times P_f^{\,Br(k,j,i+1)} + PoD(a(k, j, i+1)) \\ &\quad \times P_f^{\,Full\,Path(i+1)} \end{aligned} \tag{5}$$

where $i = m-1$ to k

The P_f of the structure, considering all the uncertainties, as well as a maintenance plan, is equal to the failure probability of the *Full Path* 0 found at the end of the computation by recursive equations.

2.3.2 CEM-based RPI method

This method, referred as *RPI/CEM*, is a two-stage numerical procedure. The first stage computes the crack growth history using non-inspection-related random variables that were randomly simulated using CEM. The second stage accounts for maintenance-related uncertainties by the RPI method together with the crack growth histories generated from stage 1. In the following, we discuss the CEM first and then discuss the integration of RPI and CEM for risk assessment.

Let $\underline{Z}$ be a set of random variables and g be a performance function. The probability of failure given probability event, $g(\underline{Z}) < 0$, can always be determined using MCSs by generating random samples for all random variables as shown in the following equation.

$$P_f = \int \cdots \int P[g(\underline{Z}) < 0]\ f_{\underline{Z}}\ \mathrm{d}\underline{z} \tag{6}$$

where

$$\begin{aligned} P[g(\underline{Z}) < 0] &= 0 \quad \text{if} \quad g(\underline{Z}) > 0 \\ P[g(\underline{Z}) < 0] &= 1 \quad \text{if} \quad g(\underline{Z}) < 0 \end{aligned}$$

In equations (6) and (7), f represents the probability density function (PDF), g represents performance function, other lowercase symbols represent realizations (randomly selected real values), P represents probability, E represents expectation, superscript c represents conditional, and other uppercase symbols represent random variables. As shown, equation (6) determines the average of n values from simulations; most of them are 0 and very few are 1. As a result, it is well known that MCS is time consuming and often computationally impossible for complex engineering systems. Therefore, the CEM approach was adopted to circumvent the computational demand of the MCS approach. The fundamental of CEM is described next. If one can group the random variables $\underline{Z}$ into two sets, $\underline{X}$ and $\underline{Y}$, equation (6) can be rewritten as shown in equation (7).

$$\begin{aligned} P_f &= \int \cdots \int P[g(\underline{x}, \underline{y}) < 0] f_{\underline{X}} f_{\underline{Y}}\ \mathrm{d}\underline{x}\ \mathrm{d}\underline{y} \\ &= \int \cdots \int P[g^c(\underline{X}|\underline{y}) < 0] f_{\underline{Y}}\ \mathrm{d}\underline{y} \\ &= \int \cdots \int P_f^c(\underline{X}|\underline{y}) f_{\underline{Y}}\ \mathrm{d}\underline{y} \\ &= E[P_f^c] \end{aligned} \tag{7}$$

As can be seen, equation (7) determines the average value of $P[g^c(\underline{X}|\underline{y}) < 0]$. In each simulation, a set of realizations $\underline{y}$, for random variables $\underline{Y}$, is randomly generated. $P[g^c(\underline{X}|\underline{y}) < 0]$ can be determined by a numerical method such as numerical integration or other fast probability integrators. With proper selection of $\underline{X}$, the variation of $P[g^c(\underline{X}|\underline{y}) < 0]$ can be well behaved, where fewer samples are needed to determine its mean value.

Combining the RPI with the CEM, equation (5) is modified, as shown in equation (8).

$$\begin{aligned} {P_f}^{Br(k,j,i)} &= (1 - PoD(a(k, j, i+1))) \\ &\quad \times {P_f}^{Br(k,j,i+1)} \\ &\quad + (1 - {P_f}^{\text{acc}}(j, i+1)) \\ &\quad \times PoD(a(k, j, i+1)) \times {P_f}^{FullPath(i+1)} \\ &\quad + {P_f}^{\text{int}}(j, i) \end{aligned} \tag{8}$$

where $i = m - 1$ to k

${P_f}^{\text{acc}}(j, i)$ in equation (8) is the cumulative probability of failure in the jth simulation at the $(i + 1)$th inspection time considering inspection effect, which is 0 using MCS. ${P_f}^{\text{int}}(j, i)$ represents the cumulative probability of failure (POF) between the ith and the $(i + 1)$th inspections without inspection effect, which is either 0 or 1 when MCS is used. ${P_f}^{\text{int}}(j, i)$ in equation (8) is calculated by the following equation.

$$\begin{aligned} {P_f}^{\text{int}}(j, i) &= {P_f}^{c}(\underline{X}|\underline{y_j} \text{ at } i+1^{\text{th}} \text{ inspection}) \\ &\quad - {P_f}^{c}(\underline{X}|\underline{y_j} \text{ at } i^{\text{th}} \text{ inspection}) \end{aligned} \tag{9}$$

In equation (9), $\underline{X}$ represent random variables not randomly generated, and $\underline{y_j}$ are the realizations randomly generated for the random variables $\underline{Y}$ in the jth simulation. $P_f^c(\underline{X}|\underline{y_j})$ can be determined by any numerical methods, such as numerical integration or advanced probability algorithms, for a given failure function $g(\underline{X}|\underline{y_j}) < 0$.

This procedure reduces the computational burden of using MCS alone (randomly simulates all uncertainties) or using RPI/MCS (randomly simulates non-inspection-related uncertainties). In addition, it maintains the unique capability of the reuse of stored crack growth histories without inspections to compute risk as a function of inspection time and/or inspection techniques without additional stress and life analyses.

Maintenance optimization can be achieved by repeated stage-2 analysis using the same set of crack growth histories considering various maintenance strategies. Since the RPI method uses the probabilistic results of inspection-free structures, advanced probabilistic methods for inspection-free structures, such as importance sampling methods, can be used to generate crack growth history. This particular feature results in additional computational time reduction.

As mentioned earlier, repeated risk assessments can be achieved for different maintenance strategies using the same set of random crack growth histories generated using non-inspection-related random variables. As a result, more computational time can be saved for maintenance optimization.

2.3.3 Sampling-based RPI method

For MCS or importance sampling (IS)-based RPI method, all non-inspection-related random variables are randomly generated to compute crack growth histories without inspection. As a result, equation (9) for RPI/CEM can be simplified to equation (10) for MCS or IS-based RPI.

$$\begin{aligned} {P_f}^{\text{int}}(j, i) &= 0 \quad \text{no failure between } i\text{th and } i+1\text{th inspection} \\ {P_f}^{\text{int}}(j, i) &= 1 \quad \text{one failure between } i\text{th and } i+1\text{th inspection} \end{aligned} \tag{10}$$

In addition,

$${P_f}^{\text{acc}}(j, i+1) = 0 \tag{11}$$

2.3.4 Special RPI algorithms for perfect repairs

In cases where the failure contribution of repair components is known to be secondary, equations (10) and (11) can be simplified, as shown in equations (12) and (13).

$$P_f = P_0 \frac{1}{n} \sum_{j=1}^{n} P_f^{Br(0,j,0)} \tag{12}$$

$$P_f{}^{Br(0,j,i)} = (1 - PoD(a(0, j, i+1))) \times P_f{}^{Br(0,j,i+1)} + P_f{}^{\text{int}}(j, i) \quad (13)$$

where $i = m - 1$ to 0 ; $j = 1, n$

2.4 Probabilistic methods for SHM system with on-board sensors

SHM with on-board sensors minimizes human error through automation. However, other factors that affect the sensor sensitivity and variability still exist. Therefore, the concept of PoD curves is still valid to characterize the performance of sensors, as done for traditional inspections and for risk assessment. When on-board sensors are installed in a structural component, the same sensor is performing subsequent inspections. As a result, the variation of sensor sensitivity from one inspection to the other on the same component is reduced. Inspection dependency has been studied extensively by Shook *et al.* [7–9]. If there are no other factors that may affect the sensor sensitivity, inspections on the same structural component are considered fully dependent. However, the operational condition at each inspection time may be different. In addition, it is known that on-board sensors are sensitive to surrounding operating environments such as temperature, moisture, and loads. As a result, multiple inspections of on-board SHM systems will not be completely repeatable and they are not truly fully dependent. However, as shown by Shook *et al.* [7], modeling the inspection as independent leads to unrealistic over optimistic inspection efficiency and modeling the inspections as dependent is conservative.

Clearly, there is an underlying mechanism for variations in the sensor sensitivity. The degree of dependency can be accounted for by a physics-based stochastic modeling of those random factors, such as operational temperature, moisture, and load condition, if data is available. If not, risk should be bounded by results with independent and fully dependent inspections. In this section, only dependent inspection will be addressed. Further study is required to investigate risk forecasting with partially correlated inspections. Another issue that will not be addressed here is the effect of repair quality on the risk forecasting. The reason is that, other than MCSs methods, efficient probabilistic algorithms considering repair quality and correlated inspections have not yet been developed.

Assumption of dependent inspection assumes that damage detectability of a sensor is developed at a fixed testing environment or effects on its variability by other factors are negligible. Figure 7 illustrates the inspection sequence in jth simulation. Let E_i be the probability event of damage detection at ith inspection. As shown in Figure 8, E_i is a subset of E_{i+1} when dependent inspection is assumed. The reason is that the sensor sensitivity does not change during the damage evolution process. The probability event E_i becomes a subset of E_{i+1} is due to an increase in the crack size only since there is no additional contribution from other sources of randomness to the probability space of E_{i+1}. In Figure 7, F_j is the conditional failure event for the jth simulation conditional on the jth set of realizations. The joint failure event A_j for the jth simulation can be expressed by equation (14).

$$A_j = \left(\bigcap_{i=1}^{m} \overline{E_i}\right) \bigcap F_j \quad (14)$$

where $\overline{E_i}$ is the complement of E_i, and m is the number of inspections. When failure occurs at a given time in service in the jth simulation, the conditional probability of failure $P(F_i)$ is equal to 1. If the jth simulation survives at a given time in service, $P(F_i)$ is equal 0. As can be seen from Figure 8, the joint event of $\overline{E_i}$ is the complement of E_m, as shown in equation (15).

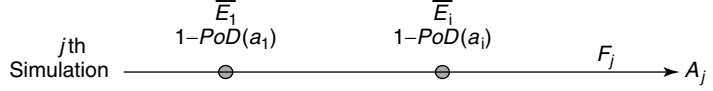


Figure 7. Inspection sequence in a given simulation. E_i: probabilistic event that crack was detected at ith inspection. $\overline{E_i}$: probabilistic event that crack was not detected at ith inspection. F_j: probabilistic event that failure occurs at time in service. A_j: probabilistic event that jth simulation has failed at time in service.

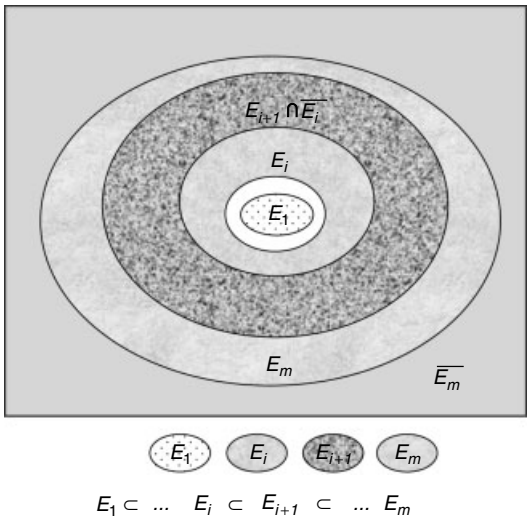


Figure 8. Dependency of continual inspection events.

$$\bigcap_{i=1}^{m} \overline{E_i} = \overline{E_m} \tag{15}$$

By substituting equation (15) in equation (14), we obtain equation (16).

$$A_j = \left(\bigcap_{i=1}^{m} \overline{E_i} \right) \bigcap F_j = \overline{E_m} \bigcap F_j \tag{16}$$

We also know that $\overline{E_m}$ and F_j are independent probability events. Therefore, the conditional probability of failure for each simulation conditional on a set of realizations is shown in equation (17).

$$\begin{aligned} P(A_j) &= P(\overline{E_m} \bigcap F_j) \\ &= P(\overline{E_m}) P(F_j) \\ &= (1 - PoD(a_m)) P(F_j) \end{aligned} \tag{17}$$

Probability of failure is then determined by equation (18).

$$P_f = \frac{1}{n} \sum_{j=1}^{n} P(A_j) = \frac{1}{n} \sum_{j=1}^{n} (1 - PoD(a_m)) P(F_j) \tag{18}$$

where n is the number of simulations.

3 PROBABILISTIC FRAMEWORK FOR RISK UPDATING

SHM is a way to assess the airworthiness of aircraft structural systems. One of the major elements of the SHM is the ability to update the risk using the information provided by either the on-board sensors or off-board scheduled or unscheduled NDIs. For fatigue damage management, probabilistic DT analysis is one of the methods to predict the risk and to characterize the uncertain damage state of structures associated with damage initiation, accumulation, inspection, detection, and other maintenance effects. However, probabilistic analysis relies on proper selection of the input parameters of uncertain variables, which are typically difficult to define.

During the design phase and early operational life, some of the input parameters for probabilistic DT analysis were quantified on the basis of expert opinions from previous experience, or were extracted from laboratory experiments. Although experiments and analyses can be performed to help characterize uncertainties, realistic conditions are often difficult to model. As a result, the risk forecasting may require further attention owing to questionable parameters that may be used in the prediction. This dilemma can be alleviated by using the new information in structural damage state found in later structural life. The measured damage information can be used to update prior knowledge of uncertainties by Bayesian updating approach, as described in the following section.

3.1 Bayesian updating formulation

Structural applications using Bayesian updating approach to improve the accuracy of risk prediction have been studied in the past [10, 11]. The Bayesian updating framework discussed in this section is illustrated using damage accumulation processes modeled through a probabilistic DT analysis with a set of random variables. The random variables include EIFS, fracture toughness, parameters for crack growth equation, extreme load, yield strength, and plate thickness. For illustration purposes, a simple example is used that assumes EIFS follows the two-parameter (α, β) Weibull distribution and α and β are quantified by random variables α_0 and β_0,

respectively. Using the example, Bayesian updating formation is described to determine the optimal values of α_0 and β_0 based on the latest crack information.

In Figure 9, $\underline{a}_{i,}(t_i)$ represents the crack found in the ith component at the inspection time t_i. The solid circle indicates the time and crack size at detection. Figure 9 also shows the PoD curve of the NDI technique used for the crack detection. The dashed line represents real crack growth history of each operational component, which is normally unknown. Baye's theorem allows an update of prior knowledge using new evidence. With this theorem, updated probability distribution function of α_0 and β_0 based on the observed crack sizes $\underline{a}_{i,}(t_i)$ is obtained, as shown in equation (19).

$$\begin{aligned} & f_U(\alpha, \beta | \underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)) \\ &= \frac{q_D(\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n), \alpha, \beta)}{\iint q_D(\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n), \alpha, \beta)\ \mathrm{d}\alpha\ \mathrm{d}\beta} \\ &= \frac{g_D(\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n) | \alpha, \beta) f_0(\alpha, \beta)}{\iint g_D(\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n) | \alpha, \beta) f_0(\alpha, \beta)\ \mathrm{d}\alpha\ \mathrm{d}\beta} \end{aligned} \tag{19}$$

In equation (19), n is the total number of components. If a crack is not found in component j, $\underline{a}_{i,}(t_i)$ should be removed from the formulation. f_0 represents the prior joint PDF of α_0 and β_0. f_U represents the updated joint PDF of α_0 and β_0, given a detected crack size at each component. It is also referred to as the *posterior joint PDF* of α_0 and β_0. q_D represents the joint PDF of detected crack sizes, α_0 and β_0. g_D represents the joint PDF of detected crack sizes, given the realizations α and β of random variables α_0 and β_0, respectively. It can be computed using various probabilistic methods; however, joint PDF is extremely difficult to obtain. To address this problem, the following approach is proposed.

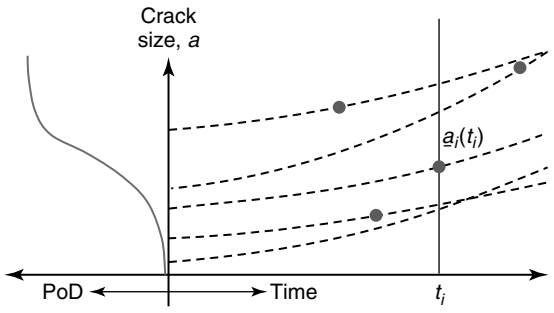


Figure 9. PoD curve and crack growth histories and detected crack sizes.

For many fatigue problems, $\underline{a}_{i,}(t_i)$ can be considered to be statistically independent events. For those cases, g_D can be simplified, as shown in equation (20).

$$\begin{aligned} & g_D(\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n) | \alpha, \beta) \\ &= \prod_{i=1}^{n} h_D(\underline{a}_i(t_i) | \alpha, \beta) \end{aligned} \tag{20}$$

h_D represents PDF of detected crack size in the ith component at inspection time t_i and is determined by equation (21).

$$\begin{aligned} h_D(\underline{a}_i(t_i) | \alpha, \beta) = c_i \times h_0(\underline{a}_i(t_i) | \alpha, \beta) \\ \times PoD(\underline{a}_i(t_i)) \end{aligned} \tag{21}$$

In the equation, $PoD(\underline{a}_i(t_i))$ is the PoD of crack size $\underline{a}_{i,}(t_i)$ for a NDI method used at time t_i. c_i is the normalization constant. h_0 represents PDF of a crack size in the ith component found at inspection time t_i, given α and β. It can be determined by various physics-based probabilistic methods. By substituting equation (20) in equation (19), equation (22) is obtained.

$$\begin{aligned} & f_U(\alpha, \beta | \underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)) \\ &= \frac{\left[\prod_{i=1}^{n} h_D(\underline{a}_i(t_i) | \alpha, \beta)\right] f_0(\alpha, \beta)}{\iint \left[\prod_{i=1}^{n} h_D(\underline{a}_i(t_i) | \alpha, \beta)\right] f_0(\alpha, \beta)\ \mathrm{d}\alpha\ \mathrm{d}\beta} \end{aligned} \tag{22}$$

In equation (22), f_U is the posterior joint PDF of α_0 and β_0 with observed crack sizes $\underline{a}_i(t_i)$. f_0 is called the *prior probability* and represents prior knowledge of input parameters, and $\prod_{i=1}^{n} h_D(\underline{a}_i(t_i) | \alpha, \beta)$ is the "likelihood function", which reflects the likelihood that the observed event could indeed take place.

The optimal estimations of random variables α_0 and β_0 are the values where the posterior joint

PDF is maximized. In many practical applications, such as parameter estimation, the denominator in equation (22) can be considered as a normalization constant. Therefore, a proportional relation can be shown in equation (23).

$$f_U(\alpha, \beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)) \propto \left[\prod_{i=1}^{n} h_D(\underline{a}_i(t_i)|\alpha, \beta)\right] f_0(\alpha, \beta) \tag{23}$$

Therefore, a better alternative is to maximize the function G defined in equation (24) for the optimal parameter estimation.

$$G = \left[\prod_{i=1}^{n} h_D(\underline{a}_i(t_i)|\alpha, \beta)\right] f_0(\alpha, \beta) \tag{24}$$

Owing to the complexity of the function G, it is not a trivial task to find the maximum value. An alternative in Bayesian estimation is to compute the posterior means for respective parameters, as shown in equations (25) and (26).

$$E[\alpha|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)] = \int_0^\infty \int_0^\infty \alpha f_U(\alpha, \beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)) \times \mathrm{d}\alpha \ \mathrm{d}\beta \tag{25}$$

$$E[\beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)] = \int_0^\infty \int_0^\infty \beta f_U(\alpha, \beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)) \times \mathrm{d}\alpha \ \mathrm{d}\beta \tag{26}$$

Since it is hard to compute the posterior joint PDF, an importance sampling scheme is introduced [12], as shown in equations (27) and (28). In the equations, k is the joint PDF to be selected for importance sampling.

$$E[\alpha|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)] = \int_0^\infty \int_0^\infty \frac{\alpha f_U(\alpha, \beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n))k(\alpha, \beta)}{k(\alpha, \beta)} \times \mathrm{d}\alpha \ \mathrm{d}\beta \tag{27}$$

$$E[\beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n)] = \int_0^\infty \int_0^\infty \frac{\beta f_U(\alpha, \beta|\underline{a}_1(t_1), \ldots, \underline{a}_i(t_i), \ldots, \underline{a}_n(t_n))k(\alpha, \beta)}{k(\alpha, \beta)} \times \mathrm{d}\alpha \ \mathrm{d}\beta \tag{28}$$

The Bayesian updating formulations derived in this section can be applied to any damage accumulation processes not just to aircraft fatigue damage evolution, that observations of damage states can be obtained. With the new estimates, designers or analysts can proceed to forecast and update the potential risk as discussed in the previous section.

4 SUMMARY

A probabilistic framework was described for risk monitoring of aircraft fatigue damage evolution at critical locations when SHM concepts are used. The risk monitoring framework includes probabilistic risk forecasting and updating of damage accumulation processes with inspections and maintenance. Four main features of this framework are (i) modeling the damage accumulation process by a fatigue crack analysis, (ii) modeling inspection dependency within an SHM system, (iii) forecasting risk using special probabilistic algorithms to account inspection and maintenance-related uncertainties, and (iv) enabling Bayesian updating using measured data from multiple components with different inspection times and NDI methods or advanced SHM systems. The fast probability integration method was described for the probabilistic risk forecasting, including failure contribution from repaired structural components for independent inspections. Probabilistic formulas for fully dependent inspection assuming perfect repair are also provided. The risk prediction of correlated inspection should be bounded by results of risk assessments from independent and fully dependent inspections. Mathematical formulations for risk updating were described. Risk updating is accomplished by the updating of uncertain statistical parameters. The updating approach consists of (i) probabilistic DT analysis without inspection-related uncertainties, (ii) damage information obtained by SHM, and (iii) Bayesian updating approach.

REFERENCES

[1] Rajesh SN, Udpa L, Udpa SS. Estimation of eddy current probability of detection (PoD) using finite element method. *Review of Progress in Quantitative Nondestructive Evaluation* 1993 **12**:2365–2372.

[2] Berens AP, Hovey PW, Skinn DA. *Risk Analysis for Aging Aircraft Fleets. U.S. Air Force Wright Laboratory Report. WL-TR-91-3066*, Vol. 1, October 1991.

[3] Wu YT, Shiao M, Shin Y, Stroud WJ. Reliability-based damage tolerance methodology for rotorcraft structures. *Proceeding of 2004 SAE World Congress; Reliability and Robust Design in Automotive Engineering*. Detroit, MI, Special Publication No. 1844, March 2004.

[4] Shiao M, Boyd K, Fawaz S. A risk assessment methodology and tool for probabilistic damage tolerance-based maintenance planning. *8th FAA/NASA/DoD Aging Aircraft Conference*. Palm Springs, CA, January 31–February 3, 2005.

[5] Shiao M Shyprykevich P. Hybrid probabilistic method for composite aircraft design. *45th Structures, Structural Dynamics and Materials Conference*. Palm Springs, CA, April 19–21, 2004.

[6] Shiao M. Risk-based maintenance optimization. *9th International Conference on Structural Safety and Reliability*. Rome, Italy, June 19–22, 2005.

[7] Shook B, Millwater H, Hudak S, Enright MP, Francis WL. Impact of Multiple On-Board Inspections on Cumulative Probability of Detection. *IGTI Conference*. Reno, Nevada. June 2005, GT2005-68585.

[8] Shook B, Millwater H, Enright MP, Hudak SJ, Francis WL. Simulation of recurring automated inspections on probability-of-fracture estimates. *Journal of Structural Health Monitoring* (accepted).

[9] Shook B, Millwater H, Hudak SJ, Enright MP, Francis WL. Comparison of continual on-board inspections to a single mid-life inspection for gas turbine engine disks. *46th AIAA Structures, Dynamics and Materials Conference*. Austin, Texas. April 18–21, 2005.

[10] Lin KY, Du J, Rusk D. *Structural Design Methodology Based on Concepts of Uncertainty*. NASA/CR-2000-209847. 2000.

[11] Shiao M. Risk forecasting and updating for damage accumulation processes with inspections and maintenance. *5th International Workshop on Structural Health Monitoring*. Stanford, CA, September 2005.

[12] Gelman AB, Carlin JS, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman & Hall/CRC: New York, 1995, pp. 307–311.

Chapter 87
Risk Monitoring of Civil Structures

Narito Kurata
Kobori Research Complex, Kajima Corporation, Tokyo, Japan

1 INTRODUCTION

Risk of buildings and civil engineering structures from natural and man-made hazards is large and growing. Natural hazards include earthquake, tsunami, floods, and hurricane. The 1995 Kobe earthquake in Japan killed over 6400 people and the number of completely destroyed buildings and houses was over 100 000. The 2004 and 2007 Niigata earthquake in Japan, the tsunami due to the 2004 Indian Ocean earthquake, and the 2005 Hurricane Katrina in New Orleans caused heavy damage. Man-made hazards include fires, crime, and terrorist attack. The 110-floor twin towers of the World Trade Center and numerous other buildings at the World Trade Center site were destroyed by the September 11, 2001 terrorist attacks. The 2007 Minneapolis bridge disaster was not a natural disaster, but an important issue of structural damage. The Interstate 35W bridge collapsed into the Mississippi River during rush hour on August 1, 2007. It was pointed out that a quarter of the nation's major bridges in the United States carried more traffic than they were designed to bear. Risk monitoring is one of the most promising emerging technologies for mitigation of these hazards [1].

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 CONCEPT OF RISK MONITORING

The word "risk" is used in a number of different ways and contexts. For example, in [2], it is defined as the chance of loss, the possibility of loss, uncertainty, the dispersion of actual from expected results, the probability of any outcome different from the one expected, or a condition in which a possibility of loss exists. Also, it is often used alongside the term *hazard*. This term is used to describe the potential of a compound to cause damage or harm.

Buildings are subjected to natural hazards such as severe earthquakes, tsunamis, floods, and hurricanes,

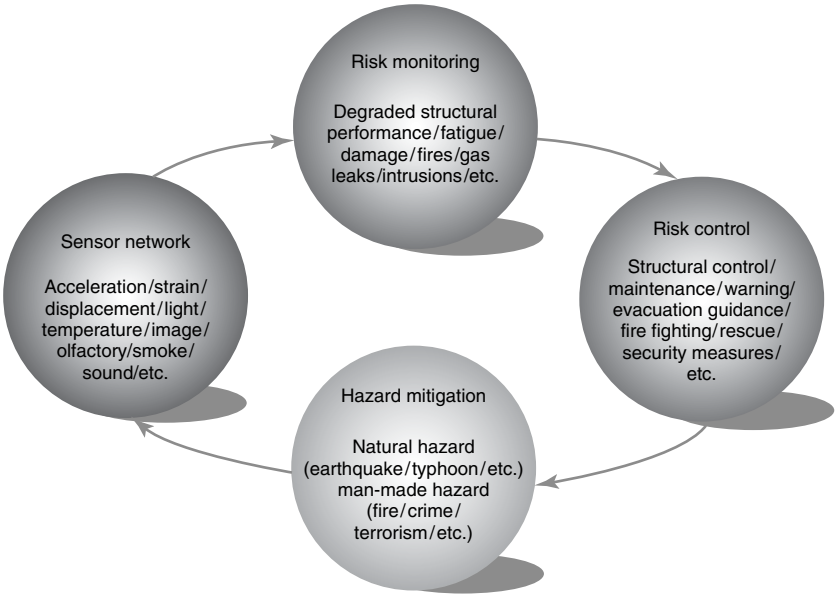


Figure 1. Risk monitoring and hazard mitigation.

as well as man-made hazards such as fires, crime, and terrorist attacks, during their long-term use. To mitigate these hazards, monitoring various risks in a building employing structural health monitoring (SHM) technologies is necessary. Figure 1 indicates a concept of risk monitoring and hazard mitigation, and Table 1 shows various kinds of hazards and possible applications/combination of sensors. The risks to buildings are defined as follows.

2.1 Risks for natural hazards

Risks for natural hazards are related to structural performance of buildings. Structural deterioration is a risk that should be monitored for maintenance of buildings for long-term use (*see* **Maintenance Principles for Civil Structures**; **Usage Management of Civil Structures**). It can be evaluated by inspection of crack, corrosion, and fatigue of structural members. Nondestructive evaluation and strain-based sensing technologies are effective (*see* **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**). Structural damage is an important risk that should be monitored for postdisaster mitigation. It can be evaluated by sensing of floor acceleration and story deformation of buildings. Information from risk monitoring of structural

Table 1. Sensor applications

Hazard	Application	Sensor
	Observation	Acceleration, wind speed
Earthquake/ tsunami/flood/ hurricane	Experiment	Acceleration, strain
	Structural control	Acceleration
	Structural deterioration	Crack, corrosion, fatigue, strain
	Structural damage	Acceleration, story deformation
	Fire detection	Temperature, smoke, acoustic, acceleration, olfactory
Fire	Gas leak detection	Olfactory
	Alarm, warning	Sounder
	Evacuation control	Temperature, smoke, acoustic, light, olfactory
Crime/terrorist attack	Surveillance	Camera, acceleration, acoustic, light
	Security alert	Sounder

damage can be used for risk control measures such as warning, alarms, and evacuation guidance.

2.2 Risks for man-made hazards

Influence on buildings by man-made hazards such as fires, crimes, and terrorist attacks is growing in recent years. Especially, there are points in common between terrorist attack and natural disaster. They are unexpected and unpredictable, and happen during short term. Early detection just after the event is required. Risks for man-made hazards, which should be monitored, include small fire, gas leaks, intrusions, an explosive substance, etc. Temperature, smoke, acoustics, acceleration, and olfaction are sensing items that are required for detection of fires, gas leaks, etc. Camera, acceleration, acoustics, and light are effective for surveillance for avoiding crimes and terrorist attacks. According to the risk monitoring results, appropriate risk control measures such as fire alarms and fire fighting, rescue, and security measures can be applied.

3 SENSING TECHNOLOGIES FOR RISK MONITORING

The general purpose of SHM includes hazard mitigation, improvement of safety and reliability of the structural system, sustainability, and life-cycle cost reduction. The SHM technology consists of sensing, signal processing, health evaluation, and system integration. In recent years, a number of conferences have been held in which SHM techniques for buildings and civil engineering structures has been presented [3–12]. Some of this work has focused on wireless sensing technology. Researchers at Stanford University have developed a wireless sensing unit for real-time structural response measurements and conducted a series of validation tests [13, 14].

"Ubiquitous computing/networking/sensing" is expected to be realized over the next 10 years. The interest in sensing technology for various uses has been growing, and new kinds of sensors have been developed by microelectromechanical systems (MEMS) technology (*see* **Microelectromechanical Systems (MEMS)**). Environmental information, such as brightness, temperature, sound, vibration, and a picture of a certain place in a building, is evaluated by the network to which a huge number of microcomputer chips with sensors are connected [15]. Figure 2 shows the flow toward a ubiquitous sensing/computing/networked society. A wireless sensor network plays an important role in such strategies and can be connected to the internet so that this information can be used to monitor future risks. Wireless sensors are easy to install, remove, and replace at any location, and are expected to become increasingly smaller (i.e., "smart dust" [16]) by using MEMS technology. They provide a ubiquitous, networked sensing environment in building and civil engineering structures. For example, the acceleration and strain at numerous locations on each structural member, temperature and light in each space, as well as images and sounds in desired regions can be obtained by the "smart dust" sensors, as illustrated in Figure 3.

The requirements of a dense array of smart sensors using wireless technology for SHM have been investigated by using the Mote platform [1, 17–21]. It is an open hardware and software platform for smart sensing and supports large-scale, self-configuring sensor networks as shown in Figure 4 (*see* **Wireless Sensor Network Platforms**; **Sensor Network Paradigms**). Ruiz-Sandoval developed an

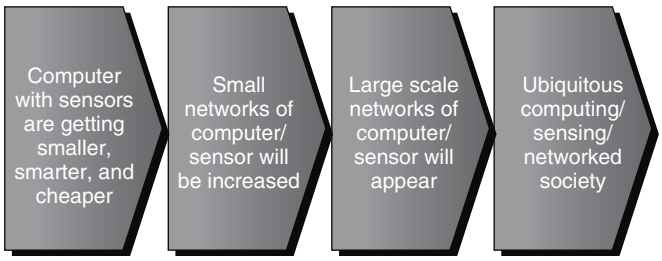


Figure 2. Toward a ubiquitous computing/sensing/networked society.

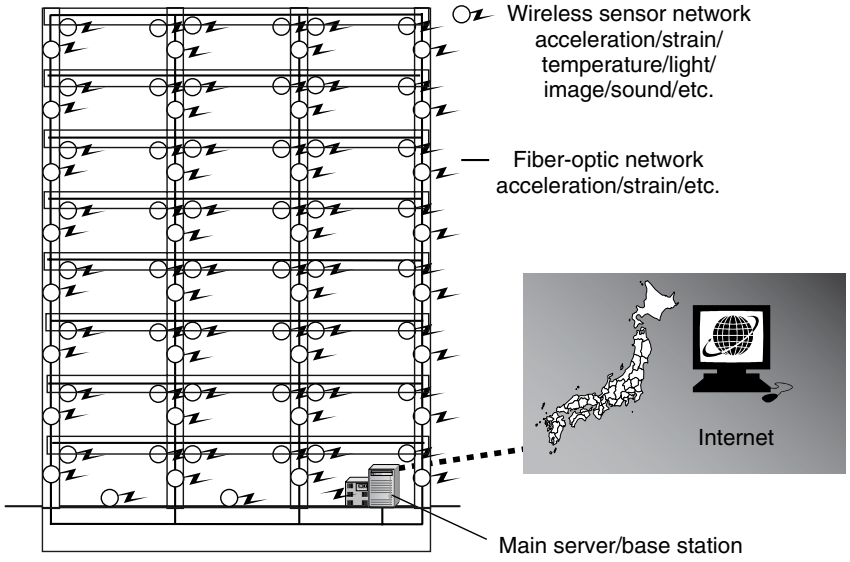


Figure 3. Example of risk monitoring system.

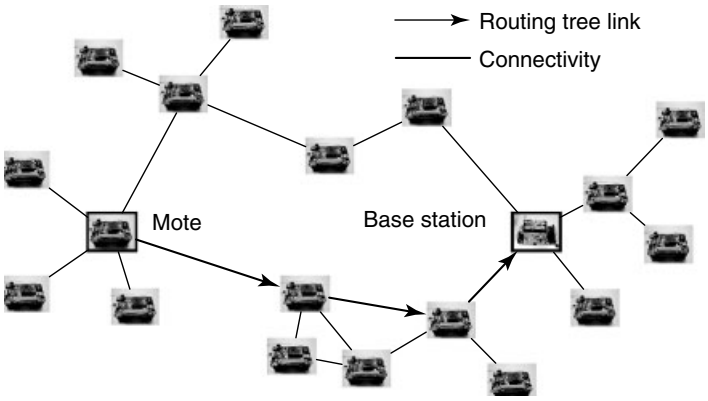


Figure 4. *Ad hoc* and multihop sensing.

agent-based framework, a hardware- or software-based computer system, which enjoys the properties of autonomy, social ability, reactivity, and proactiveness for SHM [18]. A distributed computing strategy for SHM was proposed, which is suitable for implementation on a network of densely distributed smart sensors [19]. A scalable and autonomous SHM system using smart sensors was developed and the damage detection capability and autonomous operation of the developed system were experimentally verified [20]. Tenet architecture that simplifies application development for tiered sensor networks without significantly sacrificing performance was proposed [21]. Its collection of tasklets support data acquisition, processing, monitoring, and measurement functionality. A wireless sensor system using the Mote platform with a developed sensor board was tested on the 4200-ft-long main span and the south tower of the Golden Gate Bridge [22]. Reference 23 provides a summary review of the collective experience the structural engineering community has gained from the use of wireless sensors and sensor networks for monitoring structural performance and health.

4 APPLICATIONS

4.1 Ubiquitous structural monitoring system using wireless sensor networks

One of the main purposes of SHM is for damage detection of the structure [24] (*see* **Signal Processing for Damage Detection**; **Damage Detection Using Piezoceramic and Magnetostrictive Sensors and Actuators**; **Damage Measures**). Since structural damage is a local phenomenon, a ubiquitous structural monitoring (USM), i.e., high-density distributed structural monitoring, is needed. From this point of view, a laboratory experiment has been devised to assess the performance of the "smart dust" Mote, the platform of wireless sensor networks (*see* **Wireless Sensor Network Platforms**), for SHM. Through shaking table experiments, it was recognized that the Mote was useful for the damage detection of the structure [1]. To increase the performance of structural monitoring, a high-sensitivity acceleration sensor board for the Mote has been developed and tested [17, 18]. On the basis of these activities, research on wireless sensor network architecture for USM in the next generation, which includes a design of hardware and development of communication software, has been conducted [6, 7]. Figure 5 shows an example of the concept of the USM system. Wireless sensor nodes are distributed in the room to measure the acceleration during the earthquake. By using the double integration scheme, displacement of each sensing point is evaluated. After the earthquake, information of risk monitoring by space deformation is obtained.

The authors have developed a sensor module for USM consisting of a sensor board and wireless network module as shown in Figure 6. There are several test beds for wireless network module

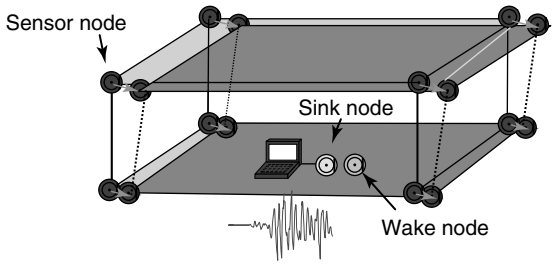


Figure 5. Example of ubiquitous structural monitoring system.

[16, 20, 25, 26]. The MEMS acceleration sensor was examined for the development of the acceleration sensor board that can be connected with such a wireless network module. Required specifications for the acceleration sensor board for the USM are shown in Table 2. It shows the performance necessary for general seismic observation and vibration measurement of buildings from a moderate earthquake to a large earthquake. Performance of various MEMS sensors is usually confirmed only in the frequency range of 10 Hz or more, though the performance in the low-frequency range from 0.1 Hz is important for the vibration measurement of high-rise building and the observation of long-period ground motion.

The developed sensor board consists of MEMS acceleration sensor, which was selected by the benchmark tests, low-pass filter, and 16-bit A/D converter (Figure 7). The sleep function of the MEMS acceleration sensor and the low-pass filter can be used to achieve low power consumption. The module for the

Table 2. Required specifications

Frequency	0.1 – 20 Hz
Measuring range	±2000 Gal
Measuring resolution	0.1 Gal
A/D resolution	16 bit

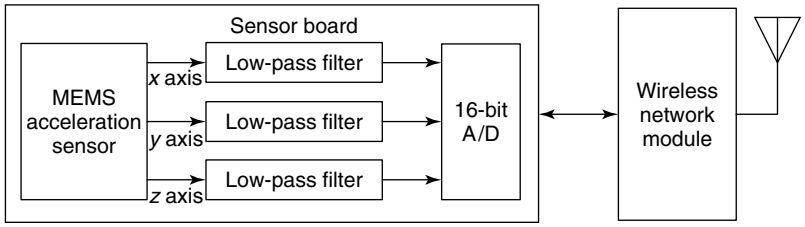


Figure 6. Sensor module for USM.

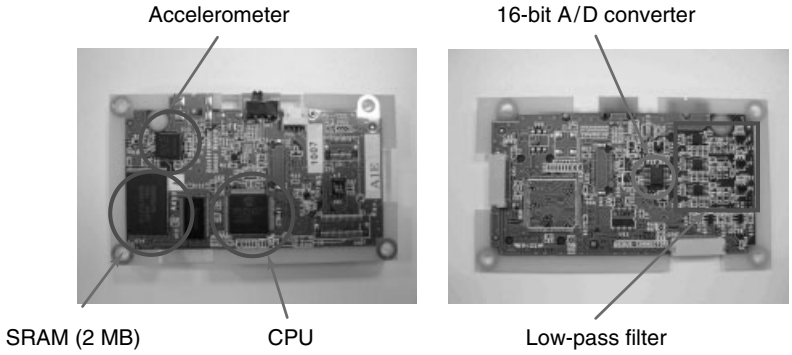


Figure 7. Developed sensor board for the USM.

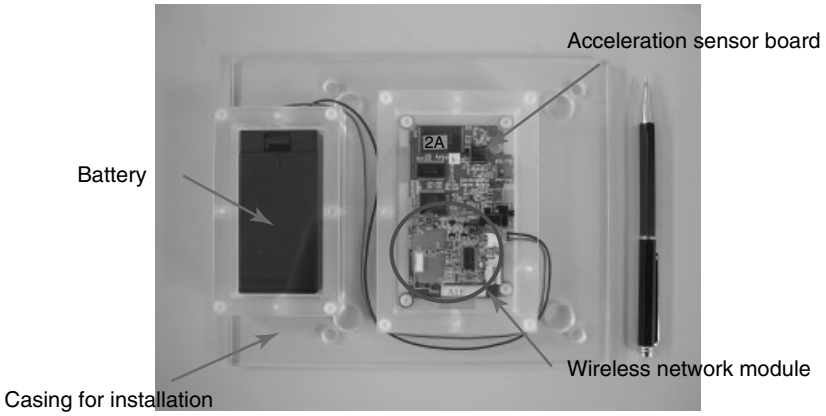


Figure 8. Developed sensor module for the USM.

wireless sensor network was connected to the developed sensor board for the test as shown in Figure 6. The developed sensor module for the USM and USM system is shown in Figures 8 and 9, respectively. A shaking table test was carried out to recognize the performance of the developed sensor module for the USM. Eight sensor nodes, a wake node, a sink node connected to the PC, and the reference sensor were fixed on the shaking table as shown in Figure 10. The shaking table is 1.8 m $\times$ 1.5 m and can shake at 0.1–100 Hz using external input waves. The input wave was the Japan Meteorological Agency (JMA) Kobe North-South direction (NS), which was observed at the Great Hanshin–Awaji disaster that occurred in 1995. Figure 11 shows the measurement acceleration time histories of the eight sensor nodes and the reference sensor. The eight sensor nodes were synchronized within 1 ms by a protocol proposed by the authors [27]. According to the packet from the wake node, sensor nodes started to measure the acceleration. After the measurement, stored acceleration data in the memory on the sensor board was transmitted to the PC through the sink node by wireless communication. The measurement result by the eight sensor nodes corresponds well to the result by the reference sensor. It was confirmed that the developed sensor module had enough basic performance for the USM.

4.2 Risk information delivery system using wireless sensor networks

A safe evacuation is expected as a countermeasure for a large earthquake [28]. There is the first disaster such as structural damage of a building and the secondary disaster such as fire. It is required

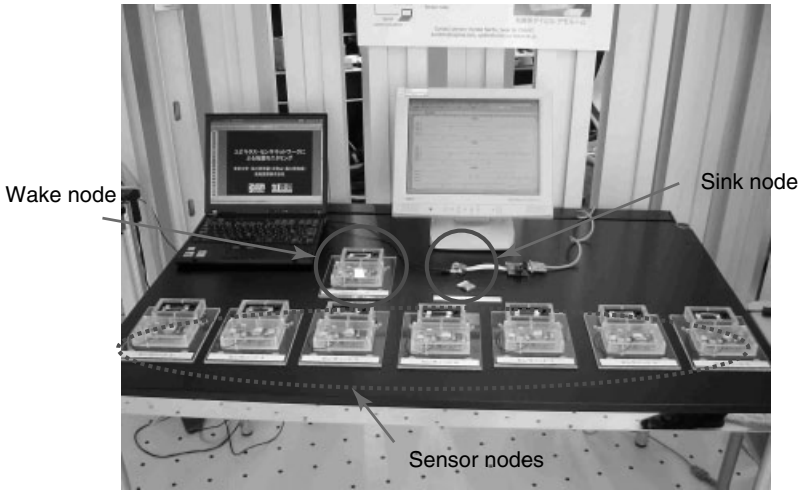


Figure 9. Ubiquitous structural monitoring system.

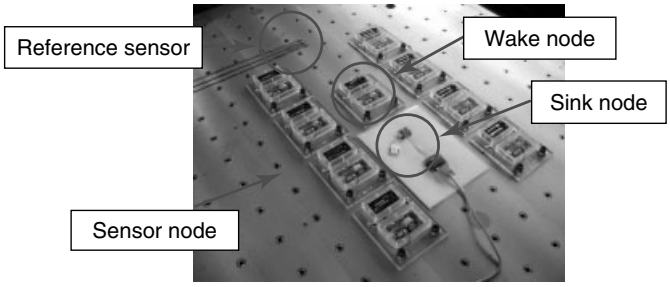


Figure 10. Developed USM system and reference sensor on the shaking table.

to offer users adequate information more immediately after earthquakes occur. From this point of view, a risk information delivery (RID) system using wireless sensor networks was proposed [28]. RID detects dangerous areas by using a sensor network constructed in the structure and judges the degree of risk by calculating the risk of dangerous areas.

The RID system is based on a client–server model and composed of RID_Client, RID_Server, stationary node (SN), and emergency node (EN) as shown in Figure 12. An accelerometer and a temperature sensor are used. The accelerometer detects earthquakes and the temperature sensor detects fire.

An SN with an accelerometer always observes the state. When SNs detect an acceleration more than constancy, SNs judge the occurrence of earthquakes and send wake-up messages to the EN. In addition, SNs aggregate risk information that ENs calculated and also calculate risk of an area where SNs belong. An EN is equipped with sensors that detect various dangers. ENs start by receiving wake-up messages from SNs through the occurrence of earthquakes. Starting ENs send SNs risk information calculated from sensors data. RID_Servers are equipped with SNs and save risk information on an area that SNs calculated in a database. Moreover, a risk map is constructed by adding risk information to map information. This map is offered to the RID_Client. The RID_Client assumes personal digital assistant (PDAs) and mobile phones are carried by people who then view risk maps (viewer) that the RID_Server constructed. Figure 13 shows the example of node arrangement.

The degree of risk is decided by making some thresholds in the value of sensors. The maximum value of degree of risk is 100, and the value

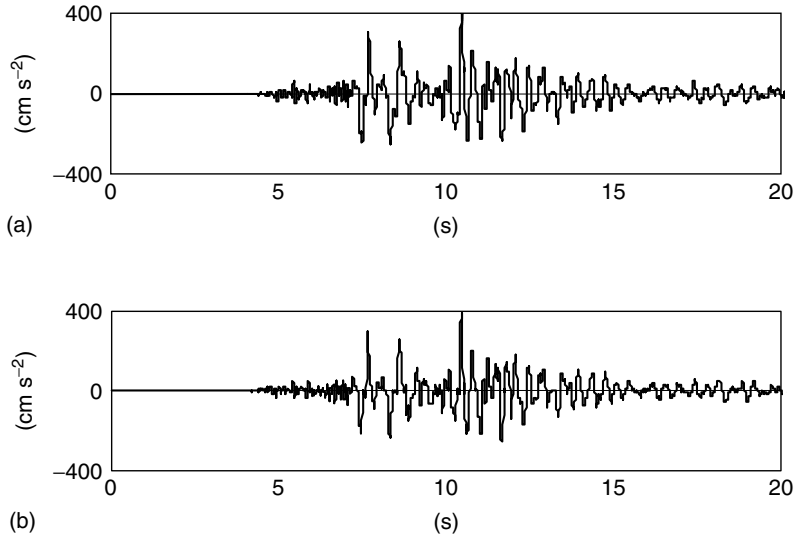


Figure 11. Measured and transmitted time histories. (a) Eight sensor modules and (b) Reference sensor.

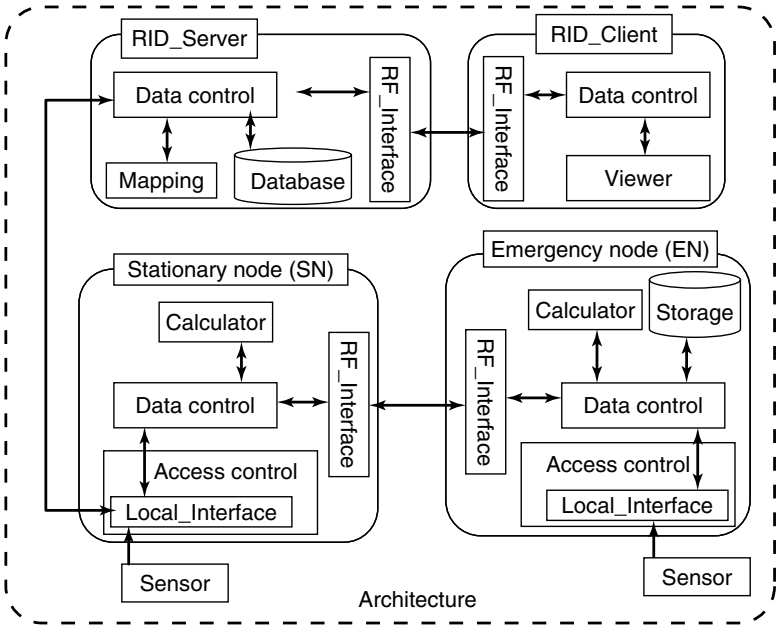


Figure 12. Risk information delivery (RID) system architecture.

changes based on the threshold. Tables 3 and 4 show examples of the degree of risk for temperature and maximum acceleration, respectively. A judgment of the degree of risk is defined by three stages. Users can stand ready to evacuate at the first phase from 0 to less than 20, need to evacuate at the second stage from 20 to less than 60, and must not act and keep themselves in safe at the third stage of 60 or more. An SN calculates the degree of risk of areas. First, an SN aggregates risk information on all the sensors of

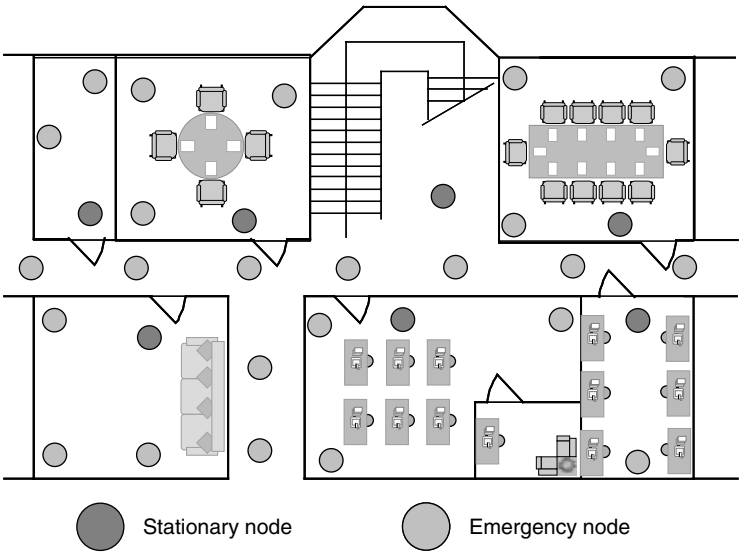


Figure 13. Example of ubiquitous structural monitoring system.

ENs. Next, the SN averages the degree of risk information on each sensor and reflects the degree of risk of the sensor with the highest risk.

Table 3. Relation between temperature and degree of risk

Temperature	Degree of risk
Below standard temperature	0
From 20 to less than 30 °C against standard temperature	50
From 30 to less than 50 °C against standard temperature	75
50 °C or more from standard temperature	100

Table 4. Relation among maximum seismic intensity and degree of risk

Maximum acceleration (Gal)	Seismic intensity	Degree of risk
0–40	∼3	0
40–110	4	10
110–240	5 lower	20
240–520	5 upper	40
520–830	6 lower	60
830–1500	6 upper	80
1500 ∼	7 ∼	100

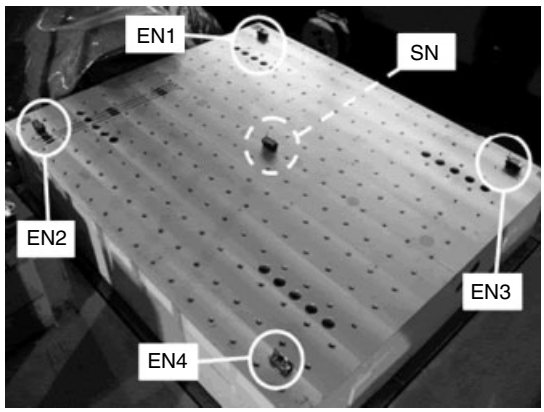


Figure 14. Arrangement of ENs and SNs.

A shaking table test was carried out to recognize the performance of the developed RID system. Both SN and EN were installed on a shaking table as shown in Figure 14. The SNs and ENs are Mica Z Motes. The sensor nodes utilize an accelerometer and a temperature sensor, respectively, to measure the degree of risk. The sampling frequency of these sensors is 100 Hz. The ENs are quaked on the shaking table and heated by a drier. All of the data, i.e., temperature, acceleration, and risk degrees of temperature and acceleration, which ENs sensed during 40 s are collected. The maximal value of accelerations is

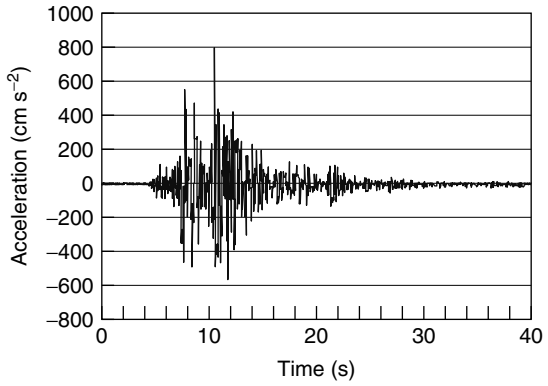


Figure 15. Result of reference sensor.

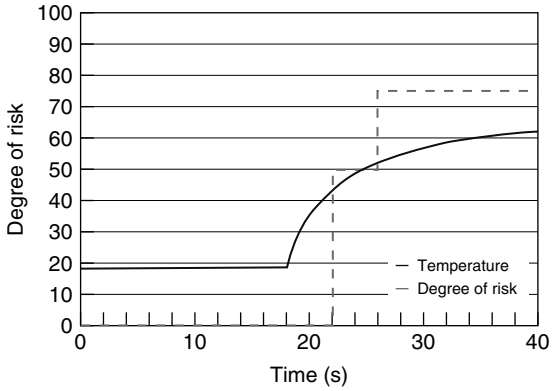


Figure 17. Result of temperature sensor of ENs.

averaged per 10 samples and assigns the risk corresponding to Table 4 using the average. On the other hand, sensed temperature data are averaged per 100 samples every 1 s. Then the RID system subtracts averaged temperature from the standard temperature that was measured when the system started and assigns the risk corresponding to Table 3 using the averaged temperature. Figure 15 shows measured data of the reference acceleration sensor for comparison. Figures 16 and 17 show measured data of acceleration and temperature on ENs, respectively. As a result of this experiment, it was recognized that the RID system could measure each degree of risk accurately. The RID system detects a dangerous part in the structure and offers risk information after an earthquake. Users can get risk information on a detailed area such as rooms and the passages by the RID.

5 CONCLUSIONS

This article provides a concept for the risk monitoring of buildings for disaster mitigation and reviewed related sensing technology. An USM system and an RID system using wireless sensor networks were introduced as applications. Risk monitoring is expected as one of the most promising and emerging

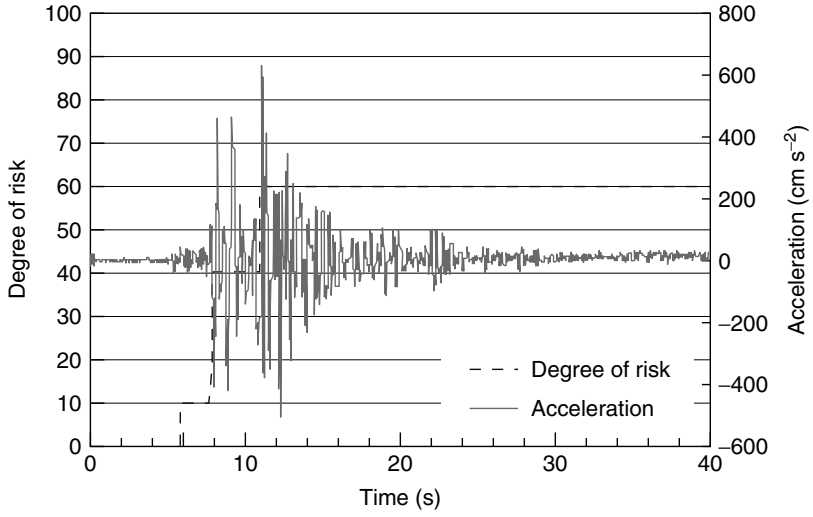


Figure 16. Result of acceleration of ENs.

technologies for the mitigation of natural and man-made hazards.

REFERENCES

[1] Kurata N, Spencer BF Jr, Ruiz-Sandoval M. Risk monitoring of buildings with wireless sensor networks. *Structural Control and Health Monitoring* 2005 **12**:315–327. Special Issue: Advanced Sensors and Health Monitoring.

[2] Drennan LT, McConnell A. *Risk and crisis management in the public sector*. Routledge: London and New York, 2007, 1–30.

[3] Casciati F, *et al. Proceedings of the US—Europe Workshop on Sensors and Smart Structures Technology*. Como and Somma Lombardo: Italy, 2002.

[4] Wu Z, *et al. Proceedings of the First International Conference on Structural Health Monitoring and Intelligent Infrastructure*, Tokyo, Japan, November 2003.

[5] Chang FK. *Proceedings of the 4th International Workshop on Structural Health Monitoring*, Stanford, CA, September 2003.

[6] Tachibana E, *et al. Proceedings of the International Symposium on Network and Center-Based Research For Smart Structures Technologies and Earthquake Engineering*, Osaka, Japan, July 2004.

[7] Spencer BF. Jr *et al. Structural Control and Health Monitoring*, 2005 **12**:225–479. Special Issue: Advanced Sensors and Health Monitoring.

[8] Mufti A, *et al. Proceedings of the 2nd International Conference on Structural Health Monitoring of Intelligent Infrastructure*, Shenzhen, China, November 2005.

[9] Chang FK. *Proceedings of the 5th International Workshop on Structural Health Monitoring*, Stanford, CA, September 2005.

[10] Skelton R, *et al. Proceedings of the Fourth World Conference on Structural Control and Monitoring*, San Diego, CA, July 2006.

[11] Mufti A, *et al. Proceedings of the 3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure*, Vancouver, British Columbia, Canada, November 2007.

[12] Chang FK. *Proceedings of the 6th International Workshop on Structural Health Monitoring*, Stanford, CA, September 2007.

[13] Lynch JP, Kiremidjian AS, Law KH, Kenny T, Carryer E. *Proceedings of the Third World Conference on Structural Control*, 2002, Vol. 2, 667–672, Issues in Wireless Structural Damage Monitoring Technologies.

[14] Lynch JP. *Decentralization of Wireless Monitoring and Control Technologies for Smart Civil Structures*. Ph.D. Dissertation, Department of Civil and Environmental Engineering. Stanford University, 2002.

[15] Morikawa H. *Ubiquitous Sensor Networks, Proceedings of US-Japan Workshop on Sensors*, Tokyo, 2005 Smart Structures and Mechatronic Systems.

[16] Pister KSJ, Kahn JM, Boser BE. *Smart Dust: Wireless Networks of Millimeter-Scale Sensor Nodes*. 1999, Highlight Article in 1999 Electronics Research Laboratory Research Summary.

[17] Spencer BF Jr, Ruiz-Sandoval M, Kurata N. Smart sensing technology: opportunities and challenges". *Structural Control and Health Monitoring* 2004 **11**:349–368.

[18] Ruiz-Sandoval, M. *Smart Sensors for Civil Infrastructure Systems*, Ph.D Dissertation. University of Notre Dame, 2004.

[19] Gao Y. *Structural Health Monitoring Strategies for Smart Sensor Networks*, Ph.D. Dissertation. University of Illinois at Urbana-Champaign, 2005.

[20] Nagayama T. *Structural Health Monitoring Using Smart Sensors*, Ph.D. Dissertation. University of Illinois at Urbana-Champaign, 2007.

[21] Gnawali O, Greenstein B, Jang KY, Joki A, Paek J, Vieira M, Estrin D, Govindan R, Kohler E. The Tenet Architecture for Tiered Sensor Networks. *Proceedings of the 4th ACM Conference on Embedded Networked Sensor Systems (Sensys '06)*, ACM Press, 2006.

[22] Kim S, Pakzad S, Culler D, Demmel J, Fenves G, Glaser S, Turon M. Health Monitoring of Civil Infrastructures Using Wireless Sensor Networks. *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*. ACM Press: Cambridge, MA, 2007; pp. 254–263.

[23] Lynch JP, Loh K. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest* 2006 **38**(2):91–128.

[24] Kurata N, Saruwatari S, Morikawa H. Ubiquitous Structural Monitoring Using Wireless Sensor Networks. *Proceedings of 2006 International Symposium on Intelligent Signal Processing and Communication Systems*, WAM1-5-4, Tottori, Japan, December 2006.

[25] Saruwatari S, Kashima T, Minami M, Morikawa H, Aoyama T. PAVENET: A hardware and software

framework for wireless sensor networks. *Transactions of SICE* 2005 **E-S-1**(1):74–84.

[26] Kling R, Adler R, Huang J, Hummel V, Nachman L. Intel Mote-based sensor networks. *Structural Control and Health Monitoring* 2005 **12**:469–479.

[27] Suzuki M, Saruwatari S, Kurata N, Morikawa H. A high-density earthquake monitoring system using wireless sensor networks. In *Proceedings of the 5th ACM Conference on Embedded Networked Sensor Systems (Sensys '07)*, Sydney, Australia, ACM Press, November 2007.

[28] Sasaki K, Ishii N, Kurata N, Tobe Y. An assistance system for safe evacuation using wireless sensor networks. International Workshop on SensorWebs, Databases and Mining in Networked Sensing Systems (SWDMNSS 2007), Braunschweig, Germany, June 2007.

Chapter 94

Value Assessment Approaches for Structural Life Management

Enrique A. Medina[1] and John C. Aldrin[2]

[1] *Radiance Technologies Inc., Dayton, OH, USA*
[2] *Computational Tools, Gurnee, IL, USA*

1 INTRODUCTION

Managing the life-cycle costs (LCCs) of an aircraft structure has become a critical challenge for engineers and managers of aircraft fleets. The importance of controlling sustainment costs is epitomized by the reality that many old aircraft in the US Air Force fleet are expected to remain in service for another 25 years. Even today, many of these aircraft exhibit aging problems such as fatigue cracking, stress corrosion cracking, corrosion, and wear [1]. For example, the annual costs of maintenance due to corrosion alone are already many hundreds of millions of dollars, and these costs are steadily increasing [2]. To address these challenges, condition-based maintenance (CBM) approaches are being implemented to perform inspections and repairs only when necessary. In the case of aircraft structures, CBM employs nondestructive evaluation (NDE) through external tests and structural health monitoring (SHM) using embedded sensors as part of maintenance plans that are based on system condition diagnostics and prognostics.

NDE consists of test methods used to examine the material integrity of an object or system of objects without impairing its future usefulness. It is utilized for detecting and characterizing discontinuities such as cracks, porosity, and corrosion in materials and structures to ensure their reliability and extend their service life. The reliability of NDE techniques is critical for aircraft maintenance programs. Issues in NDE discovered through probability of detection (POD) studies have generated an interest in determining the impact of inspection performance on total service life [3]. For inspection problems that include manual

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

scanning, complex procedures, and low frequencies of finding critical flaws, there is a potential for some critical sites to not be inspected effectively owing to the requirements of the inspection process, or there are inconsistent requirements for identifying marginal defects [4]. Although methods have been proposed to help address NDE reliability issues, the application of SHM systems with *in situ* sensors also has the potential to improve health assessment reliability by eliminating problematic human factors.

The application of *in situ* sensors for SHM has been extensively explored in recent years [5–9]. Two classes of SHM systems are considered here. The first approach is based on the acquisition of data for life prediction models. During in-service periods, distributed sensors can be used to measure the loading and environmental conditions experienced by the structure (*see* **Loads Monitoring in Aerospace Structures**; **Environmental Monitoring of Aircraft**; **Agile Military Aircraft**; **Loads and Temperature Effects on a Bridge**). Subsequently, these data can be used to improve fracture mechanics models to better predict the flaw state. This approach can be considered "indirect", since the damage state in the future is estimated using a model prediction based on input measurement data. The most significant benefit of "indirect" SHM schemes is a reduction in the uncertainty of the fracture mechanics model predictions. The second approach is based on the acquisition of NDE data using *in situ* sensors to quantify the damage state of a structure (*see* **Signal Processing for Damage Detection**; **Model-based Statistical Signal Processing for Change and Damage Detection**; **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**; **Damage Presence/Growth Monitoring Sensors**). This approach is classified as a "direct" method, requiring that the damage state be observable in distributed sensor data. There are two subclasses of SHM damage characterization systems: global health monitoring systems incorporate distributed sensors such as strain gauges and acoustic emission transducers to detect the presence and general location of damage in a structure, while local health monitoring systems detect damage in critical structural locations by using ultrasonic and eddy-current sensors.

In practice, there is a continuum of potential scenarios using NDE and *in situ* sensors to ensure structure reliability and mitigate maintenance costs. These scenarios include hybrid NDE and SHM strategies, such that SHM can be used to identify the presence of a critical damage state, while NDE can be used for final verification before an expensive teardown or repair is performed. Examples of the array of maintenance approaches for management of critical damage locations include the following:

1. fail-safe design (using life prediction based on model only);
2. scheduled NDE maintenance (NDE);
3. load monitoring with life prediction model refinement (indirect SHM);
4. damage state monitoring with *in situ* sensors (direct SHM);
5. load monitoring with CBM (indirect SHM and NDE);
6. damage state monitoring with follow-up NDE inspection (direct SHM and NDE).

This hybrid approach for fleet management is proposed through the careful selection and pairing of reliable SHM and NDE systems for a given critical location. The theme of this article is a value assessment methodology incorporating cost–benefit analysis (CBA) and probabilistic risk assessment into NDE and/or SHM system design to maintain acceptable structural health assessment reliability and minimize total service life cost.

2 VALUE ASSESSMENT PROBLEM FOR SHM

At the center of the fleet management problem exists a compromise between aircraft availability, reliability, and cost: (i) maximizing availability and total life of an aircraft may require higher costs for maintenance systems (such as SHM) or may adversely impact reliability; (ii) minimizing cost of a maintenance program for a critical location may require limiting availability (such as reducing overall life) or sacrificing reliability; and (iii) improving reliability may reduce availability owing to longer depot

maintenance periods, or may require more expensive SHM systems. An approach is needed to evaluate and establish the optimal trade-offs between these factors for several maintenance approaches including SHM. CBA provides a framework for this evaluation.

2.1 Overview of SHM benefits and costs

The primary benefit of SHM concerns integration with prognostics, where the management of high-value assets such as military aircraft is improved through the quantitative prediction of future operating capability and accurate determination of remaining life. Potential cost benefits for SHM include (i) reduction in labor cost and time for unnecessary NDE inspections, (ii) management of locations of limited accessibility to minimize costly teardowns, and (iii) availability of robust indications of impending failure of the structure to trigger safe retirement. Improvements in availability of aircraft can also be addressed using SHM, by limiting time in maintenance to only when absolutely necessary. In addition, when considered during the aircraft design phase, SHM has the potential to provide engineers with the means to reduce structure weight by avoiding conservative designs, reduce the need for costly assessments of fatigue-critical locations [5], and improve aircraft dynamic performance. Additional benefits may also be realized through the use of *in situ* sensor data to indirectly detect unsafe conditions of interest such as excessive loading or icing conditions, and to support accident investigation, potentially leading to a safer fleet over the long term.

Recent work has also explored the significant costs and challenges of SHM implementation [10, 11]. The costs associated with SHM systems can be categorized as development costs, implementation costs, in-service costs, and end-of-life costs. Development costs include any initial research and system development work for a particular application. Implementation costs are associated with the fixed initial cost for purchasing and installing the on-board SHM system and for performing validation studies to satisfy reliability requirements. Both development and implementation costs are expected to be much higher for SHM systems with respect to those of NDE techniques, given the increasingly difficult system requirements concerning inspection and reliability. In-service costs can include the additional cost of fuel due to added SHM system weight, data interpretation labor costs, SHM maintenance costs, and the cost of secondary inspection and unnecessary repair due to false calls or unnecessary calls when flaws are very small. While in-service costs of SHM systems are expected to be low in relation to those of NDE procedures, design-time consideration must be given for the possibility of such costs being significant to minimize their impact on total LCC.

Given the potential benefits and costs of SHM, a key issue for the success of any SHM implementation is the ability to establish how the SHM technology affects figures of merit such as cost, safety, and availability of aerospace structures. Without reliable value assessment studies that take into account all impacts of the technology on these figures of merit, it will likely be impossible to justify the initial investment in SHM technology to financial decision makers and, more importantly, it will be impossible to predict whether the SHM implementation will improve or degrade the ability to efficiently manage the aerospace structure.

2.2 Basic principles of value assessment methods

In this section, we review several methods for value assessment. We describe the cost–benefit analysis, cost–effectiveness analysis (CEA), and cost–utility analysis (CUA) methods, which have been widely used for economic analysis in many fields. We also describe multiobjective optimization methods, which are not yet as widely used, but can be quite useful when the diversity of the multiple objectives makes it difficult to aggregate objectives for comparing alternatives.

CBA is based on representing in monetary units all consequences of the proposed technology as either costs or benefits. The net benefit of a proposed solution, which equals the difference between its benefits and costs, is used as the sole basis for comparing alternate solutions. CBA or other types of value analyses can be costly themselves, and enough care and diligence must be exercised to obtain the intended insight for decision making. The major steps in CBA, as adapted for the SHM case from those listed in [12], are as follows:

1. Specify the set of alternative SHM technologies, including the complementary NDE processes if existent and not fixed, and the alternate parameters of the combined SHM-NDE (structural health monitoring—nondestructive evaluation) system. This can also include the alternative consisting of not using SHM.
2. Decide whose benefits and costs count. For example, the use of SHM may eventually reduce the need for inspection labor. This would be detrimental to an inspection business, and the loss of revenue and expertise could be considered a cost.
3. Catalog the impacts and select measurement units. This would include, for example, the impact on reliability, safety, development costs, and labor over the system's life.
4. Predict the impacts (benefits and costs) quantitatively over the life of the system.
5. Assign monetary values to all impacts (benefits and costs).
6. Apply interest and/or discount rates to obtain the present value of each impact.
7. Compute the net present value (NPV) of each alternative SHM-NDE combination. This should be equivalent to the LCC of each SHM-NDE alternative [13].
8. Perform sensitivity analyses. These determine how changes in the estimated impacts, the interest rate, or other parameters can affect the NPV of the various alternatives.
9. Make a recommendation based on the NPV and sensitivity analyses. Normally, the recommended alternative is that with the largest NPV.

It is sometimes impossible or inappropriate to assign monetary value to the most important benefit expected from a technology. In CEA, this nonmonetary benefit is called the *effectiveness*. The ratio of the effectiveness and the cost for a given alternative is referred to as the *cost–effectiveness criterion*, and it is used instead of the monetary NPV as a basis for comparing alternatives. For example, the number of lives saved can be considered the effectiveness of a particular SHM implementation. In CUA, alternatives are compared on the basis of the ratio of their utility and their cost, where the utility is a number that combines at least two benefits. The difficulty in appropriately combining benefits makes CUA less appealing for SHM technology value analysis.

The value analysis problem can also be posed as a multiobjective optimization problem are constraints functions [14],

$$\begin{aligned} &\text{Maximize } f(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_p(\mathbf{x})] \\ &\text{Subject to } g_j(\mathbf{x}) \leq 0 \quad j = 1, 2, \ldots, m \end{aligned} \tag{1}$$

where $\mathbf{x}$ is a vector of independent variables, the f_i are objective functions, and the g_j are constraint functions. Usually, the objective functions are competing and cannot be simultaneously optimized. A candidate solution $\mathbf{x}$ in the decision space is called *feasible* if it satisfies all the constraints. Consider two feasible solutions $\mathbf{x}_a$ and $\mathbf{x}_b$. If $f_i(\mathbf{x}_a) \geq f_i(\mathbf{x}_b)$ for all i and there exists an i for which $f_i(\mathbf{x}_a) > f_i(\mathbf{x}_b)$, then the solution $\mathbf{x}_b$ is said to be dominated by the solution $\mathbf{x}_a$, and $\mathbf{x}_b$ can be eliminated from further consideration. Eliminating all dominated solutions results in a set of *nondominated, Pareto efficient*, or *efficient* solutions from which the decision maker must choose [14]. For a solution that is efficient, it is not possible to improve any objective without making one or more of the other objectives worse.

There exist methods that can process multiple alternate solutions to yield nondominated solutions, and others that can even help the decision maker select among a set of nondominated solutions, while preserving each objective's variable type without the need to transform all impacts to financial values. These methods are generally known as *multicriteria optimization*, and can enable effective trade-off analysis when the number and types of impacts make it difficult to aggregate them. Good reviews of existing evolutionary and nonevolutionary methods for solution of multicriteria optimization problems can be found in [14] and [15], respectively. The software used in the examples of Section 4 utilizes multicriteria optimization methods for trade-off analysis.

2.3 Prior applications

Related to the value assessment problem for SHM, prior work has discussed the need to evaluate the costs and benefits of NDE technology. Hagemaier describes the uses of NDE in the *cradle-to-grave* product cycle at a large aircraft manufacturing company [16, 17]. That description starts to address how NDE capabilities, requirements, methods, and

parameters affect decisions across stages of the life cycle, but does not address the methods by which this decision making process can be systematically improved by the use of a systems and optimization approach and corresponding numerical tools. Given this need, specific methods have been developed to estimate the economic value of NDE techniques. In the area of technology value assessment, a technology benefit estimator (T/BEST) was developed by NASA (National Aeronautics and Space Administration) to quantitatively assess the value of advanced aerospace technologies and quantify the benefits for prioritization [18]. Brechling developed a methodology for the economic assessment of NDE research and development in terms of NPV, including direct and indirect benefits to society [19]. Papadakis outlined three practical financial methods for making decisions in NDE: the Deming inspection criterion (DIC), the internal rate of return (IRR), and the productivity and profit methodology [20, 21]. Wall and Wedgwood also developed a quantitative inspection value method (IVM) for application to NDE focusing on the major costs and benefits of a particular technique by evaluating the impact of sensitivity, speed, and reliability [22].

A few studies have been presented to date concerning the cost justification for SHM applications [6–9]. Kent and Murphy performed a technology assessment of health monitoring system technology including a CBA [23]. In prior work, Aldrin *et al.* have also explored the problem of cost–benefit assessment of SHM incorporating probabilistic models [10]. Kapoor, Goh, and Boller perform SHM cost estimation as part of their proposed procedures for assessment of SHM potentials [24]. Boller and Staszewski provide an example of how to estimate LCCs for a new damage monitoring technique in [[6], pp. 36–43] and the references therein. The integration of cost models with probabilistic risk assessment tools incorporating NDE and SHM is presented in the next section.

3 COMPONENTS OF STRUCTURAL LIFE MANAGEMENT

Experience has demonstrated the need to develop lifelong plans for evaluating and maintaining a structure, to avoid excessive future costs for maintenance and repairs. The concept of a systems-level approach is necessary to develop the most economical sustainment of future weapon systems throughout their lifetime. The systems-level approach implies that tasks are accomplished in teams that bring together all interested and affected personnel. For instance, a new repair to a specific location on an aircraft must take into account larger structural considerations.

A strategy for component life-cycle optimization incorporating NDE and SHM is presented first in terms of the necessary components. Refined system components are models, characteristics, and functions that are necessary for constructing the component life model. To achieve optimal component life management, the following refined system components must be well understood and integrated within a complete system approach: (i) identification of critical locations to inspect, failure analysis, and loss assessment; (ii) initial quality and flaw initiation models (including material, geometry, and process models); (iii) flaw growth models; (iv) load and environmental monitoring, which evaluates the conditions under which the component will operate; (v) maintenance scheduling and repair tracking; (vi) value analysis methodology; and (vii) inspection technique capability, which is normally measured by POD of a certain flaw type and size. Figure 1 presents a diagram of a component life management approach incorporating NDE and SHM. The refined system components have been broken up into two categories: those that specifically address modeling the inspection technique and those that indirectly impact the inspection technique design, defined as *economic service life management resources*.

3.1 Models and software tools for system life management

Several software packages incorporating modeling tools have been proposed over the years to address various aspects of the system management problem. In [9], Kacprzynksi, Roemer, and Hess propose an approach for design of health management systems that is based on extending failure modes effects and criticality analysis (FMECA) with system modeling, LCC modeling, CBA, trade-off analysis, and sensitivity analysis. The proposed result is a prognostic

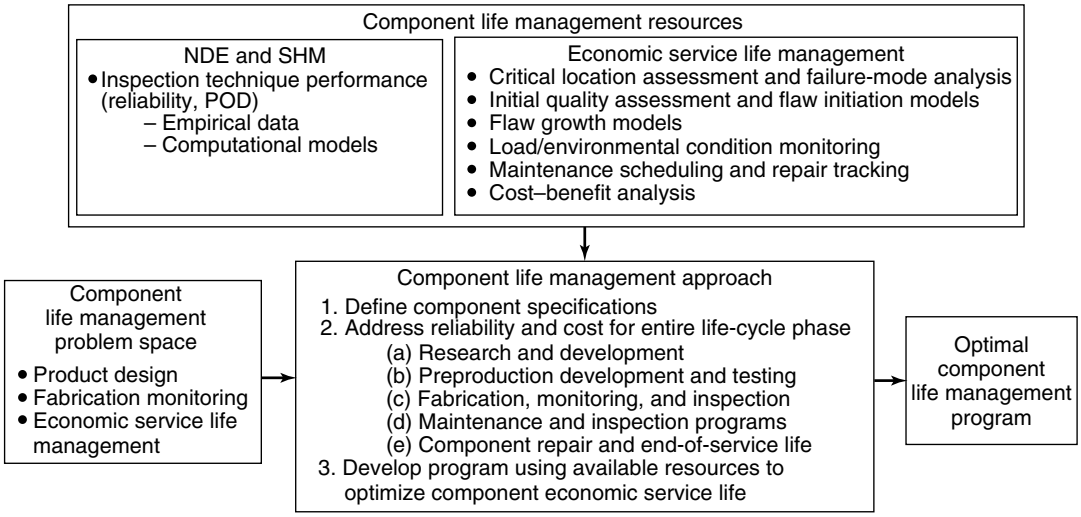


Figure 1. Component life management approach incorporating NDE and SHM.

and health management (PHM) system design tool that can enable analysis of alternative sensors, algorithms, and other health management technology, can use interactive databases, and can operate in a collaborative, web-based fashion. They present an example of an electromechanical valve of an aircraft. Related work by the same authors has focused on aircraft propulsion, industrial power generation, and other systems with rotating components. FMECA and aggregated modeling of component and system reliability are an important part of health management for that type of system because of the number of components involved and the complexity of the functions performed by the subsystem. Two of the most common functions of aerospace structures are (i) carrying of mechanical load and (ii) protection from the environment. In such applications, failure modes are often related to fatigue cracking and corrosion in multilayered metallic structures with stiffeners, and delamination and other types of damage in structures manufactured from composites. Accessing the areas of the structure where critical damage can occur can be difficult, labor intensive, and risky in terms of the possibility of damaging the structure during teardown or reassembly. One common important challenge in health monitoring of structures is the selection, placement, and installation of sensors that can reliably provide the necessary data and the design of algorithms that can process these data to detect damage trends in the presence of environmental disturbances and variations in the operational loads of the structure.

3.2 Probabilistic models for risk assessment incorporating NDE and SHM

Probabilistic models are another key component to successfully represent SHM systems with life prediction models for risk assessment and CBA (*see* **Risk Monitoring of Aircraft Fatigue Damage Evolution at Critical Locations**). This approach builds upon prior work comprising the development of a strategy and software framework for integrating NDE design and product life management tools [10, 25–27]. The model is based on prior work by Berens *et al.*, who developed a software tool (PROF) for probabilistic risk assessment of fatigue crack growth and fracture incorporating NDE [28, 29]. This model primarily addresses direct SHM systems capable of acquiring NDE-type data using embedded sensors to quantify the damage state of a structure. To model an indirect SHM system, the SHM measurement data would simply be used to refine the life prediction model.

Figure 2 depicts a diagram of a generic SHM process. First, there are two time intervals to consider: one associated with each opportunity for decision on maintenance (indexed by i) to be performed in

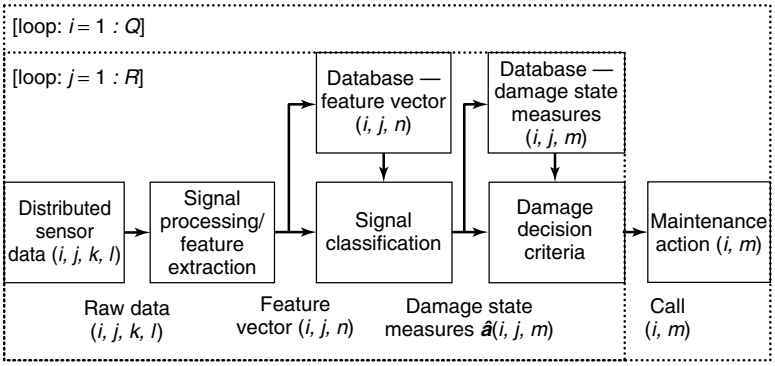


Figure 2. Flow diagram representation of an SHM system model identifying the steps of analysis from sensor data to decision on a maintenance action.

the field, and the other, an in-service SHM data-acquisition time interval (indexed by j) at which an assessment of the damage state can be performed. It is important to distinguish between these two time intervals, since data may be acquired and a damage state estimate may be obtained at a rate different from the rate corresponding to the opportunity for decisions on performing in-field maintenance in the form of secondary inspections and/or repairs. For each data-acquisition time interval (j), data can be acquired from each sensor (indexed by k) in the array for a given number of samples (indexed by l). For example, the number of samples (l) may be large for the case of acoustic emission measurements for impact damage estimation or quite small for humidity sensors for corrosion monitoring. Starting with the raw data, signal-processing and feature-extraction algorithms are applied to filter and extract features as a set of scalar values (indexed by n). Signal classification can subsequently be applied to a database of feature vectors collected over time to estimate the damage state ($\hat{\boldsymbol{a}}$) for each critical location (indexed by m). To assess the damage state and perform maintenance for each opportunity, a damage decision criterion, is applied on the basis of a maximum acceptable critical flaw size (a_{cr}). A database of damage state estimates ($\hat{\boldsymbol{a}}$) from prior decision intervals and data-acquisition periods (i, j) may be used in the decision process. The final step is the decision to perform a maintenance action such as a secondary inspection or repair.

From the perspective of quantifying the reliability of an SHM system, there is an underlying relationship that must be evaluated between the damage state estimate ($\hat{\boldsymbol{a}}$) and the actual damage state ($\boldsymbol{a}$), with special interest placed on the critical flaw size (a_{cr}) that prompts a maintenance action. This POD assessment is no different from the "$\hat{\boldsymbol{a}}$ versus $\boldsymbol{a}$" analysis procedure previously devised for NDE systems [30]. Although a model-based approach including each analysis step for the SHM process shown in Figure 2 would be ideal, it is proposed, as a first approximation, to represent the relationship between the flaw size and the POD, false call rate, and random missed flaw rate directly using a four-parameter POD model [31] given by

$$POD(a,t) = \alpha(t) + (\beta(t) - \alpha(t)) \times \left\{1 + \exp\left[-\frac{\pi}{\sqrt{3}}\left(\frac{\ln a - \ln \mu(t)}{\sigma(t)}\right)\right]\right\}^{-1} \quad (2)$$

where a is the flaw size, t is time, α corresponds to the false call rate, β is defined as one minus a random miss rate, σ controls the steepness of the POD curve, and μ is the flaw size for which the POD is 50%. Use of this four-parameter POD model has been recommended to address the fact that both hits and misses are often made for reasons that are independent of crack length. Note that α, β, σ, and μ can be functions of time to represent changes in the characteristics of the SHM system during its service life. One commonly proposed advantage of SHM systems is that they can be installed at locations of difficult access. This makes it necessary to model and evaluate the effects of time-dependent variations on the response of the SHM system.

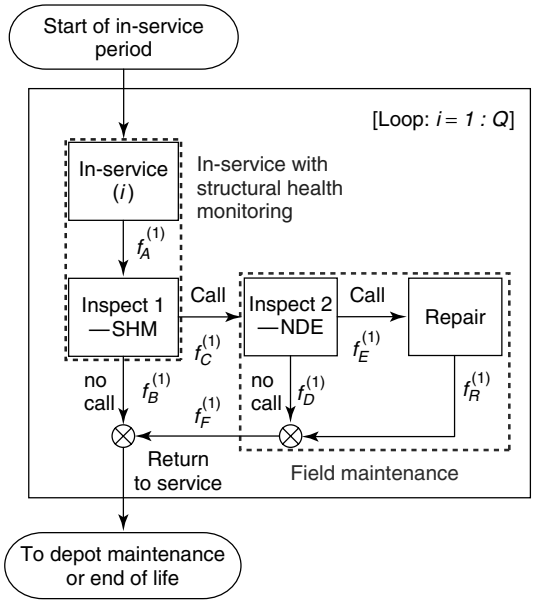


Figure 3. Flow diagram representing a model of an in-service period with SHM and an optional in-field maintenance.

Figure 3 presents a flow diagram for a basic SHM system integrated with an in-service period, with the opportunity for in-field maintenance incorporating a secondary inspection and repair process (with $j = 1$). A probabilistic analysis methodology is utilized for evaluating the model component blocks "Inspect 1—SHM", "Inspect 2—NDE", and "Repair" found in the figure.

In this formulation, $F\mathrm{p}^{(1)}$ and $f\mathrm{p}^{(1)}$ are defined as the cumulative density function (CDF) and probability density function (PDF) respectively, representing the flaw size distribution for feature type 1 (given by the superscript) at stage p (given by the subscript) in the inspect–repair subprocess. The subscripts A, B, and C are associated with flaw size distributions for the start of the SHM process, the portion associated with no call made (flaws not found), and the portion associated with a call made (flaws detected). ${P_{\mathrm{SHM}}}^{(1)}$ is defined as *the fraction of the PDF called (flaws detected) by the SHM process*, and is given by

$$ {P_{\mathrm{SHM}}}^{(1)} = \int_0^\infty POD_{\mathrm{SHM}}(a)\, f_A^{(1)}(a)\, \mathrm{d}a \tag{3} $$

where a is associated with flaw size and $POD_{\mathrm{SHM}}(a)$ is the POD function for the SHM process. The corresponding "no call" and "called" distributions resulting from SHM are respectively given by

$$ f_B^{(1)}(a) = (1 - POD_{\mathrm{SHM}}(a)) \times f_A^{(1)}(a) \tag{4} $$

$$ f_C^{(1)}(a) = POD_{\mathrm{SHM}}(a) \times f_A^{(1)}(a) \tag{5} $$

A secondary inspection given in block "Inspect 2—NDE" can also be evaluated in a similar fashion, where ${P_{\mathrm{NDE}}}^{(1)}$ is defined as *the fraction of the PDF called (flaws detected) by the NDE procedure*, and is given by

$$ \begin{aligned} {P_{\mathrm{NDE}}}^{(1)} &= \int_0^\infty POD_{\mathrm{NDE}}(a)\, f_C^{(1)}(a)\, \mathrm{d}a \\ &= \int_0^\infty POD_{\mathrm{NDE}}(a) POD_{\mathrm{SHM}}(a) \\ &\quad \times f_A^{(1)}(a)\, \mathrm{d}a \end{aligned} \tag{6} $$

The corresponding "no call" and "called" distributions resulting from the secondary NDE procedure are respectively given by

$$ \begin{aligned} f_D^{(1)}(a) &= (1 - POD_{\mathrm{NDE}}(a)) \times f_C^{(1)}(a) \\ &= (1 - POD_{\mathrm{NDE}}(a)) POD_{\mathrm{SHM}}(a) \\ &\quad \times f_A^{(1)}(a) \end{aligned} \tag{7} $$

$$ \begin{aligned} f_E^{(1)}(a) &= POD_{\mathrm{NDE}}(a) \times f_C^{(1)}(a) \\ &= POD_{\mathrm{NDE}}(a) POD_{\mathrm{SHM}}(a) \times f_A^{(1)}(a) \end{aligned} \tag{8} $$

For this example, the resulting repair distribution represents a return to the original state of the part for those flaws called both by the SHM process and the NDE technique, and can be expressed as

$$ f_R^{(1)}(a) = {P_{\mathrm{NDE}}}^{(1)} \cdot f_{R\_\mathrm{EIFS}}^{(1)}(a) \tag{9} $$

where $f_{R\_\mathrm{EIFS}}(a)$ represents the equivalent initial flaw size PDF for the original part. This process is repeated for Q iterations corresponding to each SHM manager decision and maintenance opportunity (i). Following this process, depot maintenance or end of life may be reached, depending on the design life of the aircraft. Extensions of this probabilistic model can be made to address problems where aircraft are removed from

service because of SHM calls (including availability tracking) and degradation of the SHM system using time-dependent POD models [10].

3.3 Cost models

Accuracy of cost models, one of the key elements of life-cycle management design, depends on various factors. First, the level of detail at which a cost model is developed will necessarily affect value analysis efforts. Second, not all relationships between cost and the various parameters that define health assessment technologies and processes may be taken into account. Third, the effects of random variations on cost can best be modeled using probabilistic methods. Finally, many of the numbers that are necessary for cost estimation must often be provided by the user in the form of coefficients and rates, and can be highly subjective, especially when historical data are not available or not appropriately considered. The influence of human factors on the accuracy of probabilistic cost estimates has been studied in [32], reportedly resulting in probabilistic cost models that better reflect the overconfidence in assessing uncertainty and relationships among cost elements and risks that are not usually considered. Here, we briefly describe a few cost modeling methods and tools, and warn the reader that these are only examples from a larger set of methods and tools that are currently available. As expected, the choice of the model will depend on the particular application, the desired level of detail, whether deterministic or stochastic models are desired, and how much data and knowledge are available to populate the model.

- The Integrated Cost Estimation (ICE™) software tool, by Frontier Technology (www.fti-net.com), uses a work breakdown structure for cost and provides capabilities for estimating LCC and for performing what-if studies and sensitivity analyses.
- TruePlanning®, from PRICE Systems (www.price systems.com) uses parametric models for estimating costs. Instead of building the model from the bottom up using materials and labor rates, the software uses representative data attributes from up-to-date knowledge bases.
- Crystal Ball, from Decisioneering, Inc. (www.crystalball.com) and @RISK, from Palisade Corp. (www.palisade.com), are two probabilistic modeling software tools that work with Microsoft Excel and use Monte Carlo simulation to estimate and optimize cost and risk statistics, among other variables.
- BlockSim, from Reliasoft Corp. (www.reliasoft.com), is capable of performing deterministic or probabilistic LCC analyses and other analyses (reliability, availability, etc.).

As discussed in Section 2.1, cost models for SHM systems must include all development, implementation, and in-service costs. To properly assess the true in-service cost associated with an SHM system, integrated models are needed to track all significant variable costs. These costs are a function of several variables including SHM indications, SHM false calls, flight hours (required to assess fuel cost due to added weight), and SHM monitoring and maintenance intervals.

3.4 Data and databases

A significant challenge to performing an accurate value assessment for an SHM system concerns the quality of the available data for the analysis. Examples of input parameter data for the probabilistic model that often exhibit large variability or significant expense to accurately quantify include (i) the equivalent initial flaw size distribution, (ii) the flaw growth model, (iii) the inspection POD model, and (iv) cost information. Databases have been developed and implemented for improved fleet structure management that also provide quality data to better understand the real flaw growth condition and inspection calls made over time. Two database examples include AIRCAT [33, 34] and FLEETLIFE [35, 36]. These data can be helpful in reducing uncertainty in the initial model assumptions and parameter values. Advancements in numerical fracture mechanics modeling and model-assisted POD evaluation can also be used to supplement initial model approximations [37, 38]. To practically obtain the best cost data for the analysis, partnerships must also be built and maintained with maintenance engineering, inspection services, repair teams, and the fleet managers to acquire the most accurate cost data to perform trade-off studies.

3.5 Framework for model analysis and optimization

Optimal design of an aerospace structure's life management strategy must take into consideration models for the various components described above, namely, the damage type, the damage detection process by on- and off-line transducers and algorithms, the scheduling of inspection processes, and the effects that changes in design parameters have on availability, reliability, and cost. Given such a complex set of interrelated factors, it becomes necessary to devise a method to integrate all these models to enable (i) estimation of costs, reliability, availability, and other figures of merits; (ii) sensitivity analysis; and (iii) optimization-based design of life management strategies. For example, in the case of probabilistic models for reliability and cost, Monte Carlo simulation becomes one obvious choice for estimation of outcomes when the statistics of all stochastic input variables are given. Several of the software tools described in Section 3.3 intend to provide such an integrated framework. The methods and web-based software framework described in [9], and other tools that are available in the market could also be expanded to provide these integration capabilities. During the last few years, the authors participated in the development of an integration tool specifically motivated by the need for optimal design of structural component life management strategies [10, 25–27]. That tool is used in the following section in the context of several examples.

4 APPLICATIONS

4.1 Applications—potential methods with related examples

Most of the existing applications of health monitoring in aerospace are in areas other than structures. However, there has been progress in applying systematic methods for justifying and designing SHM systems for aircraft structures. Albert *et al.* [39] present a systems engineering approach to SHM design for a nonspecific aging aircraft, and show by damage simulation and CEA, among other analyses, that implementing an SHM system on this aircraft may improve safety and decrease maintenance costs. Boller [40] studies various ways for monitoring structural health in a aging aircraft, from traditional NDE methods through loads monitoring to damage monitoring, from an LCC-benefit standpoint, and shows for a particular example that SHM makes good cost–benefit sense for certain components in the aircraft. Kent and Murphy [23] study the impacts, in terms of cost and other measures, of implementing SHM systems on several components of commercial aircraft. They found that if a 30–40% savings in maintenance costs is achieved, the initial investment can be recovered in 2–3 years for two of the structural components studied and 6–7 years for the third component because of the higher complexity of the SHM system. They also identified the need to integrate the SHM data with existing maintenance procedures to achieve best return on investment. In [24], Kapoor, Goh, and Boller present a methodology for simulating the operational and maintenance processes for a structure, including costs, and propose methods for identifying elements of the life management strategy that can be improved using SHM. Boller and Staszewski describe a procedure for establishing the LCC impact of new damage monitoring techniques in [[6], pp. 36–43] and the references therein, and show that the higher LCC impact is achieved on "highly loaded and difficult to access components".

4.2 SHM design space exploration

A methodology for optimum component life management including NDE and SHM has been proposed through the application of unified life-cycle engineering principles with integrated virtual design tools [10]. Prior work by the authors has addressed development of a software tool to enable analysis of trade-offs in NDE and SHM design in terms of structure life-cycle outcomes [10, 25]. Several important features in the software tool facilitate specialized design studies. In particular, any model factor can be selected and defined as *a variable for a parametric or optimization study*. Visualization and tabular features facilitate exploration of the design space in terms of key measures, e.g., total cost and maximum probability of failure (POF).

The continued development and integration of component models with structure service databases will provide the functionality to improve NDE and

Table 1. Parameter values for case study simulation

Fracture toughness normal distribution	
Mean	32.3 MPa-$m^{1/2}$ (29.4 Ksi-$in^{1/2}$)
Standard deviation	2.4 MPa-$m^{1/2}$ (2.2 Ksi-$in^{1/2}$)
Gumbell distribution for critical stress during flight	
α	1.26
β	21.7
Flaw growth model (Paris' law) parameters	0.0001
Repair EIFS distribution	
Au	0.01
ϕ	3
α	1.8
Parameters of POD	
50% median flaw size	1.27 mm (0.05 in.)
Curve steepness (s)	0.51 mm (0.02 in.)
False call rate	0
Random missed call rate	0
Critical crack length	12.7 mm (0.5 in.)

SHM design. CBM and prognostics programs can also benefit by (i) indicating the best service life time window for application (i.e., limit prognostics decisions in early stages of life to eliminate false calls), (ii) refining life prediction models based on monitoring the load and environmental conditions, (iii) determining precise end-of-life date from damage state measurement data, and (iv) reducing model uncertainty (in life prediction and POD models) over time.

4.3 Value assessment demonstrations using probabilistic tools

Several case studies are now presented to demonstrate the capability of the approach described in [10, 25] and to gain a better understanding of the dynamics of the SHM system model. Given the difficulty of obtaining precise data on specific cases for the equivalent initial flaw size distribution, the flaw growth model, the POD capability, and cost data, values were set based upon a combination of information acquired from maintenance personnel and values cited in the literature for related aircraft components. The general class of problem considered here was the inspection of fatigue cracks at fastener sites in multilayer aircraft wing structures as found in such aircraft as the C-130 [41]. More details on the life prediction model, including parameter descriptions, can be found in work by Berens *et al.* [28, 29]. Specific life prediction and POD model parameter values used for this study can be found in Table 1. The CDF for equivalent initial flaw size and the geometric relation between fracture toughness (Kc) over stress (σ) with respect to crack size are given in Figure 4(a) and (b), respectively. While the software allows for multiple fastener hole sizes to accommodate the actual case where holes with cracks are repaired by enlarging the diameter, the examples presented here use a single hole size. In addition, the relationship between flight hours and elapsed time was not taken into account in these hypothetical cases, but normally it will vary from aircraft to aircraft. The software also allows for the length of different in-service periods to be different and for specific inspection and repair parameters and cost factors to vary between in-service periods, but all these input quantities were kept constant for these examples.

The first study explores the effect of varying the number of in-service intervals for a fixed total service life, taking into account that during each service interval an integrated SHM system is collecting sensor data that must be analyzed upon landing. Any nonzero indication by the SHM system results in a field inspection and repair process. Figure 5 shows a screenshot of the cost report window for this example, in the case where the life of the system is divided into five in-service periods with SHM data processing and

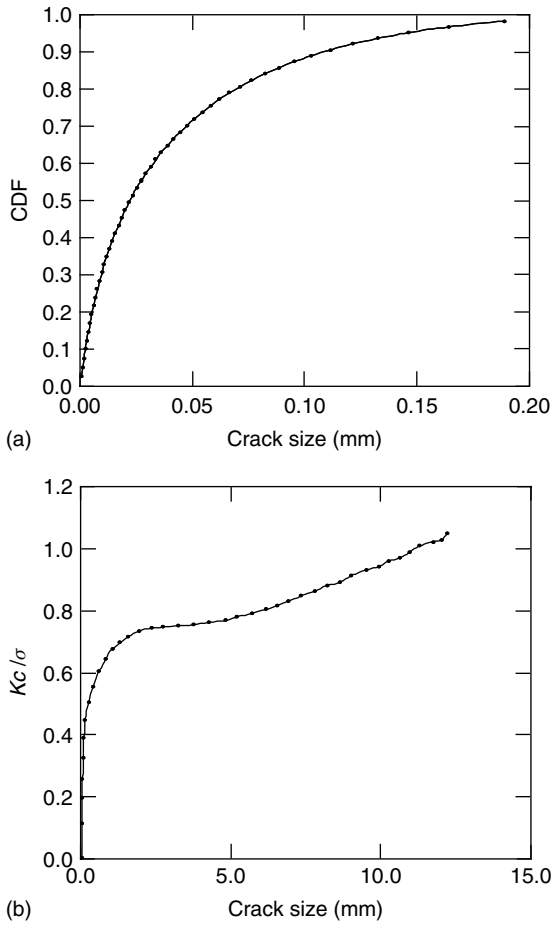


Figure 4. (a) Cumulative distribution function for equivalent initial flaw size and (b) geometric relation between fracture toughness (Kc) over stress (σ) with respect to crack size.

possible inspection and repair processes after each of the first four in-service periods.

As shown in Figure 5, the model includes initial costs of capital equipment and training, estimated at \$500 000 for this example. Variable costs for each service period are associated with the added fuel consumption due to the weight of the SHM system. For example, for the third in-service period, this cost is equal to \$2000, which is the product of the flight hours (8000) and the added cost of fuel per hour due to the SHM system weight (\$0.25). The fixed costs of preparing for SHM data processing and interpretation on the ground after the third in-service period are \$1500, while the variable costs of data interpretation depend on the amount of data collected, which, in turn, depends on the number of hours in the previous in-service period. That is, 8000 h at \$0.50/h results in \$4000 for interpreting the data from the SHM system after the third in-service period. Since the SHM system detects flaws at that time, a subsequent inspection and repair process must be performed. The fixed costs of maintaining the necessary facilities and equipment are \$20 000 and \$5000 for inspection and repair, respectively. Variable repair labor costs depend on the portion of the items being monitored that was identified by the SHM system and the subsequent field inspection as damaged. In the case at hand, after the third in-service period, 7.66% of all fastener locations being monitored were identified as requiring repair. This percentage, multiplied by the unit cost of repair (\$2 500 000), yielded variable repair labor costs of \$191 495.35 after the third in-service period.

Figure 6 shows the simulated results for probability of failure and cumulative maintenance cost as a function of time for 5, 20, and 50 in-service periods. Table 2 presents the maximum probability of failure over time and the total cost of SHM, NDE, and repair, for the same three cumulative in-service periods.

It is easily observed that for this first case study, a higher frequency of SHM calls will result in higher LCC. The source of this higher cost is twofold. First, the total cost associated with labor hours for data interpretation is increased with the frequency of SHM calls. In theory, this cost could be quite small if robust automated algorithms for data interpretation are used. In practice, given the high cost for repairs and potential for false calls due to unknown conditions not considered in the original design, secondary assessments of the SHM data by an expert inspector would be necessary. The second source for higher costs occurs over the later part of the service life, where noncritical flaws are called by the SHM system. Ideally, minimizing the frequency of calls while maintaining an acceptable level of reliability in terms of probability of failure is a fundamental design principle for minimizing LCCs. Alternatively, higher frequency rates of SHM calls can significantly improve reliability. This strategy is particularly valuable when the SHM system is designed to only detect very large flaws, the crack growth model is nonlinear, or large uncertainty is present in the crack growth model parameters.

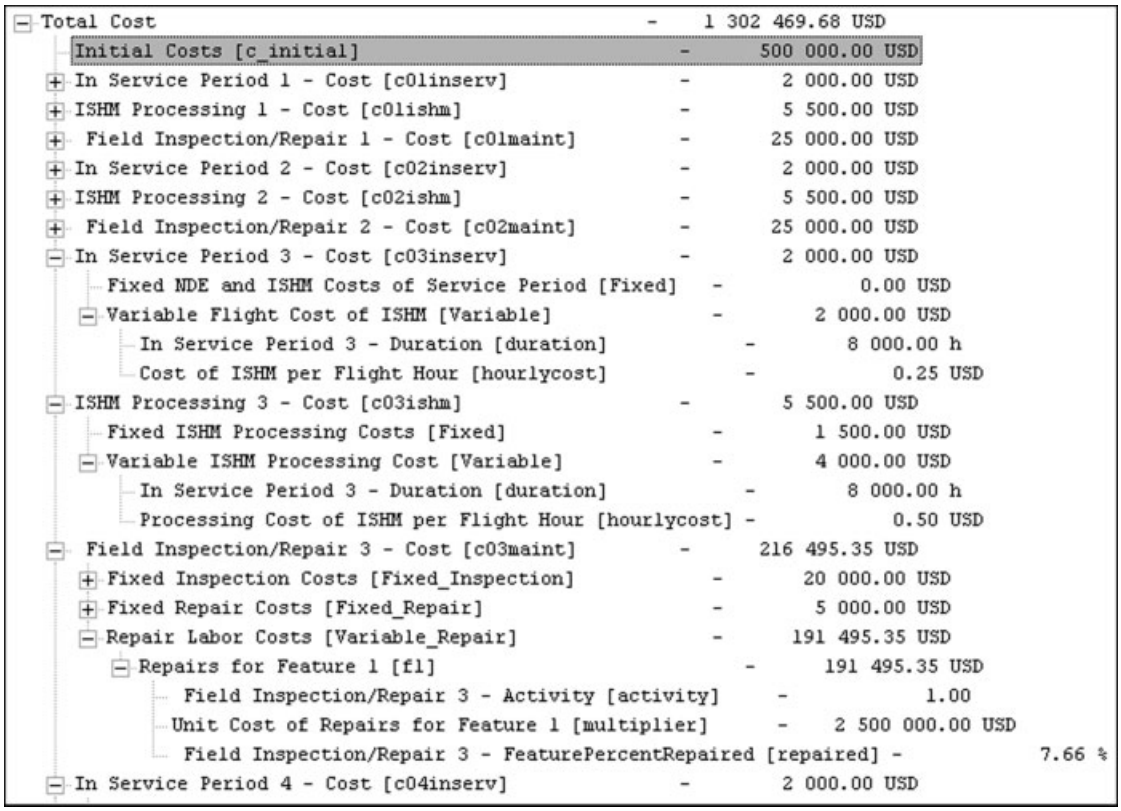


Figure 5. Cost structure and data for the first case study, using five in-service periods.

A second case study explores variations in the detectable flaw size for both an SHM system (Inspect 1—SHM) and a secondary NDE inspection technique (Inspect 2—NDE). Specifically, the 50% detectable flaw size parameters for the SHM and NDE inspection models were both varied from 0.51 mm (0.02 in) to 2.54 mm (0.10 in) as a full-factorial study. Figure 7 presents the design solution space resulting from the study in terms of maximum probability of failure and total cost. This design solution space plot provides the means to select the Pareto solutions providing the optimal trade-off between the two objectives. Furthermore, it is possible to select from this reduced solution set a design that minimizes cost while maintaining an acceptable probability of failure, typically set at 10^{-6}. Using these criteria, the optimal SHM system 50% detectable flaw size was found to be 1.52 mm (0.06 in), with the secondary NDE system 50% detectable flaw size set to any value greater than 1.27 mm (0.05 in)″.

A third case study explores the sensitivity of cost and reliability measures to SHM system false call rate. Owing to space limitations, only a discussion of the trends in the results is presented. As previously mentioned, the issue of false calls can hinder the application of SHM systems. Given the challenging problem of reliably detecting cracks using distributed sensors, false calls rates are expected to be comparable or higher with respect to NDE cases, typically on the order of 1%. However, although a 1% false call rate may be acceptable for less-frequent NDE inspections, when SHM calls are made at a more-frequent rate, a greater number of locations will be falsely called over the life of the aircraft and thus prompt some form of secondary maintenance action. This would be especially problematic given that the model predicts that most calls that are initially made are most likely false calls, thus having a negative impact on the product life management program and its sponsors. Secondary inspections were found to be

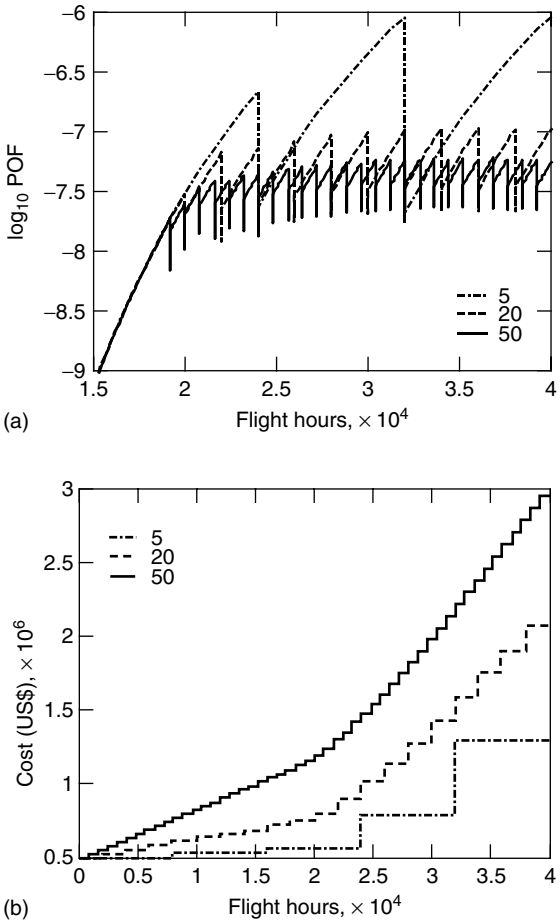


Figure 6. Plots of (a) probability of failure (POF) and (b) cumulative maintenance costs as a function of flight hours and SHM number of cycles.

Table 2. Maximum probability of failure and total cost for first case study

Number of in-service periods	Log_{10} of maximum probability of failure	Total SHM, NDE, and repair costs (US$)
5	−6.036	1 302 469.68
20	−6.969	2 066 261.90
50	−7.218	2 945 338.23

quite beneficial in mitigating cost by limiting unnecessary repairs due to false calls. However, if the cost of secondary inspections is not small, the total cost may be excessive and thus hinder practical use.

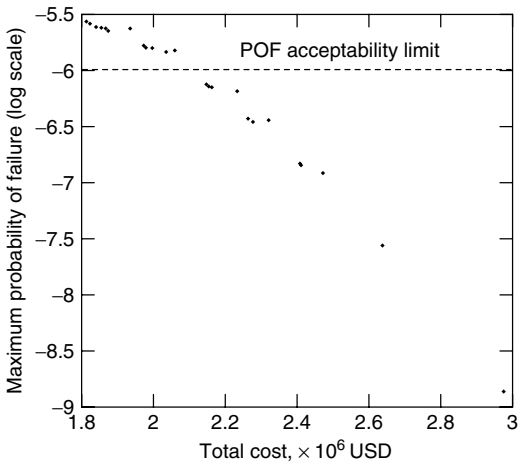


Figure 7. Design solution space for varying SHM and secondary NDE sensitivity in terms of maximum probability of failure (POF) and total cost.

This is a good example of the need to perform value assessment analysis taking into consideration all relevant technical factors of the technology in question.

5 CONCLUDING REMARKS

We have attempted to describe the state of the art in value assessment for SHM technologies. We believe the most likely current role of SHM in aerospace is as part of a hybrid structural life management strategy in which SHM and NDE technologies must work in a complementary fashion. We have presented an overview of existing methods for value analysis and identified several software tools that implement the various methods. Since SHM is a relatively new technology, not many field implementations have been found that resulted from a thorough analysis of the value of the technology. However, traditional value-analysis tools such as CBA can be utilized as long as the necessary models are in place. In addition, it may be advantageous to use more powerful trade-off analysis tools such as those offered by multicriteria optimization methods to clearly identify trade-offs among all the figures of merit for both SHM and the NDE methods that complement it. The accuracy of the data to populate all necessary models is of utmost importance, and the conclusions obtained from any of these analyses must always be qualified with references to the data that were used. This is especially

important because of the few existing applications of SHM and the fact that the real benefits of this technology are likely to be apparent only several years (or decades) after its adoption as part of structural life management. The ability of the design methods and software to facilitate design trade-off assessment and optimization, as shown in the examples, is quite valuable even when the models for the technology and its effects on reliability and cost and the necessary data and knowledge bases are still under development.

Future work should explore the sensitivity of model trends to model parameters and acquire better data on flaw size distributions and real costs of SHM implementations. Collaboration throughout the aerospace community is needed to make existing data available. Laboratory studies, aircraft teardown studies, the application of quantitative NDE techniques for flaw characterization, *in situ* sensors for damage state, loading and environment condition tracking, numerical models, and well-managed databases are all integral to building the necessary data sets and models.

ACKNOWLEDGMENTS

This work was partially supported by the U.S. Air Force Research Laboratory, Nondestructive Evaluation Branch, through contracts F33615-03-D-5204 and F33615-03-C-5226.

RELATED ARTICLES

Principles of Structural Degradation Monitoring

Design Principles for Aerospace Structures

Military Aircraft

Use of Leave-in-place Sensors and SHM Methods to Improve Assessments of Aging Structures

Usage Management of Military Aircraft Structures

REFERENCES

[1] National Research Council, *Aging of U.S. Air Force Aircraft*, Publication NMAB-488-2, National Academy Press: Washington, DC, 1997.

[2] Cooke GR, Kealy HD. *A Study to Determine the Annual Direct Cost of Corrosion Maintenance for Weapon Systems and Equipment in the United States Air Force*, Final Report on Contract No.F09603-89-C-3016. NCI Information Systems: Dayton, OH, February 1990.

[3] Lewis WH, Sproat WH, Dodd BD, Hamilton JM. *Reliability of Nondestructive Inspections—Final Report*, SA-ALC/MEE 76-6-38-1, December 1978.

[4] Brausch J, Butkus L, Kraft K, Schmidt J, Goglia J. ASIP panel session: addressing the NDI crack miss problem for safety of flight structures. *ASIP Conference*. Memphis, TN, November 30, 2005.

[5] Boller C. Next generation structural health monitoring and its integration into aircraft design. *International Journal of Systems Science* 2000 **31**(11):1333–1349.

[6] Staszewski WJ, Boller C, Tomlinson GR. *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons: Chichester, 2002.

[7] Adams DE, Nataraju M. A nonlinear dynamical systems framework for structural diagnosis and prognosis. *International Journal of Engineering Science* 2002 **40**:1919–1941.

[8] Malas J. Requirement for structural health monitoring / prognosis. *Review of Progress in Quantitative Nondestructive Evaluation* 2005 **24**:1987–2000.

[9] Kacprzynski GJ, Roemer MJ, Hess AJ. Health management system design: development, simulation and cost/benefit optimization. *IEEE Aerospace Conference Proceedings*. Big Sky, MT, 2002; Vol. 6, pp. 3065–3072.

[10] Aldrin JC, Medina E, Knopp JS. Cost benefit analysis incorporating probabilistic risk assessment for structural health monitoring. *Review of Progress in QNDE* 2006 **25**:1910–1918.

[11] Jata KV, Knopp JS, Aldrin JC, Medina EA, Lindgren EA, Transitioning from NDE inspection to online structural health monitoring—issues and challenges. *Proceedings of the 3rd European Workshop on Structural Health Monitoring*. Granada, 2006; pp. 987–995.

[12] Boardman AE, Greenberg DH, Vining AR, Weimer DL. *Cost-Benefit Analysis, Concepts and Practice, Second Edition*, Prentice Hall: Upper Saddle River, NJ, 2001.

[13] Fabrycky WJ, Blanchard BS. *Life-Cycle Cost and Economic Analysis*. Prentice Hall: Englewood Cliffs, NJ, 1991.

[14] Mollaghasemi M, Pet-Edwards J. *Making Multiple-Objective Decisions*. IEEE Computer Society Press: Los Alamitos, CA, 1997.

[15] Coello Coello CA. Recent trends in evolutionary multiobjective optimization. In *Evolutionary Multi-objective Optimization: Theoretical Advances and Applications*, Abraham A, Jain L, Goldberg R (eds). Springer-Verlag: London, 2005, pp. 7–32.

[16] Hagamaier DJ. Effective implementation of NDT into aircraft design, fabrication, and service. *Materials Evaluation* 1988 **46**:851–868.

[17] Hagamaier DJ. Cost benefits of nondestructive testing in aircraft maintenance. *Materials Evaluation* 1988 **46**:1272–1284.

[18] Generazio ER, Chamis CC. Technology benefit estimator for aerospace propulsion systems. *30th AIAA/ASME Joint Propulsion Conference*. Indianapolis, IN, 1994.

[19] Brechling VJ. *Methodology for the Economic Assessment of Nondestructive Evaluation Techniques used in Aircraft Inspection*, Final Report, DOT/FAA/AR-95/101, 1995.

[20] Papadakis EP. Financial justification for investment in nondestructive testing equipment. *Materials Evaluation* 1997 **55**:1155–1158.

[21] Papadakis EP. A cost of quality: three financial methods for making inspection decisions. *Materials Evaluation* 1997 **55**:1336–1345.

[22] Wall M, Wedgwood FA. Economic assessment of inspection—the inspection value method. *ECNDT*, (NDT.net), **3**(12):1998.

[23] Kent RM, Murphy DA. *Health Monitoring System Technology Assessments–Cost Benefit Analysis*. NASA / CR-2000-209848. ARINC: Annapolis, MD.

[24] Kapoor H, Goh WT, Boller C. Procedures for the assessment of structural health monitoring potentials. *Proceedings of the 3rd European Workshop on Structural Health Monitoring*, Granada, 2006; pp. 191–198.

[25] Aldrin JC, Medina E, Altynova M, Knopp J, Kropas-Hughes CV. Strategy and software framework for integration of QNDE and product life management design. *Review of Progress in Quantitative Nondestructive Evaluation* 2005 **24**:1682–1689.

[26] Medina EA, Aldrin JC, Allwine DA,Fisher J, QadeerAhmed M, Knopp JS. Simulation-based design and tradeoff analysis with probabilistic risk assessment for NDE and structural health monitoring. *ASNT Fall Conference*, Columbus, 2005.

[27] Medina EA, Aldrin JC, Knopp JS, Allwine DA. Value assessment tools for hybrid NDE-SHM life management strategies. *Proceedings of the 3rd European Workshop on Structural Health Monitoring*, Granada, 2006; pp. 1027–1034.

[28] Berens A, Hovey P, Skinn D. *Risk Analysis for Aging Aircraft Fleets, Volume 1—Analysis*, UDRI, Final Report for Period Sep. 1987—Jan. 1991, Contract F33615-87-C-3215, to Flight Dynamics Directorate, Wright Laboratory: Wright-Patterson AFB, OH.

[29] Berens A, West JD, Trego A. Risk assessment of fatigue cracks in corroded lap joints. *Proceedings of the NATO-RTO Air Vehicle Technology Panel Workshop on Fatigue in the Presence of Corrosion*. Corfu, 1998; pp. 21.1–21.10.

[30] Berens A. NDE reliability data analysis, *Metals Handbook*, ASM International, 1989; Vol. 17, pp. 689–701.

[31] Moore DG, Spencer FW. Interlayer crack detection results using sliding probe eddy current procedures. *10th Asia-Pacific Conference on Non-Destructive Testing*, Brisbane, 2001.

[32] Kujawski E, Alvaro M, Edwards W. Incorporating psychological influences in probabilistic cost analysis. *Systems Engineering* 2004 **7**(3):195–216.

[33] Stilley S, Pratt EM. C-130 AIRCAT—automated inspection, repair, corrosion and aircraft tracking (AIRCAT). *Aging Aircraft Conference*. St. Louis, MO, 2000.

[34] Prewett M, Waldbusser R. Two-phase fleet management using NAV-AIRCAT. *Aging Aircraft Conference*. Atlanta, GA, 2005.

[35] Giese RD, Herring GD, Johnson T. Managing the economic aspects of ASIP programs, *The Third Joint FAA/DoD/NASA Conference on Aging Aircraft*, September 20–23, 1999, Albuquerque, 1999.

[36] Rice RC. Integration of ASIP management and economic service life analysis software. *ASIP Conference*. San Antonio, TX, 2000.

[37] Schmerr LW, Thompson DO. *Review of Progress in QNDE* 1993 **12**:2325–2332.

[38] Thompson RB. Using physical models of the testing process in the determination of probability of detection. *Materials Evaluation* 2001 **59**(7):861–865.

[39] Albert A, Antoniou E, Leggiero S, Tooman K, Veglio R. *A Systems Engineering Approach to Integrated Structural Health Monitoring for Aging*

Aircraft, M.S. Thesis. U.S. Air Force Institute of Technology, March 2006.

[40] Boller C. Ways and options for aircraft structures health management. *Smart Materials and Structures* 2001 **10**:432–440.

[41] Lindgren E, *et al.* Validation and deployment of automated ultrasonic inspections for the C-130 center wing. *ASIP Conference*. Savannah, GA, 2–4 December 2004.

Chapter 88
Environmental Monitoring of Aircraft

Nicholas C. Bellinger and Marcias Martinez
Institute for Aerospace Research, National Research Council Canada, Ottawa, Ontario, Canada

1 INTRODUCTION

When fixed-wing aircraft were designed and built, most manufacturers made little allowance for environmental degradation since aircraft were not expected to remain in service beyond their design service life. However, in the mid-1980s, when a 19-year-old Boeing 737 aircraft, operated by Aloha Airlines, lost a large portion of its front upper fuselage, it became the defining event in creating awareness into the possible issues associated with aging aircraft [1]. The cause of the accident was determined to be the development and linking up of multiple cracks at numerous rivet holes in the fuselage lap joints, which is referred to as *multi-site damage* (MSD) [2]. Since then, millions of dollars have been expended on MSD-related research activities to determine the impact this cracking scenario has on the structural integrity of fuselage lap joints.

During the Aloha investigation, both environmental (corrosion) and age degradation (fretting) modes were found to be present within the failed lap joints. However, since no direct correlation was made between these modes and the crack nucleation mechanism, very little research has been carried out into the possible effects these degradation modes have on aircraft structural integrity. Thus, corrosion has only been considered a durability issue affecting the remaining (residual) strength of a component because of the increase in stress due to the reduction in cross-sectional area resulting from material thinning. To ensure that corrosion does not cause premature failure of a component, limits are set on the amount (level) of acceptable damage (thinning). Therefore, components must undergo a repair procedure when the detected damage is above these limits. Owing to these arbitrary limits, billions of dollars are being spent worldwide each year in repairing corrosion damage. In the United States alone, it is estimated that the Department of Defense spends $20 billion annually in corrosion-related repairs [3]. One reason for this high expense is that, in the past, it was not possible to predict where or when corrosion would occur, or the time required to generate sufficient damage to affect the structural integrity of the aircraft.

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

2 WHY ARE WE IN THIS STATE?

Presently, two design methodologies are being used to determine the life of fixed-wing aircraft structures; safe life and damage tolerance. There is an additional paradigm known as *the fail-safe approach* but for this article, this concept is considered to be a derivative of the damage tolerance approach. The safe-life method is used to determine the time that a component can remain in service without the possibility of a crack forming. This method is used to design "safety" critical components, such as landing gears. The safe-life methodology assumes that all materials are ideal, continuous, homogeneous, and isotropic. To take into account the inherent material scatter, safety factors are applied to the fatigue (or endurance) limit. To prevent cracking, safety critical components are designed to avoid stresses above the fatigue limit. However, owing to the fact that materials are assumed to be free from "defects" or discontinuities, components must be immediately replaced once any damage is found, which has led to the current "find-it and fix-it" maintenance philosophy. This philosophy also results in the retirement of components that are not significantly damaged, and the failure of some components that develop cracks at stress raisers, such as corrosion pits.

To reduce the risk of aircraft flying with unknown cracks, the damage tolerance methodology was developed, which assumes that all fatigue critical components contain growing cracks and that failure can occur when actual conditions are different from those modeled. In this methodology, the crack growth life is estimated, which drives directed in-service inspections, allowing components to remain in service longer, thus reducing the cost of maintaining aircraft. These inspections are carried out to prevent cracks from occurring at unknown sources of discontinuities that may occur during the life of the component. However, failures have occurred in the past and are still happening that have been caused by unexpected degradation modes, such as fretting and corrosion.

Both these methodologies were developed on the basis of a combination of basic scientific principles that were developed at the time (30–40 years ago) as well as practical experience, but this experience does not include the effect that the environment has on component life. As when they were first developed, these methods still rely on the generation of empirical material properties, such as the fatigue properties used for the strain–life method (safe life) or crack growth rate properties (damage tolerance) to determine the number of cycles, or flight hours, to a prespecified small crack length or the time required to grow a crack to a critical size. This life estimation is then monitored throughout the life of an aircraft using specified nondestructive inspection (NDI) techniques in order to reduce the risk of failure for a specific component. This reliance on both analytical and experimental techniques, to predict the life of aircraft components that do not take into account environmental or age degradation modes, is the main cause for the significant increase in the cost of aircraft maintenance. In a previous study, corrosion was shown to significantly decrease the time required to form a crack (nucleation life) without affecting the time required to grow a stable crack (long crack growth rate) [4, 5]. This early crack nucleation life raises concerns about the effect corrosion might have on the structural integrity of aircraft components, particularly if they are known to be susceptible to cyclic loading (fatigue).

3 WHAT CAN WE DO?

On the basis of past experience, corrosion prone areas in aircraft are well known and include fuselage lap joints, upper and lower wing skins, and landing gear. Some examples of the types of corrosion damage found in aircraft are shown in Figure 1, which includes pitting, crevice, and exfoliation corrosion [6]. To determine if corrosion is present, traditional NDI techniques are used to examine specific components at predetermined operating (flying) times. It is interesting to note that the time an aircraft remains on the ground, which can be significant, is not taken into account in determining inspection frequencies even though corrosion requires the presence of stagnant water to form and grow. One problem with current NDI techniques, such as ultrasonic and eddy current, is the fact that a significant amount of damage needs to be present (about 10% thickness loss) to reliably detect corrosion to avoid expensive false calls. These high levels usually result in damaged components being replaced, which can significantly increase the cost of aircraft maintenance. In addition, since some of the areas of concern are not easily accessible, a

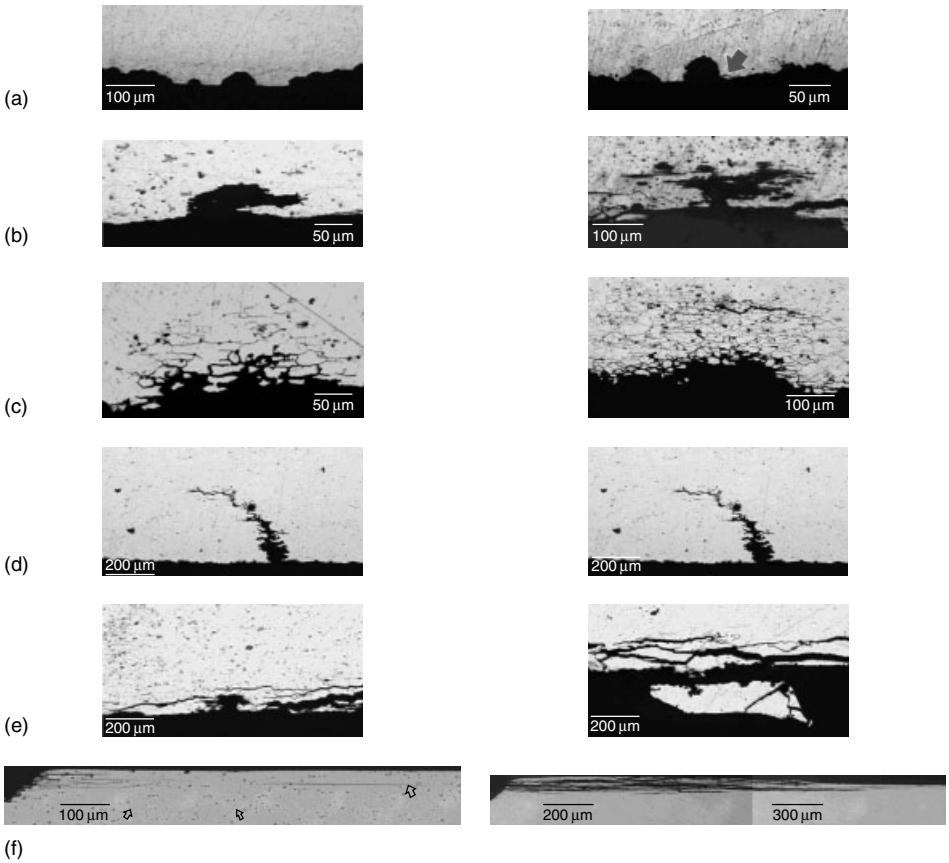


Figure 1. Examples of service-induced corrosion obtained from in-service and retired aircraft [6]. (a) Corrosion pitting in thin aluminum (2024-T3); (b) severe pitting resulting from crevice corrosion in thin aluminum (2024-T3) lap joints; (c) intergranular corrosion in thin aluminum (2024-T3); (d) environmental-assisted cracking in aluminum lap joints (2024-T3), referred to as *pillowing cracks*; (e) exfoliation corrosion in thin aluminum (2024-T3) lap joints; and (f) exfoliation corrosion in thick section aluminum (7178-T6) upper wing skins.

significant amount of time may be required just to gain access to the corrosion, assuming it is actually present. Corrosion, quite simply, is unpredictable and therefore often unexpected, thus increasing downtime for maintenance, delaying return to service, and adding to costs.

To reduce this cost, inspections need to move from opportunistic (as they are now) to directed. To accomplish this, one needs to know what to look for (in this case, what type of corrosion), where to look, how to look (inspection technique to use), and when and how often to look. Basically, this means that the state of an aircraft must be known prior to commencing a scheduled maintenance cycle. One possible approach is to assess the state of an aircraft by placing direct corrosion sensors (DCS) and/or corrosive environment sensors (CES) at key locations. These sensors could be used to monitor the state of the component to determine when corrosion damage is present, which, in turn, would allow inspections for corrosion to be tailored to individual aircraft, possibly delaying repairing the damage until a more appropriate time, resulting in a significant cost saving.

Many factors affect the deterioration of metallic structures, such as changes in material properties, time/temperature/chemical environment, and in-service operations, including field use and storage [7, 8]. It is for this reason that corrosion sensors are important for monitoring and establishing the rate and degree of deterioration of aircraft structures. See

Part 5 of this encyclopedia for more information regarding the different sensor types.

In this section, we have subdivided the different corrosion sensors into three categories. The first category, which we refer to as *corrosive environment sensors* for the most part, monitors environmental properties, such as pH levels of the atmosphere or fluid, oxygen levels, and temperature or humidity. This category also includes those sensors that measure the corrosivity of the environment by using a surrogate metal to determine the need for maintenance. These sensors do not necessarily provide information on the health of the structure but determine if the conditions are present to cause corrosion. Therefore, to use these types of sensors, signal processing has to be carried out to determine a correlation between the corrosiveness of the environment and the actual corrosion damage incurred by the structure [9]. Refer to **Data Preprocessing for Damage Detection** for more details on data processing, as well as **Statistical Pattern Recognition** and **Artificial Neural Networks**. Also, these sensors do not detect or determine the location of the corrosion on a structure and thus prior knowledge of which components are susceptible to corrosion is necessary.

The second category of sensors contains those that monitor the health of the structure. This is done by sensors that are adhered to the surface or embedded within the structure. We refer to these types of sensors as *direct corrosion sensors*, because they attempt to directly measure the damage that is present on the structure by measuring the changes in various parameters. These parameters depend on the type of sensor but include electrical impedance, resistance measurements, and wave propagation through the structure. Algorithms must be developed to relate the measured parameter to the level of damage present on the structure. One system that has been developed by Honeywell International (USA) and Avonwood Developments LTD (UK) [10] measures the resistance value along a prescribed path that contains a sealant and protective coating. When this protective system has broken down and moisture has penetrated, a conductive path is created resulting in a low resistance value indicating the presence of galvanic corrosion.

With the availability of smart materials, new types of sensors have been developed. Piezoelectric (PZT) sensors are being used as DCS. PZT arrays have been placed in structures in strategic locations creating sensor arrays. The piezoelectric materials are capable of generating elastic waves that propagate through the structure, which are then sensed by other PZT sensors within the array. These types of arrays are able to identify cracks and corrosion damage on the structure due to changes in the characteristics of the waves. The results can be interpreted to determine location and amount of damage generated in the structure [11].

We refer to the third and last category as *multifunctional corrosion sensors* (MFCS). These sensors have the capability of monitoring chemistry and characteristics to establish the corrosive nature of the environment as well as measure corrosion damage that is present on the structure. By combining the different sensors, already developed in conjunction with wireless/wired transmission of data and the miniaturization of high-speed processors, a network of sensors can be created to produce "smarter" structures that can be an integral part of any structural health monitoring program.

4 HOW SHOULD WE PROCEED?

With the introduction of different types of sensors, a link must be made between the parameters being monitored, the level of corrosion damage present on the component (if it is not measured directly) and the effect this damage has on the structural integrity of the component. To develop this link, numerous experiments will need to be carried out under similar service conditions to generate the algorithms and data required to relate the data generated by the sensors to the life assessment models. In addition, statistically significant databases need to be generated to determine the various distributions that take into account the large scatter associated with environment degradation. These distributions, which have to be generated for each material system that is susceptible to environmental degradation, take into account the large variation in the size of corrosion damage as well as the very large scatter in corrosion growth rates. To avoid the use of large safety factors, Monte Carlo simulations are used to account for the scatter associated with the various databases.

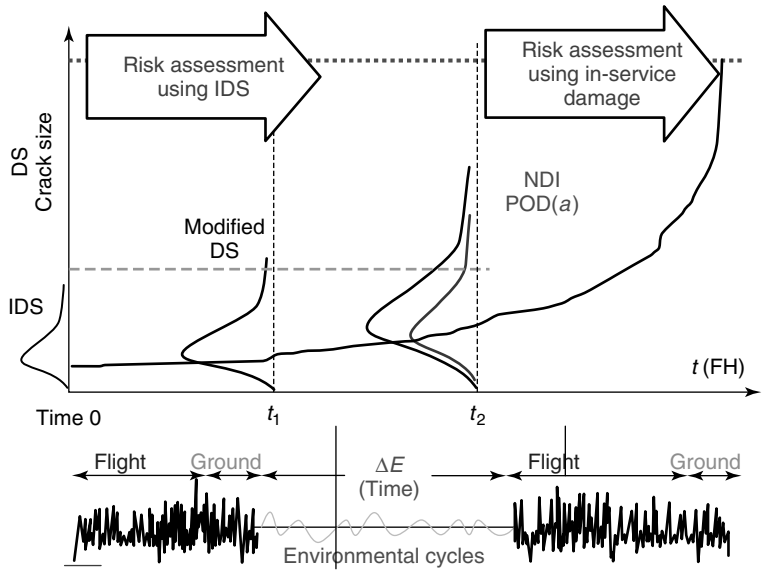


Figure 2. HOLSIP considers the progression of discontinuity states (cyclic operation spectra and time on ground spectra both with environment). FH, fight hours.

Since both existing life prediction methods do not take into account environmental degradation, new physics-based life management methodologies must be developed so that the causes, or reasons, behind potential failures are well understood (physics-of-failure). This, in turn, will allow manufacturers and operators to predict potential failures prior to an aircraft entering service. It can also be used to schedule maintenance by determining if the environmental degradation present on a component would result in unsafe operation of the aircraft, thus ensuring that repairs are planned.

One such methodology being developed is known as the *holistic structural integrity process* (HOLSIP), which is able to determine the total life of a component by including all four phases of life: nucleation, short/small crack growth, long crack growth, and unstable fracture [12–15]. The ultimate goal of this process is to ascertain the basic fatigue response of a structure from the as-manufactured state, including the early stages of damage formation. It is also being developed to assess the criticality of structures subjected to cyclic loading as well as age and environmental degradation, such as fretting, corrosion pitting, general corrosion, and corrosion pillowing. This is accomplished by accounting for the time in the air as well as the time while moving or at rest on the ground, as shown in Figure 2 [16].

Presently, HOLSIP contains physics-based models that take into account both cyclic and age/environmental effects by predicting the progression of initial discontinuity states (IDS) through the life of a component [12]. These models have been verified using results obtained from both coupon and in-service aircraft [16]. The term *IDS* is used to describe the as-produced or as-manufactured state of the material and is determined using standard metallurgical procedures to obtain the size and shape of the intrinsic discontinuities. IDS is a geometric and material characteristic that is a function of composition, microstructure, phases and phase morphology, and the manufacturing process used to process the material. However, since it is well known that not all discontinuities lead to the formation of cracks, a limited number of cyclic tests need to be carried out to determine the fatigue critical discontinuities that result in the formation of cracks. It is recognized that this process can be time-consuming and costly and thus procedures are being developed that use Monte Carlo simulations to be able to analytically predict the critical fatigue subset.

The effects of environmental degradation and fatigue, which can act not only concurrently but also independently, are characterized within HOLSIP

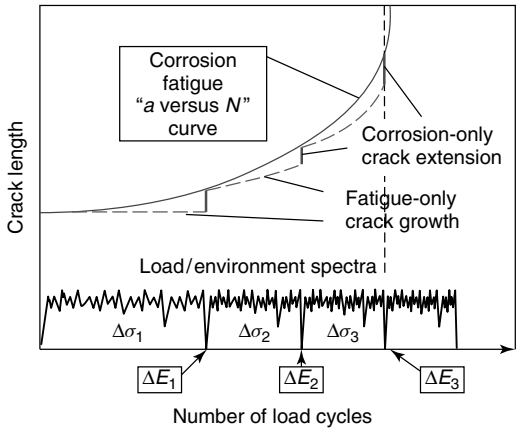


Figure 3. Corrosion fatigue analysis procedure in HOLSIP. [Reproduced with permission from Ref. 17. © USAF—ASIP, 1999.]

by changes in the crack-tip stress intensity. Fracture mechanics principles and special environmental parameters are used to formulate the required stress intensity factor solutions that are used to calculate the crack growth increment [16]. The data obtained from the various corrosion or environmental sensors could be used to more accurately calculate these special parameters. Over the past few years, the National Research Council Canada (NRC), in conjunction with other organizations, has carried out extensive research to develop special parameters to represent the effects of corrosion (corrosion-related thickness loss and corrosion pillowing stress) and corrosion–fatigue interaction (local geometry stress risers, topography change, and corrosion-induced sustained stress) on the stress intensity factor. In HOLSIP, the total crack growth is determined through the summation of the cyclic damage and time-dependent damage as shown in Figure 3, which is used to grow an initial discontinuity state to final instability.

To be able to account for not only the typical fatigue-associated uncertainties within a material, manufacturing, and loading but also age degradation uncertainties such as corrosion growth rate, probabilistic techniques must be included within the HOLSIP framework [15]. An additional probabilistic method is also required to interpret the NDI results for age degradation damages, which is not available in the current damage tolerance analysis methodology. An in-house probabilistic damage tolerance analysis (ProDTA) methodology has been developed as shown in Figure 4, which incorporates the HOLSIP-based lifing models and enhanced probabilistic techniques to analyze structural risk with age degradation. This type of analysis can be used to determine the repair action that would be more cost effective, since some repairs can result in more damage than was originally present [18]. The various input parameters that are required to carry out a quantitative risk assessment are shown in Figure 5. As can be seen from this figure, a significant amount of information is needed. The probabilistic analysis procedures employed in ProDTA are shown schematically in Figure 6. The analysis starts from the IDS distribution for the as-produced and as-manufactured material. Two methods were developed to grow the IDS distribution to the next inspection/repair time, t_1. The

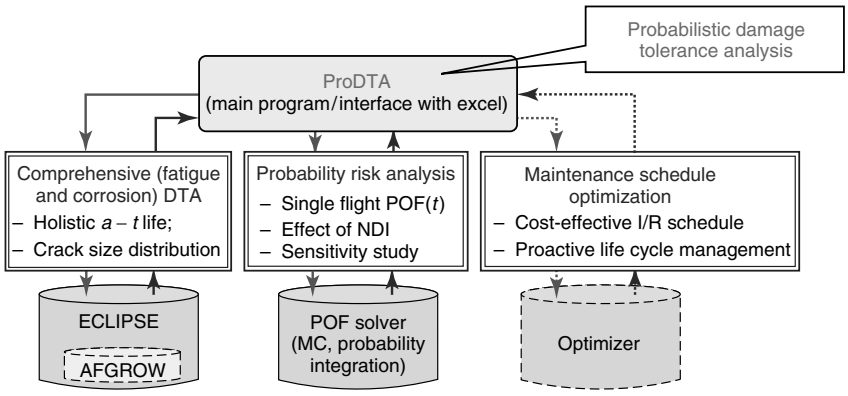


Figure 4. Schematic of probability of failure software [15]. POF, probability of failure. [Reproduced with permission from Ref. 15. © ICAF, 2005.]

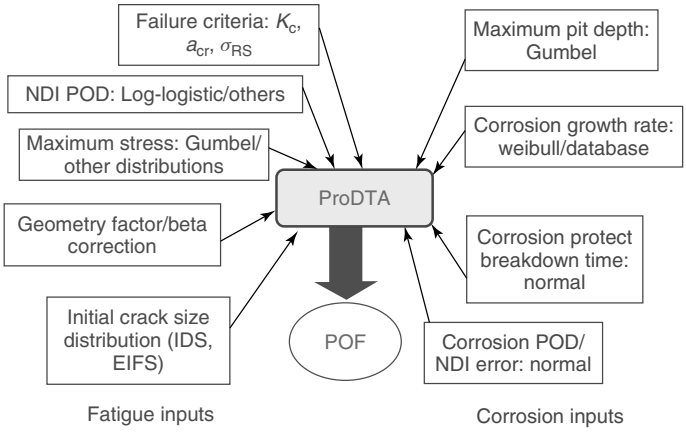


Figure 5. Input parameters required for age and environmental degradation risk assessment. EIFS, equivalent initial flaw size.

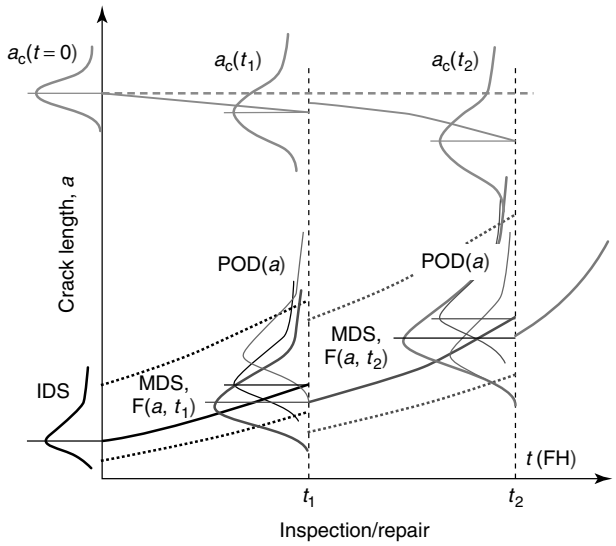


Figure 6. Analysis procedures employed in ProDTA.

first method is to project the crack size distribution based on a single master crack growth curve. The second method is to grow the crack size distribution using a Monte Carlo analysis, which allows ProDTA to use more random variables, especially age degradation parameters, such as corrosion growth rate, thickness loss, pit depth, and corrosion protection breakdown time [15]. When inspection/repair actions take place at a given time (t_1), the crack size distribution would be modified on the basis of a probability of detection function POD(a) and a repaired crack size distribution. The modified crack size distribution is grown again to inspection/repair time, t_2, and the above procedure is repeated till the end of the life cycle. For an analysis that includes age and environmental degradation, both crack size distribution and critical crack size distribution are varied with time t, and joint distribution functions of more age-degradation-related random variables, such as corrosion growth rates, pit depth, and corrosion protection breakdown time. Given the number of fatigue and environmental degradation random variables (more

than three), it is impossible to obtain a closed-form solution and, therefore, mixed techniques of Monte Carlo simulation and probability integration have been developed in ProDTA.

It needs to be emphasized that the accuracy of the risk assessment results is strongly linked to the quality of the data available. By placing the appropriate sensors at specific locations within an aircraft, the data required to carry out a risk assessment can be obtained, which can then be used within ProDTA. This data can include corrosion growth rates by determining the thickness loss versus time as well as the time required to breakdown the corrosion protection system, which is one of the main parameters that is very difficult to determine. As data is accumulated over time through the use of these sensors, statistical distributions for various parameters can be generated for a particular material system, which will allow for a combined mixed Monte Carlo simulation and probability integration to be carried out.

REFERENCES

[1] Miller D. Corrosion control on aging aircraft: what is being done. *Materials Performance* 1990 **29**:10–11.

[2] Wildey JF. Aging aircraft. *Materials Performance* 1990 **29**:80–85.

[3] Under Secretary of Defense (Acquisition, Technology and Logistics), *Status Update on Efforts to Reduce Corrosion and the Effects of Corrosion on the Military Equipment and Infrastructure of the Department of Defense*, www.dodcorrosionexchange.org (date accessed May 2005).

[4] Eastaugh GF, Straznicky PV, Krizan DV, Merati AA, Cook J. Experimental study of the effects of corrosion on the fatigue durability and crack growth characteristics of longitudinal fuselage splices. *Fourth Joint DoD/FAA/NASA Conference on Aging Aircraft*. St. Louis, MO, May 2000.

[5] Komorowski JP, Bellinger NC, Gould RW, Forsyth D, Eastaugh G. Research in corrosion of ageing transport aircraft structures at SMPL. *Canadian Aeronautics and Space Journal* 2001 **47**(3):289–299, Special Edition 50th Anniversary of IAR.

[6] Bellinger NC, Forsyth DS, Komorowski JP. Damage characterization of corroded 2024-t3 fuselage lap joints. *Proceedings of the Fifth Joint NASA/FAA/DoD Conference on Aging Aircraft*. Kissimmee, FL, 10–13 September 2001.

[7] Bruhn EF. *Analysis and Design of Flight Vehicle Structures*. Tri-State Offset: Cincinatti, OH, 1965.

[8] Hertzberg RW. *Deformation and Fracture Mechanics of Engineering Materials*. John Wiley & Sons: Toronto, 1976.

[9] Krebs LA. A brief history of corrosion sensing methods. *Conference on Corrosion 2003*. San Diego, CA, 16–20 March 2003; pp. 7.

[10] Braunling R, Dietrich P. Corrosion and corrosivity monitoring system. *Proceedings of SPIE on Smart Structures and Materials*. San Diego, CA, 2005; Vol. 5765.

[11] Giurgiutiu V, Zagrai A, Bao JJ, Remdond JM, Roach D, Rackow K. Active sensors for health monitoring of aging aerospace structures. *International Journal of the Condition Monitoring and Diagnostic Engineering Management* 2003 **6**:3–21.

[12] Brooks CL, Honeycutt K, Prost-Domasky S. Case studies for life assessments with age degradation. *Proceedings of the Fourth Joint DoD/FAA/NASA Conference on Aging Aircraft*. St. Louis, MO, 15–18 May 2000.

[13] Brooks CL, Prost-Domasky S, Honeycutt K. *Fatigue in the Presence of Corrosion*, RTO-MP-18. NATO: Brussels, 1998.

[14] Komorowski JP, Forsyth DS, Bellinger NC, Hoeppner DW. Life and damage monitoring-using NDI Data Interpretation for corrosion damage and remaining life assessments. Published in the *Proceedings of the RTO Specialist's Meeting on Life management for aging air Vehicles*, Paper No. 13. Manchester, 08–11 October 2001.

[15] Liao M, Forsyth DS, Bellinger NC. A new probabilistic damage tolerance analysis tool and its application for corrosion risk assessment. Published in the *Proceedings of the 23rd Symposium of the International Committee on Aeronautical Fatigue*. Hamburg, June 2005.

[16] Liao M, Bellinger NC, Komorowski JP, Rutledge R, Hiscocks R. Corrosion fatigue prediction using holistic life assessment methodology. Published in the *Proceedings of Fatigue 2002*. Stockholm, June 2002.

[17] Brooks CL, Prost-Domasky S, Honeycutt K. Correlation of life prediction methods with corrosion-related tests. *Proceeding of the 1999 USAF ASIP Conference*. San Antonio, TX, December 1999.

[18] Bellinger NC, Gould RW, Komorowski JP. Repair issues for corroded fuselage lap joints. *Journal of Aerospace—SAE Transactions* 1999 **108**, Section 1: 902–908.

Chapter 91

Use of Leave-in-place Sensors and SHM Methods to Improve Assessments of Aging Structures

Dennis Roach

Sandia National Laboratories, Albuquerque, NM, USA

 ISBN: 978-0-470-05822-0.

1 SMART STRUCTURES VERSUS NONDESTRUCTIVE INSPECTION

The costs associated with the increasing maintenance and surveillance needs of aging structures are rising. The application of distributed sensor systems can reduce these costs by allowing condition-based maintenance practices to be substituted for the current time-based maintenance approach. Through the use of *in situ* sensors, it is possible to quickly, routinely, and remotely monitor the integrity of a structure in service [1]. This requires the use of reliable structural health monitoring (SHM) systems that can automatically process data, assess structural condition, and signal the need for human intervention. Prevention of unexpected flaw growth and structural failure can be improved if onboard health monitoring systems could continuously assess structural integrity. Such systems would be able to detect incipient damage before catastrophic failures occur.

A "smart structure" is one that is sufficiently instrumented so that the data can be synthesized to form an accurate real-time picture of the state of the structure in all its critical aspects. In this case, the absence of disbonds and delaminations indicates that the doubler

is able to perform its duty. The absence of cracks indicates that the structure is able to continue to operate safely. The current state of nondestructive inspection (NDI) involves the manual application of visual, ultrasonic eddy current, X-ray, and penetrant NDI methods. While the data presented in this report indicates that manual inspections provide a reliable health monitoring approach, less labor intensive and more frequent structural assessments can be performed via distributed sensor systems. Such health monitoring systems utilize a network of leave-in-place sensors that can assess a structure on a frequent, or even continuous, basis.

Nondestructive inspection (NDI): The examination of a material to determine geometry, damage, or composition by using technology that does not affect its future usefulness.

- involves a high degree of human interaction
- local, focused inspections
- requires access to area of interest
- time-based monitoring—applied at predetermined intervals
- portable and applied to numerous areas.

Structural health monitoring (SHM): Also known as "Smart structures"; the use of NDI principles coupled with *in situ* sensing to allow for rapid, remote, and even real-time condition assessments; the goal is to reduce operational costs and increase life of structures.

- allows for greater vigilance in key areas—address damage tolerance needs;
- overcomes accessibility limitations, complex geometries, depth of hidden damage;
- eliminates costly and potentially damaging disassembly;
- minimizes human factors with automated sensor deployment and data analysis;
- supports adoption of condition-based maintenance.

2 USE AND ADVANTAGES OF *IN SITU* STRUCTURAL HEALTH MONITORING

Prevention of unexpected flaw growth and structural failure could be improved if onboard health monitoring systems that could continuously assess structural integrity exist [2–4]. Reliable, SHM systems can automatically process real-time data, assess structural condition, and signal the need for human intervention. Such systems would be able to detect incipient damage before catastrophic failures occur. The replacement of our present-day manual inspections with automatic health monitoring would substantially reduce the associated life-cycle costs. SHM systems using distributed sensor networks allow for condition-based maintenance practices to be substituted for the current time-based maintenance approach. Other advantages of onboard distributed sensor systems are that they can eliminate costly, and potentially damaging, disassembly, improve sensitivity by producing optimum placement of sensors with minimized human factor concerns in deployment, and decrease maintenance costs by eliminating more time-consuming manual inspections.

Whether the sensor network is hardwired to an accessible location within the structure or monitored in a remote, wireless fashion, the sensors can be interrogated easily and often even in a real-time mode. It is anticipated that the sensors will most likely be examined at discrete intervals, probably at normal maintenance checks. The important point to note is that the ease of monitoring an entire network of distributed sensors means that structural health assessments can occur more often, allowing operators to be even more vigilant with respect to flaw onset. Figure 1 depicts a sensor network deployed on an aircraft to monitor critical sites over the entire structure.

Multisite fatigue damage, hidden cracks in hard-to-reach locations, disbonded joints, erosion, impact, and corrosion are among the major flaws encountered in today's extensive array of aging structures and mechanical assemblies. Furthermore, the extreme damage tolerance and high strength-to-weight ratio of composites have motivated designers to expand the role of advanced materials in industrial structures. These developments, coupled with new and unexpected phenomena, have placed greater demands on the application of advanced NDI and health monitoring techniques. In addition, innovative deployment methods must be employed to overcome a myriad of inspection impediments stemming from accessibility limitations, complex geometries, and the location and depth of hidden damage. Recent requests for real-time

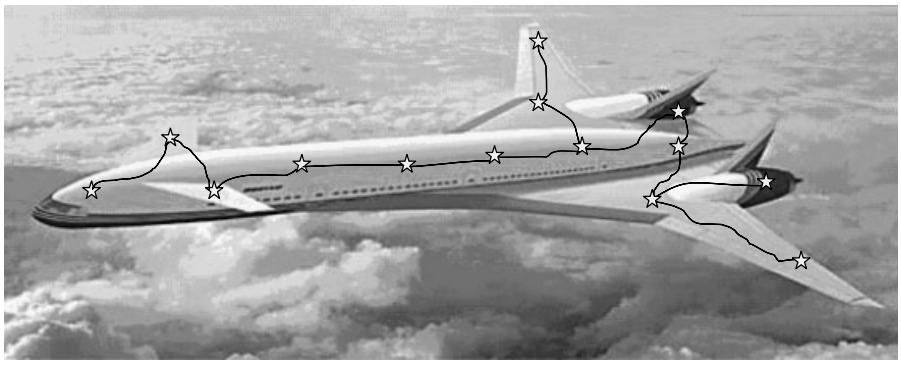


Figure 1. Depiction of a distributed network of sensors to monitor structural health.

monitoring of structures have produced a niche for active sensor systems.

In addition to mature microelectromechanical systems (MEMS) devices such as acoustic emission sensors, accelerometers, strain gauges, and pressure sensors, recent advances in microsensors have produced miniature eddy current (EC), ultrasonic, piezoelectric, fiber-optic (FO), and other devices that lend themselves more directly to damage detection. Technology exists to colocate the processing electronics with *in situ* sensor networks to produce real-time transmission of data and real-time diagnostics of structural health. When combined in a systems approach that includes sensors to monitor electronics, hydraulics, and avionics, it is possible to produce a prognostic health management (PHM) architecture that can assist in maintenance scheduling and tracking. This article focuses on developments and testing of mountable sensors and how they can be integrated into such a health management system. Specific example applications are discussed along with issues that must be addressed to realistically deploy leave-in-place sensors. Successful field testing is presented to quantify the performance of real-time health monitoring systems and to highlight their use in guiding condition-based maintenance activities.

The costs associated with the increasing maintenance and surveillance needs of our aging infrastructure are rising at an unexpected rate. The application of distributed sensor systems may reduce these costs by allowing condition-based maintenance practices to be substituted for the current time-based maintenance approach. In the near future, it may be possible to quickly, routinely, and remotely monitor the integrity of a structure in service. A series of expected maintenance functions will already be defined; however, they will be carried out only if their need is established by the health monitoring system [3, 4].

3 COMPARATIVE VACUUM MONITORING

Comparative vacuum monitoring (CVM) has been developed on the principle that a small volume maintained at a low vacuum is extremely sensitive to any ingress of air [5]. Figure 2 shows top-view and side-view schematics of the self-adhesive, elastomeric sensors with fine channels on the adhesive face along with a sensor being tested in a lap joint panel. When the sensors are adhered to the structure under test, the fine channels and the structure itself form a manifold of galleries alternately at low vacuum and atmospheric pressure. When a crack develops, it forms a leakage path between the atmospheric and vacuum galleries, producing a measurable change in the vacuum level. This change is detected by the CVM monitoring system shown in Figure 3. Embedded sensors may be formed using a load-bearing elastomer. This material is able to withstand the high loading stresses that result during the riveting process.

These sensors can be attached to a structure in areas where crack growth is known to occur. On a preestablished engineering interval, a reading is taken from an easily accessible point on the aircraft. Each time a reading is taken, the system performs a self-test. This inherent fail-safe property ensures the sensor is

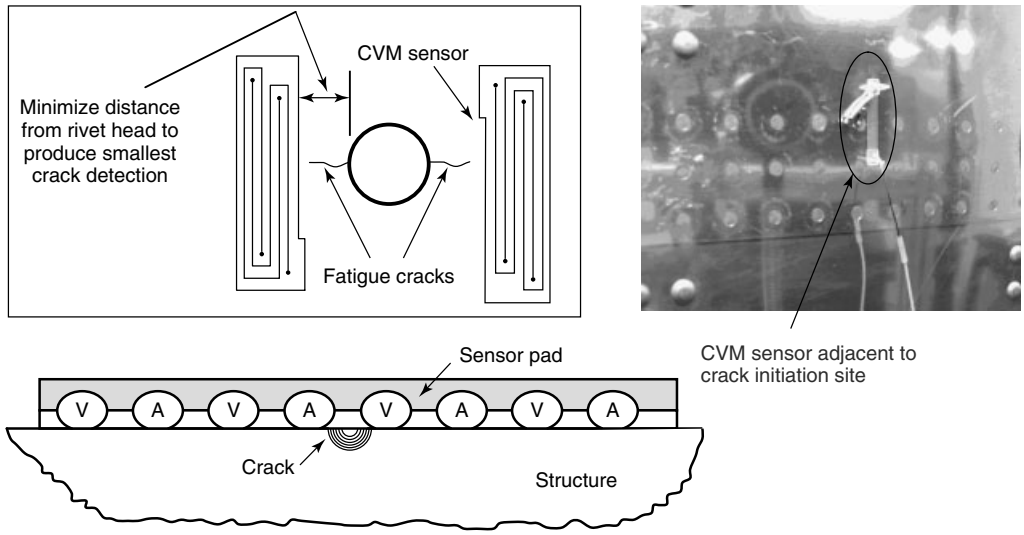


Figure 2. Schematics depicting operation of CVM sensor and polymer sensor mounted on outer surface of a riveted lap joint.

attached to the structure and working properly. Since the sensor physics is based on pressure measurements, there is no electrical excitation involved. This can be important in areas where electrical signals can create interference (near avionics) or where electrical connections may pose a hazard (fuel tanks).

Figure 3 also shows sample CVM sensors mounted on an aircraft structure as part of a performance validation effort. A series of 26 sensors have been mounted on the structure in four different DC-9, 757, and 767 aircraft of the Northwest Airlines and Delta Air Lines fleet. Some of the sensors were installed over two years ago. Periodic testing was used to study the long-term operation of the sensors in actual operating environments. This environmental durability study complements the laboratory flaw detection testing described below as part of an overall CVM certification effort.

Sandia Labs, in conjunction with Boeing, Northwest Airlines, Delta Airlines, Structural Monitoring Systems, the University of Arizona, and the Federal Aviation Administration (FAA), completed validation testing on the CVM system in an effort to adopt CVM as a standard NDI practice [5, 6]. Fatigue tests were completed on simulated aircraft panels to grow cracks in riveted specimens (see Figure 2) while the vacuum pressure within the various sensor galleries was simultaneously recorded. The fatigue crack was propagated until it engaged, and fractured, one of the vacuum galleries such that crack detection was achieved (sensor indicates the presence of a crack by its inability to maintain a vacuum). In order to properly consider the effects of crack closure in an unloaded condition (i.e., during sensor monitoring), a crack was deemed to be detected when a permanent alarm was produced and the CVM sensor did not maintain a vacuum even if the fatigue stress was reduced to zero.

3.1 CVM validation—data analysis using one-sided tolerance intervals

The CVM sensor is based on the principle that a steady-state vacuum, maintained within a small volume, is sensitive to any leakage. A crack in the material beneath the sensor allows leakage, resulting in detection. The data analyzed here consists of cracks that were fatigue cycled in various metal specimens with the direction of growth aligned with the CVM mounted sensors. The data captured is that of the flaw length at the time for which the CVM provided sustainable detection. Thus, the specific flaw would not be considered detectable prior to reaching the stated length, but is considered to continue to be detectable upon further growth. With these assumptions, there exists a distribution on the flaw lengths at which detection is first made. In this context,

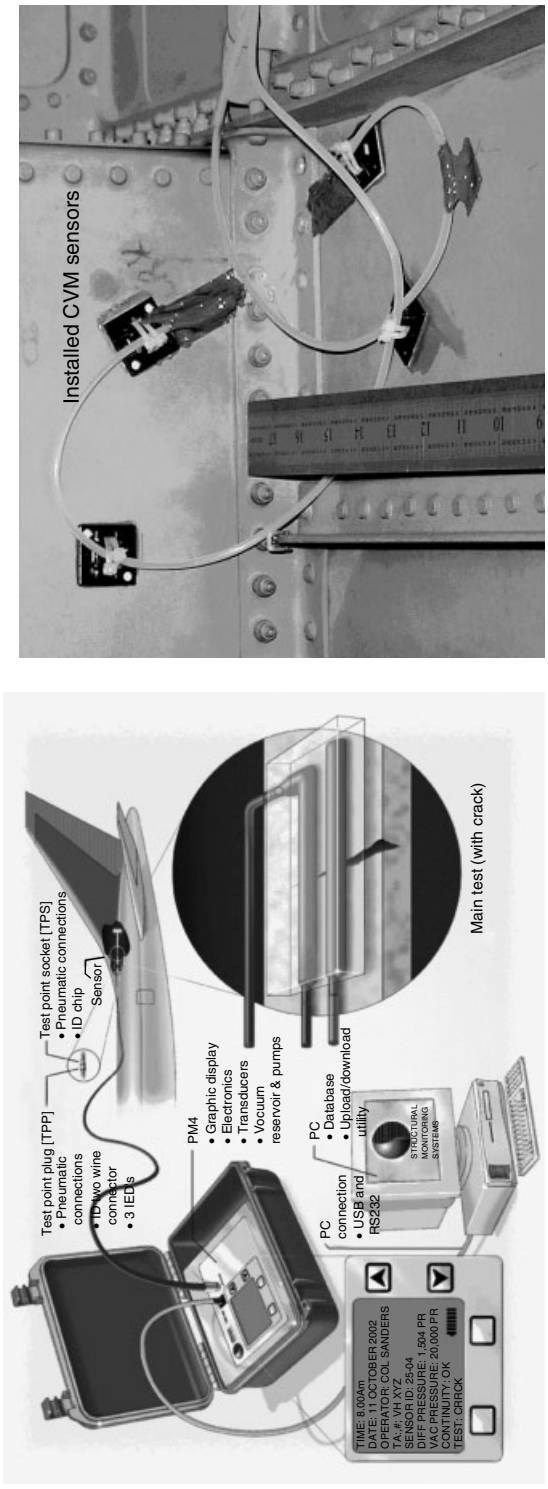


Figure 3. Schematic showing hardware set-up for crack detection via CVM system and sample aircraft test installations of sensors.

the probability of detection (POD) for a given flaw length is just the proportion of the flaws that have a detectable length less than that given length. That is, the reliability analysis becomes one of characterizing the distribution of flaw lengths, and the cumulative distribution function is analogous to a POD curve. Assuming that the distribution of flaws is such that the logarithm of the lengths has a Gaussian distribution, we calculate a one-sided tolerance bound for various percentile flaw sizes. To do this it is necessary to find factors $K_{n,\gamma,\alpha}$ such that the probability γ is such that at least a proportion $(1-\alpha)$ of the distribution will be less than $X - K_{n,\gamma,\alpha}$, where X and S are estimators of the mean and the standard deviation computed from a random sample of size n. The data captured is the crack length at CVM detection. From the reliability analysis a cumulative distribution function is produced to provide the maximum likelihood estimation (POD). This stems from the one-sided tolerance bound for the flaw of interest using the equation:

$$POD_{95\%\ \text{confidence}} = \mathrm{X} + (\mathrm{K}_{n,0.95,\alpha})(\mathrm{S}) \qquad (1)$$

where X: mean of detection lengths; K: probability factor (~sample size, confidence level desired); S: standard deviation of detection lengths; n: sample size; and $1-\alpha$; detection level.

3.2 CVM performance testing on thin aluminum structures

This test program produced a statistically relevant set of crack detection levels for 0.040″- and 0.100″-thick panels in both the bare and primed configurations. Figure 4 shows the fatigue test setup used to grow cracks and a close-up photo of the CVM sensors monitoring cracks initiating from a center hole. Figure 5 shows a photo of a fatigue crack as it engages the first vacuum gallery of a CVM sensor. The pressure rise, corresponding to a rupture in the gallery and a leakage path to atmospheric pressure, is shown on the right side of Figure 5. The large increase in the pressure corresponds to crack detection. In actual field measurements, the plot shown in Figure 6 would be produced. One signal (lower curve) corresponds to vacuum levels produced when there is no crack indication and the other signal (upper curve) occurs when a vacuum is not achievable. This latter signal is produced when the CVM detects a crack. Such a curve with a pressure level of at least 300 Pa must be produced when the structure is unloaded.

Results to date have revealed crack detection lengths—permanent alarm after the fatigue crack engages the CVM sensor—for the bare and primed 0.040″-thick panels. Tables 1 and 2 summarize some

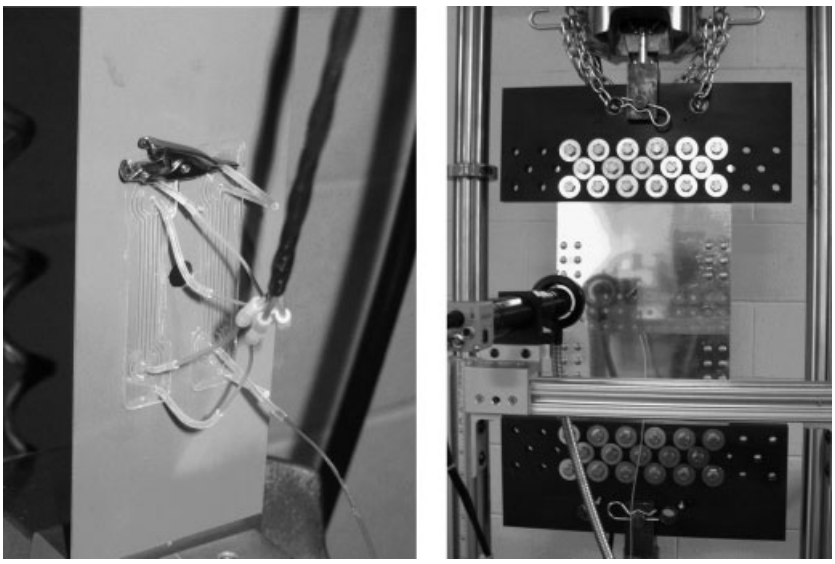


Figure 4. CVM sensors monitoring crack growth on aluminum test specimens.

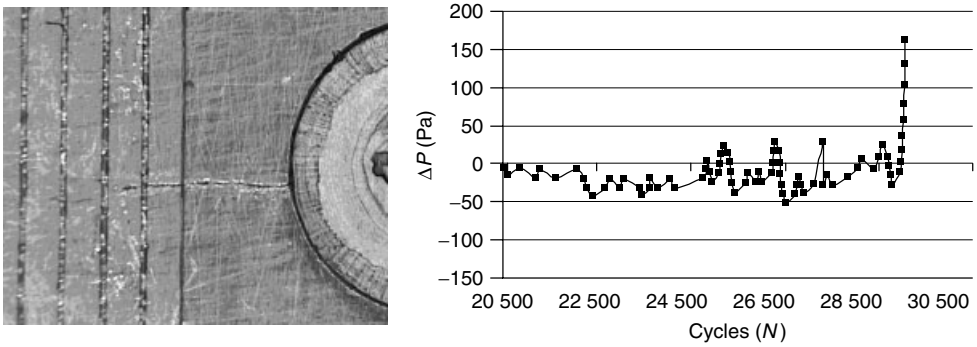


Figure 5. Fatigue crack crossing into CVM galleries; differential pressure shown as a function of cycle number (pressure increase caused by crack reaching a gallery is clearly indicated).

of the results. Table 2 lists CVM crack detection results for 0.040″-thick aluminum plate in both a coated (primer) and an uncoated condition. Crack detection lengths ranged from 0.002″ to 0.020″. Fatigue tests have shown that pressure levels in excess of 300 Pa were measured during fatigue testing; however, the compressive residual stresses at the tip of a fatigue crack could allow a vacuum to be produced when the specimen was unloaded. The numbers presented in Tables 1 and 2 correspond to *permanent* alarm levels for cracks engaging CVM sensors and the structure in an unloaded condition.

The 90% POD level for crack detection on 0.1″-thick aluminum—calculated from equation (1)—is also listed in Table 1. Owing to the limited number of data points, the reliability calculations induce a penalty by increasing the magnitude of the K (probability) factor. As a result, the overall POD value (95% confidence level) for CVM crack detection in 0.100″-thick aluminum skin is 0.023″. This POD curve is plotted in Figure 7. As the number of data points increases, the K value will decrease and the POD numbers could also decrease. In this particular instance, it was desired to achieve crack detection before the crack reached 0.100″ in length so that this goal was achieved. Table 3 summarizes the 90% POD levels (95% confidence level) for CVM crack detection for the array of thin-walled aluminum plates tested.

3.3 CVM performance on thick steel structures

The results cited above are valuable for thin-walled structures such as those used in aircraft, automotive, and some pipeline construction. However, many civil structures use thick steel members. Earlier studies revealed that the thickness of the plate can affect CVM performance, so a second round of tests included CVM crack detection in thick-walled structures. It should be noted that aircraft use thinner materials and have crack detection requirements of 0.050–0.100″ in length. Civil structures contain thicker materials and have higher safety factors. Thus, these structures can tolerate longer cracks and their crack detection requirements are in the range of 0.5–1″ in length. CVM sensors can be fabricated with different gallery sizes in order to accommodate various sensitivity requirements.

Figure 8 shows the installation of a CVM sensor on a 0.375″-thick steel (ASTM 572) plate. The seeded fatigue crack along the edge of the specimen is visible. These test specimens were then exposed to tension–tension fatigue tests in order to propagate the crack into the CVM sensor. Figures 9 and 10 show the overall test setup along with the equipment used to monitor the CVM sensors.

Compressive stresses around the tip of a fatigue crack create a tight tip when the load is removed. As a result, the initial engagement of a crack with a CVM sensor may induce a high-pressure reading (crack detection) when the structure is under load; however, the compressive residual stresses at the tip of a fatigue crack could allow a vacuum to be produced when the specimen was unloaded. Therefore, crack detection can be achieved much earlier if the sensors can be monitored while the structure is in use. In the case of real-time monitoring for the steel plate test series, CVM crack detection results for the unloaded and

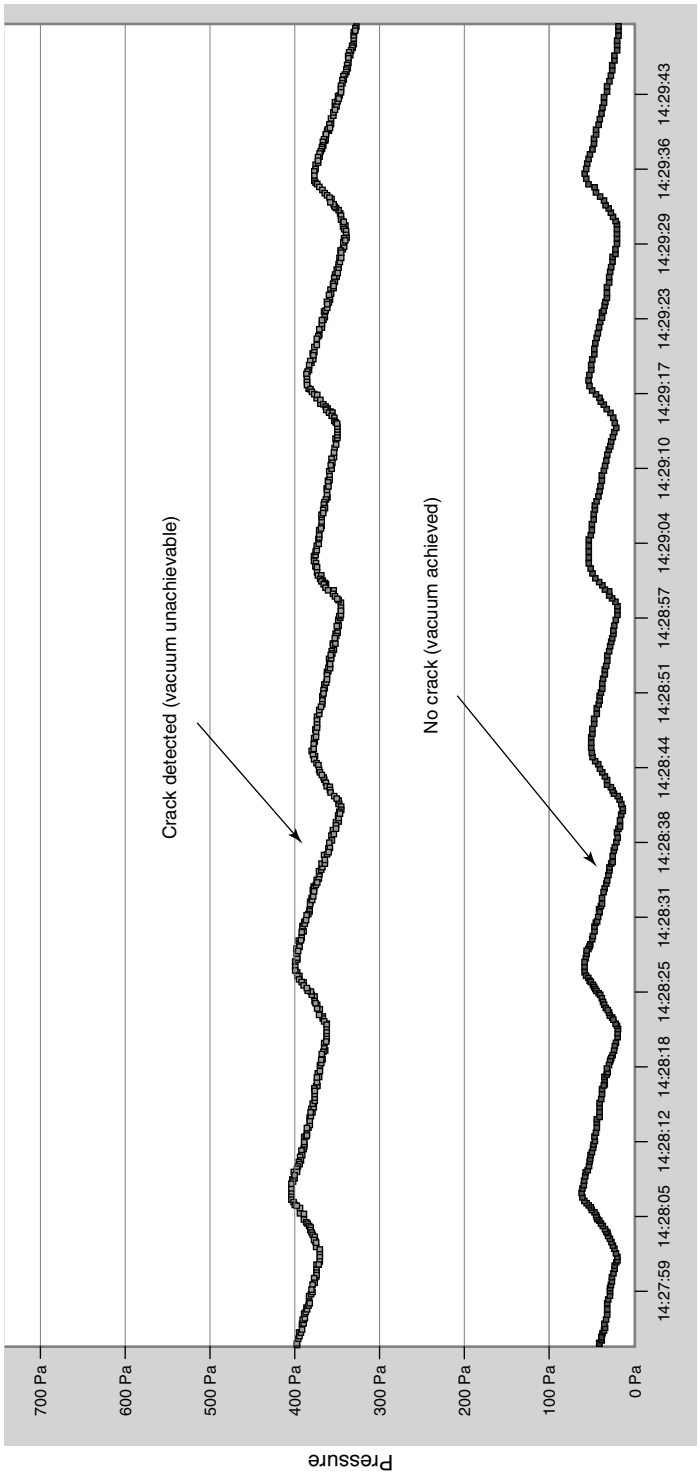


Figure 6. Typical pressure versus time plots from CVM sensors indicating "no crack" and "crack detection".

Table 1. CVM crack detection values from 0.100″ aluminum plate

Panel	Fastener crack site	Number of fatigue cycles	Crack length at CVM detection (growth after installation in inches)	PM-4 read -out (pasm)	PM-4 indicate crack (Y or N)	90% POD level	False calls
1	1L	3505	0.007	2123	Y	0.023″	0
1	1R	3205	0.007	1938	Y		
1	2L	5350	0.010	2251	Y		
1	2R	5550	0.011	1954	Y		
1	3L	6650	0.009	4526	Y		
1	3R	7099	0.016	7099	Y		
2	1L	3100	0.011	1786	Y		
2	1R	3400	0.014	1707	Y		
2	2L	5300	0.005	2383	Y		
2	2R	5300	0.016	2204	Y		
3	1L	4475	0.019	1790	Y		
3	1R	4825	0.013	1904	Y		
3	2L	7025	0.008	2100	Y		
3	2R	7878	0.010	4302	Y		

Table 2. Cracks lengths detected by CVM sensors on 0.040″-thick skins

Unpainted 0.040″-thick skin			0.040″-skin with primer coating		
Panel number	Fastener crack site	Crack length at CVM detection (growth after installation in inches)	Panel number	Fastener crack site	Crack length at CVM detection (growth after installation in inches)
4017	8R	0.003	4018	5R	0.002
4017	6R	0.030	4018	6R	0.007
4017	5R	0.007	4018	7R	0.010
4017	7R	0.002	4018	5R(2)	0.009
4011	7R	0.009	4018	6L	0.005
4011	7L	0.005			
4014	7R	0.004			
4015	7L	0.002			

loaded steel structure are summarized in Tables 4 and 5, respectively.

For the loaded structure, CVM crack detection occurred when the fatigue cracks ranged from 0.040″ to 0.070″ in length (Table 5). This would correspond to the ability of the CVM sensor to monitor cracks in real time while the structure is in use. For the unloaded condition, CVM crack detection occurred when the fatigue cracks ranged from 0.060 to 0.380″ in length (Table 4). Note that the data spread is much larger for this condition owing to the varying state of the compressive stresses at the end of the fatigue crack. However, regardless of whether the sensor monitoring is completed during a loaded or unloaded condition, the results indicate that CVM sensors could reliably detect fatigue cracks well before they reach 0.5″ in length.

4 PIEZOELECTRIC TRANSDUCERS (PZT)

Prime candidates for sensors based on active-material principles utilize thin piezoelectric wafers

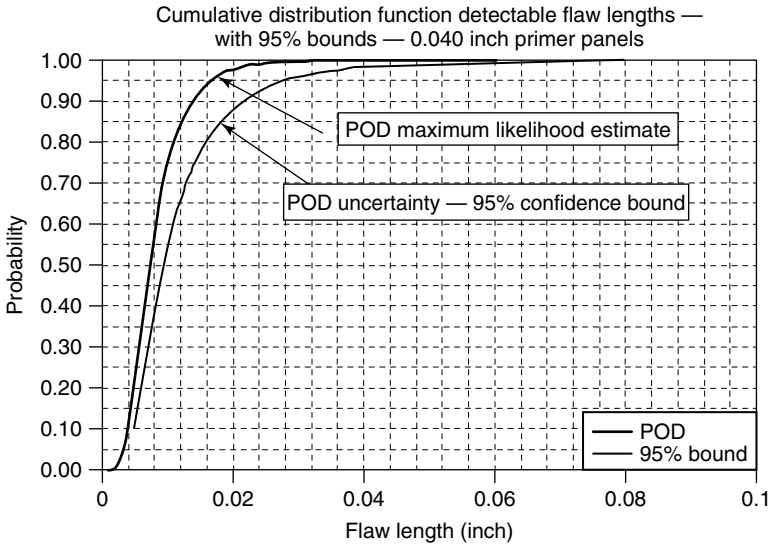


Figure 7. Typical probability of crack detection curves generated by CVM data and data analysis using one-sided tolerance intervals.

Table 3. Summary of raw crack POD levels for CVM deployed on different materials, surface coatings, and plate thicknesses

Material	Thickness	Coating	90% POD for crack detection
2024-T3	0.040″	Bare	0.049″
2024-T3	0.040″	Primer	0.021″
2024-T3	0.071″	Primer	0.042″
2024-T3	0.100″	Bare	0.272″
2024-T3	0.100″	Primer	0.090″
7075-T6	0.040″	Primer	0.026″
7075-T6	0.071″	Primer	0.033″
7075-T6	0.100″	Primer	0.023″

of 0.125″–0.25″ diameter with thicknesses of 0.010–0.030″. They can be easily attached to existing aging structures without changing the local and global structural dynamics. Piezoelectric transducers (PZT) sensors can also be embedded inside composite structures to closely monitor for internal flaws. These sensors can act as both transmitters and receptors. As transmitters, piezoelectric sensors generate elastic waves in the surrounding material. As receptors, they receive elastic waves and transform them into electric signals. It is conceivable to imagine arrays of active sensors, in which each element would take, in turn, the role of transmitter and receptor, and thus scan large structural areas using ultrasonic waves [7]. The structural interrogation strategies using active piezoelectric sensors are twofold:

1. For local area detection, the electromechanical (E/M) impedance method is applied to detect changes in the point-wise structural impedance resulting from the presence and propagation of structural damage.
2. For large-area detection, wave propagation techniques using Lamb and Love wave methods are used to identify zones in the monitored area that have undergone changes in their structural integrity.

In the high-frequency E/M impedance approach, pattern recognition methods are used to compare impedance signatures taken at various time intervals and to identify damage presence and progression from the change in these signatures. In the Lamb/Love wave approach, the acousto-ultrasonic methods identifying changes in transmission velocity, phase, and additional reflections generated from the damage site are used. Both approaches can benefit from the addition of artificial intelligence neural network algorithms that can extract damage features based on a learning process.

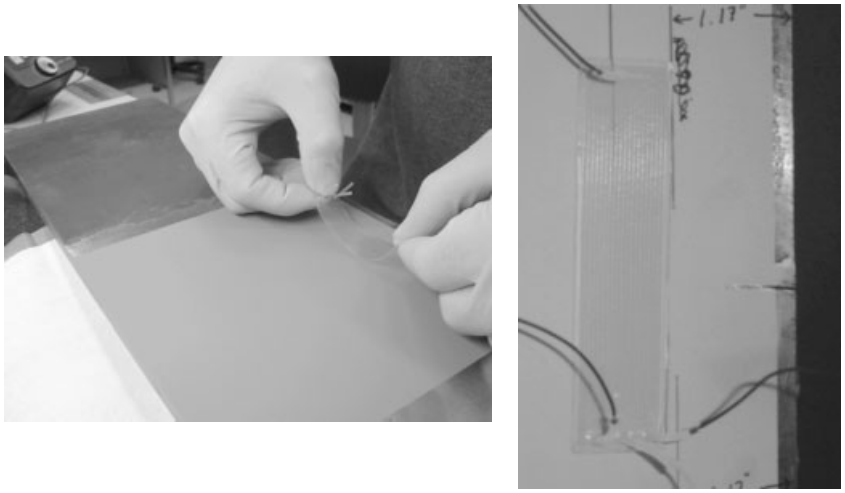


Figure 8. Installation of CVM sensor on primed steel surface and close-up of fatigue crack approaching sensor.

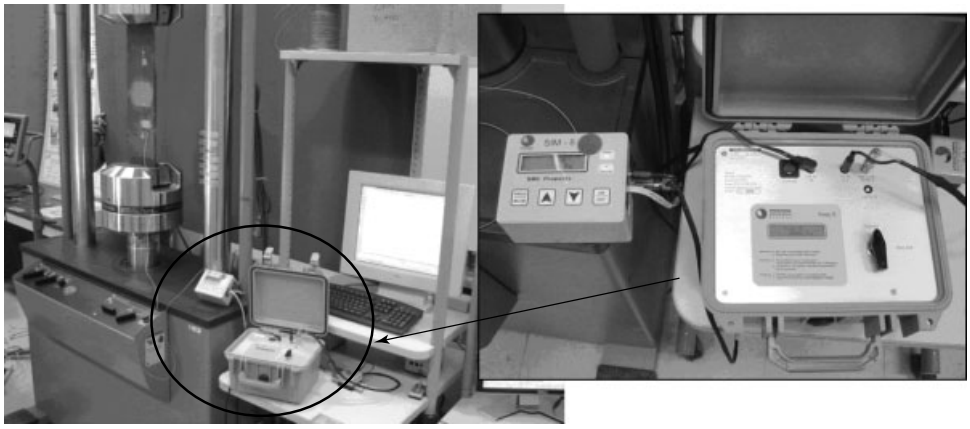


Figure 9. Overall setup for monitoring crack growth with CVM sensor system and close-up of sensor interrogation equipment.

Mountable PZT networks and lamb wave interrogation methods

This SHM approach uses a built-in network of PZT embedded in a thick dielectric carrier film. The SHM system included the PZT network connected to portable, diagnostic hardware and software developed by Acellent Technologies, Inc. The system performs *in situ* monitoring, data collection, signal processing, and real-time data interpretation to produce a two-dimensional image of the structure being interrogated. The Acellent software instructs the actuators to generate preselected diagnostic signals and transmit them to neighboring sensors. Multiple diagnostic wave types can be generated including 3-peak, 5-peak, and 10-peak narrowband frequency waveforms and chirp, random, and user-defined excitations. The software links each sensor with its neighbors to form a web, or network, covering the structure. The system then collects the total set of responses from each of the sensor sets as each PZT takes its turn as the actuator. Changes in the Lamb waves generated within the structure are used in concert with triangulation methods to detect the presence of structural anomalies and to determine the size and location of the flaws.

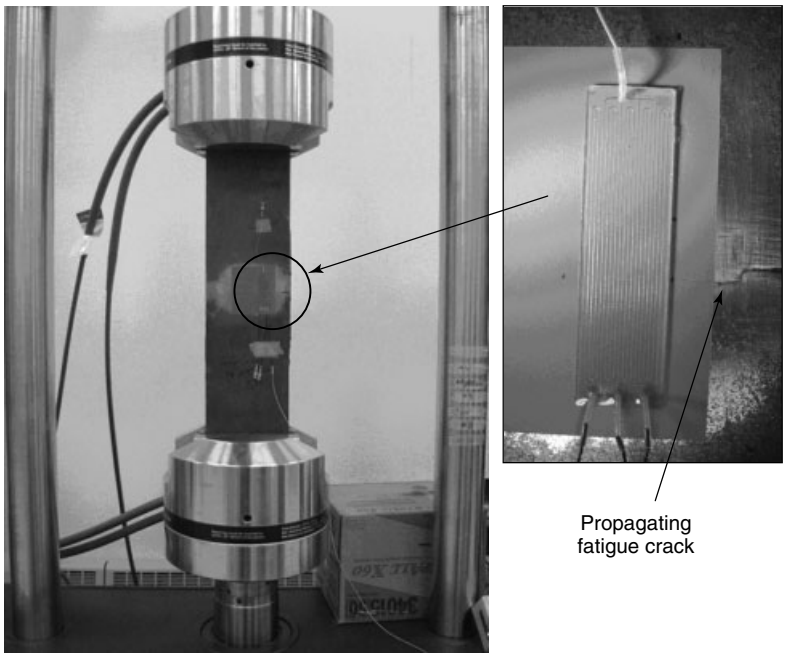


Figure 10. Fatigue test of steel specimen to propagate crack into CVM sensor (inset).

Damage identification through elastic wave propagation

The wave propagation approach uses the pitch–catch method for detecting damage in a structure. Acousto-ultrasonic methods are used to identify changes in wave transmission. Figure 11 shows some of the wave motion from sensors (1) and (9) when they are used as the source of excitation for the structure. The mechanical vibration is introduced into the structure by the PZT element and travels by wave motion through the test piece at the velocity of sound, which depends on the material. If the pulses encounter a reflecting surface, some or all of the energy is reflected and monitored by adjacent PZT sensors in the network. The reflected beam, or echo, can be created by any normal (e.g., in multilayered structures) or abnormal (flaw) interface. Figure 11 highlights the interaction of the ultrasonic testing (UT) waves with a flaw within the structure. The degree of reflection depends largely on the physical state of the materials forming the interface. Cracks, delaminations, shrinkage cavities, pores, disbonds, and other discontinuities that produce reflective interfaces can be detected. Complete reflection, partial reflection, scattering, or other detectable effects on the ultrasonic waves can be used as the basis for flaw detection.

Validation testing of PZT sensor network

In the first test series, PZT were built by embedding PZT materials in a thin dielectric carrier film. The spacing of the active PZT elements and the shape of the film were determined by the need to monitor local and global damage in the composite laminate and steel substructure. The PZT sensor network was laid out in custom configurations to detect damage in critical regions and provide an image of the structure in the area of the sensor network. The network of PZT sensors was deployed to assess bonded joints and crack growth in a composite doubler repair installation. Figures 12 and 13 show schematics and photos of the boron–epoxy laminate repair on a metal parent structure along with the set of PZTs distributed over the structure to be monitored. It is to be noted that the network of sensors/actuators is embedded in a custom polyamide film to allow for accurate placement of the network and eliminating the need for each sensor to be installed individually. The test specimen, containing engineered disbonds and a central crack, was subjected to constant-amplitude

Table 4. Permanent (no load) crack detection produced by CVM sensors on steel plate

CVM setup					CVM crack detection with no load			
Test specimen	Sensor	Initial crack length (in.)	Initial sensor location (distance from specimen edge) (in.)	Baseline CVM pressure reading (no crack engagement condition) (Pa)	Cycles at permanent CVM crack detection (no load)	CVM pressure reading at crack detection (no load) (Pa)	Total crack length at permanent CVM crack detection (no load) (in.)	Crack growth for CVM crack detection (engagement with CVM sensor) (in.)
SYN FAT 24	1	1.10	1.10	1580	7626	5300	1.48	0.380
SYN FAT 24	2	1.48	1.52	1435	9797	10 710	1.60	0.080
SYN FAT 24	3	1.90	1.92	1460	10 768	19 693	2.02	0.100
SYN FAT 19	4	1.17	1.22	1488	135 000	2900	1.54	0.315
SYN FAT 19	5	1.55	1.63	1500	143 358	2300	1.81	0.175
SYN FAT 19	6	1.81	1.87	1500	146 000	4950	1.93	0.060
SYN FAT 22	7	0.94	1.15	1740	180 000	2580	1.48	0.330
SYN FAT 22	8	1.48	1.53	1363	188 500	2580	1.70	0.170
SYN FAT 22	9	1.70	1.76	1530	192 000	3427	1.83	0.075
SYN FAT 21	10	1.00	1.09	1510	84 000	3000	1.50	0.410
SYN FAT 21	11	1.50	1.53	1433	91 500	2500	1.81	0.275
SYN FAT 23	12	1.45	1.50	1457	5000	2500	1.81	0.310
SYN FAT 23	13	1.81	1.84	1570	8500	2400	1.98	0.135

Table 5. Initial (under load) crack detection produced by CVM sensors on steel plate

CVM setup					CVM crack detection with no load			
Test specimen	Sensor	Initial crack length (in.)	Initial sensor location (distance from specimen edge) (in.)	Baseline CVM pressure reading (No crack engagement condition) (Pa)	Cycles at initial CVM crack detection (under load)	CVM pressure reading at crack detection (under Load) (Pa)	Total crack length at initial CVM crack detection (under load) (in.)	Crack growth for CVM crack detection (engagement with CVM sensor) (in.)
SYN FAT 24	1	1.10	1.10	1580	2137	16 500	1.15	0.050
SYN FAT 24	2	1.48	1.52	1435	9451	19 600	1.57	0.050
SYN FAT 24	3	1.90	1.92	1460	10 698	12 250	1.99	0.070
SYN FAT 19	4	1.17	1.22	1488	115 000	11 610	1.29	0.065
SYN FAT 19	5	1.55	1.63	1500	139 843	17 000	1.68	0.050
SYN FAT 19	6	1.81	1.87	1500	145 800	19 000	1.92	0.045
SYN FAT 22	7	0.94	1.15	1740	150 839	7000	1.20	0.050
SYN FAT 22	8	1.48	1.53	1363	184 412	17 800	1.59	0.060
SYN FAT 22	9	1.70	1.76	1530	191 315	17 000	1.80	0.045
SYN FAT 21	10	1.00	1.09	1510	44 800	3000	1.15	0.060
SYN FAT 21	11	1.50	1.53	1433	88 100	19 000	1.60	0.070
SYN FAT 23	12	1.45	1.50	1457	2000	11 000	1.56	0.060
SYN FAT 23	13	1.81	1.84	1570	6400	20 000	1.88	0.040

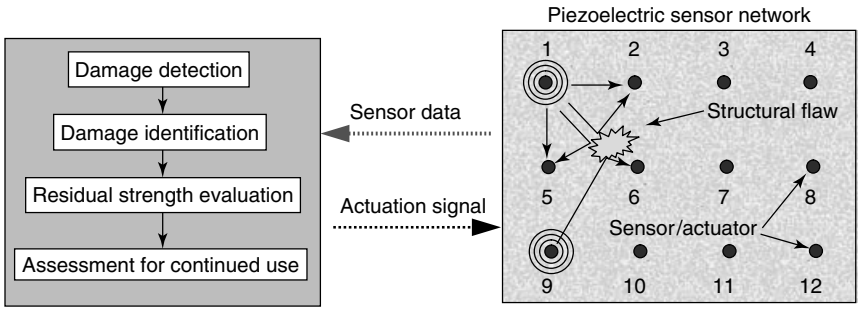


Figure 11. Flaw detection using the wave propagation method.

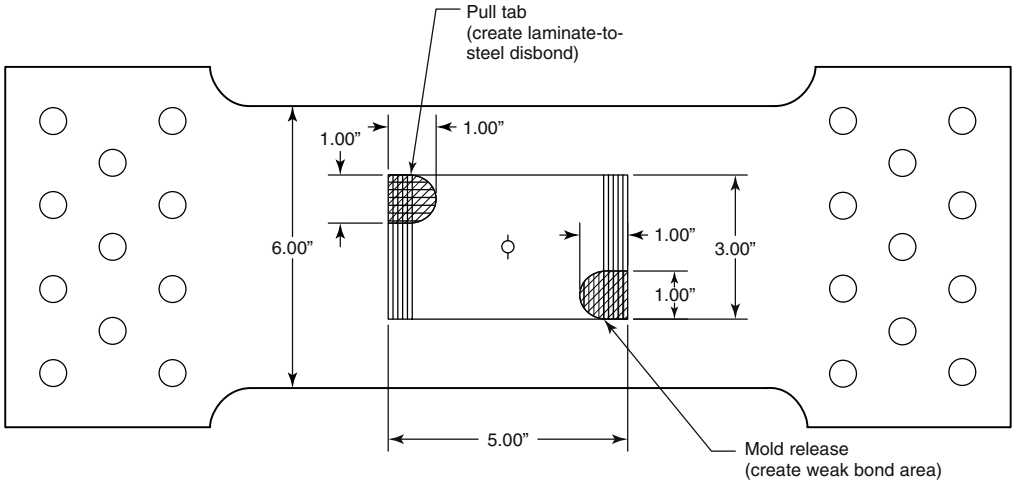


Figure 12. Composite doubler repair test coupon with disbond and fatigue crack flaws for evaluation of PZT and fiber-optic sensors.

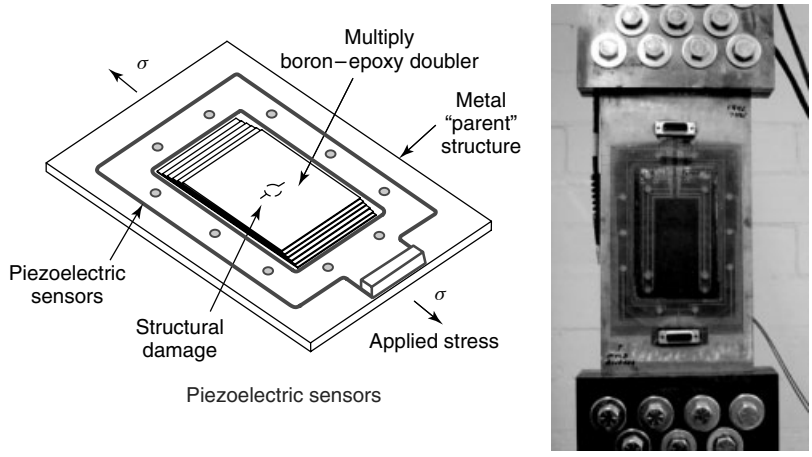


Figure 13. Set of piezoelectric sensors used to monitor crack growth and disbonds in a composite doubler bonded to a metal plate.

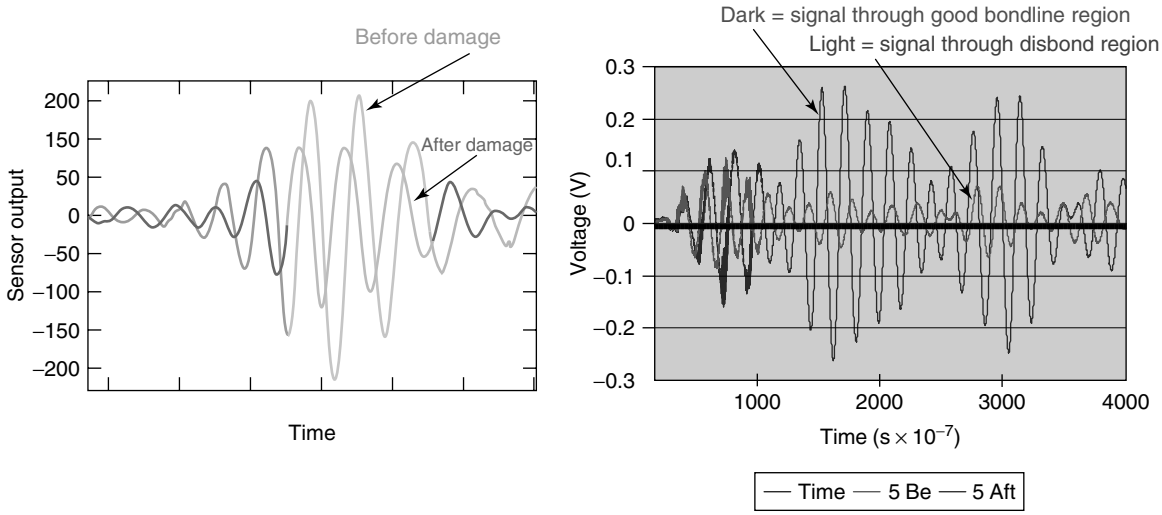


Figure 14. Sample signals observed by PZTs during 50-kHz Lamb wave interrogation showing the attenuation corresponding to disbonds in the structure.

fatigue loads with maximum stresses in excess of 80% of yield levels for the ASTM A36 steel plate (thickness = 0.188″). A time-varying electrical signal was input to the actuators/sensors. This caused a propagating stress wave to emanate from the actuator and travel through the material for detection by the neighboring sensors. These signals were then compared with previously recorded baseline test signals to identify the location and extent of damage or other structural anomaly.

Similar to conventional UT, the PZT data analysis can include one or more of the following measurements: time of wave transit (or delay), path length, frequency, phase angle, amplitude, and angle of wave deflection (reflection and refraction). In this test series, the pitch–catch method was used to study the transmission of sound waves as they traveled from each actuator to all other receiving sensors. The sum total of received beams was then analyzed to define the presence and location of flaws. In order to optimize flaw detection, a series of excitation frequencies were used: 50, 200, 350, and 500 kHz. Overall test results revealed that disbond flaws were most strongly detected with the lower, 50 kHz, excitation while the crack growth was monitored best with the highest, 500 kHz, excitation. Figure 14 shows raw PZT response data produced during the Lamb wave interrogation method. Signal attenuations, corresponding to disbonds between the laminate and parent skin, are apparent. When all of the signals are analyzed with the Acellent imaging software and flaw locations are determined by using the time base and triangulation methods, a two-dimensional image of the disbond flaws was produced. Figure 15 shows the engineered disbonds in the test specimen along with the image produced by the PZT sensor network. It is to be noted that both disbond flaws were clearly imaged even though one is a weak bond produced by a mold release agent and the other is a complete disbond produced by a Teflon insert.

Crack detection was achieved and crack growth was monitored using the same approach. PZT data was acquired at discrete intervals during the crack growth process. In addition, EC and microscopic inspections were conducted to measure the crack lengths at each cycle count. Figure 16 shows PZT response signals before and after crack growth occurred in the sensor path. A set of images produced by the PZT network are shown in Figure 17. The crack growth (two fatigue cracks emanating from a central hole) can be clearly seen. The PZT crack growth data was analyzed further to produce crack length predictions. The data analysis software contains an algorithm that allows for system learning. After inputting several crack lengths to match with

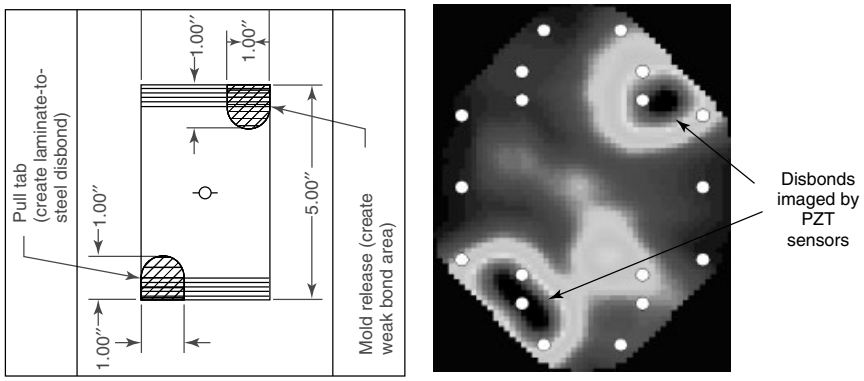


Figure 15. Shaded image of disbond flaws produced by the PZT sensor network.

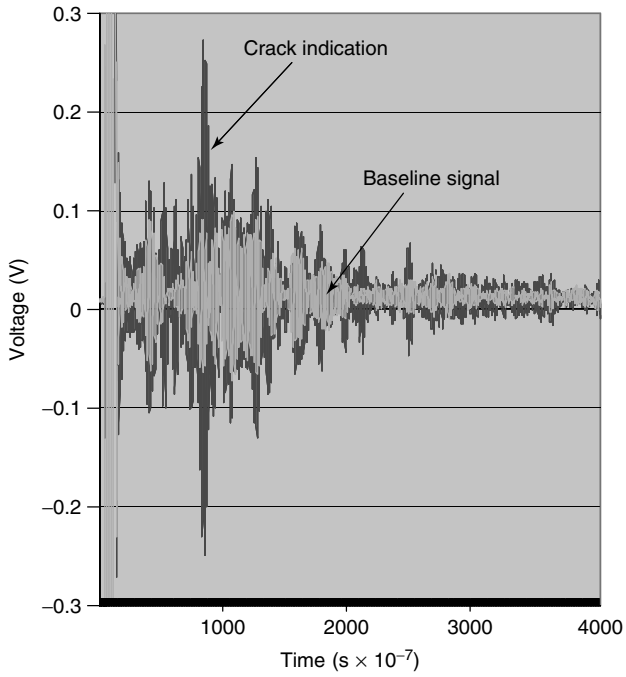


Figure 16. Sample PZT signals showing the indication of a fatigue crack with a 500-kHz excitation.

the PZT data at discrete fatigue intervals, it was possible for the system to predict all subsequent crack lengths using the PZT data alone. Table 6 compares the crack lengths predicted by the PZT sensor network with the crack lengths determined from EC and microscopic measurements. The PZT predictions were all within 5% of the actual crack lengths for data taken at maximum load (34 kips) and, for the most part, within 10% of actual values for PZT data taken in the unloaded condition.

4.1 PZT evaluations on steel specimens with edge cracks

Similar tests were conducted using a custom array of PZT sensors to monitor a 0.375″-thick ASTM 572 steel plate. The test specimen and PZT network are shown in Figures 18 and 19. This specimen is the same as the general composite doubler performance specimens described earlier. While assessing the crack mitigation capabilities and durability of the

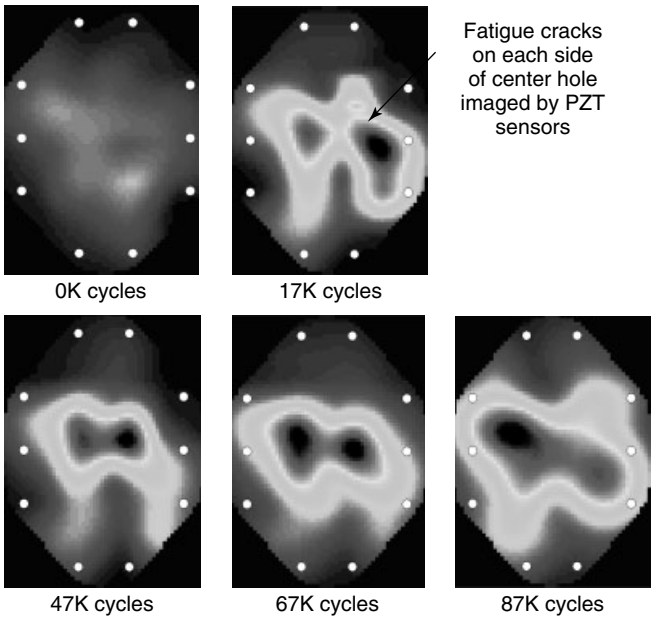


Figure 17. Shaded PZT images showing crack growth.

Table 6. Comparison of crack lengths predicted by PZT sensors with actual crack lengths measured using eddy current and microscopic methods

Composite doubler with PZT health monitoring			
Fatigue cycles	Measured total crack length	Estimated crack length from PZT sensor data (0 lbs load)	Estimated crack length from PZT sensor data (34 kips load)
Specimen 1—unflawed composite doubler			
0	0.00		
26 218	0.32	PZT learning data	PZT learning data
47 000	0.70	PZT learning data	PZT learning data
67 000	1.50	1.274	1.385
87 000	2.44	1.956	2.367
Specimen 2—composite doubler with disbond flaws			
0	0.00		
19 252	0.16	PZT learning data	PZT learning data
29 274	0.32	PZT learning data	PZT learning data
38 064	0.48	PZT learning data	PZT learning data
51 576	0.80	PZT learning data	PZT learning data
60 438	1.08	0.981	1.099
66 439	1.34	1.35	1.349
76 444	1.76	1.567	1.762
82 446	2.02	1.909	2.08

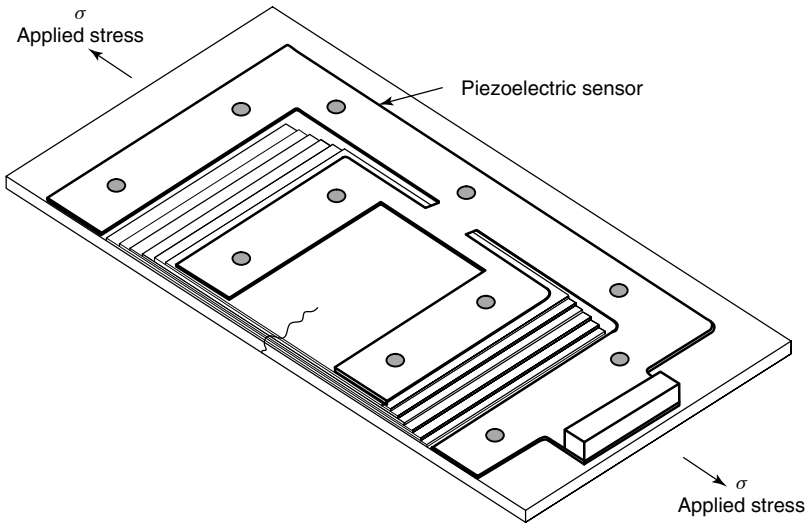


Figure 18. General layout of PZT network over and around composite doubler repair for edge doubler test series 2.

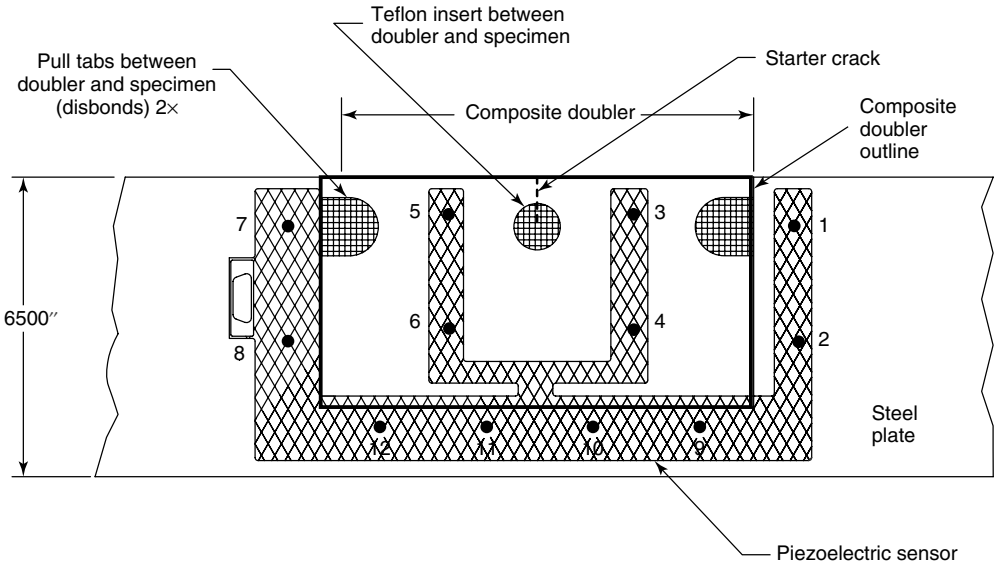


Figure 19. Layout of PZT sensor network relative to composite doubler and engineered flaws.

composite repair, it was possible to determine the crack and disbond detection capability of the PZT system. The three engineered disbond flaws and the center fatigue crack placed in the specimen prior to repair are indicated in Figures 18 and 19. Disbond flaws were placed in each end of the patch-to-steel bondline at the critical load transfer region, and a third disbond was placed over the fatigue crack to degrade the patch protection around the crack. These flaws were added to (i) determine the damage tolerance of composite doubler repairs in poor installation conditions and (ii) study the ability of sensor networks to detect and accurately track flaw growth.

Figures 20 and 21 show the installation process used to place the PZT network on the structure. A two-part epoxy was used to bond the PZT film to

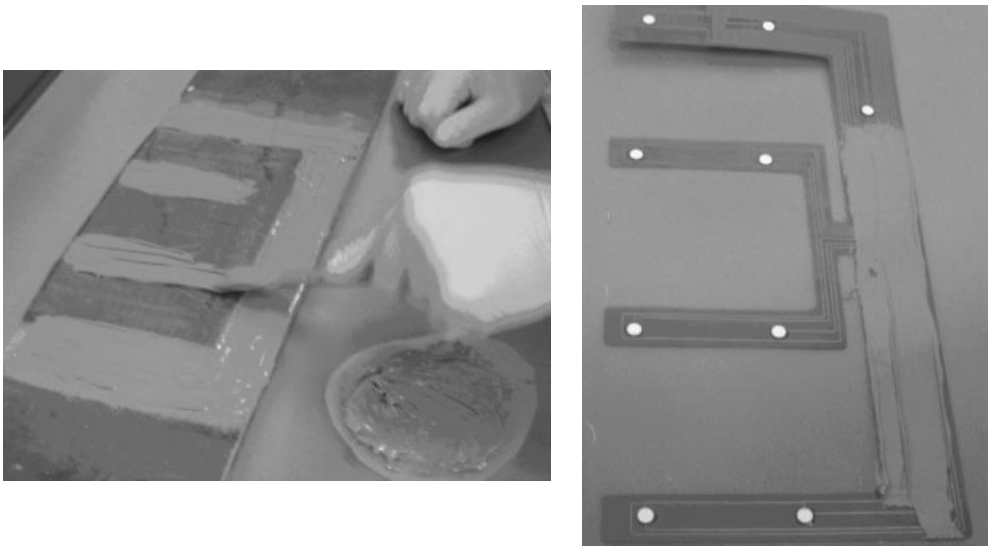


Figure 20. Installation of PZT sensor film—application of EA9394 epoxy to the test specimen and sensor array.

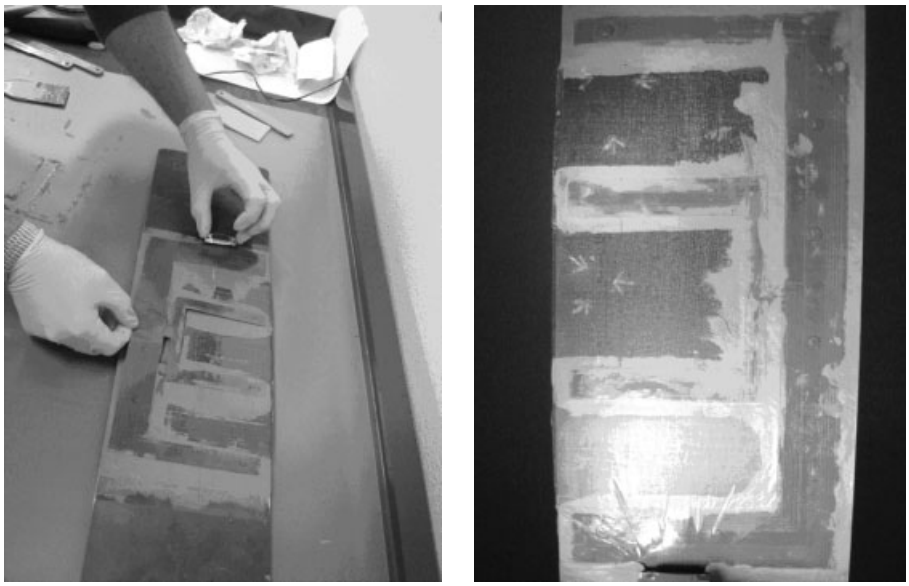


Figure 21. Application of sensors to specimen and view of the PZTs after the bonding process.

the steel and it was cured at 140 °F while applying vacuum pressure to the film. Since the PZT transducers must act as excitation sources, as well as signal receivers, it is necessary for them to be bonded to the structure they are monitoring. When the PZTs are driven with a voltage, they expand and contract at the driving frequencies. The strong bond allows the PZTs to transfer this strain energy into the structure, which results in the generation of the Lamb waves described above.

Figure 22 shows the repaired test specimen in the fatigue test machine with the PZT network installed over the composite doubler. It is to be noted in Figure 18 that there are essentially two concentric

Figure 22. Steel plate with composite doubler, engineered flaws, and PZT sensor network undergoing fatigue tests.

sets of inner and outer PZT sensors (seven PZTs on the outer loop and four PZTs on the inner loop). This was used to evaluate the maximum PZT spacing that could be deployed while still obtaining the desired flaw detection sensitivity. In the realm of SHM, it is desirable to accurately interrogate large areas without having to use an extensive number of sensors. With the PZT network shown in Figure 18, it was possible to set up a series of different sensor-receiver paths and to independently determine the flaw detection capability of each of these paths. Figure 23 shows the various paths that were set up to study flaw detection versus sensor spatial resolution. The "all-to-all" path set uses every sensor as a transmitter (exciter) and receiver, while the "in-to-in" path set uses only the inner four PZT sensors. Similarly, the "in-to-out" path set looks at the wave travel from each inner PZT to each outer PZT and the "out-to-out" path set uses only the sensors in the furthest outer loop. It can be seen that different path sets emphasize different area coverage and provide a different density of coverage.

Figures 24 and 25 show the detection of the three engineered disbonds that were imaged by the PZT network. The optimum images are shown here as produced by the densest "all-to-all" PZT path set. While flaw detection could be achieved with the other path sets, they were not as accurate in flaw placement—as calculated using the signal triangulation method described above—or in relative flaw sizing. Obviously, the "in-to-in" path set was not able to image the outer edge disbonds since this data set did not involve wave travel through the outer disbond regions and thus did not include any information from these flaws.

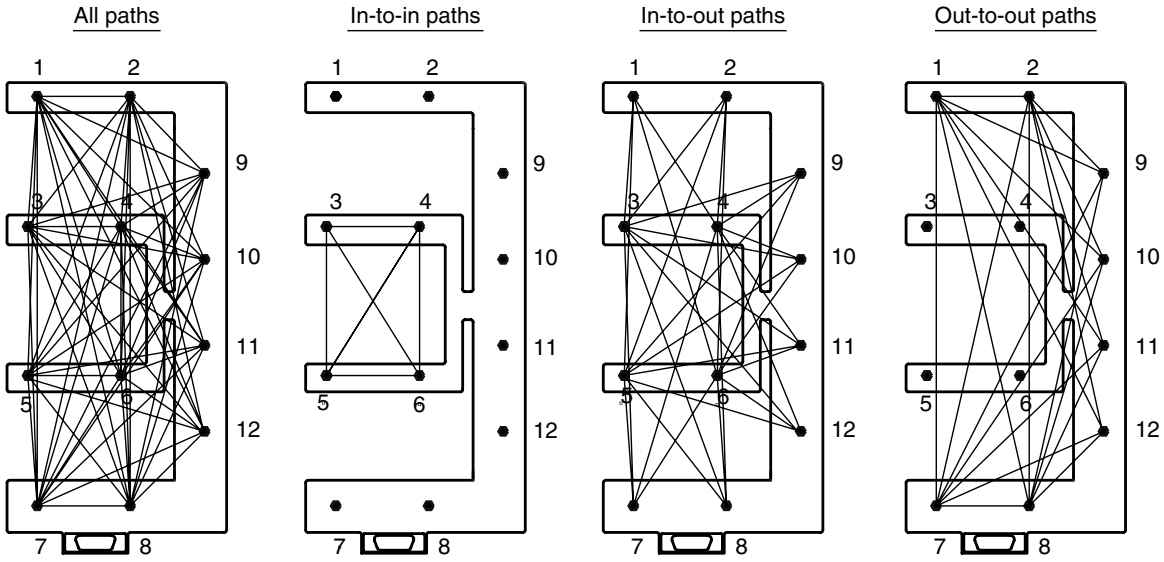


Figure 23. PZT sensor-receiver paths used for data analysis.

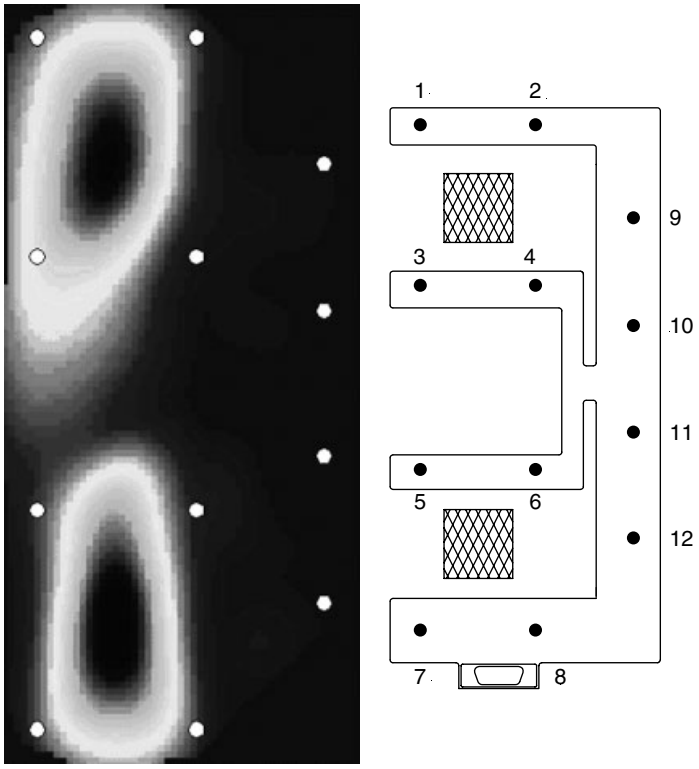


Figure 24. Disbond detection with general location of edge disbonds shown.

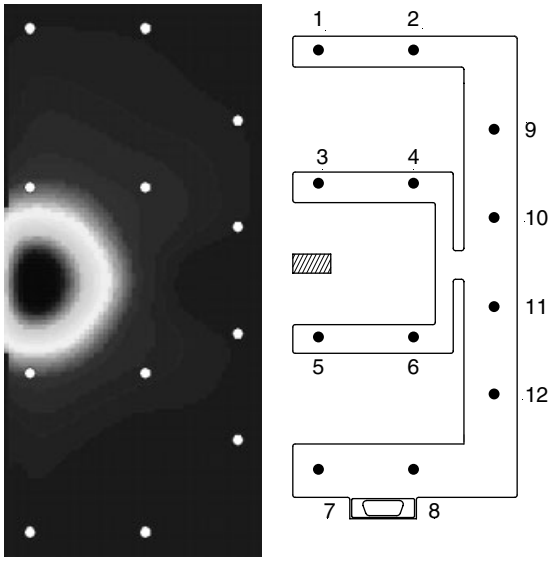


Figure 25. Disbond detection with general location of central disbond shown.

Crack detection was achieved and growth was monitored using the same set of PZT paths. PZT data was acquired at discrete intervals during the crack growth process. EC and microscopic inspections were also conducted to measure the crack lengths at each cycle count. The variation in PZT signal, for the same path, before and after crack growth is highlighted in Figure 26. This data is then used to construct a two-dimensional image of the crack within the PZT network. Figure 27 shows the set of images produced by the PZT sensors at various crack lengths. The corresponding number of fatigue cycles and crack lengths associated with each measurement may be noted. These optimum images were produced by the "all-to-all" path set. Some level of crack detection was achieved using the next densest "in-to-out" path set but the images were not as clear. The "out-to-out" path set did not encounter the crack region sufficiently to properly image the crack, while the "in-to-in" path set produced a sparsely populated data set that was unable to detect the fatigue crack.

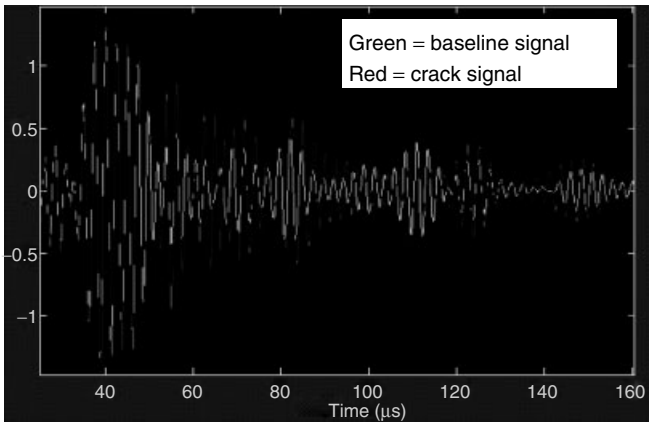


Figure 26. Changes in PZT signals produced by crack growth.

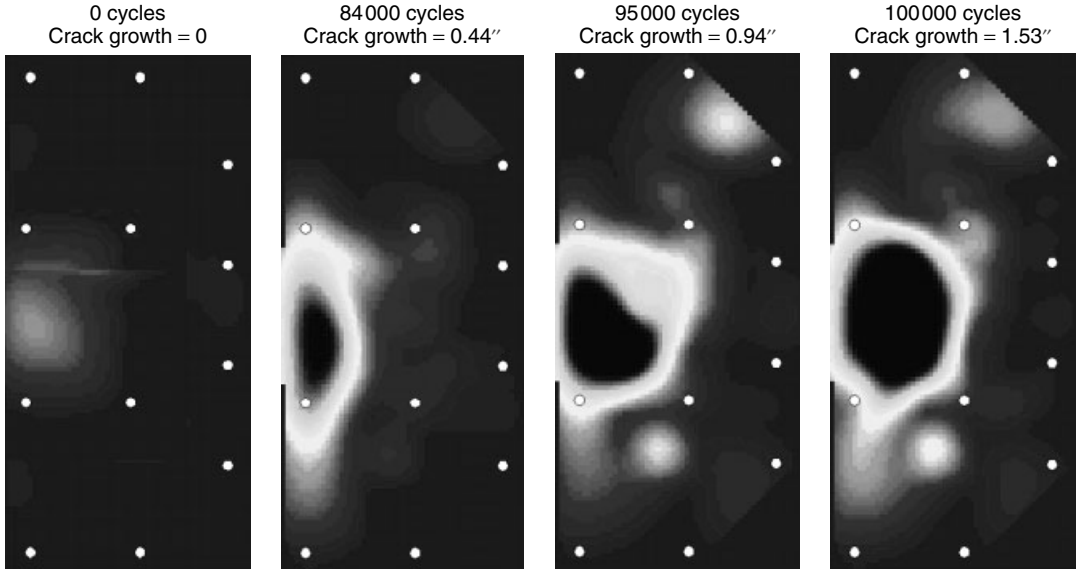


Figure 27. Series of images produced from PZT data corresponding to different crack lengths.

5 MOUNTABLE EDDY CURRENT SENSOR FOR *IN SITU* HEALTH MONITORING

A wireless, integrated, mountable, battery-operated, noncontact EC sensor was developed by Sandia Labs to provide the same accuracy as the large, manually applied NDI equipment and transducers. The sensor can be mounted on a wide array of structures for general surface and subsurface crack detection. It can produce a strong enough magnetic field to produce deep crack detection for inspecting the second and third layers in complex joints or for detecting cracks hidden beneath nonconducting layers and coatings. Figure 28 shows a photo of the two-coil sensor inspecting for cracks beneath a composite doubler. The use of a rectangular planar coil, coupled with embedded processing, conditioning electronics, and wireless communications provides a stand-alone solution to perform accurate, remote, *in situ*, real-time, noncontact, SHM.

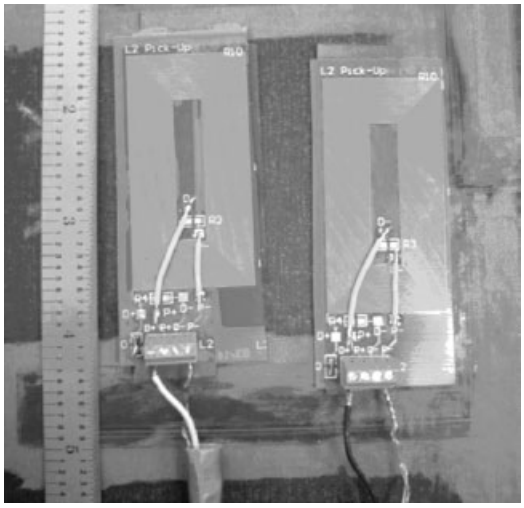


Figure 28. Dual/differential coil version of Sandia eddy current sensor for surface and subsurface crack detection (*in situ* health monitoring methodology).

5.1 Quantification of crack detection sensitivity

A series of crack detection tests were completed in order to quantify the crack detection performance of this mountable EC sensor. In order to construct a good basis of comparison to known crack detection results, the inspection data generated by the EC sensor were compared with results obtained using conventional inspection methods. The EC pencil probe and spot probe methods were deployed along with optical measurements. The EC pencil probe (higher frequency NDI) is intended for surface crack detection. The pencil probe and optical inspections were applied to the metal side of the test specimen shown in Figures 18 and 29. This is not the normal inspection surface but these measurements provided the greatest accuracy and were used as the "referee" for comparison purposes. The EC spot probe and the EC sensor, both with greater depth-of-penetration capabilities, were applied through the composite doubler in the inspection scenario expected in the field.

Optical microscope
Crack on steel side of specimen imaged at 24× magnification with load on the test specimen to open crack tip; resolution of 0.0005″; used as referee for determining the actual crack length.

Eddy current pencil probe
Higher frequency with less penetrating power but smaller diameter for better small crack resolution; applied to steel surface directly (nondoubler side of test specimen).

Eddy current spot probe
Larger coil producing lower frequencies for deeper penetration; truly represents capability in the field for inspections on repaired structure; inspections conducted through composite doubler that produces lift-off, thus requiring lower frequency.

Mountable EC sensor
Larger coil producing lower frequencies for deeper penetration; truly represents capability in the field for inspections on repaired structure; inspections conducted through composite doubler that produces lift-off, thus requiring lower frequency.

Figure 29 shows the orientation of the two-coil sensor relative to the fatigue crack. The coil labeled number 1 is the reference coil and it remains stationary in a known unflawed area. The coil labeled number 2 is used to detect the crack in a differential setup that compares the signals from the unflawed (coil #1) and flawed (coil #2) regions. This comparison ensures that the signal generated by the presence of a crack is significantly different from the signals generated in the unflawed region. Good signal-to-noise (S/N) levels are needed to avoid false calls. Figure 29 also shows the sensor inspecting for cracks beneath various thicknesses of lift-off. In addition to the composite doubler, Lucite Test Blocks, ranging in thickness from 0.135 to 0.905″, were placed beneath the EC sensor. This produced additional lift-off effects, which evaluated the penetration power and sensitivity of the EC sensor though extreme lift-off scenarios (refer to later text for discussion on crack detection through various thicknesses of nonconductive lift-off).

The graphs in Figures 30–32 provide summary results of the direct comparison between the Sandia-developed EC sensor and the conventional handheld transducers often used to conduct manual inspections. A photo comparing the EC spot probe, pencil probe, and mountable sensor is provided in Figure 33. Figures 34 and 35 show each of the EC transducers being deployed to detect fatigue cracks beneath the

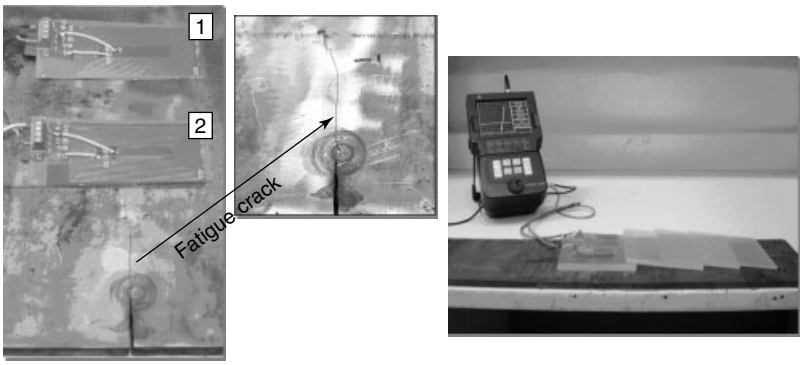


Figure 29. Mountable EC sensor applied to bare steel side and composite doubler side of the fatigue crack test specimen.

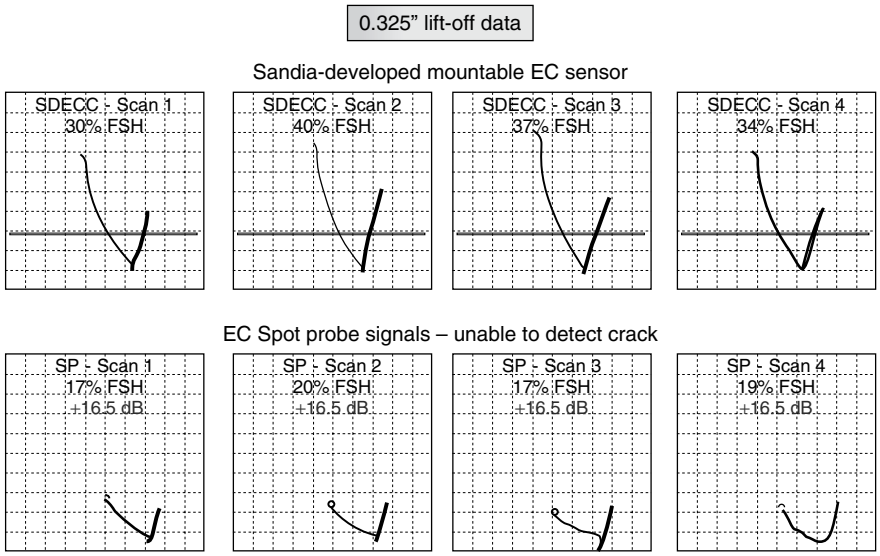


Figure 30. Comparison of eddy current signals generated by conventional EC spot probe and signals produced by the Sandia EC sensor.

composite doubler. The EC signal graphs clearly illustrate that the EC sensor can provide significant response even when subjected to the impediment of large lift-offs. Conversely, the spot probe faltered because it could not maintain its sensitivity while trying to interrogate a structure beneath thick lift-off layers. The pencil EC probe pictured in Figure 33 can interrogate with only high frequencies, so it was not able to produce any usable inspection signal in these tests. Of equal importance to this comparison is the fact that hand-deployed probes cannot be permanently mounted to produce *in situ*, and potentially real-time, health monitoring.

Inspection performance results for the final sensor design are shown in Figures 31 and 32. These impedance plots show the overall sensitivity of the mountable sensor as it detects cracks located through increasingly thick lift-off layers up to 0.98″ thick. It is to be noted that these plots also include the signals produced by the sensor when placed in an unflawed region. Such signals from unflawed regions correspond to the noise in the sensor, so it is desirable for the sensor to produce extremely small signals in unflawed regions. By comparing the signal levels at cracked and uncracked regions, it is possible to assess the sensitivity of the EC sensor using S/N

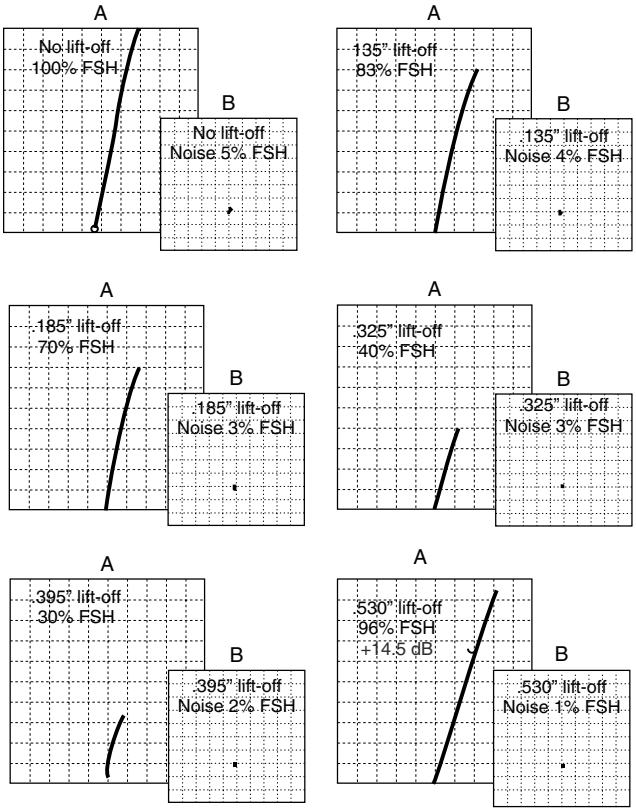


Figure 31. Crack detection performance of Sandia mountable eddy current sensor for lift-off layers up through 0.5″ thick.

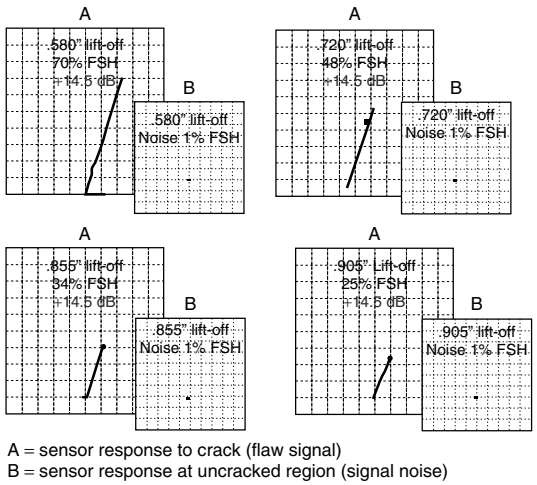


Figure 32. Crack detection performance of mountable eddy current sensor for lift-off layers up through 0.9″ thick.

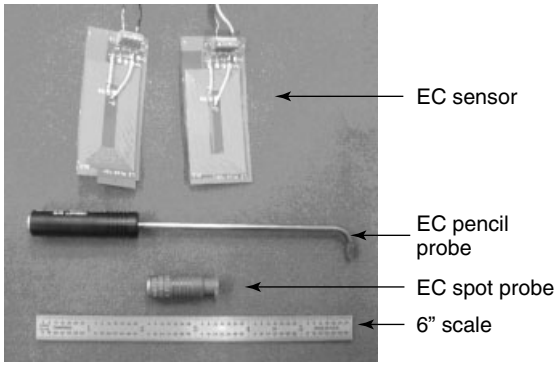


Figure 33. Handheld and mountable eddy current transducers used to track crack growth.

levels. Normally, flaw calls can be made if S/N values exceed 3 (S/N ratio of 3 : 1). The results in Figures 31 and 32 show that even through extreme

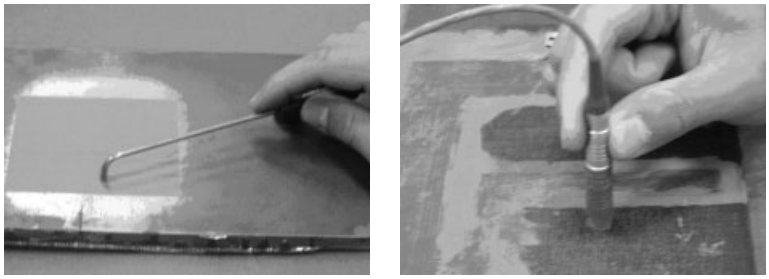


Figure 34. Deployment of EC pencil probe (steel side) and EC spot probe (composite doubler side) on fatigue crack specimen.

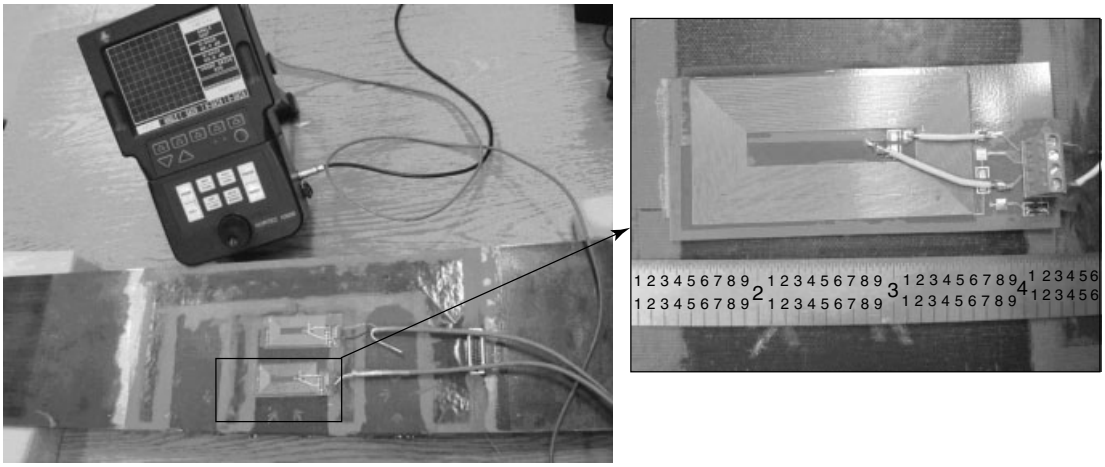


Figure 35. Deployment of EC sensor on composite doubler side of fatigue crack specimen.

lift-off conditions (inspection impediment), the sensor produced an S/N level of 15 : 1 even for the most extreme case of the 0.9″-thick lift-off.

The final assessment of the EC sensor quantified its ability to not only detect a fatigue crack but also to accurately determine its length. Figure 34 shows how the EC pencil probe was applied to the exposed metal side of the test specimen. The optical microscope was also used on the side with the exposed fatigue crack. As mentioned above, this is not the normal inspection surface but these measurements provided the greatest accuracy and were used as the "referee" for comparison purposes. The EC spot probe and the EC sensor, both with greater depth-of-penetration capabilities, were applied through the composite doubler in the inspection scenario expected in the field (see Figures 34 and 35).

Tables 7 and 8 compare the crack lengths as determined by the various methods described above, while Figure 36 plots the crack length predictions for each of the four methods. The Sandia EC Sensor was able to accurately track the crack length even when inspecting through the composite doubler. The maximum deviation from the referee measurements was less than 4%.

6 REMOTE-FIELD EDDY CURRENT

There are two fundamental approaches that have been developed to improve the performance of EC NDI. The first approach attempts to cancel, or compensate for, the large surface signal that is directly coupled to the pickup coil without first entering the part being inspected. This portion of the overall EC signal has no useful information and serves to increase the noise floor. It can be removed by altering the probe

Table 7. Comparison of crack lengths measured by different methods on SYN-FAT 21 Specimen

SYN-FAT-21 specimen				
No. of cycles	Crack length (Optical microscope) (in.)	Crack length (eddy current pencil probe) (in.)	Crack length (eddy current spot probe) (in.)	Crack length (eddy current mountable sensor)
0	1.000	1.063	1.063	1.000
73 000	1.200	—	1.240	1.250
81 000	1.410	—	1.400	1.410
84 000	1.495	1.500	1.480	1.510
87 700	1.600	—	1.630	1.650
91 500	1.805	1.800	1.840	1.825
96 500	2.130	2.130	—	—
97 200	2.200	2.195	2.045	2.020
99 000	2.400	2.390	2.550	2.450
100 200	2.590	2.590	2.750	2.620

Table 8. Comparison of crack lengths measured by four different transducers showing ability of mountable EC sensor to accurately track crack growth

SYN-FAT-21 Specimen				
No. of cycles	Crack length (optical microscope) (in.)	Crack length (eddy current pencil probe) (in.)	Crack length (eddy current spot probe) (in.)	Crack length (eddy current mountable sensor)
0	1.45	1.45	1.41	1.48
5000	1.81	1.81	1.75	1.85
8500	1.975	1.97	1.92	1.93

design or by applying special data analysis to the captured signal. The self-nulling probe (SNP) and the remote-field eddy current (RFEC) method are examples of the probe alteration approach [8, 9]. The sensor and the drive coil of the SNP are separated by a ferromagnetic cylinder called a *flux focusing lens* to minimize the sensor signal that arrives directly from the drive coil. The RFEC approach blocks the unwanted, directly coupled signal through the use of shields and magnetic circuitry. This produces a signal that is dominated by the indirect coupling path and contains the information of interest.

The inspection of multilayered structures requires detection of fine cracks and shallow corrosion pits that are embedded deep below the inspection surface. If there are air gaps between the layers, ultrasonic and thermography methods are not good candidates since it is difficult for them to penetrate beyond the first layer. Conventional EC techniques are often limited by depth-of-penetration effects wherein deeper penetration requires larger EC coils that produce a corresponding loss in sensitivity. A relatively new breakthrough in the EC arena involves the use of "remote fields" to interrogate conductive structures. The RFEC method is especially suited for detecting deep flaws.

Figure 37 depicts the physics behind the RFEC approach. When a coil is excited by an alternating current and is placed on a structure, the energy diffuses along two different paths. The interaction between the two fields results in what is often referred to as the *remote-field EC effect*. Studies have shown that the energy diffusing via the direct path attenuates very rapidly. The signal received by the pickup coil that is located a certain distance away from the excitation coil is primarily due to the energy

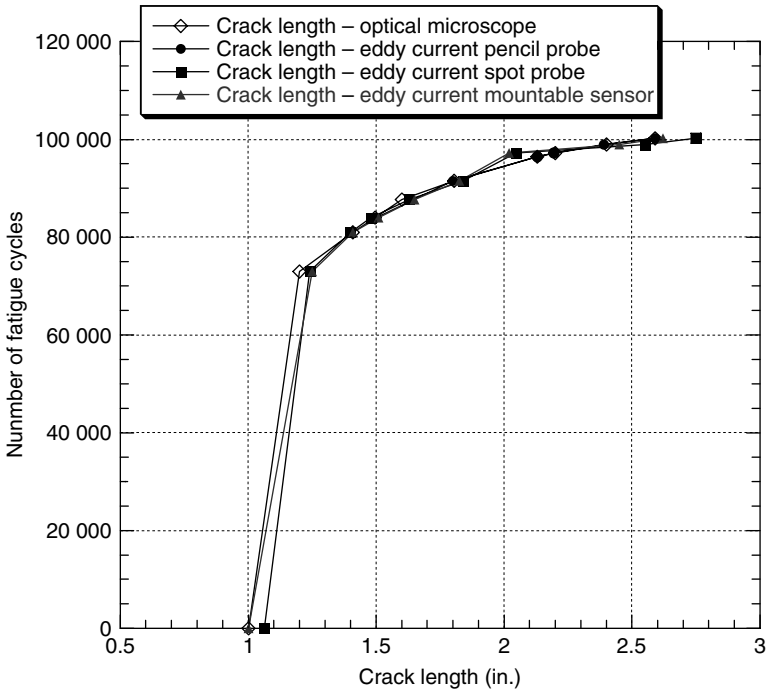


Figure 36. Comparison of crack lengths measured by four different methods: (a) microscopic measurement (referee), (b) conventional eddy pencil probe (referee), (c) conventional eddy current spot probe, and (d) eddy current *in situ* sensor.

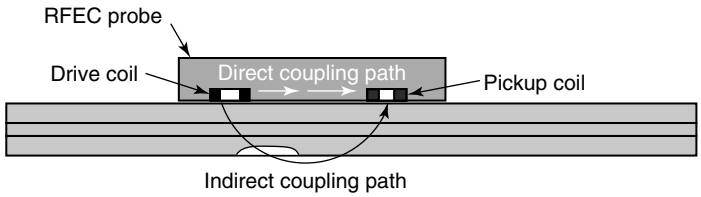


Figure 37. Schematic of remote-field eddy current approach.

diffusing via the indirect path. This portion of the energy passes through the structure twice before arriving at the pickup coil. Thus, the RFEC probe is designed to focus on the indirect coupling path, so that the signal measured by the pickup coil carries the information of the whole wall thickness. It produces deeper penetration for detection of deep, subsurface flaws. Traditional EC approaches depend on changes in the electromagnetic flux to signal the presence of a flaw. Typical values of such changes are very small and may be less than 0.01%. Therefore, it is very difficult to separate this change from the quiescent signal generated by an unflawed structure. This limits the maximum gain that one can employ when using these instruments. RFEC is based on the measurement of the voltage induced in a pickup coil by the flux that has passed through the test object twice (see Figure 25). This setup creates a significant change in voltage so that a stronger flaw signal is produced. The end result is a very high flaw-signal/quiescent-signal ratio. This allows higher gain settings to be used so that deeper flaws can be detected.

A super-sensitive EC system has been developed to accommodate the RFEC probes. The system is capable of providing an extremely high gain, up to 135 dB, for amplifying the remote-field probe signal. This allows the inspection device to provide higher S/N ratios, thus increasing its flaw detection

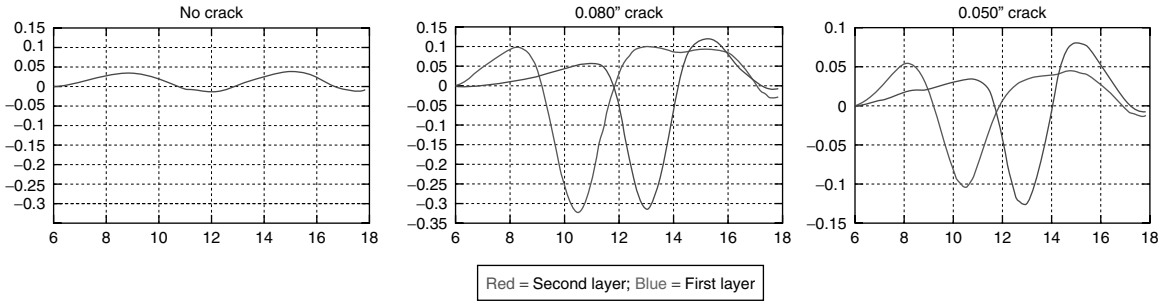


Figure 38. First- and second-layer crack detection in 0.040″-thick plates with cracks hidden under raised head fasteners.

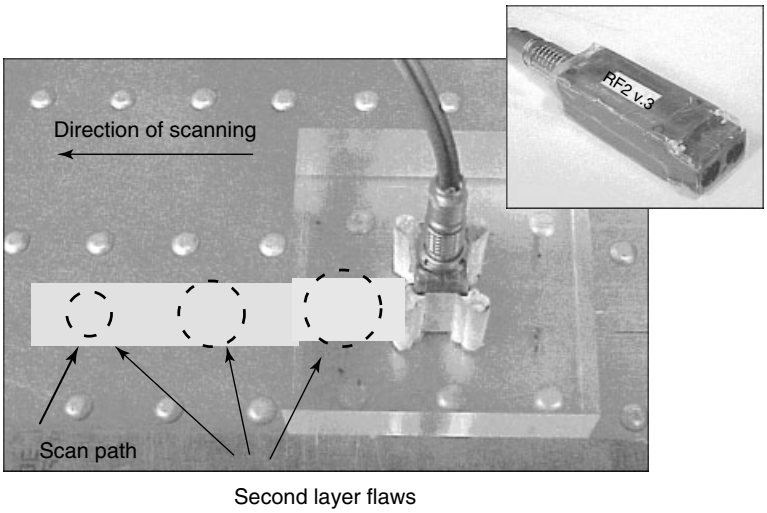


Figure 39. RFEC probe inspecting for corrosion in a riveted aircraft joint.

sensitivity. Sample crack detection signals from a 2-kHz RFEC probe inspection are shown in Figure 38. These plots demonstrate that the voltage amplitude-based data presentation requires minimal signal interpretation. Experimental measurements have shown the ability of the RFEC method to detect a 0.03″-long fatigue crack that is located 0.446″ below the inspection surface.

When scanning a multiple-layer aluminum sample with a total thickness of 0.29″, the RFEC system was able to detect wall thinning as small as 1% of the total thickness. Figure 39 shows a handheld RFEC probe scanning a riveted aluminum joint containing various levels and sizes of corrosion. The dotted lines show the locations of the corrosion sites that were engineered into the specimens. Sample signals from the hidden, second-layer corrosion regions are shown in Figure 28. The RFEC system produced S/N ratios that were well in excess of 5 to clearly indicate the presence of corrosion. It is to be noted that the size of the second-layer corrosion detected with the signals shown in Figure 40 was 0.125″ in diameter with depth ranging from 0.0008 to 0.004″ (2–10% corrosion thinning in 0.040″-thick plate).

6.1 RFEC deployed with in situ sensors

Efforts are under way to deploy this inspection method using a permanently mounted, *in situ* sensor called the *remote-field eddy current magnetic carpet probes* (RFEC-MCP). This novel leave-in-place magnetic carpet probes (MCP) can be installed to detect and monitor fatigue cracks or corrosion

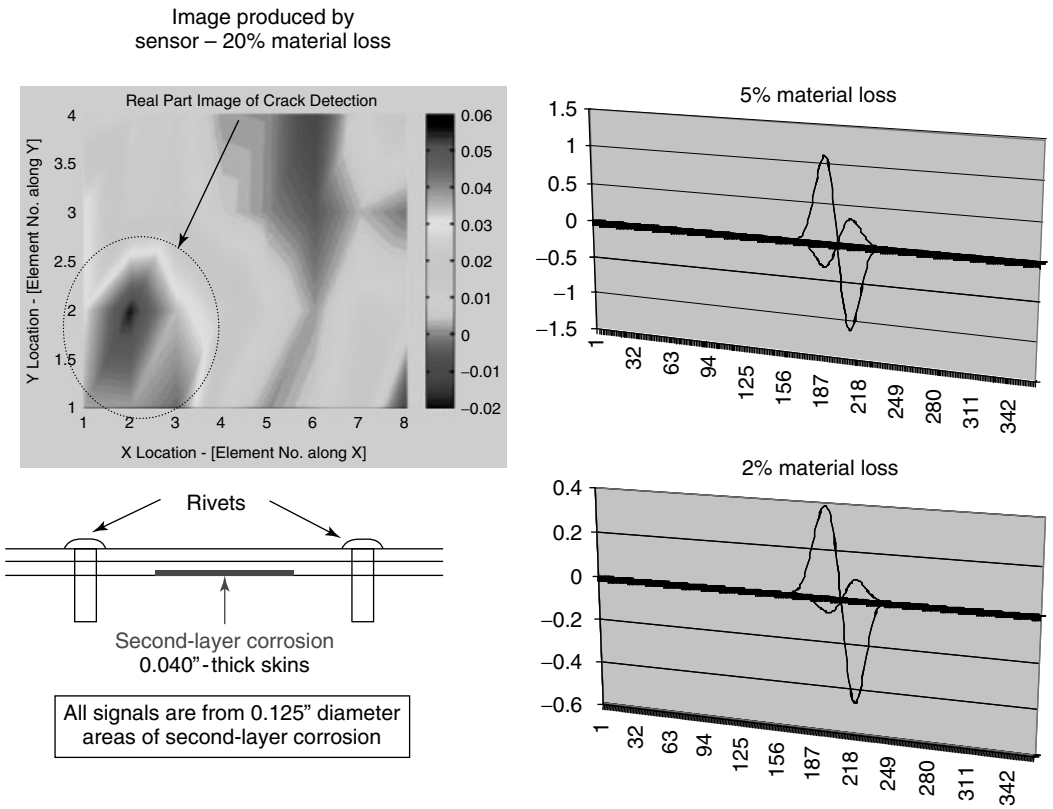


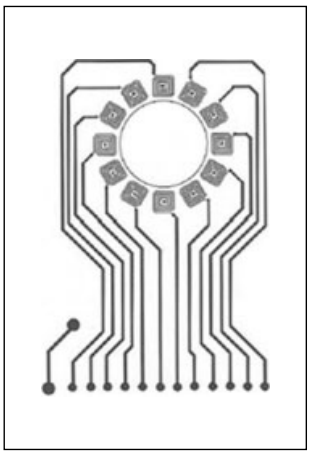
Figure 40. Large signals produced by RFEC probe indicating the presence of corrosion.

in inaccessible regions on structures. Networks of surface-mounted and embedded EC sensor arrays permit rapid, large-area damage monitoring with the high resolution of localized probe testing. The RFEC-MCP is made using a piece of thin, flexible printed circuit board (PCB) with densely populated printed coil sensors on the PCB as shown in Figure 41. A sensor for detecting cracks emanating from a hole and a sensor for monitoring larger areas for corrosion are shown.

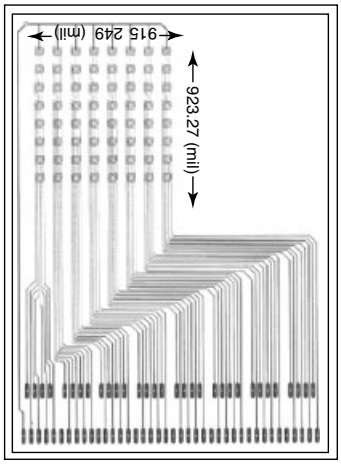
The sensors work on the same principal as the RFEC method described above to provide deep penetration for locating subsurface flaws. Multiple-phase AC currents are applied to each sensor on the PCB. This generates a traveling magnetic wave that is stronger than those generated by conventional EC excitation. Each sensor on the PCB alternately acts as a transmitter and receiver so that a uniform magnetic field is set up over the entire MCP area and the probe can detect cracks in any orientation. In addition, the RFEC-MCP method is insensitive to the geometry variations of a component while competing ultrasonic-based sensors must carefully account for the effects of complex geometries on the transmission of UT waves. Initial tests with a prototype corrosion sensor showed great promise and a second design iteration is currently in process to further improve sensitivity.

7 OTHER SHM SENSORS

Breakthroughs in microelectronics are being used to miniaturize conventional handheld probes into mountable, *in situ* devices. Miniature flaw detection sensors, combined with easily fieldable diagnostic electronics, are the backbone of any global SHM system. An extensive amount of research is ongoing to develop and mature a wide variety of new sensors that are applicable to *in situ* SHM. The sensors



For monitoring a fastener hole
(crack in any direction)



For monitoring an area

Magnetic carpet probe

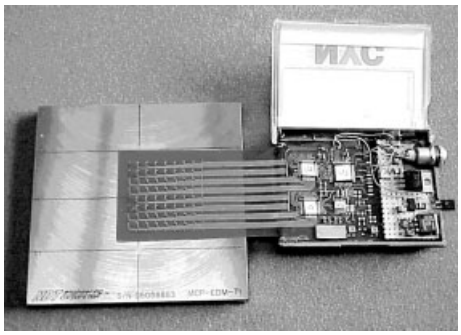


Figure 41. Sample RFEC printed circuit board designs and prototype corrosion sensor with colocated electronics.

may directly detect flaws, such as those discussed here, or other physical parameters, such as acceleration, force, pressure, flow, displacement (capacitive proximity sensors), temperature, and the presence of chemicals, from which flaws or faulty operation can be identified. Other leave-in-place sensors that may produce direct indications of structural flaws include magnetoresistors; polycrystalline silicon for measuring an array of physical properties including transmission of ultrasonic waves and variations in electric fields, FO sensors [10–13], capacitive micromachined ultrasonic transducers (cMUT) [14], flexible EC array probes [15], nickel-foil magnetostrictive sensors (MsS) [16], optical fiber long-period grating (LPG) chemical sensors [17], adhesive bond degradation sensors (ABDS) [18], meandering winding magnetometer (MWM) probes [19, 20], microwave antenna, acoustic emission transducers, spectroscopic instrumentation and optoelectric sensors such as injection lasers, photovoltaic diodes, and photoconductors.

Following are brief descriptions of other sensors being pursued for SHM.

1. Fiber-optic sensors

Rapid growth in the optoelectronics and telecommunications industries has resulted in the evolution of highly sensitive FO sensors. FO sensors have been developed for a wide variety of applications including the measurement of rotation, acceleration, vibration, strain, temperature, pressure, electric and magnetic fields, moisture, and humidity. FO sensors are lightweight, low profile (typically 145 μm in diameter with a polyimide coating) and corrosion

resistant, and multiple sensors are easily multiplexed into a single fiber. These factors make them ideal for embedding in or surface mounting on composite and metallic structures without affecting structural performance. Other advantages of FO sensors include high sensitivity, wide bandwidth, electromagnetic interference (EMI) resistance, low power requirements, and environmental ruggedness. FO sensors can also be configured to monitor crack growth and corrosion in civil and aerospace structures. Omnidirectional FO sensors can be used to measure large strains (up to 150% strain), small displacements (10-μm range), and crack growth in any material. The major disadvantages include high cost, mechanical frailty (during handling stage), and unfamiliarity to the end user. The introduction of low-cost laser diodes, the alternative use of light-emitting diodes (LEDs) as light sources, and the development of inexpensive, single-mode optical fibers have greatly reduced the costs associated with deploying FO sensors. Information about the environment to which an FO is exposed can be inferred by analyzing the guided light transmitted through the optical filaments. In this approach, the entire length of the fiber acts as a continuous sensor. The fiber can be mounted in a serpentine path or FO tentacles can be created to provide full coverage over a large area of concern. The presence of a crack can be determined by monitoring changes in the magnitude and phase of the returned light in the fiber. Figure 42 shows a close-up of the fiber bragg grating (FBG) sensors, and associated end connectors, as they enter and exit the bondline of the test specimen. Typical data-acquisition equipment used to monitor the FBG sensors is also pictured.

2. Capacitive micromachined ultrasonic transducers (Stanford, General Electric)
This is an electrostatic transducer that utilizes semiconductor fabrication techniques to allow for mass production and low cost (see Figure 43 for example). An applied voltage produces a displacement of the membrane (generates ultrasonic wave) due to electrostatic attraction. When it is used to receive input signals (monitor ultrasonic waves), the displacement of the membrane causes a change in capacitance of the device. Advantages of this device include completely integrated electronics and a large operation bandwidth.

3. Flexible eddy current array probes (General Electric)
This probe allows the advantages of EC arrays to be implemented in a conformable PCB (see Figure 44). The sensor incorporates magnetoresistors and spiral coils in two-dimensional arrays. This allows for simultaneous, multifrequency inspections and real-time imaging of cracks and corrosion. Successful demonstrations of GE flexible EC sensors have been completed on the AANC (FAA Airworthiness Assurance center) testbed aircraft.

4. Meandering winding magnetometer probes
JENTEK sensors has developed mountable, electromagnetic sensors for fatigue and corrosion monitoring of aircraft structures. The MWM and the segmented-field, spatial wavelength interdigitated dielectrometer

Figure 42. Fiber-optic sensors in adhesive bondline and FO monitoring equipment.

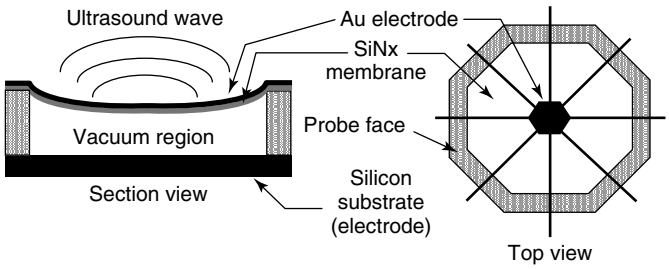


Figure 43. Schematic of capacitive micromachined ultrasonic transducer (cMUT).

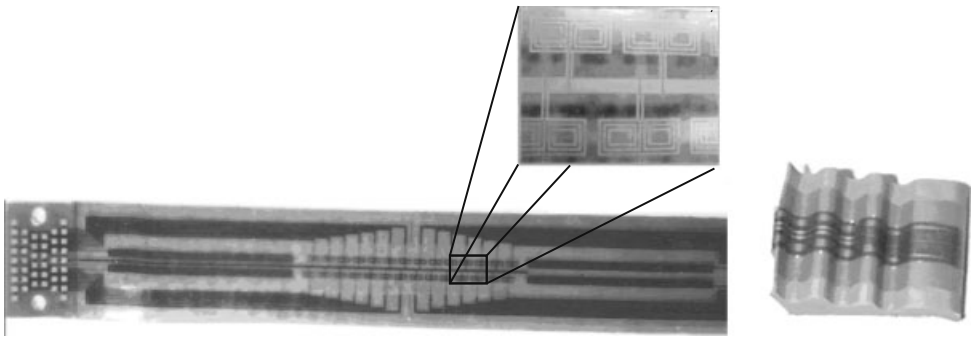


Figure 44. General electric's flexible eddy current array.

(IDED) have been produced and demonstrated for a wide variety of health monitoring applications. The MWM is an inductive EC sensor, and the IDED is a capacitive sensor. MWMs are suitable for metals, graphite fiber composites, reinforced carbon composites, and low observable coatings (nonconducting, magnetizable coatings using magnetic particle suspensions), while the IDED is suitable for characterization of glass fiber composites, corrosion protection coatings, sealants, glass, paint, and wood, as well as for detection/monitoring of corrosion products and moisture ingress, and monitoring of cure states of epoxies and adhesives. The corrosion monitoring technology is relatively new and requires additional research before field implementation, while the MWM-array fatigue monitoring technology is already a commercial product to support component fatigue testing. It is being transitioned to the fleet for specific applications. Through the combined use of IDED and MWM sensors, it is possible to detect not only metal loss, as with EC-based NDI methods for corrosion, but also corrosion products and moisture ingress within the joint itself. Example geometries for MWMs are shown in Figure 45.

5. Nickel-strip magnetostrictive sensors (Southwest Research Institute)

This guided-wave probe consists of a thin nickel substrate (foil) and a thick MsS coil in a PCB that is placed on the nickel strip. The strip is mounted on the structure being monitored. It generates pulses of guided waves and detects the signals being returned from within the structure. Data comparisons with baseline signals (undamaged configuration) are used to identify defects. A single probe can work over long distances allowing for economical SHM of large structures.

6. Adhesive bond degradation sensor (Lockheed-Martin, Tetra Tech)

This sensor is used in conjunction with FO lines. The ABDS is being developed to detect bondline decay by measuring water ion concentration in the adhesive. These ions are by-products of decay at the adherend-to-adhesive interface. Features of light reflection from the ABDS are altered in response to the concentration of water ions. Testing is under way to calibrate the adhesive bond degradation sensor in order to quantitatively relate the level of water ions to the decay in bond strength.

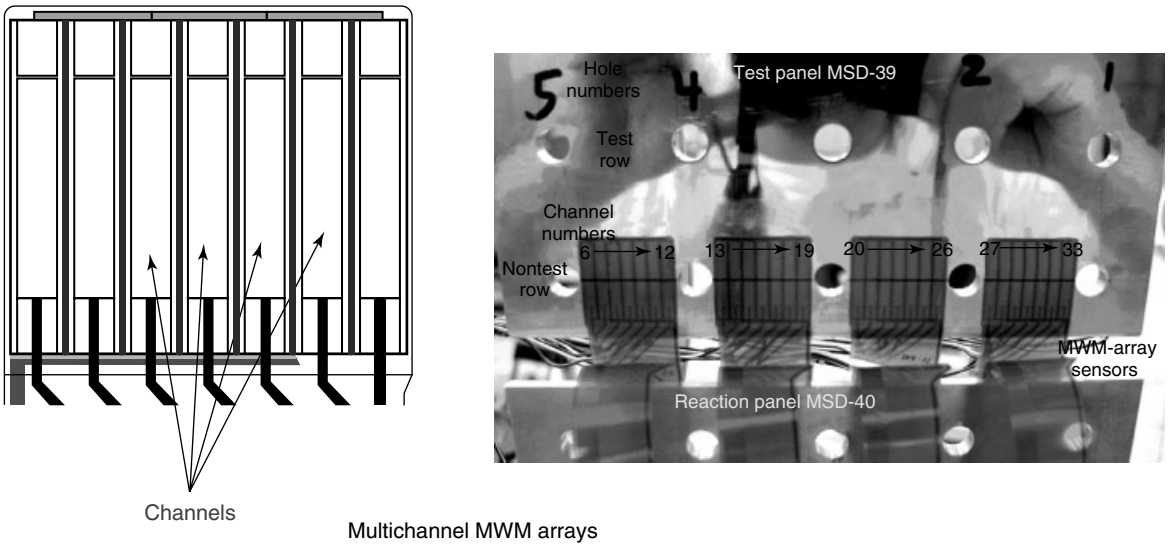


Figure 45. Seven-channel MWM array and surface-mounted MWM arrays monitoring for cracks around fastener holes.

7. Long-period grating (LPG)-based chemical sensors (Luna Innovations, Boeing, Navy)
This is an optical fiber sensor that can detect precursors to corrosion such as moisture, pH levels, and metal-ion corrosion by-products. By applying special coatings that change refractive index with absorption of target molecules on the LPG surface, the optical fiber sensor becomes the transducer for chemical measurement. For detection of moisture, for example, the LPG sensor is coated with a poly-ethylene oxide (PEO) hydrogel that swells in the presence of moisture.

Alternatively, sacrificial corrosion sensors—for installation on the ends of FO lines—are in the development stage. These sensors are designed to deteriorate in the same manner and at the same rate as the parent structure they are monitoring. Characteristics of light reflected back through the FO line can be calibrated to indicate different levels of corrosion.

8 DEPLOYMENT OF HEALTH MONITORING SENSOR NETWORKS

Distributed sensor networks can be deployed in any of the three approaches listed below. These options are listed in the order of increasing complexity; however, less labor is required to monitor the systems as they become more complex.

1. In situ sensors only
The sensors are the only items permanently installed on the structure. At the desired inspection intervals, power, signal conditioning, and data-acquisition electronics are manually transported to the structure to be monitored. The sensors are linked to the monitoring electronics via an electrical connector and flaw detection is completed by an inspector at the site.

2. Sensor network with *in situ* data acquisition
In this system, miniature, packaged electronics are also placed *in situ* with the sensor network. The electronics contains the necessary power, memory, and programmable circuitry for automated data logging. The data is periodically downloaded to a laptop through manual hookups at the site.

3. Sensor network with real-time data transmission to a remote site
This approach is similar to item (2) above with the addition of a telemetry system that allows for continuous, wireless transmission of data to a web site. The web site can be programmed to interrogate critical aspects of the data and use preset thresholds to

provide continuous green light/red light information regarding the health of the structure. The web site can even be programmed to automatically send an e-mail to operation personnel if the condition monitoring process indicates the need for repairs or other maintenance.

8.1 Principles of improved maintenance through SHM

The costs associated with the increasing maintenance and surveillance needs of our aging infrastructure are rising at an unexpected rate. Through the use of SHM concepts, it is possible to quickly, routinely, and remotely monitor the integrity of a structure in service. The ease of monitoring an entire network of distributed sensors means that structural health assessments can occur more often or even continuously, allowing operators to be even more vigilant with respect to flaw onset. SHM systems allow the operators to track structures when their use has changed substantially (e.g., increase in bridge load, duration of daily operation and increase in spin rate of turbines) or immediately after an abnormal event. SHM systems can even automatically determine that an abnormal event has occurred and acquire data between programmed monitoring intervals.

SHM addresses the need for reliable SHM systems that can automatically process data, assess structural condition, and signal the need for human intervention. The design, evolution, and deployment of SHM systems and concepts for field applications depend greatly on the structure's inspection needs and the type of maintenance program currently used. More rigorous and scheduled maintenance programs, such as those associated with aviation and nuclear industries, readily lend themselves to transitions into SHM methods. However, unexpected phenomena and advanced aging concerns may hasten the introduction of SHM into other civil structures that are more difficult to closely monitor on a regular basis.

The critical element of the SHM system is the array of sensors. These sensors must be accurate, reliable, repeatable, low cost to mass produce, easy to install, and easy to monitor. Interpretation of data is an important feature that brings numerous human factor issues into play. The sensors must be integrated into automated data-acquisition equipment to convert the data into usable information about structural integrity. The implementation goals for SHM must consider the need for local versus global coverage. In order to maximize the global health monitoring feature of the SHM system, the sensors must be either low cost, to accommodate many installations, or capable of monitoring large areas. Finally, the sensors must have a fail-safe mode of operation (auto fault detection) so that abnormal data is not incorrectly interpreted as damage.

Some of the key issues that must be addressed in order to transition SHM into the field include (i) the goals for the SHM system and its targeted applications, (ii) required performance characteristics and the means to assess that performance, (iii) use of SHM in lieu of existing inspections or use of SHM as a basis for further optimizing structural design (design credits accumulated through integrated SHM), (iv) minimizing system costs, (v) eliminating risks associated with system use, (vi) integration of SHM system with existing subsystems, (vii) use of energy harvesting (solar, vibration) hardware to allow for long-term, unattended operation, and (viii) adapting existing maintenance programs to take advantage of condition-based maintenance opportunities that can stem from the use of SHM.

To date, SHM systems have not been widely deployed in the field, thus limiting the successful operational data needed to instill confidence in their use. However, vibration monitoring (accelerometers), acoustic emission sensors, and crack monitoring (CVM sensors) have all made significant inroads and demonstrate the potential for the SHM approach in areas with high probability for damage. While these limited applications have not allowed for comprehensive cost–benefit analyses to be performed, key savings associated with their use have been identified. These include the remote and rapid acquisition of data and the elimination of setup tasks such as the disassembly or removal of equipment to access an inspection site. In addition to the operating benefits, there are recurring and nonrecurring costs—design savings, training, equipment costs, and increased structure availability—that must be considered. The idea of condition-based maintenance is a critical consideration when determining overall cost–benefits. SHM systems using distributed sensor networks can reduce costs by allowing condition-based maintenance practices to be substituted for the

current time-based maintenance approach. A series of expected maintenance functions have been already defined; however, they will be carried out only if their need is established by the health monitoring system.

One cost–benefit analysis for commercial transport aircraft showed that implementing SHM systems could result in a significant reduction in both maintenance and operations and support (O&S) costs. An example is given where implementation of SHM systems could result in a 30–40% reduction in maintenance requirements, resulting in recovery of the initial implementation costs in only two to three years. For larger and more complex structures, the time required to recover the initial implementation costs may be longer, but is still well within the life of the vehicle. A similar justification can be made for future launch vehicles. For such vehicles to be successful, launch costs must be reduced an order of magnitude below those of current launch vehicles. To meet turnaround goals, the time required to assess the structural condition needs to be significantly reduced. An automated SHM system could potentially assess the health of the entire structure within hours of a completed mission and recertify the structure for flight.

The basic steps required to mature an SHM system from concept to implementation are (i) laboratory evolution to prove the principle of the monitoring physics, (ii) sensor and data-acquisition system development, (iii) application on representative structural components in the lab, (iv) structured validation exercises to statistically determine all performance characteristics (e.g., probability of flaw detection, probability of false calls, repeatability of data), (v) exposure of SHM system to all potential impediments to operation, (vi) development of a field implementation plan, (vii) beta site testing in the field over a representative time frame, and (viii) concurrence from operators, regulators, and researchers that the SHM system is acceptable for use in a particular set of applications (certification).

For now, pilot programs are being pursued to accumulate operational history, further mature SHM technology, and identify the obstacles to its use in more complex and widespread applications. In the future, SHM systems should be integrated with PHM systems to allow for improved and more proactive structural health management. When SHM is combined in a systems approach, which includes sensors to monitor electronics, hydraulics, machinery operations, avionics, etc., it is possible to produce a PHM architecture that can assist in maintenance scheduling and tracking. Such a comprehensive SHM system will reach beyond safety considerations and allow for optimized designs and longer life spans while maximizing performance and operational availability at minimum costs.

9 CONCLUSIONS

Detection of unexpected flaw growth and structural failure could be improved through the use of onboard health monitoring systems that could continuously assess structural integrity. Such systems would be able to detect incipient damage before catastrophic failures occur. Local sensors, such as the ones described in this article, can be used to directly detect the onset of crack, corrosion, or disbond flaws. Global SHM, achieved through the use of sensors such as accelerometers coupled with structural dynamics assessments, can be used to assess overall performance (or deviations from optimum performance) of large structures such as bridges and buildings. Whether the health monitoring approach is local or global, the key element in a SHM system is the calibration of sensor responses so that damage signatures can be clearly delineated from sensor data produced by unflawed structures.

This article focused on local flaw detection using embedded sensors. While some of these leave-in-place sensors are able to produce wide area inspections, their use is predicated on the identification of primary flaw regions to be monitored. The replacement of our present-day manual inspections with automatic health monitoring would substantially reduce the associated life-cycle costs. The ease of monitoring an entire network of distributed sensors means that structural health assessments can occur more often, allowing operators to be even more vigilant with respect to flaw onset. When accessibility issues are considered, distributed sensor systems may also represent significant time savings by eliminating the need for component teardown. In addition, corrective repairs initiated by early detection of structural damage are more cost effective since they reduce the need for subsequent major repairs. Aerospace structures have one of the highest payoffs for SHM

applications since damage can quickly lead to expensive repairs and aircraft routinely undergo regular, costly inspections.

In general, SHM sensors should be low profile, lightweight, easily mountable, durable, and reliable. To reduce human factors concerns with respect to flaw identification, the sensors should be easy to monitor with minimal need for users to step through additional data analysis. For optimum performance of the *in situ* sensor-based approaches, the signal processing and damage interpretation algorithms must be tuned to the specific structural interrogation method. For example, in the high-frequency E/M impedance approach, pattern recognition methods can be used to compare impedance signatures taken at various time intervals and to identify damage presence and progression from the change in these signatures. These approaches can benefit from the use of artificial intelligence and neural network algorithms that can extract damage features based on a learning process.

Recent failures of aircraft and civil structures have compelled the engineering community to take a fresh look at the fail-safe, safe-life, and damage tolerance design philosophies. The effect of structural aging and the dangerous combination of fatigue, corrosion, and other environmentally induced deterioration is now being reassessed. The end result of these assessments has been a greater emphasis on the application of sophisticated health monitoring systems. Recent advances in onboard SHM sensors have proven that distributed and autonomous health monitoring systems can be applied to reliably detect incipient damage. Such systems have wide use in aerospace, automotive, civil infrastructure, and other industrial applications.

REFERENCES

[1] Roach D. Health monitoring of aircraft structures using distributed sensor systems. *DoD/NASA/FAA Aging Aircraft Conference*. Palm Springs, CA, March 2006.

[2] Bartkowicz TJ, Kim HM, Zimmerman DC, Weaver-Smith S. Autonomous structural health monitoring system: a demonstration. *Proceedings of the AIAA/ASME Structures, Structural Dynamics, and Materials Conference*. Salt Lake, April 1996.

[3] Beral B, Speckman H. Structural health monitoring for aircraft structures: a challenge for system developers and aircraft manufacturers. *4th International Workshop on Structural Health Monitoring*, Stanford, CA, September 2003.

[4] Roach D. Use of distributed sensor systems to monitor structural integrity in real-time. *Quality, Reliability, and Maintenance in Engineering*. Professional Engineering Publishing: Oxford, 2004.

[5] Roach D, Kollgaard J, Emery S. Application and certification of comparative vacuum monitoring sensors for in-situ crack detection. *Air Transport Association Nondestructive Testing Forum*. Fort Worth, TX, October 2006.

[6] Wheatley G, Kollgaard J, Register J, Zaidi M. Comparative vacuum monitoring as an alternate means of compliance. *FAA/NASA/DOD Aging Aircraft Conference*. New Orleans, LA, September 2003.

[7] Kumar A, Roach D, Hannum R, In-situ monitoring of the integrity of bonded repair patches on civil infrastructures. *SPIE Smart Structures and Materials Symposium*. San Diego, CA, February 2006.

[8] Sun Y, Roach D, New advances in detecting cracks in raised-head fastener holes using rotational remote field eddy current technique. *American Society of Nondestructive Testing Research Symposium*. Albuquerque, NM, October 2005.

[9] Sun Y, Ouyang T. Application of remote-field eddy-current technique to aircraft corrosion detection. presented at *Tri-Service Corrosion Conference*. San Antonio, TX, January 2002.

[10] Roach D, Wanser K, Griffiths R. Application of fiber optics to health monitoring of aircraft. *Advanced Aerospace Materials and Processes Conference*. Anaheim, CA, June 1994.

[11] Udd E, Schulz WL, Seim JM, Haugse E, Trego A, Johnson PE, Bennett TE, Nelson DV, Makino A. Multidimensional strain field measurements using fiber optic grating sensors. *SPIE Proceedings* 2000 **3986**:254.

[12] Froggatt M, Moore J. Distributed measurement of static strain in an optical fiber with multiple Bragg gratings at nominally equal wavelengths. *Applied Optics* 1998 **37**:1741.

[13] Udd E, Kreger S, Calvert S, Kunzler M, Davol K, Usage of multi-axis fiber grating strain sensors to support nondestructive evaluation of composite parts and adhesive bond lines. *4th International Workshop on Structural Health Monitoring*. Palo Alto, CA, September 2003.

[14] Barshinger J, *GE Nondestructive Evaluation and Introduction to Capacitive Micromachined Ultrasonic Transducers*, General Electric presentation to Sandia Labs, April 2005.

[15] Plotnikov Y, Nath S. Real time imaging of subsurface flaws using the pulse eddy current array probe. *American Society of Nondestructive Testing Conference*. Atlanta, GA, November 2004.

[16] Kwun H, Kim S, Light G. Long-range guided wave inspection of structures using magnetostrictive sensor. *Journal of Korean Society of Nondestructive Testing* 2001 **21**:282–288.

[17] Elster J, Trego A, Jones M, Tulou P, Fitz-Patrick B, Perez I. Corrosion monitoring in aging aircraft using optical fiber-based chemical sensors. *FAA/NASA/DOD Aging Aircraft Conference*. Williamsburg, VA, September 2000.

[18] Jackson A, Schaafsma D. Aircraft fail-safe self monitoring system. *FAA/NASA/DOD Aging Aircraft Conference*. Williamsburg, VA, September 2000.

[19] Goldfine N. Magnetometers for improved materials characterization in aerospace applications. *Materials Evaluation* 1993 **51**:396–405.

[20] Goldfine N, Washabaugh A, Schlicker D, Sheiretov Y, Huguenin C, Lovett T (JENTEK Sensors), Roach D (Sandia Labs). Corrosion and fatigue monitoring sensor networks. *5th International Workshop on Structural Health Monitoring*, September 2005.

FURTHER READING

Giurgiutiu V, Redmond J, Roach D, Rackow K. Active sensors for health monitoring of aging aerospace structures. *International Journal of Condition Monitoring and Diagnostics Engineering*, February 2001.

Chapter 90
Military Aircraft

Mark M. Derriso[1], Steven E. Olson[2] and Martin P. DeSimio[2]

[1] *US Air Force Research Laboratory, Wright Patterson Air Force Base, OH, USA*
[2] *University of Dayton Research Institute, Dayton, OH, USA*

1 OVERVIEW

Military organizations are continuously evaluating different methodologies to reduce cost, increase availability, and maintain safety of current and future air vehicle systems. Recently, emphasis has been placed on the development of integrated systems health management (ISHM) technologies. These methods would be able to determine the health of an entire vehicle based on combined assessments of various vehicle subsystems. As necessary, the health management system could also interact with flight control systems to ensure mission success. We give a brief overview of ISHM methods in the first part of this article.

Since the vehicle structure is one of the most critical subsystems, structural health monitoring (SHM) techniques must be developed to enable implementation of the ISHM methods. To date, however, few fielded applications of SHM systems exist for aerospace structures (*see* **Commercial Fixed-wing Aircraft**; **Flight Demonstration of a SHM System on a USAF Fighter Airplane**; and; **Experience with Health and Usage Monitoring Systems in Helicopters**; **Validation of SHM Sensors in Airbus A380 Full-scale Fatigue Test**; **Aerospace Applications of SMART Layer Technology**). In the second part of this article, we outline the underlying reasons for such limited use, with a specific focus on applications for military aircrafts. The need for a business justification to field an SHM system is discussed, along with the technical feasibility of implementing such systems.

For at least certain aerospace applications, a strong business case can be made for implementing such a system. The SHM community has also made considerable efforts to address the technical challenges of implementation. Assuming that the business and technical challenges can be overcome, there are various

 ISBN: 978-0-470-05822-0.

near-term and far-term aerospace applications. We discuss potential applications of SHM systems in the last part of this article.

2 INTEGRATED SYSTEMS HEALTH MANAGEMENT

Recently, military organizations have focused their attention on maturing ISHM technology for potential implementation on current and future air vehicle systems. We define ISHM as any system that collects, processes, and manages health data to assess the current condition of the vehicle and determine its ability to perform a given mission. By knowing the condition of each air vehicle system in the fleet, we can effectively deploy systems for the appropriate mission, even in a degraded state.

As an example, consider a case where System "A" is damaged and currently operating at 80% of its full capability until repaired. We can still use System "A" to perform a mission that requires less than 80% of its full potential before being serviced. We can utilize the system with a high level of confidence because its health is known. As a result, we can increase the availability of air vehicle systems without increasing the risk of mission success.

Ideally, we will be able to create fully integrated health management systems, enabling affordable combat support, maximizing mission capability, and abating secondary damage. In order to achieve this, we must be able to accurately assess the condition of each individual vehicle subsystem so that the overall health and capability of the entire vehicle can be determined. One of the most critical components of a military vehicle is the structural subsystem, since a catastrophic failure could occur if the structure is not functioning properly. Therefore, the development of SHM systems is essential.

2.1 SHM system development

The process required to develop an SHM system for a specific aerospace component is illustrated graphically in Figure 1. For the first step in the process, we must identify the component or system of interest. Although this seems like an obvious step, it is a common pitfall to start SHM development with sensor system design.

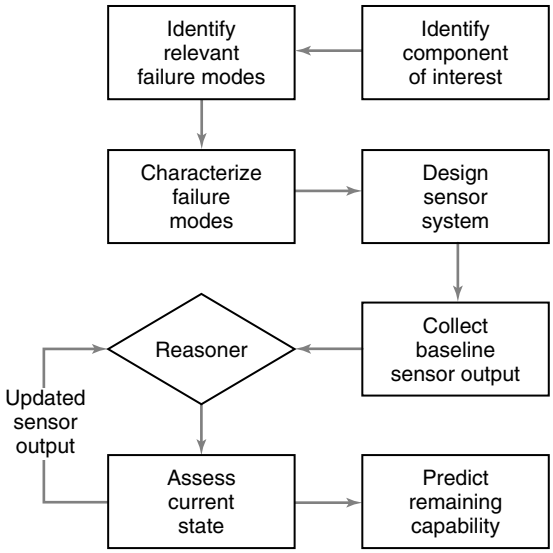


Figure 1. Process required for SHM system development.

For the next step, we need to identify the relevant failure modes of interest. These failure modes depend on the operating conditions and are unique to a specific component. For example, metallic components may experience failures due to cracking or corrosion. Composite parts may experience disbonds or delaminations. Components of both materials are susceptible to impact damage or fastener failures. We make an underlying assumption in this step that the SHM system will not be able to detect every single failure mode.

Once the component and relevant failure modes are identified, we must characterize the failures by investigating the physics of the contributing damage mechanisms. If possible, it is also useful to identify any precursors of damage. The definition of what constitutes damage is dependent on the component and application. For example, in an aircraft spar, failure may occur owing to widespread cracking. The precursors to such failure might include damage due to cracks that are small with respect to the component, but large with respect to the microstructure of the material. We identify precursors of damage to better understand which damage mechanisms we can sense.

At this point, we can design an appropriate sensor system. For most engineering applications, we cannot directly measure damage. Therefore, the

sensor network typically must use a combination of direct, indirect, and virtual sensors, in both on-board and off-board applications. Direct sensors measure the specific damage mechanism of interest, whereas indirect sensors measure another physical property that can be correlated to damage. Virtual sensors marry existing sensor data and numerical models to estimate damage. The application and the operational environment dictate the types and placement of the sensors we use. Sensor authentication techniques should be used to verify the integrity of the deployed sensors.

With a sensor system instrumented on the component, we must next develop a baseline for the sensor responses. The baseline should be completed, if possible, on components with minimal operational time. If we cannot do so, any previous operation of the components will add uncertainty to the output of the SHM system. The baseline sensor responses are fed into a component-level reasoner to provide a reference of the healthy state of the component.

After the component has been used for a given amount of time, updated sensor data is collected and fed back into the component-level reasoner. The reasoner takes the output of the sensors and utilizes the response to determine the current state of the component. The assessment can occur in real time, using on-board sensors or, intermittently, using off-board sensors. The interval of the updating differs between components and is determined by the importance of the sensed data and whether on-board or off-board sensors are utilized. In most cases, there will be some variability in the sensed data due to time-varying, nonlinear effects. Efforts should be undertaken to identify these sources of variability to avoid incorrect reasoner state identifications, including missed detections and false alarms. Both physics-based modeling and data-driven modeling are usually required to develop the reasoner. Physics-based models capture the anticipated damage mechanisms and failure modes, and data-driven models have the potential to capture unanticipated damage mechanisms.

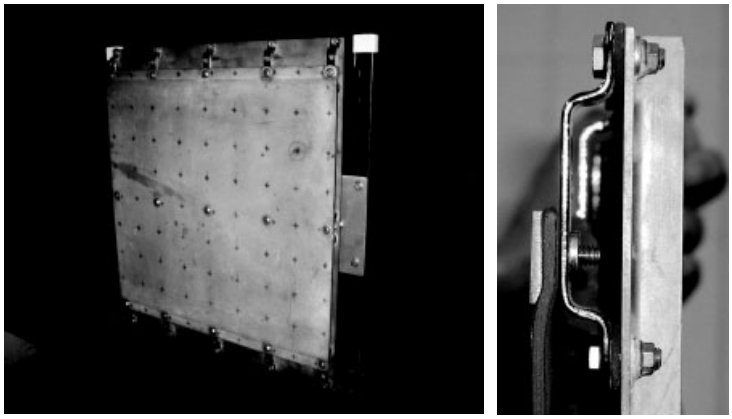
If possible, we would like to develop true prognostic systems that are capable of predicting the remaining capability of the component of interest. To predict the remaining capability, we require information regarding the loading profiles that the component is experiencing. These loading profiles can either be calculated from health and usage monitoring data, if available, or can be estimated using a realistic approximation of the mission profiles the component will experience. Using the mission loading profiles and advanced life prediction tools, we can typically predict the remaining capability.

2.2 SHM system development example

To illustrate SHM system development, we present an example for a specific aerospace structure (*see* articles in **Part 8** for more examples). To meet future requirements in the areas of control of space and global engagement, military organizations are working toward developing reusable space vehicles. In current conceptual models, these vehicles incorporate metallic or composite panels over the fuselage as part of a thermal protection system (TPS). If the panels become damaged, the vehicle's fuselage could be exposed to the environment, compromising vehicle integrity and ultimately jeopardizing mission safety and effectiveness.

As shown in Figure 1, the first step in the ISHM development process is to identify the item of interest. TPS structures of future military space vehicles are a good example of a structural system for which prognosis systems would be extremely beneficial. The TPS is one of a space vehicle's more vulnerable components owing to the likelihood of space debris strikes and the certainty of extreme thermoacoustic loading during launch and reentry. The TPS must be in good condition prior to launch owing to its critical role in protecting the vehicle's primary structure and subsystems.

We show a representative TPS panel in Figure 2. A carbon–carbon panel provides the foundation of the TPS. The panel is approximately 0.61 m square and 3.175 mm thick and is bolted to 15 evenly spaced brackets. The brackets are bolted to a 2.54 mm thick ribbed backing structure with 16.94 mm thick ribs. The backing structure represents the vehicle's fuselage. Figure 2 also shows a side view of the representative TPS article, illustrating the connection details of the panel, brackets, and backing structure. The brackets are fabricated from inhibited silicon carbide–carbon with a chromalloy coating. The ribbed backing structure is titanium.



(a) (b)

Figure 2. (a) Representative thermal protection system component and (b) side view showing fastening details.

For the next step in the process, we must identify the failure modes of interest. Damage modes identified for TPS structures include loose or missing fasteners, impact damage, delaminations or disbonds, and cracks. To improve maintainability and reduce turnaround times, military space vehicles will likely utilize mechanically attached TPS structures. In addition, it has been noted that approximately 70% of all mechanical failures occur in fasteners [1]. As a result, for this example, we focus our attention on detecting fastener failures.

The third step in the ISHM development process is to characterize the failure modes. The various TPS failure modes can have widely varying effects on the structural dynamics. Damage such as cracks have a highly localized effect on the structural dynamics, and the use of elastic wave damage detection techniques, such as Lamb waves (*see* **Fundamentals of Guided Elastic Waves in Solids**), may be suitable. As discussed above, our focus is on detecting fastener failure. Damage such as fastener failure induces a more global effect on the structural dynamics and, therefore, modal-based damage detection techniques (*see* **Modal–Vibration-based Damage Identification**) have been investigated. It may be difficult to identify precursors that indicate imminent bolt damage. However, we are exploring the capability to assess the level of bolt torque or preload so that any degradation is predicted as early as possible.

Designing a sensor system that can capture the relevant degradation mechanics is the next step in the process. During operation, the outer surface of the TPS panel will be at elevated temperatures, which precludes the use of most sensors on the TPS panel itself. As a result, we will most likely make measurements on the backing structure or brackets, where the temperatures are considerably lower. For our current studies, we have utilized four piezoelectric transducers-0.635 mm diameter, 0.25 mm thick lead zirconate titanate (PZT) discs-adhered in the valleys between the ribs of the backing structure. The transducers are attached on the side of the backing structure representing the inner fuselage of the aerospace vehicle, where temperatures should remain well within the operating range of the piezoelectric materials. This configuration represents realistic actuator/sensor placement, since piezoelectric devices fail at the high temperatures encountered on the outer surface of the vehicle. Piezoelectric sensors offer the advantages of light weight, low power consumption, operation over a wide range of frequencies, and the ability to operate as both actuators and sensors (*see* **Piezoelectricity Principles and Materials**).

We next need to obtain initial sensor input. For the representative TPS article, different structural conditions are obtained by loosening one bolt at a time from the carbon–carbon panel by a one-quarter turn. When all bolts are fully fastened, the structure is considered healthy. A bolt loose condition corresponds to a damaged state. For any of the damaged states, it would be advantageous to be able to both detect the presence of damage and to locate the damaged fastener. For the experimental

testing, an excitation signal is sent to one piezoelectric using a swept frequency sinusoid, ranging from 0 to 7000 Hz over 1.0 s. The vibration responses are recorded from the remaining three piezoelectrics. Over 100 rounds of data have been collected over a period of approximately four months, where each round corresponds to measurements at each of the different healthy and damaged structural conditions. We have used this data to train the component-level reasoner discussed in the following paragraph. The data also provides the baseline response of the healthy structure.

For the representative TPS article, we use a component-level reasoner based on statistical pattern-recognition techniques. Statistical pattern recognition considers the conditional probability density functions (pdfs) of measurements associated with each class to be identified (i.e., either healthy or one of the various damage states). The pdfs are determined from the training data collected during the previous step and used to form discriminant functions. We predict the state of the structure based on the largest discriminant function given a particular measurement. For most practical pattern-recognition tasks, we need multidimensional measurements to provide acceptable classification accuracy. Typically, we compute a number of potentially useful features and identify a subset of the most useful ones using a feature selection process. Features for the representative TPS article are based on the root-mean-square (RMS) energy over specified time periods (roughly corresponding to specific frequency ranges since the structure is excited using a swept sinusoidal signal) at the various sensors. For the representative TPS article, the accuracy of the component-level reasoner has been investigated using independent evaluation data collected in the laboratory. We have demonstrated an efficient reasoner with detection and localization accuracies of 99.9 and 99.3% respectively, and with a probability of missed detection of 0.1% and a probability of false alarm of 0.3%.

The last step in the process is to predict the remaining capability. To make an accurate prognosis of the remaining capability, we must know both the current state of the structure and the loading profiles that the structure will experience. The component-level reasoner discussed above should provide a reasonably accurate assessment of the current state of the TPS structure. Since the exact loading profiles that the structure will experience are uncertain, a prediction must be made on the basis of available loads measurements (or estimates) and anticipated mission profiles. When appropriate, the utilization of health and usage monitoring systems (HUMS) data will be considered.

To incorporate the SHM system into the ISHM framework, we need to output the prognosis data to a system-level reasoner. The system-level reasoner incorporates data from the prognoses of various components to determine the overall state of the vehicle. System-level reasoners typically involve several levels of integration. For example, prognoses of each of the TPS panels on the vehicle are integrated with each other and other structural components to provide an estimate of the state of the entire vehicle structure. The estimate of the state of the vehicle structure is further integrated with estimates of the state of the vehicle engines or other subsystems to provide a comprehensive estimate of the vehicle condition.

3 SHM SYSTEM IMPLEMENTATION

In the preceding section, we discussed how an SHM system was designed for an aerospace component. In fact, the potential use of SHM systems for aerospace structures has been investigated for more than a decade. However, there are few fielded applications [2]. In this section, we investigate the underlying reasons for such limited use. To be fielded, a framework must be established for implementing SHM systems. This framework would define the architecture of the SHM system and would include integration of the SHM system with other vehicle subsystems and overall fleet management methods. However, prior to establishing a framework, we must demonstrate a business justification for implementing SHM systems as well as technical feasibility. We discuss issues related to the business and technical challenges in the following paragraphs.

3.1 Business justification

From a business standpoint, the benefits resulting from an SHM system must outweigh the costs associated with its implementation. As we discuss below,

the potential benefits are largely in the areas of improved maintenance and repair. Additional performance benefits are anticipated once SHM systems are incorporated into the ISHM framework. Associated costs include both financial costs, such as those associated with development, installation, and operation, as well as physical costs, such as added weight or power requirements. We provide a brief cost–benefits analysis discussion in the latter portions of this section.

3.1.1 Benefits of SHM implementation

The potential benefits of implementing an SHM system are largely in the areas of improved maintenance and repair, and increased performance. As an air vehicle system increases in age, the operating and support costs also increase. We know that roughly 60% of the total life cycle cost of a system is used to sustain it. Similarly, aircraft maintenance and repairs represent about a quarter of a commercial fleet's operating costs [3]. One of the biggest cost drivers in a military air vehicle fleet is unscheduled maintenance. The problem is exacerbated since many military aircraft currently in use today are older. For example, the average US military aircraft is roughly 25 years old. To maintain safety, we need to perform increasing numbers of inspections, which results in a decrease in fleet availability.

To improve the current aircraft maintenance procedures, military organizations are exploring the use of condition-based maintenance practices. Using these practices, maintenance is performed on the basis of the current condition of the system instead of that based on a set inspection schedule. By migrating to condition-based maintenance practices, we hope to improve maintenance agility and responsiveness for quicker turnaround times, increased operational availability, and reduced life cycle total ownership cost [4].

3.1.2 Costs associated with SHM implementation

Implementation costs include the costs associated with developing, installing, and operating an SHM system. Even with mature SHM technologies, we will almost certainly need some development to adapt a particular technique for each unique application and to verify its functionality. Once an SHM system has been verified and approved for fleet-wide use, we also have an associated cost per vehicle to install the system. Additionally, there are operating costs associated with the SHM system, such as the costs of archiving SHM system data for historic use, upgrading the system with improved damage detection algorithms, or long-term maintenance or periodic calibration of the SHM system.

In addition to the monetary costs, there are other costs such as added weight or power requirements. Weight is a critical concern for aerospace vehicles. Therefore, the added weight of any SHM system including all hardware (sensors, cabling, processors, etc.) must be considered. Power requirements are an additional concern. Advanced avionics and other subsystems may require a significant portion of the available power on future air vehicles. Military space vehicles have their own unique limitations based on the available on-board power. We must consider all costs associated with SHM system implementation, including both the monetary and nonmonetary costs.

3.1.3 SHM implementation cost–benefit analysis

Few cost–benefit analyses, for military or commercial aircraft, have been discussed in the open literature. In one of the limited published studies, Kent and Murphy [5] analyzed the use of SHM systems for maintaining commercial transport aircraft. Their study showed that implementing SHM systems could result in a significant reduction in both maintenance and operating costs. They give as an example the implementation of an SHM system that could result in a 30 to 40% reduction in maintenance requirements. The initial implementation costs would be recovered in only two to three years. For larger and more complex structures, the time required to recover the initial implementation costs may be longer, but is still well within the life of the vehicle.

We can make a similar justification for military space vehicles. For such vehicles to be successful, launch costs must be reduced an order of magnitude below those of current launch vehicles. The key to reducing launch costs is to reduce the turnaround time. To meet turnaround goals, we need to significantly reduce the time required to assess the structural condition. More specifically, we need an automated

SHM system that can assess the health of the entire structure within hours of the completed mission and recertify the structure for flight. Thus, incorporation of SHM will be critical to the success of future launch vehicles.

It is likely that near-term applications will involve retrofitting existing air vehicles with SHM systems to detect damage in well-documented problem areas. We discuss potential aerospace applications of SHM in Section 3. These near-term applications have a high payoff in which the benefits of applying the SHM system significantly outweigh the associated costs. However, successful fielded applications are required to demonstrate utility and build confidence before SHM systems are widely accepted.

3.2 Technical feasibility

Regarding technical feasibility, we must address all issues related to SHM system development, implementation, and operation. Farrar and Worden [2] introduce some fundamental challenges of SHM systems, including the following: (i) small-scale damage must be detected in relatively large-scale structures; (ii) SHM systems must work in an unsupervised learning mode; and (iii) the robustness and redundancy of an SHM system must be ensured. The SHM community has made considerable efforts to address these challenges. We discuss these efforts below, along with examples from recent work.

3.2.1 Detecting small-scale damage in large structures

Typical aerospace structural damage includes cracking, corrosion, or fastener (e.g., rivet) failure, all of which occur in only a small fraction of the total area of the structure. Although the damage area is small, the consequences of reduced integrity in that area are large. Thus, SHM systems must be capable of detecting small-scale damage in relatively large-scale structures. Sensors play a critical role in addressing this challenge because they provide the raw data for classification algorithms. Sensor issues include identifying the types to use, determining the number required, and specifying the layout. We address each of these issues below.

The primary factor we must address in sensor selection is identifying sensing techniques that can detect the damage of interest. Considerable research has been performed over the past several decades to investigate various sensors (*see* articles in **Part 5**). Adams [6] gives a good overview of some of the available sensors and the advantages and disadvantages of the various types. We have investigated the use of many of these sensor types for SHM applications. A short list of sensor types considered includes piezoelectric, vacuum tube, acoustic emission, photoelastic, strain gauge, fiber-optic, and magnetostrictive sensors. Figure 3 shows examples of several of these sensor types. In addition to being capable of detecting the damage of interest, other concerns related to identifying appropriate sensor types include geometric constraints, operational and survival environments, and the coverage area.

The coverage area relates directly to the issue of defining the number of sensors required. Nondestructive evaluation (NDE) techniques to detect cracks or corrosion usually interrogate a relatively small area, typically only the area under the sensor. For example, scans using ultrasonic or eddy-current techniques

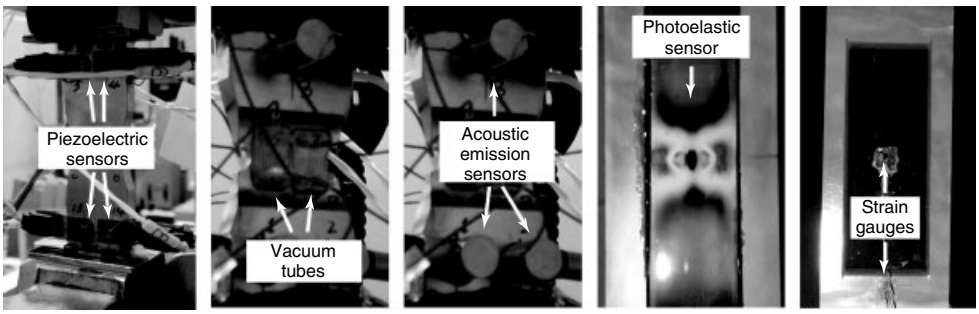


Figure 3. Examples of different sensor types investigated for SHM applications.

are performed by moving a sensor to various locations to investigate the region of interest. Obviously, these NDE techniques are not readily transferable to SHM systems where sensors would be permanently attached to, or embedded within, a structure. As discussed previously, there is a weight cost related to attaching sensors. We must consider the total added weight of any SHM system, including the associated cabling and other hardware. Even with the incorporation of wireless technology, the number of sensors should be minimized.

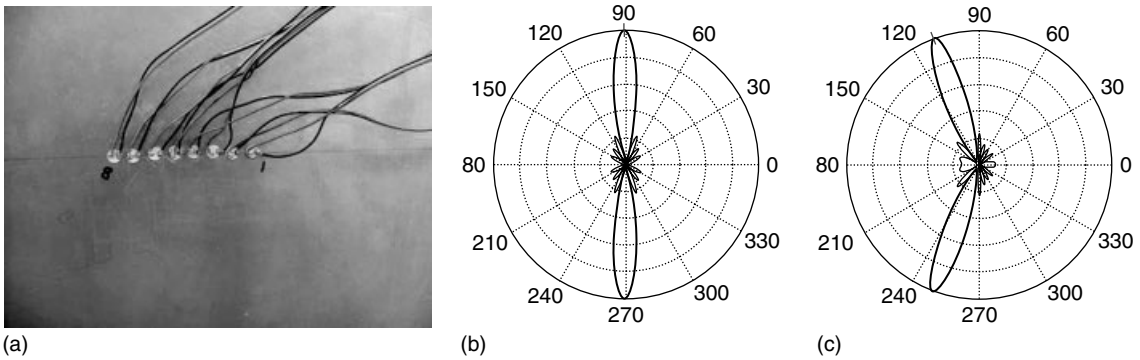
Techniques we have studied to minimize the required number of sensors include increasing the area of coverage for a given sensor (or set of sensors) and designing SHM systems into the structure to aid in state identification. One technique investigated to increase the area of coverage is to utilize elastic wave techniques with a phased array of piezoelectric sensors [7], such as that shown in Figure 4. In phased array techniques, we adjust the timing of signal data from a set of sensors to focus energy in a specific direction. By altering the phasing of individual sensors, we can sweep the "look angle" over a relatively large region of the panel.

An alternative approach for reducing the number of sensors is to include SHM in the structural design process. One possibility is to include noncritical failure modes of the structure as precursors to any potentially catastrophic failure modes. These "structural fuses" could be used as indicators for appropriate maintenance actions. Another possibility is to "design in" SHM where the structural design itself aids in identifying the structural health. McClung and Grandhi [8] demonstrated such a technique for a simple plate structure using a modal-based technique. In addition to reducing the number of sensors required, "designing in" SHM systems may improve performance and reliability, permit lighter designs with reduced reliance on structural redundancy for fail-safety, or reduce certification times.

The positioning of sensors is another factor relating to the detection of small-scale damage in large-scale structures. Sensor locations can be optimized to improve the detection of small-scale damage, or the critical damage areas can be better defined to alleviate the need to inspect large regions. Guratzsch [9] developed a methodology to optimize sensor placement. Although the optimization is computationally intensive even for the simple bolted plate structure examined, the methodology demonstrates the potential to improve damage detection by optimizing sensor placement.

In many cases, existing air vehicles have known *hot spots* where we anticipate a particular type of damage might occur or has consistently been observed in the field. Knowledge of the damage location and type enhances the damage detection capability. We can focus the sensors used for the SHM system on the particular area of interest and can design the system to detect expected damage types. When the "hot spot" is located in an area difficult to inspect by conventional NDE methods, such SHM solutions can provide substantial savings in time and money.

In summary, damage detection technologies are being matured which will enable widespread adoption of SHM systems. Considerable research has been performed related to detecting small-scale damage in



(a) (b) (c)

Figure 4. (a) Phased array of piezoelectric sensors and (b) and (c) gain plots showing the potential to focus energy in particular directions.

relatively large-scale structures. However, laboratory and academic successes need to be transitioned to actual applications.

3.2.2 Working in unsupervised learning mode

For many practical SHM system applications, data from all the relevant damage states of the structure is not available. Therefore, we face a second fundamental challenge in that the system must operate in an unsupervised learning mode. As an example, consider military space vehicles. It is likely that we will produce only a limited number of vehicles, and it may not be possible to collect damaged data from any of these vehicles. The vehicles will also operate in extreme environments, where the precise conditions may not be well quantified or even understood. Some potential SHM techniques suitable for working in an unsupervised learning mode, as well as the limited research in this area, are discussed below.

We have explored several techniques to account for the lack of damaged data during SHM system design, and to compensate for temporal or environmental changes. These techniques include detecting anomalies or outliers in the data or utilizing innovative "cancellation" techniques to eliminate changes in response signals that are not a result of damage. For example, Worden *et al.* [10–12] have utilized a novelty detection technique to identify damage in an aircraft wing. Kim and Sohn [13] have developed techniques to highlight nonlinear changes in piezoelectric sensor responses due to damage. It should be noted that both of these techniques require a threshold value to be defined beyond which damage is indicated. Since damaged response data may not be available, alternative means of establishing these threshold values are required. We should consider the use of physics-based models, historical data, or similarity to laboratory test results.

In general, we have found that greater attention has been focused on developing appropriate damage detection hardware than on developing algorithms to process the measured data. A literature review [14] found similar trends, as there is significantly less literature dealing with algorithm development than other SHM system considerations. Although maturing SHM hardware is necessary, algorithm development is critical if SHM systems are to be capable of autonomously providing accurate structural state assessments under varying conditions and in near real time.

3.2.3 Ensuring SHM system robustness and redundancy

Another fundamental challenge is ensuring the robustness and redundancy of an SHM system. Robustness relates both to the performance of the SHM system, and also to the survivability and applicability of the system. In terms of redundancy, the SHM system should still function if individual hardware fails or, at the very least, indicate that the system is not functioning correctly. SHM system robustness and redundancy are discussed below.

To be useful in practice, an SHM system must provide a high probability of damage detection at a low false alarm rate. The high probability of detection is required to ensure that critical damage is detected in a timely fashion. Low false alarm rates are needed to establish confidence in the SHM system and promote the adoption of such systems. Target values need to be established for each particular structural application. The use of physics-based modeling, including both theoretical and numerical calculations, will aid in improving SHM system accuracy. As an example, finite element simulations are currently being performed to assist in the creation of elastic wave techniques to detect damage for particular applications [15].

SHM systems also must be suitable for the operational and survival environments that a vehicle will experience. Improved air vehicle capabilities are constantly being demanded. As a result, vehicle structures are being subjected to, and must be capable of operating in, increasingly harsh environments. Since the objective of SHM is to evaluate the structural state of a component over its life, the SHM system should have a longer life than the vehicle or, as a minimum, longer than the monitored components. Ground and flight testing are required to verify the life of each of the components of an SHM system. As part of a "hot-spot" monitoring program, flight testing has been performed with piezoelectric sensors installed on an F-16 vehicle, and ground testing has been performed to evaluate the survivability of piezoelectric sensors when exposed to various environmental and chemical conditions. Figure 5 shows

(a) (b) (c) (d)

Figure 5. (a) Environmental test coupon undergoing (b) vibration testing, (c) fluid immersion testing, and (d) deicing fluid exposure testing.

a typical coupon used for the ground-based environmental testing, along with some of the tests performed.

After an SHM system has been developed and approved for fleet implementation, identical systems are manufactured and packaged for installation on each individual vehicle. Therefore, the system must be sufficiently robust to function similarly, regardless of the specific system hardware or the individual vehicle. The system should not be tied to a particular aircraft tail number or to a particular sensor serial number. Studies are currently underway to investigate the effects of materials and manufacturing variability on SHM systems. The primary objective of these studies is to improve the damage detection capability in the presence of anticipated variability; however, the studies also aid in quantifying the robustness of SHM systems with regard to installation on individual vehicles.

Lastly, an SHM system should detect failures in its own components and, if possible, still assess the structural state. For example, an individual sensor may fail in an SHM system. Through the use of substitute damage detection algorithms and a reduced set of responses, perhaps an assessment of the structural state can still be made at a lower confidence level. Less desirable alternatives are to "turn off" the SHM system if its hardware fails or to not provide a structural state assessment if the system cannot make a confident assessment based on the sensor response data [16]. In either case, operators must be notified of the health status of the SHM system itself. Since improved algorithms may become available or hardware may fail, the potential to upgrade or repair an SHM system during planned maintenance should be explored.

4 AEROSPACE APPLICATIONS OF SHM

For at least certain aerospace applications, a strong business case can be made for implementing an SHM system. However, the technical feasibility of these systems must be addressed. As discussed in the previous section, the SHM community has made considerable efforts to address these technical challenges. Assuming that all business and technical challenges can be overcome, we have identified potential near-term and far-term SHM aerospace applications. These potential applications are discussed below.

4.1 Near-term SHM aerospace applications

It is anticipated that near-term applications will involve retrofitting existing air vehicles with SHM systems to detect well-defined damage. These applications will likely focus on "hot spots", where the area and type of damage has been defined on the basis of problems observed in the field or anticipated to occur. In addition, the operational and survival environments will be known for such applications. These applications will have a high payoff in which the benefits of applying SHM systems significantly outweigh the associated costs. Candidate applications include critical components with short inspection intervals, components that are difficult to inspect using traditional NDE techniques, or components located where there is a high likelihood of secondary damage during inspection (e.g., where wing de-skinning is required). Initial near-term SHM system applications may also

involve only ground-based measurements with results processed in near real time, and early SHM systems may not be incorporated into an ISHM framework.

4.1.1 Structural "hot spots"

In certain instances, air vehicles have known areas with structural problems. To preserve safety and reliability, maintainers must inspect these problem areas at predefined intervals. In some cases, the problem resides in an inaccessible location, such as the upper or lower wing spar. Inspection therefore requires deskinning the wing. Figure 6 shows an example of such a structural component. SHM systems are being developed to detect and quantify cracks and corrosion in known problem areas of aircraft. The SHM systems will provide the aircraft maintainer with a diagnostic tool to enable early identification and assessment of structural degradation. This will allow us to identify and repair structural damage earlier than possible when relying on periodic inspections. Earlier repairs are typically less extensive and expensive, and can be phased into the aircraft flight schedule more easily.

4.1.2 Bonded repair integrity

Bonded repair technology is currently used on commercial and military air vehicles to extend the life of damaged structures. Figure 7 shows an example of a composite bonded repair patch applied to an aircraft fuselage. Laboratory tests have proved that a bonded repair could extend the life of a damaged structure by as much as a factor of 8. However, the nonrepaired inspection intervals of the damage under the patch are often still performed because the condition of the

Figure 7. Composite repair patch applied to an aircraft fuselage.

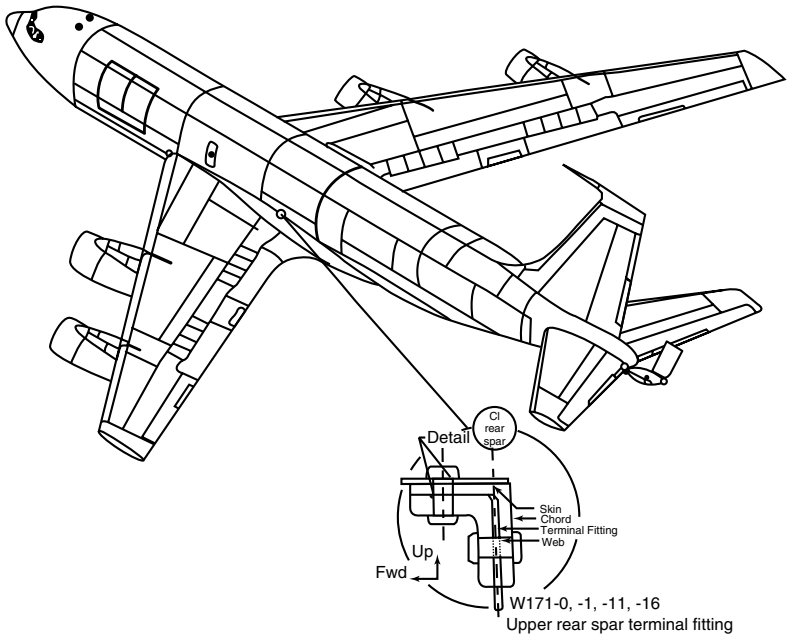


Figure 6. Example of a structural hot spot in an inaccessible location.

repair is unknown. Since inspections are performed at these nonrepaired intervals, the full benefits of the bonded repair technology are not achieved. A possible solution to this problem is to use an SHM system to determine whether or not the integrity of the repair is degrading. SHM systems are being developed to detect structural crack growth, bondline integrity, disbonding, and patch integrity of a composite bonded repair patch (*see* **Design, Analysis, and SHM of Bonded Composite Repair and Substructure**). With the incorporation of such techniques, we can perform condition-based maintenance rather than continuing to perform inspections at the nonrepaired intervals. The life of the damaged aircraft structure will be enhanced, structural safety and availability will be maintained, and maintenance costs will be reduced.

4.2 Far-term SHM aerospace applications

In addition to the near-term applications, SHM techniques offer unique benefits for future aerospace applications. Future vehicles will likely incorporate advanced materials and advanced structural concepts. In addition, these vehicle concepts may be exposed to harsher environments. We discuss the advantages of incorporating SHM systems into future air vehicle structures below. At the end of this section, we also discuss the possibility to "design in" SHM technology rather than retrofitting to existing structures.

4.2.1 Advanced materials

Historically, most current air vehicle structures are fabricated from metallic materials. The basic behavior of these materials is well characterized, although the fatigue behavior is still somewhat less understood. Demands for increased performance have led to greater use of composite materials. For example, carbon–carbon composite TPSs, such as the one shown in Figure 2, may be designed to perform at temperatures above 538 °C. In addition to the capability of some composites to perform well at high temperatures, composite materials typically offer advantages such as greater specific strength or stiffness than conventional metallic components. However, the basic behavior of composites is not nearly as well characterized as metals. As a result, there may be an increased need to monitor composite components to prevent catastrophic failure. In addition, advanced structural concepts are being proposed, which include electronics and/or transducers embedded inside composite materials. With such inclusions, it is expected that the performance of the material system is degraded. The degree of any degradation, however, is not well defined. Incorporating SHM systems into composite air vehicles structures may prove extremely beneficial, particularly since we may not have a good understanding of any potential structural degradation due to materials-related issues (*see* **Lamb Wave-based SHM for Laminated Composite Structures**).

4.2.2 Advanced structural concepts

In addition to the use of advanced materials, advanced structural concepts are being considered for future air vehicle applications to improve performance. For example, two advanced concepts under investigation are morphing structures and conformal array antennae. We provide a brief discussion of these techniques, along with the potential benefits derived from incorporating SHM technology, below.

Morphing refers to the ability to change a structure's geometry (e.g., reshape a wing profile) to enable efficient flight over a range of conditions without sacrificing performance. Morphing offers the potential to expand an aircraft's flight envelope and may offer significant vehicle mission-level benefits. Figure 8 shows an example of a morphing structure. Such structures rely on an accurate assessment of the current position or state of the structure. The state of the structure is assessed using sensors, and actuators are used to deform the structure into the desired geometry. In a sense, morphing structures are already very similar to SHM systems in that the state of the structure is sensed to detect the current shape or condition. We may be able to incorporate SHM directly into a morphing system, where a single sensing system is used to control structural morphing and to detect any structural damage.

Conformal array antennae refer to curved antennae composed of multiple individual antenna elements arranged on a curved surface. For improved radar performance, large, high-gain antennae are desired.

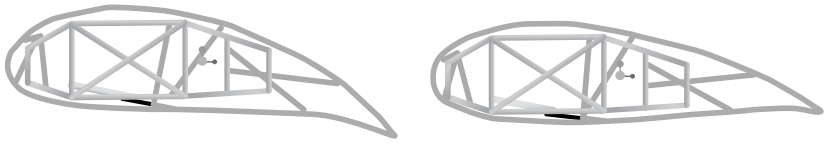


Figure 8. Example of a morphing wing concept to change from (a) a highly cambered wing design to (b) a lower camber design for reduced aerodynamic drag.

However, it is difficult to install such antennae on most current airframes without severely increasing the aerodynamic drag. One solution under consideration is to incorporate the conformal array antennae into load-bearing structures, such as the aircraft skins. Figure 9 shows an example of a conformal array antenna. We can fabricate such antennae from composite materials with the individual antennae elements embedded in the structure. As discussed in the advanced materials section above, however, the performance of the material system would likely degrade. An SHM system would prove useful in verifying the integrity of the structure. In addition, a modified SHM system could assess the health of the antennae to detect damage, such as broken electrical connections, which may inhibit the antennae performance but would not affect the structural integrity.

4.2.3 Extreme environments

Improved air vehicle capabilities are constantly being demanded. As a result, vehicle structures are often subjected to extreme environments and must be capable of operating in these environments. Examples of such structures include TPS for reusable launch vehicle applications and exhaust-washed structures. Both of these types of structures are exposed to high thermoacoustic loading. We discuss the incorporation of SHM systems into these types of structures below.

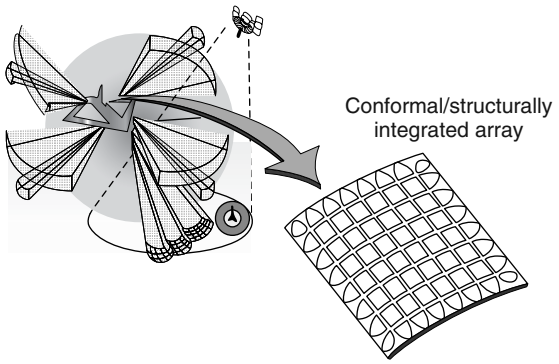


Figure 9. Example of a conformal array antenna concept.

As discussed previously, military organizations are developing reusable launch vehicles that utilize TPS panels. The TPS must be in good condition prior to launch because of its critical role in protecting the vehicle's primary structure and subsystems. For future launch vehicles to be successful, launch costs must be reduced an order of magnitude below those of current launch vehicles. The key to reducing launch costs is to reduce the turnaround time [5]. To meet turnaround goals, the time required to assess the structural condition needs to be significantly reduced. More specifically, an automated SHM system is needed that can assess the health of the entire structure within hours of the completed mission and recertify the structure for flight. Thus, incorporation of SHM for TPS structures will be critical to the success of such vehicles.

For improved aerodynamic performance and reduced observability, the engines of current and future air vehicles are highly incorporated into the airframe design. As a result, engine exhaust impinges on structures immediately aft of the engines, exposing these structures to high velocity, hot exhaust gases. Unique structural concepts and mounting arrangements have been investigated to reduce the thermomechanical loading; however, these structures still must operate in an extreme environment. It is critical that the structural integrity of exhaust-washed components is maintained, as damage could significantly degrade the vehicle's aerodynamic performance by increasing drag or affecting airflow over control surfaces. In addition, structural damage may result in the vehicle being easier to observe, or result in other problems, such as unanticipated corrosion due to hot exhaust gases inside various vehicle structures.

4.2.4 Inclusion of SHM in design process

To apply SHM systems to existing structures, we are required to retrofit the SHM solution to a previous design. This is often difficult since there are a number of constraints, such as geometric restrictions, which limit the design space for the SHM system. Incorporating SHM systems in the design process not only eliminates these constraints but may also offer other advantages. For example, we could design structures to be sensed by possibly including noncritical failure modes of the structure as precursors to any potentially catastrophic failure modes. These noncritical failure modes would serve as indicators for appropriate maintenance actions. We anticipate that the performance and reliability of SHM systems will benefit from inclusion in the design process.

In addition to improvements in the performance and reliability of the SHM systems, incorporating SHM in the design process offers potential improvements to the overall air vehicle system designs. For example, using the current structural design philosophy, incorporating SHM in the design process may permit reductions in factors of safety or lighter designs. SHM incorporation also may result in new structural design philosophies such as reliability-based, or probabilistic, designs and a reduced reliance on structural redundancy for fail-safety, with corresponding weight and cost savings. Finally, SHM incorporation may allow us to use new certification processes, which potentially could lead to reduced certification times.

REFERENCES

[1] Simmons WC. Bolt failure studies at Aberdeen proving ground. *Proceedings of the International Conference and Exposition on Fatigue, Corrosion Cracking, Fracture Mechanics, and Failure Analysis*. Salt Lake City, UT, 1986.

[2] Farrar CR, Worden K. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A* 2007 **365**(1851):303–315.

[3] Roach D, Rackow K. Health monitoring of aircraft structures using distributed sensor systems. *Proceedings of the 9th Joint FAA/DoD/NASA Conference on Aging Aircraft*, Atlanta, GA, 6–9 March 2006.

[4] Derriso MM, Calcaterra JR, Olson SE. Integrated systems health management: enable technology for effective utilization of air vehicle systems. In *Materials Damage Prognosis*, Larsen JM, Christodoulou L, Calcaterra JR, Dent M, Derriso MM, Hardman WJ, Jones JW, Russ SM (eds). TMS: Warrendale, PA, 2005.

[5] Kent RM, Murphy DA. Analyzing the cost/benefit of the use of a structural health monitoring system. In *Structural Health Monitoring: The Demands and Challenges*, Chang F (ed). CRC Press: Boca Raton, FL, 2001.

[6] Adams DE. *Health Monitoring of Structural Materials and Components: Methods with Applications*. John Wiley & Sons: Chichester, 2007.

[7] Olson SE, DeSimio MP, Derriso MM. Beamforming of lamb waves for structural health monitoring. *Journal of Vibration and Acoustics* 2007 **129**(6):730–738.

[8] McClung AJW, Grandhi R. Structural health monitoring to detect the location of fastener failure in thermal protection systems. In *Structural Health Monitoring 2005: Advancements and Challenges for Implementation*, Chang F (ed). DEStech Publications: Lancaster, PA, 2005.

[9] Guratzsch RF. *Sensor Placement Optimization Under Uncertainty for Structural Health Monitoring Systems of Hot Aerospace Structures*, Ph.D. Dissertation. Vanderbilt University, May 2007.

[10] Worden K, Manson G, Allman D. Experimental validation of a structural health monitoring methodology: Part I. Novelty detection on a laboratory structure. *Journal of Sound and Vibration* 2003 **259**(2):323–343.

[11] Manson G, Worden K, Allman D. Experimental validation of a structural health monitoring methodology: Part II. Novelty detection on a Gnat aircraft. *Journal of Sound and Vibration* 2003 **259**(2): 345–363.

[12] Manson G, Worden K, Allman D. Experimental validation of a structural health monitoring methodology: Part III. Damage location on an aircraft wing. *Journal of Sound and Vibration* 2003 **259**(2): 365–385.

[13] Kim SB, Sohn H. Instantaneous reference-free crack detection based on polarization characteristics of piezoelectric materials. *Smart Materials and Structures* 2007 **16**(6):2375–2387.

[14] Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR. *A Review of Structural Health Monitoring Literature: 1996–2001*, Report LA-13976-MS. Los Alamos National Laboratory, 2003.

[15] Olson SE, Leonard MS, Malkin MC. Analytical modeling to develop SHM techniques for aircraft 'hot spots'. *Proceedings of the 6th International Workshop on Structural Health Monitoring*. Stanford, CA, 11–13 September 2007.

[16] DeSimio M, Olson S, Derriso M. Decision uncertainty in a structural health monitoring system. *Proceedings of the SPIE 12th Annual International Symposium on Smart Structures and Materials*. San Diego, CA, 6–10 March 2005.

Chapter 100

Operational Loads Monitoring in Military Transport Aircraft and Military Derivatives of Civil Aircraft

Len Meadows[1], Steve Reed[2] and Mike Duffield[2]

[1] *Transport and Surveillance Aircraft, Air Vehicles Division, Defence Science and Technology Organisation (DSTO), Fisherman's Bend, VIC, Australia*
[2] *QinetiQ, Farnborough, UK*

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

The principal focus of military organizations is the conduct of military operations. In the conduct of these operations, there is an expectation that the risks to military personnel and to the general public will be constrained to an acceptable level. Most modern military aviation organizations have developed an extensive framework of airworthiness management in order to ensure that military aircraft operations meet these risk expectations. An essential element of this airworthiness management framework is that of structural usage monitoring or individual aircraft tracking. In the context of this article, structural usage monitoring encompasses the collection of operational structural usage data, the analysis of that data to produce information on fatigue life consumption, and the provision of that information to decision makers responsible for the ongoing airworthiness of military aircraft. Operational loads monitoring or measurement (OLM) is considered by many military airworthiness agencies as an essential element in the ongoing airworthiness management of an aircraft type. OLM is primarily a substantiation activity that may encompass the following:

- the substantiation of design and qualification fatigue usage spectra;

- identification of local stresses or strains in a structural feature;
- substantiation of structural monitoring or fatigue monitoring system;
- identification of additional monitoring requirements;
- capture of fatigue test spectra;
- identification of highly damaging activity or maneuvers;
- provision of data for investigations of structural issues.

The philosophy and conduct of OLM programs is outlined in **Loads Monitoring in Aerospace Structures** and sensor technologies are described in **Operational Loads Sensors**; **Nondestructive Evaluation/Nondestructive Testing/Nondestructive Inspection (NDE/NDT/NDI) Sensors—Eddy Current, Ultrasonic, and Acoustic Emission Sensors**; **Eddy-current *in situ* Sensors for SHM**; **Fiber-optic Sensor Principles**; and **Directed Energy Sensors/Actuators**. Furthermore, although military usage papers dominate the loads monitoring literature (*see* **Fatigue Monitoring in Military Fixed-wing Aircraft**; **Agile Military Aircraft**; **Flight Demonstration of a SHM System on a USAF Fighter Airplane**; **Health and Usage Monitoring Systems (HUM Systems) for Helicopters: Architecture and Performance**; and **Aerospace Applications of SMART Layer Technology**), there is a growing realization of the importance of such programs in the civil arena and this is reflected in **Commercial Fixed-wing Aircraft**; **History of SHM for Commercial Transport Aircraft**; **Video Landing Parameter Surveys**; and **Landing Gear**. Within this article, the regulatory basis of structural usage monitoring and OLM are described and examples of the application of OLM for military transport aircraft and military derivatives of civil aircraft across several air forces are presented.

2 STRUCTURAL USAGE MONITORING REGULATORY REQUIREMENTS

The requirement for structural usage monitoring is articulated in key policy and guidance documents of many military aviation organizations including [1, 2]. Structural usage monitoring requirements on a particular aircraft type have their genesis with the airworthiness standards and structural life assessment methodology used in the aircraft design and the ongoing airworthiness management strategies for the aircraft type. Key structural elements will degrade with use through the process of fatigue. In the design phase of the aircraft, predictions are made of this structural degradation and strategies formulated to constrain the risk of structural failure to an acceptable level through programmed replacement or inspection and repair. Inherent in these structural lifing predictions are many assumptions including the intended role and configuration of the aircraft, the operational usage it will be subjected to, and hence the expected loading spectra at the key structural elements. The lifing of some structural elements may be relatively insensitive to variations in operational usage while others may be quite sensitive to variations in a complex array of loading actions.

The principal aim of structural usage monitoring is the confirmation that the individual aircraft remains within the extent certification basis and the implicit level of risk associated with that certification basis. In order to meet this aim, a structural usage monitoring system needs to capture in-service data that the lifing predictions are sensitive to. This data then needs to be substituted for the assumed data used in the original lifing calculations and those predictions repeated, thus providing a more realistic ongoing assessment of the degradation of the structure. The ensuing ongoing airworthiness management program for the aircraft then needs to be adjusted for this updated assessment when a variance is found.

3 COMPLEMENTARY REGULATORY REQUIREMENTS—STRUCTURAL CONDITION MONITORING AND OPERATIONAL LOADS MONITORING

Structural usage monitoring is but one element of the overall ongoing fatigue management construct for military aircraft. Mechanisms also need to be in place to evaluate and ensure that structural usage monitoring is effective and adequately sensitive to

changes in role and configuration. This is typically accomplished via a combination of structural condition monitoring and OLM.

As the structure degrades through the process of fatigue, this degradation becomes evident through the manifestation of fatigue cracks. On multiload path structures typically found in transport-type aircraft, structural inspections are targeted at the principal structural elements that have been predicted to degrade. The timing of these inspections is based on a combination of the need to detect the cracking before it compromises on the ability of the structure to withstand the expected loading and on the statistical probability of the cracks being present. The ongoing structural usage analysis has a direct impact on the predictions of when the cracking is likely to be present as well as the rate at which these cracks are predicted to grow under in-service conditions. Feedback from the structural inspections, for both positive and negative results, is essential to confirm that the predictions from the structural usage analysis are reasonable. In setting the initial inspection thresholds and intervals, there has to be a compromise between ensuring an acceptable level of safety and minimizing the economic and capability burden of performing the inspections. When inspections are first carried out, there is thus some probability of finding a crack as well as a reasonable probability of not finding a crack. However, cracking should not be initially prevalent; if it is then there has been a breakdown in the process, possibly in the monitoring of the in-service usage, or in the life assessment resulting from that usage.

Mil-Std1530C [2] refers to structural usage monitoring as individual aircraft tracking, as it is mandated to occur on every aircraft in the fleet. Tracking each individual aircraft comes at a cost and as such there is always pressure to minimize this cost. The simpler the system, and the smaller the list of items to be monitored, the lower the cost in terms of system design, data processing, data archiving, and reporting. The principal goal is thus to field a system that is as simple as possible while meeting the data gathering needs of the aircraft. These systems also need to be in place before the aircraft enters service, and so system design development takes place during the design development of the aircraft itself. In this environment there is a significant risk that not everything that affects the life of the aircraft or its principal structural elements will be comprehensively understood when the design specification of the structural usage monitoring system is effectively frozen. In recognition of this situation, a number of major military airworthiness authorities have mandated that a more comprehensive monitoring approach be taken on a subset of the fleet. In the United Kingdom, OLM is required, while in the United States the equivalent program is that of the loads/environment spectra survey. In an OLM program, approximately 10% of the aircraft fleet is fitted with sensors and recording systems to gather external loads and environmental data. These results are subsequently used to confirm that the structural usage monitoring program provides adequate coverage, is sufficiently sensitive to changes in configuration and role, and that the algorithms used to convert the gathered usage data into life assessments are of acceptable fidelity. These programs are also useful in collecting detailed external and internal loading data for the investigation of specific problems with an aircraft type, for use in assembling detailed loading spectra for tests, and for gathering environmental data such as gust exceedance data or investigating wake turbulence during in-flight refueling operations.

4 CERTIFICATION ISSUES

Structural usage monitoring and OLM are requirements that are being demanded by military airworthiness agencies, and as the outputs are used in making executive decisions affecting ongoing airworthiness, the structural usage and OLM systems (including the data analysis and information reporting function) are subject to certification within the overall airworthiness management construct for that aircraft type. Australian regulations [3] define certification as "the end result of a process which formally examines and documents compliance of a product against predefined standards to the satisfaction of the certification authority". Despite the need for structural usage and OLM being mandated by many military airworthiness authorities, there is no agreed standard that this monitoring is required to conform to. Without defined standards, certification is somewhat problematic.

Guidance material on conduct of OLM programs and the substantiation of structural usage monitoring

systems is evolving as these systems become more common place. At present, in [1, 4] we have a modicum of guidance material for the certification or substantiation of usage monitoring systems using an OLM program. Australian regulations [5] have attempted to provide some working-level guidance from recent experiences. Rather than focusing on certification to a particular standard, [5] proposes a generalized specification for OLM and structural usage monitoring along with some validation requirements aimed at assessing that the system is fit for its intended purpose. As a comprehensive set of requirements has yet to receive codification, each program is inevitably reliant upon the detailed knowledge resident in the few individuals who did it last time, if they are still available. Consequently, each of these programs varies significantly in content and has significantly different degrees of success. Recently, the United Kingdom has produced a guidance document [6] developed within the UK Military Aircraft Structures Airworthiness Advisory Group, covering the philosophy and the conduct of OLM programs.

The following sections contain brief OLM case studies from Royal Australian Air Force (RAAF) C130J-30, KC30B tanker, and the Wedgetail airborne early warning aircraft, and from the Royal Air Force (RAF) Dominie TMk1 OLM program.

5 OPERATIONAL LOADS MONITORING OF A MILITARY TRANSPORT AIRCRAFT—THE RAAF C-130J-30

When the RAAF contracted to acquire 12 Lockheed C-130J-30 aircraft in the late 1990s, there was an extremely loose functional requirement included in the acquisition to fit an "OLM system" to two aircraft. This particular aircraft purchase was on an accelerated timescale. There was subsequently no dialogue between the end users of the data from these systems and the system designers until the aircraft were presented for military-type certification. At that time, there were no detailed requirements to satisfy, no agreed path to certification, and no clear certification standard that had to be met. There was high risk that a successful outcome would not be achieved for all parties. That risk was realized.

The implied requirement for the end use of the OLM data at that time was to provide ongoing validation of the structural usage monitoring system for the aircraft. The structural usage monitoring system delivered with the aircraft was also a first-of-type design for this model, with very little disclosure of the software. The manner in which the operational loads data could be used to validate the structural usage monitoring system was thus initially poorly defined, as the structural usage monitoring system was still under development. The resultant onboard elements of the OLM system comprised a suite of nine strain gauge bridges feeding a recorder with a 1-MB capacity. The locations of the strain gauges are shown in Figure 1.

The lack of communication and the accelerated timescale had a profound impact on the design of this system. Lockheed sensibly chose gauge locations that mirrored previous loads gathering programs that they had experience with and were known to respond reasonably well to the dominant loading actions. To meet the schedule, Lockheed sourced a readily available recorder. The memory limitations of this recorder then dictated what data could be captured in a typical flight. The result was a capability to capture time-stamped peak valley data only from the strain gauge bridges with no flight parameter data. Initially, the peak valley discriminate level was set at a level thought to be reasonable and which would capture significant loading events. Figure 2 shows the output from the strain gauge bridges for a flight with the initial discriminate level of 2100 psi (~14.5 MPa).

When the outputs of a series of flights were compared, the only perceived use of this data was to rank the flights in the order of comparative severity (with very coarse granularity). Subsequent attempts to reduce the discriminator to obtain greater fidelity resulted in the memory being filled to capacity in something less than 1 h on a typical mission with any low-level flying content. A complicating factor with this system design was that the recorder was not a standalone system and received data from the mission computer to zero the bridges at system startup and to obtain the time-stamp information. To make any changes to the software meant opening up the aircraft mission computer software code at enormous expense. The inherent limitations with the system and the projected costs to bring the system to a minimum level of functionality were

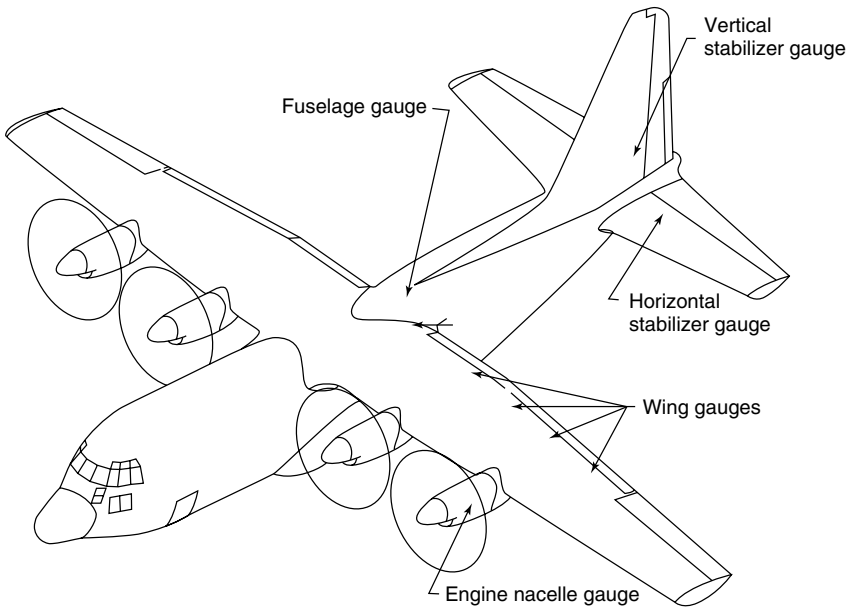


Figure 1. RAAF C-130J-30 original OLM gauge suite.

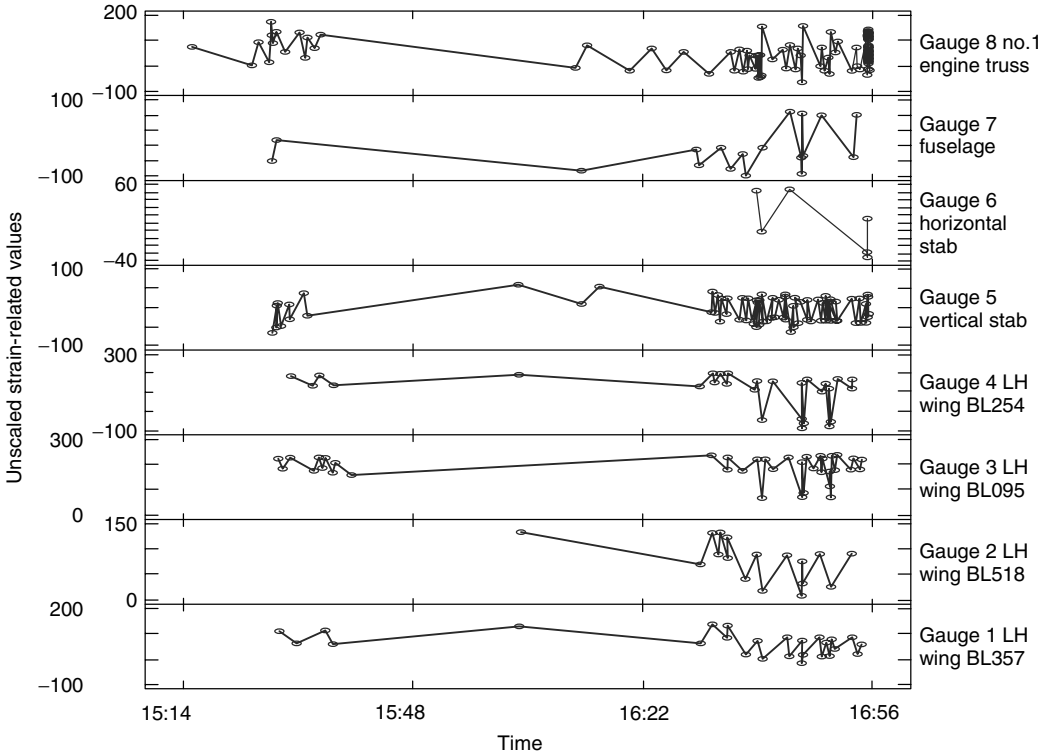


Figure 2. Strain gauge bridge outputs original RAAF C-130J-30 OLM.

considered to be cost prohibitive and the system was abandoned. However, the need for an OLM still existed.

At the same time that the RAAF was purchasing the C-130J, the RAF was also purchasing the aircraft. The Air Force decided to share experiences and collaborate on a number of issues to their mutual benefit. Two significant areas relevant to this discussion was the decision to conduct a collaborative full-scale fatigue test program of a new build C-130J wing and to conduct separate but complementary OLM programs from which data could be harvested to both validate the structural usage monitoring systems and to build the load sequence for the fatigue test. The fatigue test load sequence was considered to be the first priority.

The OLM system that has ensued was designed by Marshall Aerospace and is built upon the highly successful ACRA Control KAM-500 recorder, several modules of software from nCode International and from Lockheed, and a significant number of strain gauges and flight parameters. The RAAF version of the OLM system has 189 strain gauges, many of which are wired into loads bridges, and harvests 99 parameters directly from the aircraft data buses. All data are continuous time histories. Recording rates for the strain gauges are 40 or 80 samples per second (depending on the location and the frequencies of interest). The parameter data captures data at the same frequency that this data appears on the aircraft buses (frequencies from 0.25 to 20 Hz). Figure 3 shows the basic layout of the wiring looms and gauges. As the focus is on the building of the fatigue test spectrum, the wing receives the greatest coverage of strain gauges. Gauges are also located at "hot spots", which represent key areas targeted by the current safety-by-inspection program and which have had significant fatigue cracking in-service on previous model C-130 aircraft.

In order to convert the harvested strain bridge outputs into loads, each bridge has been calibrated *in situ* through the application of loads onto the aircraft. One hundred and thirty unique load cases were applied to the aircraft including a combination of fuselage pressurization and wing loading to achieve this (Figure 4).

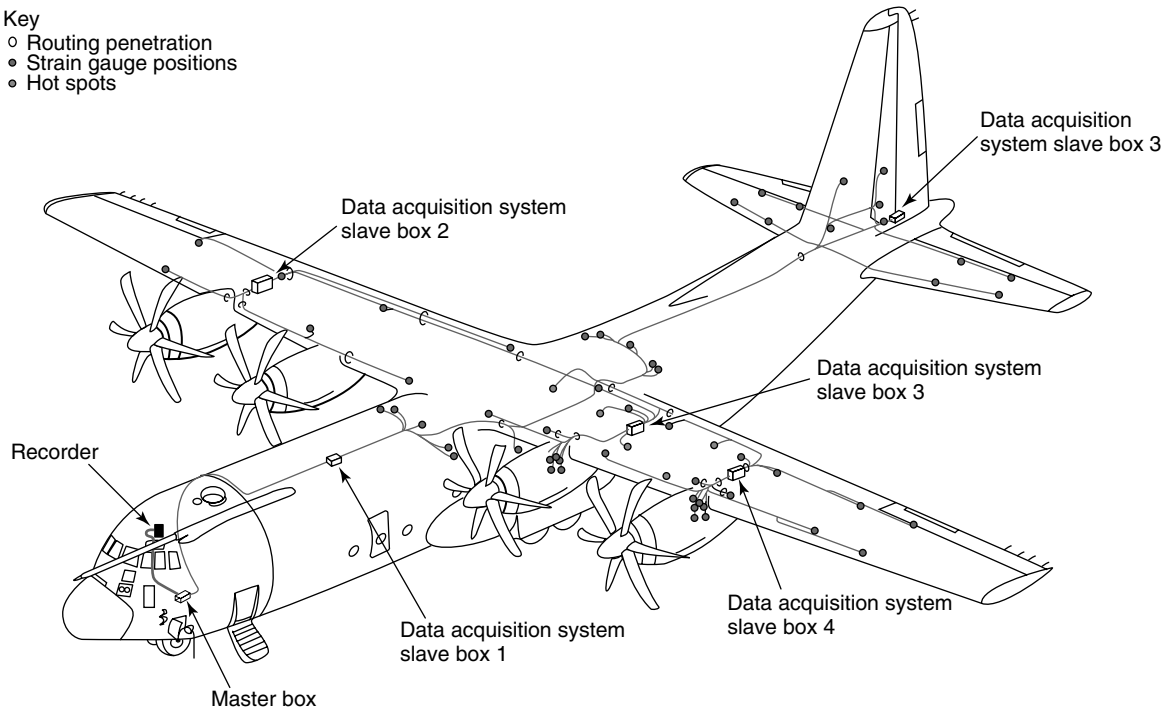


Figure 3. Layout of RAAF C-130J OLM.

Figure 4. RAAF C-130J OLM calibration.

From these calibration cases, load equations have been developed to calculate external loads at various stations from the strain gauge outputs. The sensibilities of these equations were then validated by a combination of comparisons of loads generated in a series of maneuvers in dedicated flight tests with computational fluid dynamic calculations and with original design loading data. Figure 5 shows a sample plot from this process, which led to a review of the loads equations at the wing root.

Data husbandry in a program of this nature becomes extremely important. Figure 6 shows a sample plot from four of the channels for a 3-h flight. A flight of this duration will typically produce 1 GB of data, when the data are processed into engineering units and then into loads. For the two RAF aircraft fitted with the OLM system, the data collecting period has been less than three years and already the data generated has exceeded 1 TB.

Data checking, cataloging, and archiving demand a significant amount of automation and oversight. Failure to identify and resolve data integrity problems in a timely fashion can very quickly lead to many flights where data are corrupted and the opportunity to harvest important data is lost.

The initial phase of the C-130J-30 OLM program has been a success, thanks to the efforts of a significant number of personnel not the least of which are those service men and women who download the data from the aircraft and forwarded it on for processing. From the pool of data available, a significant number of unique flights were chosen to assemble into a test load sequence representing compromise spectra of RAF and RAAF flying based on usage monitoring statistics from those respective fleets.

5.1 Challenge of military derivatives of civil transports

Civilian aviation regulations do not mandate the need to monitor the structural usage of aircraft in other than the most simplistic terms, often only requiring the recording of flying hours, flight cycles, and landings. The military and civilian requirements with respect to structural usage monitoring differ for a large number of reasons that are beyond the scope of this article; however, increasingly, the transport aircraft fleets of military organizations are an amalgam of both purpose-built military aircraft and civilian aircraft adapted for military applications. Some of these aircraft are used in roles very similar to their civilian counterparts, but others have special equipment fitted and are operated in roles and flight regimes that are fundamentally different from that for which they were designed and for which they had received civil airworthiness certification. This poses a number of challenges for military aviation organizations both in initial airworthiness certification and in ongoing airworthiness management. When talking with the

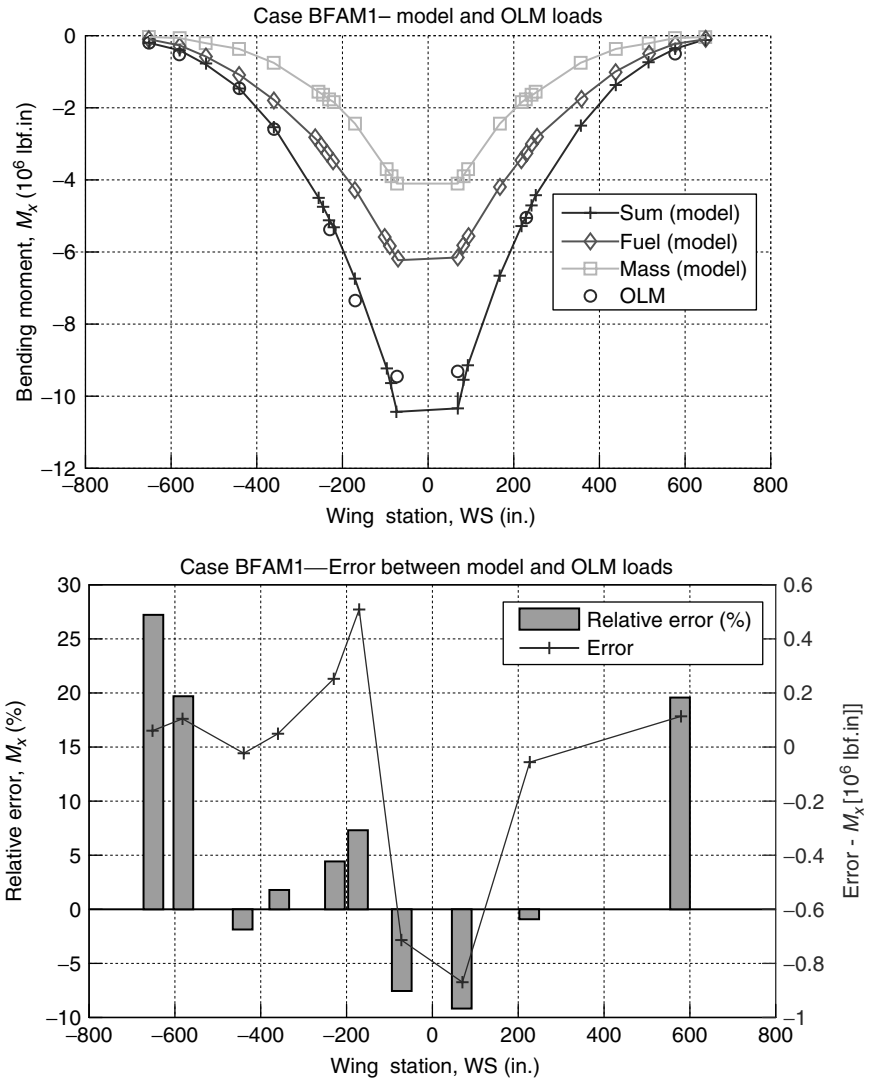


Figure 5. RAAF C-130J sample CFD-OLM comparison.

manufacturers of civil transport aircraft about certification of structural usage monitoring systems or OLM systems, the initial reaction is generally one of confusion as these things are outside of their experience. This confusion is increased when they ask what standard they have to meet for certification and the answer is that there is no standard. Concerns expressed by civilian manufactures include cost and schedule risk associated with embarking on system design outside of their experience as well as exposure of proprietary design data.

Australia currently has two programs where derivatives of civilian transports are being purchased and adapted for military use. One program is the Boeing B737-based airborne early waning and control aircraft named *Wedgetail*, the second is the EADS CASA KC30B, which is an A330-based air-to-air refueller. Only small numbers of these aircraft are being purchased, but nonetheless the RAAF has insisted on a structural usage monitoring system being installed on each aircraft and an OLM system on one aircraft of each type. Both these aircraft are still very much in

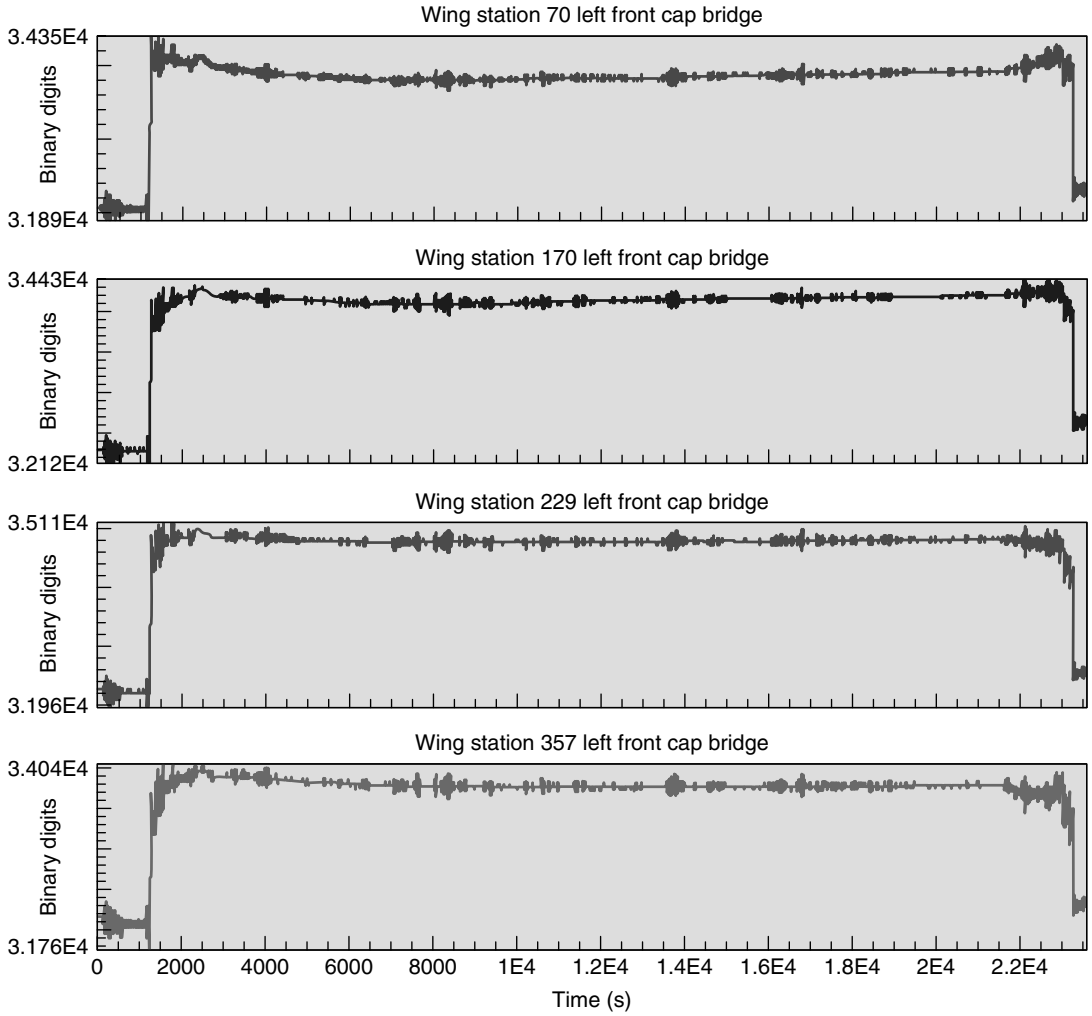


Figure 6. RAAF C130J sample plot from typical flight.

development; however, the system for the Wedgetail is now in the validation phase.

The Wedgetail system includes loads bridges located on the wing, horizontal tail, vertical fin, and multi element scanning array (MESA) antenna. These bridges have been calibrated by applying known loads and measuring the responses. Although it is not possible to discuss specifics of this program at this time owing to the proprietary nature of the data, the system has been successfully tested both on the ground and in flight with the outputs actually proving to be useful to Boeing in the development of that aircraft. The principal purpose of this system when it enters service will be to provide ongoing validation of the structural health monitoring system.

The KC30B OLM has a more comprehensive suite of strain gauge bridges than that of the Wedgetail. The intended use of this system is to provide data from which neural networks (*see* **Artificial Neural Networks**) can be trained for use in the structural usage monitoring system.

The link between the OLM systems and that of the structural usage monitoring systems for both the C-130J and the Wedgetail is that the OLM system is used to validate the structural usage monitoring system. In the case of the KC30B, the OLM system

acts as the foundation on which the structural usage monitoring system is built and provides the means to adjust that system as the aircraft configuration and role evolves.

6 OPERATIONAL LOADS MONITORING OF A MILITARY DERIVATIVE OF A CIVIL AIRCRAFT—THE RAF DOMINIE TMk1

The Dominie TMk1 is a military variant of the Hawker Siddeley HS125 business jet. The fleet is operated by the Royal Air Force in the navigator/weapons system operator training role. Although the aircraft are mid-1960s vintage, they have, in recent years, been fitted with an avionics upgrade and provide an up-to-date aircrew training environment. In order to provide realistic training for fast-jet aircrew, it is frequently necessary to operate the aircraft in low-level sorties over both land and sea.

Although the HS125 fatigue design philosophy was originally fail safe, the civil fleets moved to a damage tolerance clearance regime, following a detailed audit of the primary structure using fracture mechanics methods and civil-type usage spectra. However, the RAF Dominie was excluded from this analysis owing to uncertainties about the usage of the aircraft in the military training role. Therefore, a requirement to capture OLM data for the wing of the Dominie and compare the RAF stress usage spectrum with the damage tolerance stress spectrum was identified.

Funds available for such programs are increasingly tight and, consequently, a highly optimized OLM program was necessary. All data measurements and processes were subject to a rigorous justification to ensure the absolute minimum-cost option was developed. Design and installation of the data acquisition and recording systems was undertaken during 2004 and flight recording began in January 2005 and was complete in May 2006. Nearly 700 flying hours of data were captured from the instrumented aircraft. These data have provided the essential information needed to compare the military usage of the Dominie with the civil datum.

Within this section of the article, the conduct of the Dominie wing OLM program is described. Airborne and ground calibration methods are illustrated and the use of data confidence assurance techniques is detailed. Typical usage data, illustrating the effect of military profiles on the wing stress spectra of a civil-type aircraft, are presented. In particular, the change in gust environment with flight altitude over land or sea and the associated spectra severity is clearly detailed in the data. The output from the OLM data analysis program was stress cycle files for each flight; the United States Air Force (USAF) -developed "AFGROW" fracture mechanics software was then used to generate a direct comparison between Dominie and civil usage in terms of the crack growth lives for the critical wing section. This section concludes with a summary of the analysis process and presentation of results.

6.1 Instrumentation

Twelve electrical resistance strain gauge full bridges were bonded to the lower surface of the wing at approximately the 37.5% chord line (Figure 7), in locations coincident with the wing critical structural locations as identified by the civil damage tolerance analysis. A normal accelerometer (NZ) was mounted close to the center of gravity and an inclinometer was installed nearby to provide an indication of the aircraft's bank angle. Equivalent airspeed (EAS), pressure altitude, air temperature air data computer information, and data bus time were extracted from the aircraft ARINC429-10 data bus. Also, connections to the existing weight on wheels (WOW) switch, in the port undercarriage bay, and the No. 2 engine-driven generator were installed; this signal was used as a recording trigger signal. An ACRA KAM-500 data-acquisition unit, fitted with a Compact Flash® memory unit was installed in the aircraft. All data were sampled at 64 samples per second (S/s) with antialiasing filters set at 16 Hz.

6.2 Strain gauge calibration

Funding constraints precluded the use of a Skopinski-type [7] load calibration for the wing strain gauges. Hence, alternative methods had to be developed. A boundary element model was used to provide a transfer function from strains measured as close to

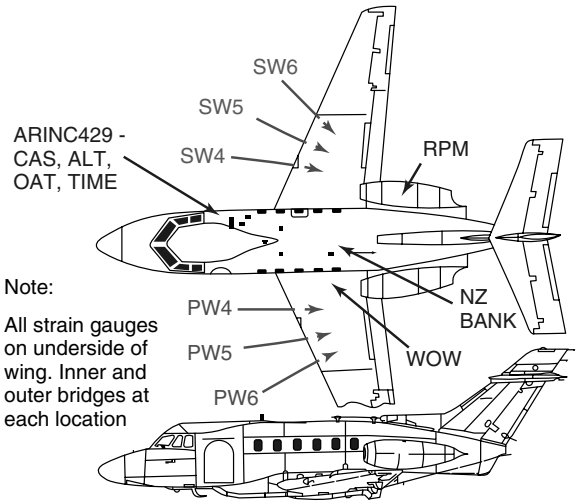


Figure 7. Dominie OLM instrumentation schematic.

the critical sections as stress concentrations would allow, to the equivalent field stresses required for input to the fracture mechanics model. Additionally, airborne calibration cases, cross-checked with ground cases, were used to identify the stress offset and hence baseline the strain data. The aircraft was flown at a series of steady-state airborne calibration points. For each point, the fuel masses were recorded, as this was not contained within the OLM instrumentation. Theoretical stress values for these flight conditions at the critical locations were produced. From this, flight datum stress offsets were calculated for the critical locations.

Confidence in the validity of this method was obtained by analyzing the stress values for 32 ground conditions for the critical locations at various aircraft masses. The offsets determined from setting the stresses to the airborne flight conditions were then compared with the theoretical values for these ground conditions. It was accepted that the ground and air conditions were not totally independent but the ground condition was dominated by the fuel load and the airborne cases were dominated by aerodynamic effects.

6.3 Data analysis

The Compact Flash® memory card and the cards were removed once per week and dispatched for analysis, with an accompanying copy of the flight record sheet. The data were subject to an array of anomaly and consistency checks before the strain data were reduced to cycle files for input into the AFGROW [8] program for comparison with civil spectra. This process is illustrated schematically in Figure 8.

6.4 Results

6.4.1 Data captured

Data recording began on January 14, 2005 and the recording program was terminated on May 8, 2006. During the period, 60 Compact Flash® downloads containing 319 sorties of data (equating to approximately 692 flying hours) were received, realizing the program data capture aim of over 300 sorties.

6.4.2 Data quality

All strain channel data were processed and identified as satisfactory. Several of the flights had been bisected due to the recording trigger going off-line as a result of the engine being shut down in flight; however, data loss was a matter of seconds and the separated elements of the flights were concatenated successfully.

The NZ data was the only channel where significant data anomalies were identified. Furthermore,

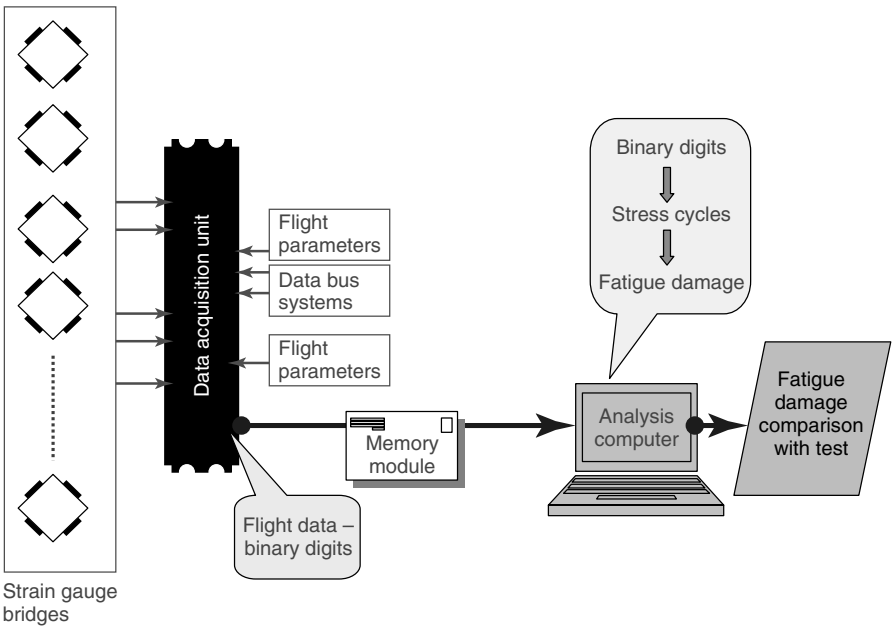


Figure 8. Dominie OLM data analysis process schematic.

different anomalies were present in different periods of data capture. Eventually, NZ data from 35 of 319 sorties were excluded from the analysis, although this was not highly significant as the NZ data were not the prime data source in this program.

6.4.3 Comparison of OLM data with fleet usage

By necessity, the OLM data captured represent a relatively small proportion of the data across the fleet and hence the usage of the OLM aircraft was compared with the fleet as a whole to ensure representativeness. RAF sortie types are described by Sortie Profile Codes (SPCs) and a comparison between the SPCs flown by the OLM aircraft and the fleet is reproduced at Figure 9. It is clear from this chart that the distribution of OLM SPCs is very similar to that of the SPCs flown by the fleet over the same period.

6.4.4 Strain data flight datum rechecks

Rechecks of the strain gauge datum values were undertaken periodically during the program to identify whether any drift was apparent in the strain gauge outputs. This was undertaken by matching flight conditions to those used in airborne datum offset calculation. Stress values within 3% of the original datum were identified for each of the six recheck sorties. Given a 2% acceptable variation in the flight parameters used to define the flight condition, this was considered to show a good agreement.

6.4.5 Strain data confidence check with neural network prediction

A further measure to provide confidence in the wing stress outputs was undertaken using a structural health and usage neural network (SHAUNN) (*see* **Artificial Neural Networks**). In essence, the SHAUNN was given stress and flight parameter time history data from three relatively severe flights early in the program. Using these data, the SHAUNN learned the mapping relationships between flight parameters and stresses. Once trained, the SHAUNN model was used to predict stresses from the flight parameters using sorties throughout the program. If the accuracy in predicting stresses from flight parameters remained fairly constant for flights captured throughout the program, then confidence in the OLM-derived stress values should remain high. The Pearson correlation coefficient achieved from the

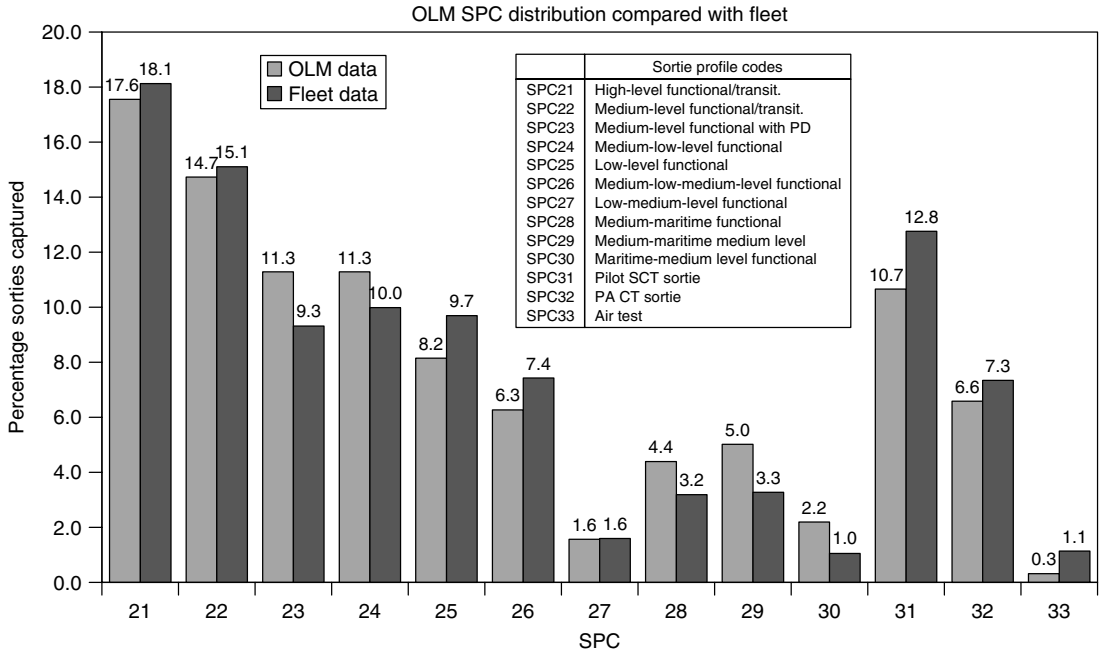


Figure 9. Dominie OLM SPC distribution compared with fleet.

training data was $R = 0.998$ and this was also the average correlation coefficient for the 80 sorties tested throughout the program. An example of the predicted and expected (from the SHAUNN model) stress time histories for an OLM sortie for location PW5 is reproduced in Figure 10. Additionally, a simple fatigue damage calculation was undertaken using the stress output from the SHAUNN and the stress from the strain gauges and the damage ratio (SHAUNN damage/strain gauge damage) was calculated. A ratio of 0.997 was achieved for the 80 test sorties.

6.4.6 Usage types

Although Dominie operations are divided into SPCs, analysis of the flight data suggested that the sorties could be naturally divided into four groups. These groups were termed *type* 1–4 and identified as "civil",

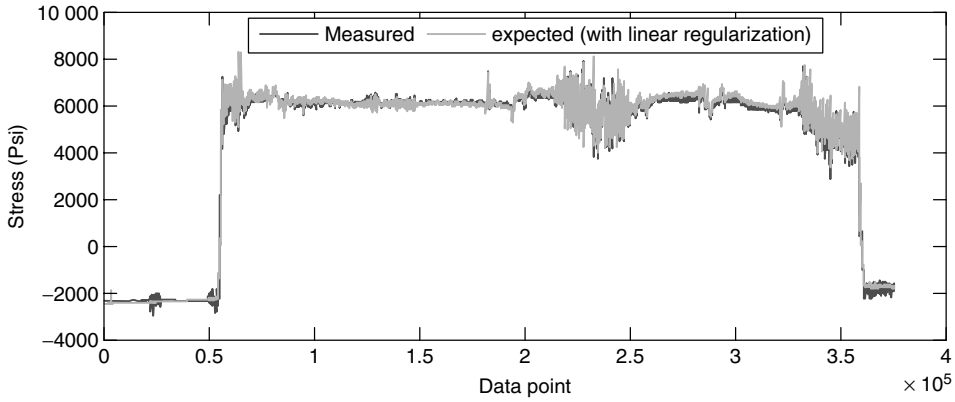


Figure 10. Dominie example SHAUNN stress prediction ($R = 0.998$).

"low level", "maritime", and "training" and particular SPCs were mapped to each group. Civil sorties were, as the name suggests, typical of civilian-type flights, generally consisting of taxi, takeoff, climb-to-cruise altitude (e.g., 31 000 ft), cruise, descent, landing, and taxi back. The majority of the flight was conducted at medium-to-high altitudes, above the severe gust regime (>10 000 ft), and rarely was more than one roller (or touch-and-go) landing undertaken. A typical altitude, airspeed, normal acceleration, and wing stress profile for a civil flight is reproduced in Figure 11. Civil-type flights were generally benign and the fatigue damage was dominated by the ground-to-air cycle (GTAC). Even a cursory glance at Figure 11 reveals that there is an obvious correlation between EAS and the mean stress in the wing, for this type of sortie. Gust loading was evident in the wing stress profile during the climb and descent phases of the flight and these regions were matched with responses in the NZ data.

Low-level sorties (Figure 12) were found to have a very different character. The majority of the sorties was flown at low level over land (often 1500 ft or below). Hence, the aircraft was operating in a severe turbulence environment and the effect of gust loading for these sorties flown over the land was immediately evident. The mean stress in the wing still followed the airspeed plot but the gust loading, which followed the NZ plot, was superimposed upon it. Maritime sorties (Figure 13) often had altitude and airspeed profiles very similar to low-level sorties. However, because the sorties were flown over the sea, the gust regime was relatively benign and this is illustrated in the wing stress and NZ plots. Finally, the training sorties (Figure 14) were primarily pilot and pilot's assistant training missions. These sorties generally contained a large number of roller landings; in the example presented here there were eight roller landings. Air tests were also included in this grouping as the profiles were similar.

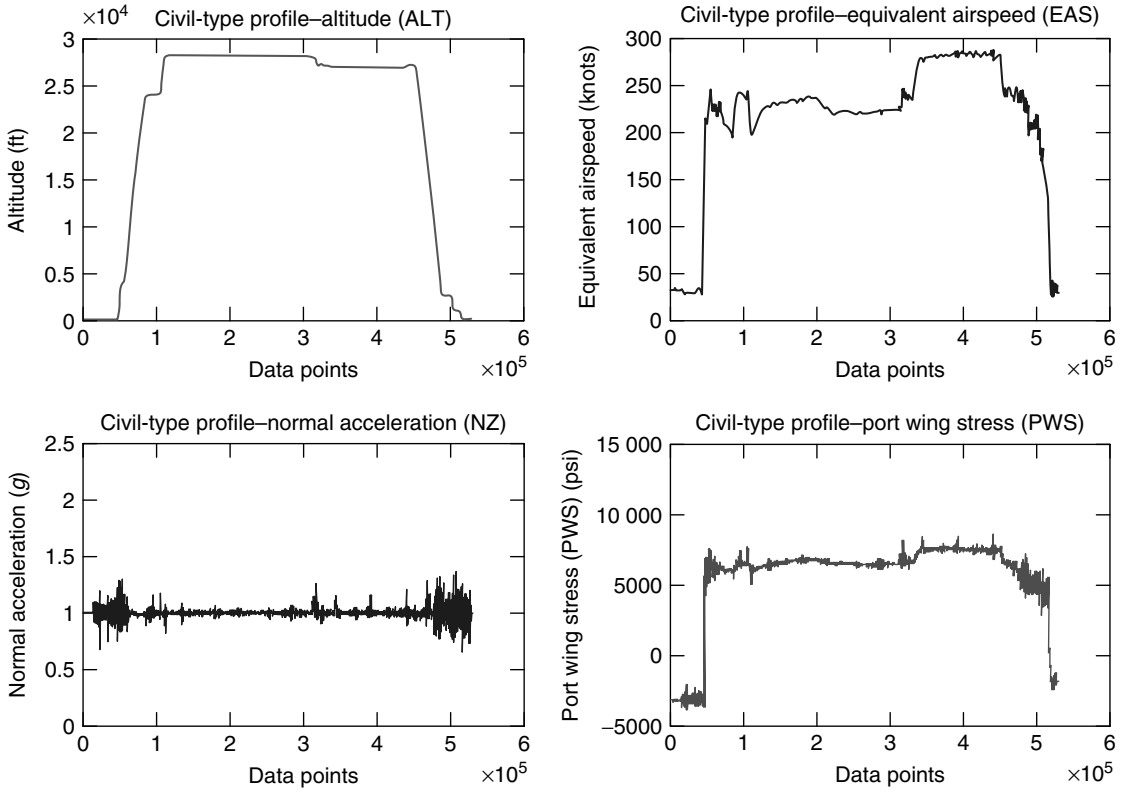


Figure 11. Dominie example "civil" flight profile.

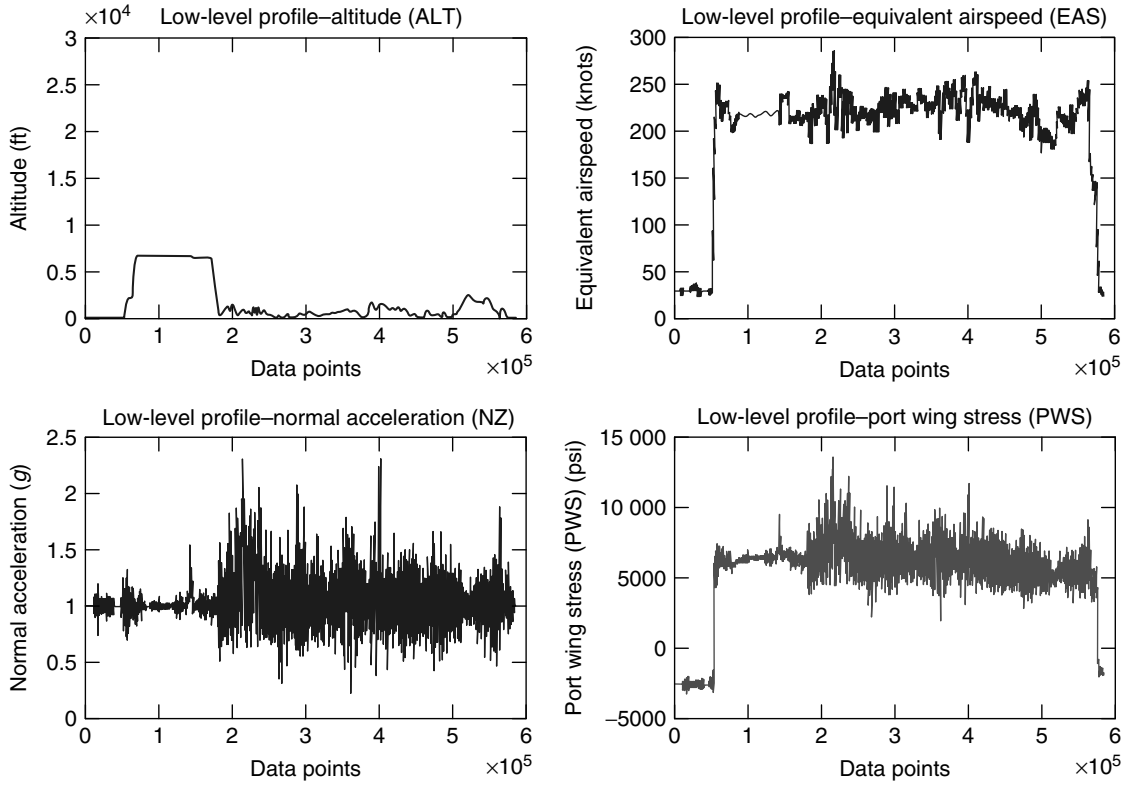


Figure 12. Dominie example "low-level" flight profile.

6.4.7 Crack propagation analysis

The output from the OLM data analysis program was stress cycle files for each flight to be used directly as input to the USAF-developed AFGROW [8] fracture mechanics software. AFGROW was chosen as an up-to-date industry standard of software for the analysis owing to its well-known qualities and array of options for modeling the crack propagation characteristics of airframe structures. A discussion of the critical wing features, use of AFGROW, and the preprocessing and postprocessing of flight data is not possible within the space constraints of this article; however, a more detailed explanation can be found in [9]. For the purposes of this article, general comparisons have been drawn.

As previously discussed, the aim of this program was to compare the RAF usage with the assumed spectra used for the civil damage tolerance analysis. Hence, the Dominie stress spectra were initially analyzed using the same Palmgren–Miner cumulative damage method. The result derived from the RAF spectrum generated from 319 OLM flights gave an equivalent fatigue life of 126 993 flights. Using this stress–life approach, the RAF usage spectrum was calculated to be around 7.5% more damaging than the spectrum used for the civil damage tolerance analysis. This result suggested that the crack propagation predictions for the RAF usage spectrum and the civil damage tolerance spectrum should be similar and that the associated damage tolerance assessments of the structure would be compatible. However, the fatigue calculation did not include the interaction between high and low loads, nor did it discriminate between the events that produce damaging stress cycles. The fatigue calculations and the results from the crack propagation analysis for the civil damage tolerance analysis spectrum and the OLM-derived Dominie spectrum are summarized in Table 1. The critical feature is illustrated in Figure 15. All the results are shown directly as computed, not as an indication of

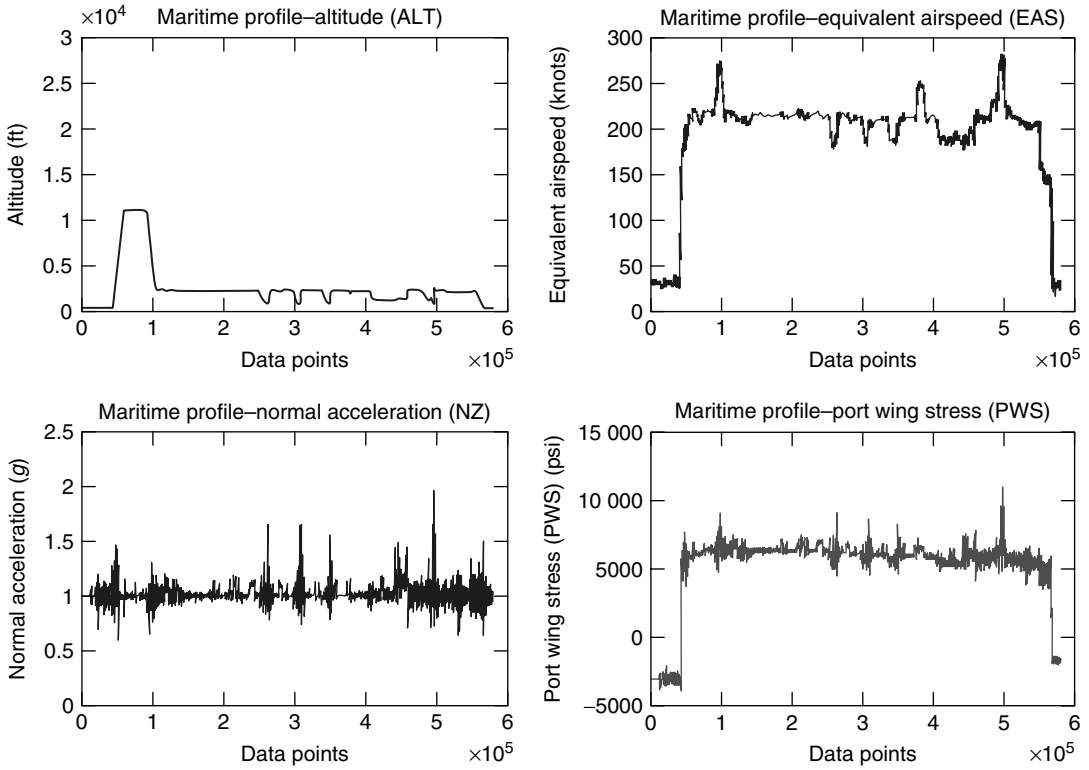


Figure 13. Dominie example "maritime" flight profile.

any presumed precision in the calculations but purely to emphasize the comparative values.

For the purposes of the AFGROW calculations, the Dominie spectra were concatenated in the same order that the 319 OLM flights had been achieved. For each flight, in turn, the ground-to-air cycle (GTAC) had been identified and was applied first; the residual gust and maneuver (RGM) and roller landing cycles were then applied in the order that they were identified by the Rainflow cycle counting process. To ensure that the crack propagation calculations were not compromised by this simplified representation of the stress sequence within each flight, the whole concatenated spectrum was reversed and the calculations repeated. As can be seen in Table 1, the results were effectively indistinguishable. As the concatenated spectra were repeated many times for each phase of crack growth, this observation suggests that the interaction effects between the GTAC and the smaller cycles were not unduly biased by the slightly artificial ordering of the stress cycles.

6.4.8 Comparison between the civil datum and the Dominie OLM stress spectra

The total peak stress spectra for the civil datum and the Dominie OLM data are illustrated in Figure 16. This shows that the two spectra are effectively coincident over the range of peak stress values between around 8 ksi (∼56 MPa) and 13 ksi (∼90 MPa). The associated stress cycles are those that account for the bulk of the fatigue damage. That damage is made up of two components for the civil spectrum, namely, the GTAC damage and the RGM damage which, on average, account for 87.7 and 12.3% of the total damage per flight, respectively. For the Dominie, in addition to the GTAC and RGM components of the damage sum, there is an additional damage component due to the roller landings not present in the civil spectrum. The relative significance of the associated stress cycles may be adjudged by reference to Figure 16, which shows the two comparative GTAC spectra and the Dominie roller landing spectrum.

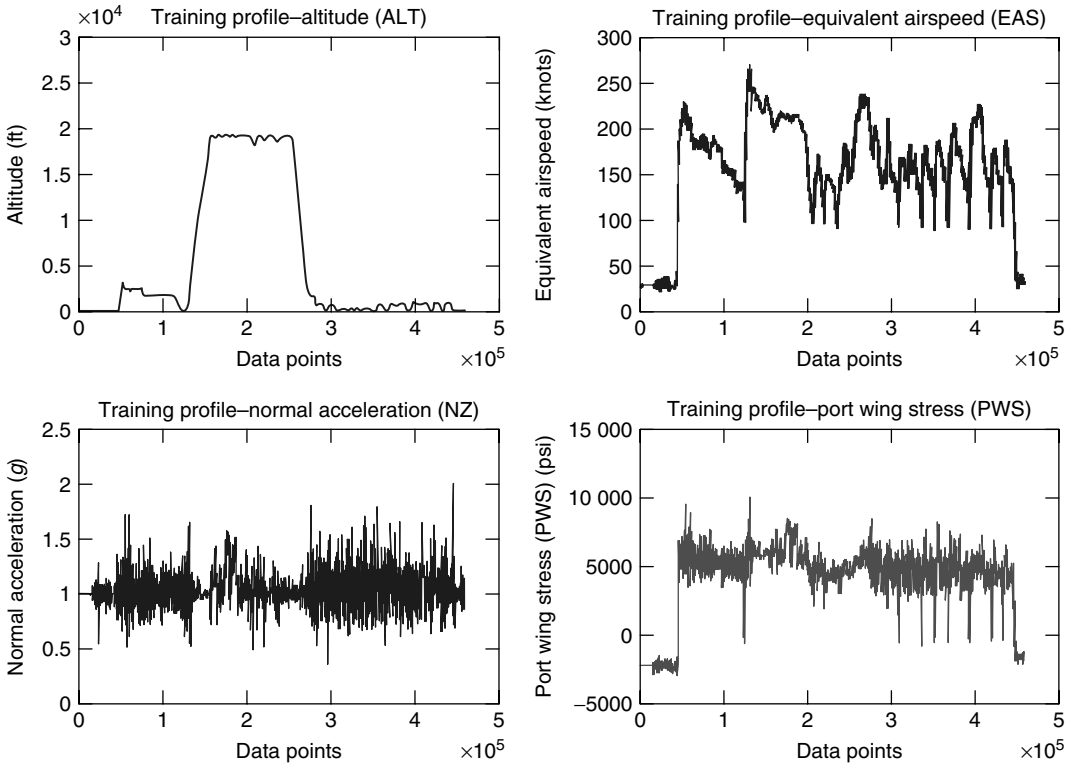


Figure 14. Dominie example "training" flight profile.

Note that the peak stresses for the Dominie GTAC spectrum are slightly less than those for the civil spectrum but the associated stress ranges are greater. This results in only a marginal overall difference in the average GTAC damage rates per flight for the Dominie and the civil aircraft. The damage rates per flight due to the RGM spectra are also very similar although slightly higher for the Dominie owing to the very large numbers of relatively small stress cycles generated during low-level navigation flights; these account for the extended "tail" of the total spectrum shown in Figure 16. However, these differences balance out such that, in the absence of the roller landings, the total damage rates per flight would be

Table 1. Summary of Dominie wing "AFGROW" calculations

Stress spectrum	Nominal fatigue life (flights)	Crack growth predictions (flights)			
		Failure of spar	Skin crack to stringer land	Total	Skin crack critical length (in.)
Civil, flight-by-flight (999 flights)	136 576	48 675	5274	53 949	
Dominie OLM, concatenated spectrum (319 flights)	126 993	24 822	3355	28 177	3.43
Dominie OLM, reverse-concatenated spectrum (319 flights)		24 880	3277	28 157	

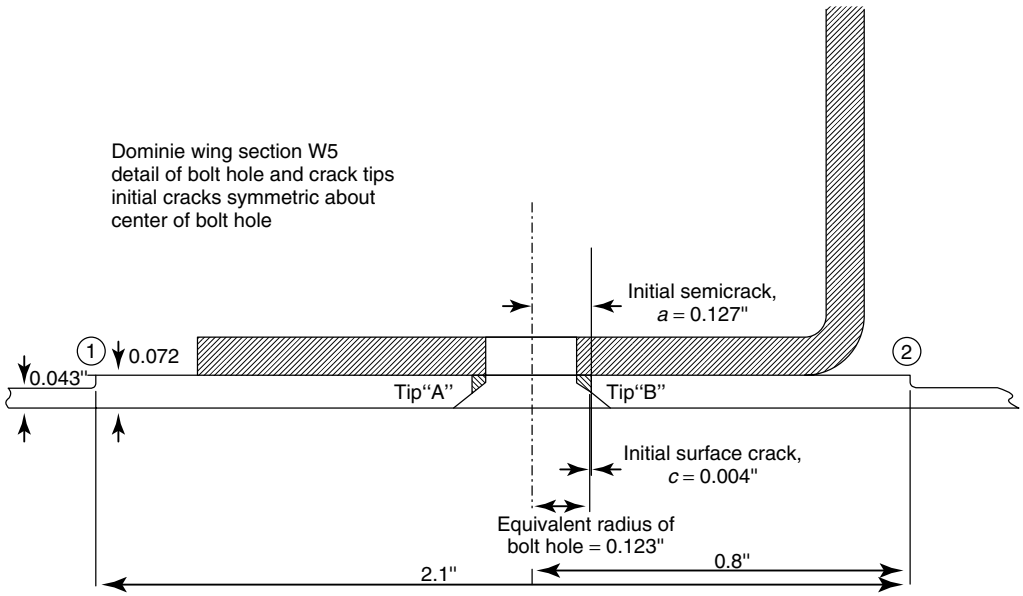


Figure 15. Dominie idealized wing section used for "AFGROW" calculations.

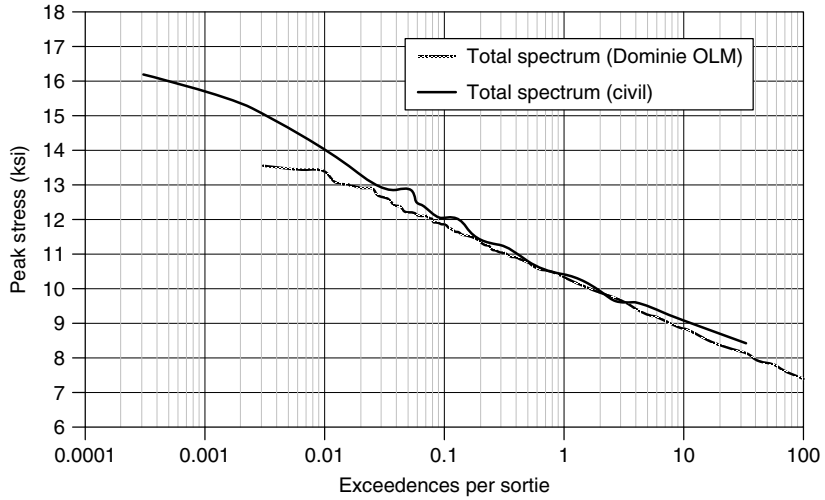


Figure 16. Dominie total peak stress spectra.

almost identical. Clearly, the contribution of the roller landings is significant and accounts for the overall difference between the fatigue damage calculations for the Dominie versus the civil datum. The roller landing stress spectrum shown in Figure 17 is also noteworthy because, prior to the OLM program, there was no evidence available to quantify the damaging effects of such events. This had led historically to the incorporation of very conservative assumptions into the fatigue monitoring model for the Dominie.

7 CONCLUSIONS

The fatigue and crack propagation calculations summarized in this section of the article are as

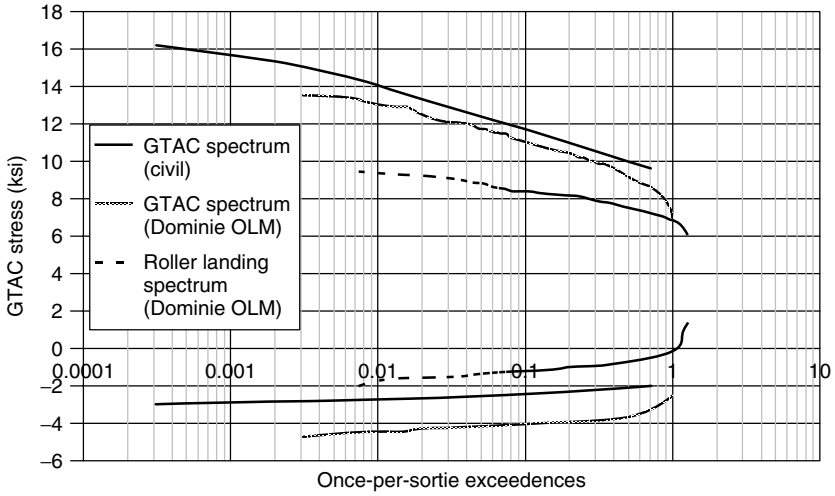


Figure 17. Dominie GTAC and roller landing stress spectra.

calculated using the conventional fatigue and fracture mechanics methodologies described; no attempt has been made to qualify the corresponding damage tolerance criteria for the wing structure. The results are intended purely to illustrate and explain the comparative content and severity of the stress spectra describing the usage of a civil-designed and qualified aircraft as flown in a military training role. Intense flying at low level and the use of the aircraft for pilot continuation training are two particular aspects of that usage that were not allowed for in the civil qualification; it has been shown that these have a considerable effect on the crack propagation predictions. This is a clear indication of the value of OLM, which, in this case, has enabled these aspects to be identified and quantified where previously there were no reliable data available. For instance, the stress levels in the wing structure during a roller landing were wholly unknown and the statistics describing the gust intensities within the low-level environments routinely encountered by the Dominie were very sparse.

ACKNOWLEDGMENTS

The success of the Dominie OLM was due to the hard work and cooperation of the team from QinetiQ, SERCo, No 55 Squadron at RAFC Cranwell, the UK MoD Training Aircraft Integrated Project Team, and the UK MoD Structural Integrity Branch.

REFERENCES

[1] Def Stan 00–970 UK Ministry of Defence, Defence Standard 00–970, Part1/2, Design and Airworthiness Requirements for Service Aircraft. Issue 5, 2007.

[2] Department of Defense Standard Practice, *Aircraft Structural Integrity Program (ASIP)*, MIL-STD1530C (USAF) 1 November 2005.

[3] *ADF Airworthiness Manual*. Australian Air Publication 7001.048, 12 September 2006.

[4] JSSG2006 Department of Defense, *Joint Service Specification Guide Aircraft Structures*, 30 October 1998.

[5] *Airworthiness Design Requirements Manual*. Australian Air Publication 7001.054, 6 June 2007.

[6] Reed SC, Holford DM, *Guidance for Aircraft Operational Loads Measurement Programmes*. UK Military Aircraft Structures Airworthiness Advisory Group (MASAAG), Paper 109, May 2007.

[7] Skopinski TH, Aitken Jr WS, Huston WB. *Calibration of Strain Gauge Installations in Aircraft Structures for Measurement of Flight Loads*, NACA Report 1178, 1954.

[8] Air Vehicles Directorate, Air Force Research Laboratory, United States Air Force, *AFGROW Version 4.0008.12.11* , 20 June 2003.

[9] Reed SC, Duffield MJ and Engelhardt ME, Operational loads measurement of a civil designed aircraft in a military role. *Presented at the 24th International Committee on Aeronautical Fatigue Symposium (ICAF 2007)*. Naples, NA, May 2007.

Chapter 157

Open Systems Architecture for Condition-based Maintenance

Robert L. Walter IV[1], David Boylan[1] and Daniel Gilbertson[2]

[1] *Applied Research Laboratory, Pennsylvania State University, University Park, PA, USA*
[2] *Boeing Phantom Works, St. Louis, MO, USA*

1 INTRODUCTION

Throughout this structural health monitoring (SHM) encyclopedia, there are numerous examples of technologies that integrators and vendors can implement: miniaturized sensors, embedded computing, and diagnostic and prognostic algorithms. Their challenge is to integrate all these technologies into one or several systems. It may seem that many integration challenges have been addressed in other systems, and this is true; but what are the best practices for communicating information among data acquisition (DA), data processing, state detection (SD), health assessment (HA), prognostics, and advisory generation functions? Since SHM systems may perform one, many, or all of these functions, how can vendors specify how systems from other vendors interact with their systems?

The International Standards Organization (ISO) Technical Committee 108/Subcommittee 5 has defined six generic functional blocks that are relevant to equipment and SHM in ISO-13374 titled "Condition Monitoring and Diagnostics of Machines" [1]. These six blocks are DA, data manipulation (DM), SD, HA, prognostics assessment (PA), and advisory generation (Figure 1). ISO-13374 also defines the information that is input and output from each of the functional blocks. Although the standard has the word "machines" in the title, the concepts in ISO-13374 are well suited to structures. So, for a diagnostic algorithm, ISO-13374 states that the algorithm should be included in the HA functional block. Its inputs are from the DM block (data, timestamp, and data quality), and its outputs are health grade, diagnosed

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

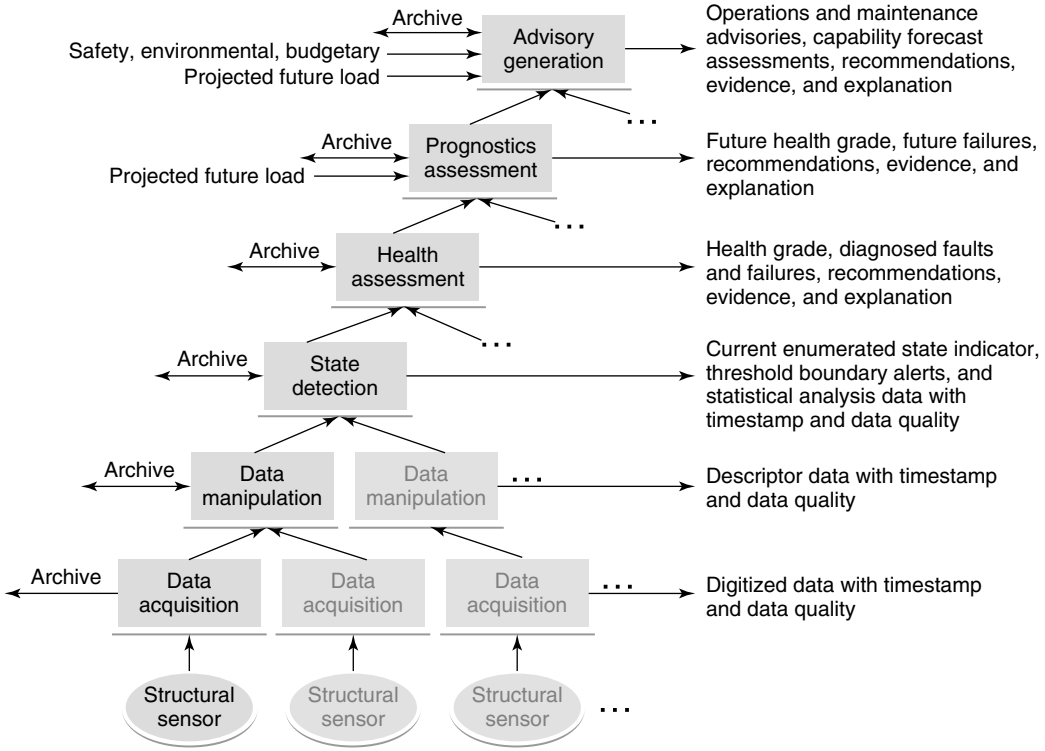


Figure 1. Six functional blocks of ISO-13374.

faults and failures, recommendations, evidence, and explanation. The other five functional blocks have a similar list of inputs and outputs. However, ISO-13374 is not sufficient to define a system architecture. What is missing is a definition of how the functional blocks interact among themselves.

A coalition of vendors has developed the open systems architecture for condition-based maintenance (OSA-CBM) standard to satisfy the guidelines of ISO-13374 and has transitioned the stewardship of the OSA-CBM standard to the Machinery Information Management Open Systems Alliance (MIMOSA) standards body. The coalition of vendors developed the OSA-CBM standard to align to the MIMOSA Open Systems Architecture for Enterprise Application Integration (OSA-EAI), and together they formulated a standards-based maintenance information management system that can find a wide variety of applications.

This article describes how to implement OSA-CBM in SHM applications in an open architecture to foster simpler systems integration.

2 KEY FEATURES OF AN OPEN ARCHITECTURE FOR CBM

The goal of implementing an open architecture for SHM is to allow vendors to add, upgrade, and interchange components in such a way as to be technology independent, yet at the same time, support protection of vendors' proprietary technologies. OSA-CBM accomplishes this as follows:

1. Promotes technology independence
The core OSA-CBM model only defines the interfaces and the class structure of messages going across those interfaces. Thus, OSA-CBM can be implemented with many technologies and various programming languages. Secondary specifications will specify implementations for certain major middleware technologies, such as extensible markup language (XML), to aid interoperability. Even so, only the structure of the messages will be enforced.

2. Protects proprietary algorithms, yet exposes inputs and outputs
Although the messages between modules are standardized in OSA-CBM, the internals of each module are not. Modules can, and are expected to, contain proprietary algorithms that will not be visible to anyone else in the OSA-CBM system. The inputs and outputs of the module are all that must be known. The rest is a black box.

3. Updates existing SHM systems
Naturally, many legacy SHM systems exist with proprietary formats. An implementer can simply make a wrapper around these systems to convert their data into OSA-CBM format, and then reap all the benefits of OSA-CBM.

4. Allows replacement DA systems
In an OSA-CBM system, if the hardware is changed, only the DA layer that supports the hardware must be updated. The rest of the system is not affected.

5. Supports dynamic updates of algorithm parameters
OSA-CBM allows a standardized way to access and modify processing parameters of algorithms within a layer. This process is discussed later.

6. Eases the addition of new processing algorithms
A new algorithm for processing data may be added without substantially affecting the present systems operations. Only processing load needs to be considered. This allows a simple path for the addition of new processing methods for continual improvement.

7. Permits replacement of entire software modules
Since OSA-CBM is modular and the structure of data between modules is standardized, existing modules can be updated or replaced without recoding the pieces around it.

8. Allows changes of load plan
In a prognostics or advisory generation capability, some plan for the future must be known to the algorithm. OSA-CBM allows a user to dynamically update the expected load on a system to examine the effects this would have on the predicted future health of a system.

3 BRIEF HISTORY OF OSA-CBM

An industry-led team developed and demonstrated OSA-CBM, completing its work in June 2001. The team's participants—Boeing, Caterpillar, Rockwell Automation, Rockwell Science Center, Newport News Shipbuilding, and Oceana Sensor Technologies—covered a wide range of industrial, commercial, and military applications of CBM technology. Other team contributors include the Penn State Applied Research Laboratory (ARL) and MIMOSA. The focus of the team was to design, develop, and demonstrate a software architecture that facilitates interoperability of CBM software modules. The OSA-CBM standard is documented using the unified modeling language (UML) and is available for download on the MIMOSA web site at http://www.mimosa.org.

4 OSA-CBM CONCEPTS

OSA-CBM categorizes data along two "dimensions". Along what can be considered the "vertical" dimension, the data classes are separated into "layers" based on the ISO-13374 standard, from DA all the way to advisory generation. At the same time, the "horizontal" dimension separates information into three categories: data, configuration, and explanation. Typically, layers are considered the dominant dimension, and each layer contains data, configuration, and explanation classes.

4.1 Vertical dimension

The OSA-CBM Standard consists of six layers. ISO 13374 defines and OSA-CBM implements the six stages (Figure 2) of CBM data processing as follows:

- **Data acquisition (DA)**

At the "bottom" of the OSA-CBM stack, data is collected from a sensor or other data source and is transformed into OSA-CBM format. Simple error-correction or data normalization can occur at this

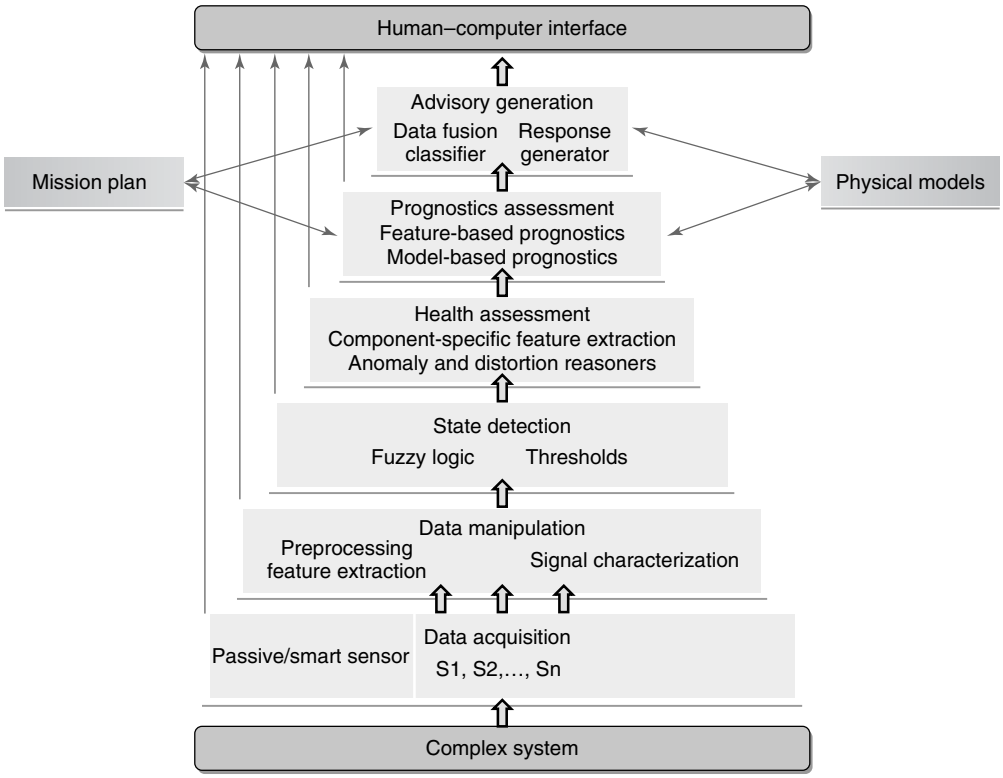


Figure 2. Vertical dimensions of OSA-CBM.

stage, but the output from the DA layer is considered essentially "raw".

- **Data manipulation (DM)**

Input to the DM layer nearly always comes from the DA layer, though the DM and DA layers can be combined to save space or processing time. In the DM layer, algorithms are applied to the data to make it more meaningful. Examples of DM processing algorithms are statistical measures (e.g., mean and root mean square), Fourier transforms, and time synchronous averaging. Output from the DM layer is considered processed data.

- **State detection (SD)**

The SD layer determines system states and conditions. Data from the DM and DA layers and possibly other SD outputs is monitored. One function of the SD layer is to monitor signal levels for simple threshold crossings and raise an alert whenever the data moves out of nominal operating range. States and conditions may be real-valued continuous state, integer-valued discrete states, or enumerated states.

- **Health assessment (HA)**

The HA layer utilizes data from the "lower" three layers (DA, DM, and SD) and produces two outputs: an estimation of health (0–1, 1 is perfect health), and a diagnosis of faults it believes to exist in the monitored object. The HA layer attempts to provide a complete assessment of the operating condition of the object being monitored, along with confidence and severity of the diagnoses. While the SD layer provides an observed set of symptoms, the HA layer interprets the observations into estimated or deduced system component health. It proposes system component faults and modes that would result in the observed system states and conditions.

- **Prognostics assessment (PA)**

The PA layer uses data from the lower four layers (DA, DM, SD, and HA) as well as the anticipated

future load to estimate remaining useful life and predict future health. The PA layer may also produce a prognosis of faults that it believes will occur in the future, as well as when they are likely to occur. PA layer modules may be constructed to accept expected future usage and return back computed life span, given the expected usage. This may be very useful with planning systems.

- **Advisory generation (AG)**

At the "top" of the OSA-CBM stack, the AG layer examines the current health and estimated prognosis and suggests a course of action such as *remove structure from service.* The AG layer also considers external constraints, such as budgetary, safety, and environmental concerns. The advisories produced can include operational readiness assessments, maintenance requests, suggestions for operational parameter changes, and capability forecasts.

These layers can be considered vertical because of two trends that occur as data travels up through the layers. First, the size of the data decreases. Data at the DA layer can contain hundreds of thousands of points, in the case of accelerometer data with a high sampling rate. One step upward, the DM layer typically reduces the size of the data to thousands of points in a Fourier transform or to single statistical values. Further up, the data gets smaller in size but more meaningful. Conversely, the complexity of processing tends to increase in higher layers. The algorithms in the DA, DM, and SD layers are relatively simple: data format transformation, straightforward algorithmic processing, and threshold comparisons. On the other hand, diagnostics, prognostics, and AG require more complex processing, often by intelligent reasoners.

4.2 Horizontal dimension

Each of the six layers can be considered to have three types of information associated with it (Figure 3):

- Data is the information gathered and processed by an OSA-CBM system. Data is usually exchanged on a timed interval, starting with data collection at the DA layer and propagating up to the higher layers. Sensor signals, extracted features, alerts, HAs, prognoses, and advisories are all considered data.
- Configuration is unchanging or infrequently changing information about the data (metadata). Configuration is typically accessed once at start-up

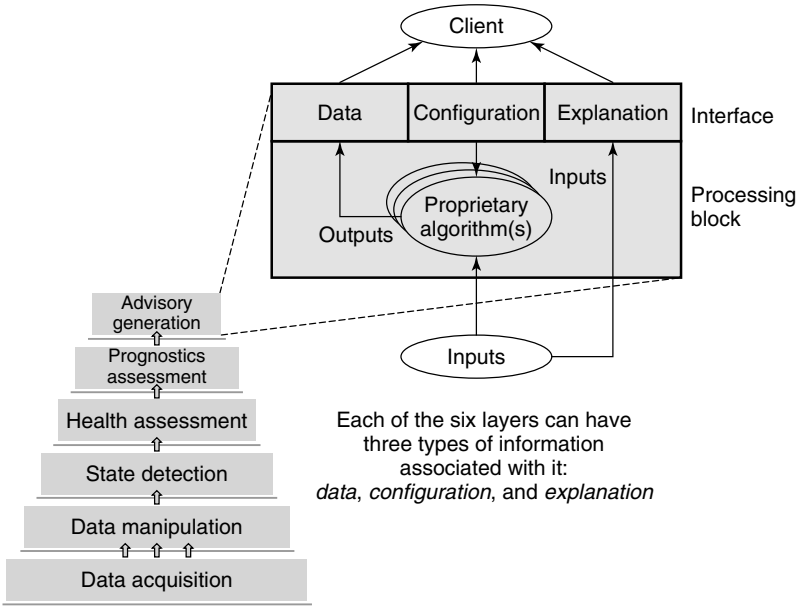


Figure 3. Horizontal dimension of OSA-CBM.

and used to aid in processing or understanding the data. Examples of configuration information are engineering units, algorithm descriptions, module descriptions, and structural information.

- Explanation is the ability to provide a description of the data used by an OSA-CBM layer to generate an output. There are several forms of explanation, such as a description of the location where data was gathered or a copy of the data itself. For example, if a DM layer computes a mean on data from an accelerometer, then the explanation for the DM layer would be the accelerometer signal. Modules are not required to support the explanation interfaces.

OSA-CBM supports three additional forms of information that cannot be easily standardized. Instead, a format is provided for storing and changing this application-specific information, but the specific parameters for each application are user defined. First, OSA-CBM provides the functionality to change processing parameters called *control parameters*. This is not to be confused with configuration: control parameters can be changed at any time in a layer's lifetime, while configuration changes very rarely. Examples of potential control parameters are data refresh frequency, sample size of DA, threshold limits, and fast Fourier transform (FFT) block size. Control parameters allow modules to be configurable and reusable without recompilation.

Second, the application parameters allow a user (human or other OSA-CBM module) to change parameters in a module. Application parameters are not meant to change the behavior of algorithms, but instead change other application-specific parameters. This interface is meant to be a catchall for needs outside the purview of OSA-CBM. One example is program initialization data such as specifying the source web service for module input.

Third, data classes do not contain mechanisms for reporting errors; instead, the error parameters allow a module to report processing errors. Note that while control and application parameters allow the user to both request and change their values, the error parameters only allow requests.

Although these specific parameters are user defined, a human operator can request a list of the parameters, their values, and a description of each parameter. In this way, a human can set parameters of a module without knowing what they are, ahead of time, but a machine cannot. To allow automated parameter changing, a list of possible parameters must be provided beforehand.

4.3 OSA-CBM interfaces

The OSA-CBM standard describes four types of interfaces that can be implemented at each layer. The concept behind OSA-CBM is not to map every interface into every technology in which OSA-CBM is implemented. Rather, it is to select the best interface for the desired type of data transfer.

The synchronous interface is designed to be a simple function request with data returned directly by the call. This type of interface maps into the internet data services concept where a connection is made, information about the request is given to the service, the service processes it, and then the service returns the required data. It maps to programming languages as a simple class method call with information returned by the method.

The asynchronous interface is a publish–subscribe type interface that establishes a connection between a user of data and a provider of data. It is useful for event-based systems and higher speed systems where the cost of connection becomes important. This interface functionality can be set up in three different ways: push all data, return data only on request, and push only data with an alert status. Since a connection is maintained, the server of data can push all new data events to subscribers of that information. There may be a situation where the user of data only requires knowing about data items that have some type of alert status. Functionality may be set up to return only those data event sets that have triggered an alert. This aids the construction of an efficient processing system.

The data service interface is designed as a consumer of data. One application would be a maintenance work order processor, which might take an advisory generation layer data event. Another potential use might be a data event recording or storing system. When a module has data that should be recorded, it could send that data to an OSA-CBM data service that is designed for storing data. An OSA-CBM data service performs a well-known process at a known or discoverable location.

Table 1. OSA-CBM parameters

Parameter	Description
MonitorIdlist	Indicates a specific desired subset of data which is provided by a module
ConfigRequest	Indicates a specific subset of configuration information
ControlChange	Indicates a desired module functionality change. For the PA layer and AG layer it may indicate an expected usage model against which to make predictions
ControlRequest	Gets information about the value of specific control settings
ControlInfo	Contains the values of the desired control settings
AppNotify	Sets up application-specific parameters
AppRequest	Requests application-specific information
AppInfo	Responds with application specific information
ErrorRequest	Requests error information
ErrorInfo	Responds with error information

The data event server interface is designed to be a method for passing single data events. It is meant to be a light way for implementing a monitoring system, typically at an embedded level. It only provides for moving individual data events, sending none of the other OSA-CBM information components such as configuration.

The functions in the interfaces provide for connectivity (asynchronous interface) and for moving each of the different types of information, i.e., data events, configuration, explanation, control, application-specific, and error information. The most commonly used parameters in the interfaces are given in Table 1.

5 RELATIONSHIP TO OSA-EAI

The MIMOSA standards body also maintains OSA-EAI standard. Since the OSA-CBM standard is implemented as an architecture for processing information, the OSA-EAI standard is a combination of a relational database schema specification for archiving information and XML interface specifications for exchanging condition, reliability, and maintenance information between systems. One may, therefore, take a legacy system and apply a MIMOSA OSA-EAI compatible XML interface without changing the legacy system to develop a MIMOSA compliant system. The OSA-CBM and OSA-EAI standards were harmonized in 2004 such that data processed in OSA-CBM can be archived in OSA-EAI compliant data archives and communicated with external systems using OSA-EAI XML schemas.

There are five key classes in the OSA-EAI standard that enable integration between the two standards: site, segment, asset, agent, and measurement location (Figure 4). A "site" can be a structure, vehicle, manufacturing plant, facility, maintenance depot, etc. A site is composed of one or more structures and substructures called *segments* that perform a function that will be monitored. Serialized "assets" may be installed in segments over time. An "agent" is a person or intelligent software capable of performing work, making diagnoses, etc. A "measurement location" identifies the physical position on an asset or segment where actual measurements are taken.

In OSA-CBM, the data events from the lower three layers (DA, DM, and SD) are mapped to measurement locations. This means the identifiers for the lower three layers in OSA-CBM have a one-to-one correspondence to measurement locations in OSA-EAI. Similarly, the data events from the higher three layers (HA, PA, and AG) are mapped to agents. Agents in OSA-CBM terms are software agents or processing centers; every module in OSA-CBM is an agent. Since OSA-CBM maps to OSA-EAI, the output of any OSA-CBM module can easily be stored in an OSA-EAI compliant database and communicated between systems using OSA-EAI XML schemas.

However, although OSA-CBM identifiers map directly into OSA-EAI, any system can use these identifiers directly without necessarily installing an OSA-EAI database.

6 BEST PRACTICES

At the time of this writing, the only technology that is publicly offered is the XML mapping. However, this mapping is not useful for all applications. Many condition-monitoring systems, especially in the lower

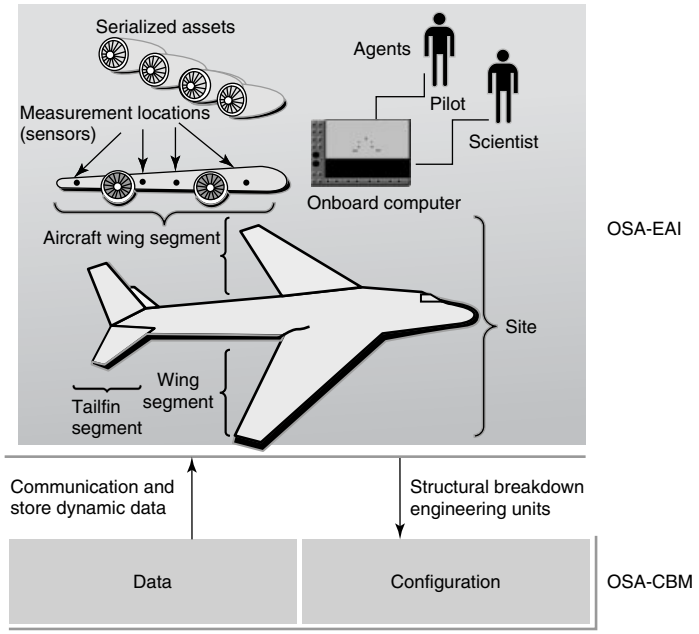


Figure 4. OSA-EAI site, segment, asset, agent, and measurement location.

three layers, process a continuous high-rate stream of data. XML is too inefficient for these applications. A programming language mapping to C or C++ could be used for a single program performing the processing. In this case, the information stays within the internal data structures, but this is done in such a manner that results can be fed to a generic module, which feeds higher layer systems. A binary mapping is also possible to construct a high-speed method of moving the information via a byte transmission mechanism such as TCP/IP. In these higher rate systems it is also suggested to use the asynchronous interface approach to eliminate communication connection costs.

When designing an OSA-CBM system, the first step is to develop a block diagram of the processing centers and indicate the layer in which a processing center occurs. For example, a DA module fetches data and serves it to a DM module, which performs some math processing. The DM module sends data to an SD module, which determines states and conditions of the system. Once these different modules are identified, the next step is to determine the signals that will flow in the monitoring system and the type of data events that will contain that information. The OSA-CBM UML can be used to identify the specific types of data events, which will be required. This system design should be recorded in an OSA-EAI database using site, segment, measurement location, and agent as main foundational tables.

One can work bottom up (starting from the sensors), top down (starting from the high-level goals), or use a combination of both. Assuming that the structure already has a set of sensors, each sensor is a potential source for a DA layer. To simplify the system, multiple sensors can also be combined into a single DA layer. Working bottom up, every sensor is monitored via DA layer module(s), and the algorithms of the higher layers are written to make use of the available data. This method can lead to waste, however, if some sensors are not necessary for health monitoring. Perhaps a better design method evaluates the available sensors for their usefulness in diagnostics or prognostics before wrapping them in a DA layer. This is an example of a combination of top-down and bottom-up design.

The DM layer is an important module because it serves to drastically shrink the size of the data while maintaining or even adding meaning to the raw data. This is especially necessary when the DA data is high bandwidth, such as vibration data. The DM

layer applies algorithms to DA data, often statistical measures, and produces "features" that can be used by the higher layers. These features communicate specific aspects or trends of the raw data. Ideally, data used by the SD, HA, PA, and AG layers from the DM and DA layers should be single values only, not waveforms or data sequences. Thus, a temperature sensor typically has a low enough bandwidth to be used directly by any layer, but a high-frequency vibration sensor should be processed first in the DM layer.

The remainder of the layers can be implemented as requirements dictate. The SD layer should monitor critical signals and report faults to operators and maintainers. HA, PA, and AG layers may be added as algorithms are developed for diagnostics, prognostics, and automated reasoning. Since an OSA-CBM system is inherently modular, different layers can be developed by companies or divisions that specialize in one aspect of the CBM cycle.

Developing an OSA-CBM system requires additional design decisions because of factors such as performance needs, memory size, processor power, bandwidth, and storage space. As noted earlier, the size of the data at the lower layers can be on the scale of megabytes, but the processing is simple. In higher layers, data becomes much smaller but the complexity of processing increases significantly. Similarly, an XML implementation is likely to be more maintainable than a binary implementation, but the bandwidth requirements are higher. If the implementation uses a powerful computer, these issues are not problems, but some implementations require computers with limited processing, power, and bandwidth capabilities. Some design decisions that can be made to address the specific requirements of an implementation are as follows:

- **Arrangement of layers**

An implementation does not have to provide every layer. One can take advantage of this by combining the DA and DM layers, such that the large data sets produced by the DA layer do not need to be transformed into OSA-CBM format and can thus save processor time, memory space, and bandwidth.

- **Physical location of layers**

Since the OSA-CBM architecture is modular, it is possible to spread the processing across multiple computers. Thus, a lightweight, limited processing power computer can be placed on the monitored item, while a more powerful processor or computer could handle the HA, PA, and AG layers. Another situation is the isolation of lower layer hard-real-time system-critical processing components from higher layer soft-real-time complex processing components.

- **Programming language implementation**

OSA-CBM does not specify what programming language to use—only the structure of the objects communicated. Ideally, this means that any programming language could be utilized, but not all languages have the same support for XML and object-oriented programming. Java or the Microsoft .Net framework is recommended for most applications, but C++ could be used for implementations with severe performance needs. New technology mappings may be added to the OSA-CBM specification by members of MIMOSA.

- **Communication implementation**

As of OSA-CBM version 3.1L, the only standardized communication format for OSA-CBM is XML. However, some implementations would be hindered by the overhead that XML adds to a message. If this is the case, the objects could be stored and sent in a binary format to save bandwidth. However, this binary format must be agreed upon in an interface design document between all parties involved until a binary format standard is added to the OSA-CBM standard.

7 GETTING STARTED

The SHM community is rich with varying technical capabilities and will greatly benefit from integrating those capabilities into a single open architecture—one which allows many proprietary technologies to be integrated into one solution. The MIMOSA OSA-CBM and OSA-EAI standards enable that vision. The authors encourage SHM developers and systems integrators to visit the MIMOSA web site at http://www.mimosa.org to download the standards, read primer documents, and download examples.

ACKNOWLEDGMENTS

We would like to thank Lynn Wang and Monica Houston for developing the illustrations.

RELATED ARTICLES

Finite Elements: Modeling of Piezoceramic and Magnetostrictive Sensors and Actuators

Modeling for Detection of Degraded Zones in Metallic and Composite Structures

Damage Measures

Sensor Network Paradigms

Wind Turbines

Large Rotating Machines

Gas Turbine Engines

Prognostics and Health Management of Electronics

Fatigue Monitoring in Nuclear Power Plants

REFERENCES

[1] ISO-13374, *Condition Monitoring and Diagnostics of Machines*. Technical Committee 108/Subcommittee 5, March 15, 2003.

Chapter 153
Fatigue Monitoring in Nuclear Power Plants

Wilhelm Kleinöder and Christian Pöckl
AREVA NP GmbH, Erlangen, Germany

Encyclopedia of Structural Health Monitoring. Edited by Christian Boller, Fu-Kuo Chang and Yozo Fujino © 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05822-0.

1 INTRODUCTION

The design of components in nuclear power plants is based on nuclear codes and standards like ASME [1], RCCM [2], or KTA [3]. These codes and standards cover not only design but also material specifications, manufacturing, and examination. For safety-related components of the pressure-retaining boundary, stress and fatigue analyses are requested during design in order to fulfill the code requirements. These analyses are made with conservative load assumptions for the expected service life of the nuclear power plant.

To verify these load assumptions, a large measurement program was carried out in the early 1980s. This was in a German nuclear power plant, where, during the commissioning phase, detailed information of the real component loadings during plant operation was recorded. The measurement program revealed that cyclic loading is much more important than it was anticipated in the design analysis and that the amplitude and the number of cycles depend closely on the operation manner. Many unexpected loadings were recorded as follows:

- higher thermal stresses;
- higher number of cycles;
- unconsidered load events by project design;
- stratified flow due to low flow rates or caused by leaking valves.

However, more detailed analyses based on these measured loads showed that the fatigue limits were not exceeded.

The advantages of monitoring the real operating conditions and using the acquired data as an input for fatigue analyses based on realistic loading data became obvious. This gave a strong development impetus toward a sophisticated fatigue monitoring system. As a consequence, the German nuclear power plants as well as many plants in other countries were equipped with FAMOS (*Fa*tigue *Mo*nitoring *S*ystem—the AREVA product for online monitoring of fatigue). The main objectives of FAMOS can be summarized as follows:

- determining the fatigue status of the most highly stressed components;
- identifying operating modes that are unfavorable to fatigue;
- establishing a basis for fatigue analysis based on realistic operating loads; and
- using the results for lifetime management and lifetime extension.

Today, FAMOS is installed in more than 20 nuclear power plants (NPPs). It is also part of the monitoring systems implemented in the new European pressurized water reactor (EPR) [4]. The following sections give an overview of the procedures that are necessary for the implementation of such a monitoring system.

2 FATIGUE LOAD MECHANISMS

There are different mechanisms causing material damage. First, so-called primary loads do cause primary stresses, which, in the extreme, may lead to single failure by plastic fracture and whose intensity is not decreased by plastic deformation. Internal pressure and dead weight loads belong to this category. Pressure vessel safety against single failure by plastic fracture is proven usually by pressure tests. Secondly, for cumulative fatigue damage, the variable load effects are crucial because these cause the stresses and deformations to alternate. Fatigue accumulation, however, does not generally relax by plastic deformation. On the contrary, recurrent loading into the plastic regime may possibly induce accumulation of plastic deformation, so-called ratcheting. The degradation effect on the material appears first at the microscale, if the stress intensity fluctuations exceed a certain value, thereby causing irreversible plastic deformation. For pressure vessels, fatigue loading is of major concern because single and simple component tests do not exist to remedy the problem. The damage induced is not obvious and cannot be measured up to date.

Great attention was paid to the design and stress analysis of nuclear power plant components. Safe operation is established if the loads do not exceed the presumed values. However, possible uncertainties in the flow conditions and thus deviations from presumptions considered in the project phase and in the design reports must not be neglected. These deviations were observed, for instance, at some horizontal pipeline segments of NPPs, wherein special modes of operation gravity forces separate hot and cold media into a layered flow of very low flow rate. The physical phenomenon is well known as *temperature stratification*. This also appears at nozzles with intermittent inflow. The gap at nozzle thermal sleeves may also cause negative loading effects.

The monitoring system FAMOS addresses just these effects. For this purpose, thermocouples are installed on the outer surface of the parts. The real, local temperature profiles are recorded, which otherwise would not be known. The main attention is given to the acquisition of the local temperature history, but it is also necessary to collect data on other loadings (internal pressure, member forces), because the resulting degradation effect is caused by the combined loading. This is why additional parameters are fed into the system. These also serve to later characterize and analyze the operational state of the facility.

3 PHENOMENA CAUSING FATIGUE

Major loads acting on primary circuit components consist of thermal constraints, internal pressure, and thermal transients, besides member forces and moments. The loads vary during operation of the equipment, thus causing material fatigue. To determine the state of material, fatigue is a very complex task. The reason lies in the uncertainties of boundary conditions, e.g., transient distribution of temperature and flow in the pressure-retaining components.

The next sections describe some processes, which can cause cyclic loading of material during the operation of primary circuit.

4 TEMPERATURE TRANSIENTS

Temperature transients in pressure vessels and piping systems are very often a result of sudden opening of a valve or starting of a pump. Cold water flows into a system, which was originally in a hot, steady-state condition. The thermal field is axially symmetrical. If the flow rate is low and the cold and hot water interface is not perpendicular but skew, then stratification can evolve.

Thermal stresses are a result of thermal gradients in the wall. The amplitudes of thermal gradients depend on temperature differences, on the rate of temperature changes, and on the heat transfer coefficients (Biot number). The stress in the wall is roughly proportional to the temperature difference between the inner surface and the mean temperature of the wall.

The temperature on the outer surface can be measured directly by using a surface-mounted thermocouple. The estimation of the temperature on the inner surface of the component is much more complex. For high Biot numbers, it is possible to approximate the temperature of the material with the temperature of cooling water, measured via the operational instrumentation and Instrumentation & Control (I&C) system.

From fatigue load point of view, the mentioned type of loading is dangerous in two cases. When the operation with transient loads is a relatively long-term operation, the numbers of load cycles are high. This load phenomenon is pictured in Figure 1, which shows the measured temperatures at feedwater inlet of a steam generator. The plant condition is zero power output, so-called hot standby. The water level in the steam generator, however, has to be maintained either by feeding continuously at very low flow rates, thereby inducing thermal stratification, or by feeding intermittently by nominal flow rate, thus accepting thermal shocks. The latter case can be seen in Figure 1. During the feeding process,

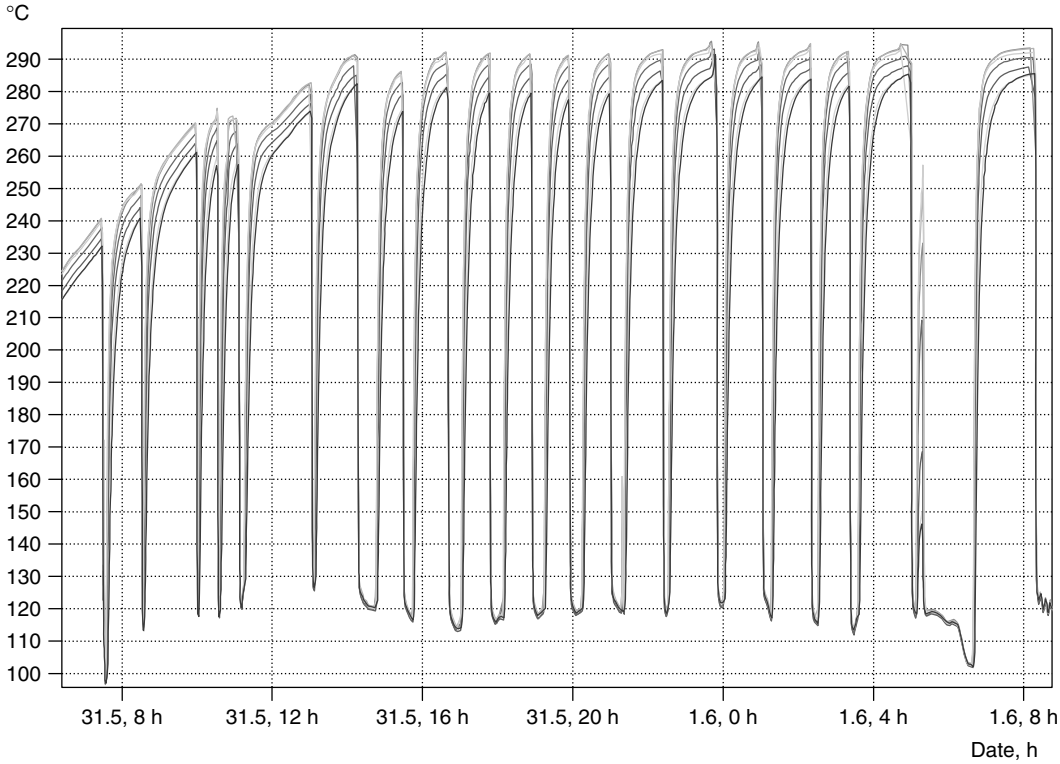


Figure 1. Intermitted feedwater injection induces thermal shock on structure (7 thermocouple signals according to in Figure 7b).

the feedwater nozzle gets a thermal shock which decreases the nozzle temperature suddenly. By the time the feedwater nozzle temperature increases again to the temperature of the steam generator.

Another process of high accumulation of fatigue load can be important for complex nozzle geometries. When a thin-walled part is connected to a massive one, then the latter does not allow the deformation of the thinner part, thus inducing constraint stresses.

5 THERMAL STRATIFICATION

Thermal stratification can be characterized by a thermal field, which is symmetrical to the vertical axis of a cross section of a pipe positioned horizontally. Stratified flow occurs when a low flow rate creates two layers. The warmer fluid will stay in the upper part of the pipe above the heavier (colder) fluid without any appreciable mixing in between. Figure 2 shows an example of thermal stratification during the startup of an NPP. The upper line represents the temperature at the pressurizer (DH) and the lower line the temperature of the main coolant line (HKML). The remaining lines are the temperatures $T1$ till $T7$ as indicated for a measurement section shown in Figure 7.

When the temperature varies in the vertical direction, the pipe tends to bow [5]. The stresses induced in the material decrease with the supports of the piping system being flexible. When the temperature interface is sharp, the gradients of temperature in the metal are high (Figure 3). The stress is highest in cases when the interface of cold and hot fluid is in the middle of the pipe. When the sharp line between cold and hot fluid moves vertically, the local temperature gradients (and therefore local stresses) also change. In such a situation, the cyclic load occurs, though the temperatures of cold and hot layers of fluid are both constant. Therefore, it is important to measure not only the temperatures on the top and bottom of the pipe but also the temperatures at several points along the circumference of the pipe (Figures 2 and 7).

6 TURBULENCE PENETRATION

This phenomenon exists in a piping system with a hot turbulent flow that penetrates into a branch line that contains a stagnant cold fluid. The intensity of

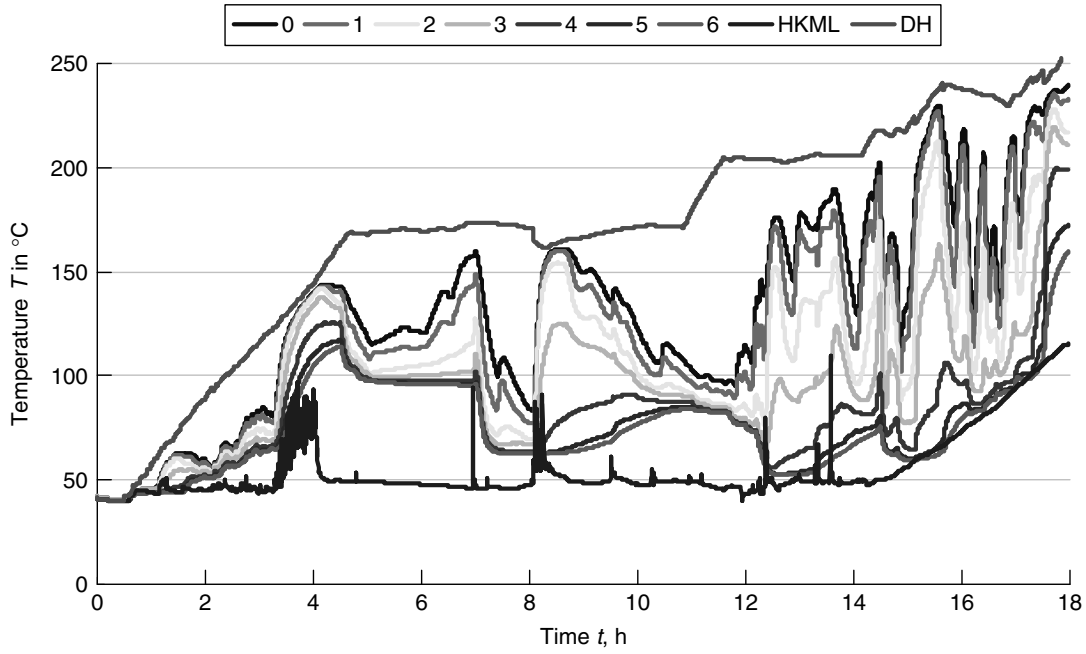


Figure 2. Example of thermal stratification in surge line during startup of NPP.

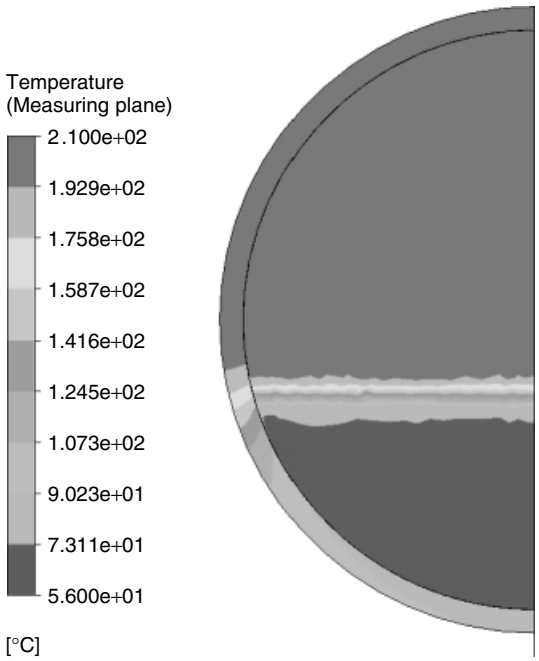


Figure 3. Simulation of stratification with a computational fluid dynamics (CFD) model.

the turbulence decays exponentially from the header pipe into the branch, but the temperature remains fairly constant with main flow temperature along the length of several diameters (Figure 4). If the turbulence penetration depth just reaches a horizontal line, thermal stratification can occur within it. The interface is not stable because of turbulence. This results in fluctuating interfaces and additional fatigue load.

This phenomenon exists in every pipe routing for three different cases: horizontal branch, up-horizontal branch, or down-horizontal branch line. The FAMOS way to find out locations like that is to look for this typical pipe routing with horizontally or vertically bent-down branches followed by a horizontal line with a closed valve. Procedures for estimation of the turbulence penetration are developed in [6].

7 THE "FATIGUE MONITORING MANUAL"

The "Fatigue Monitoring Manual" contains the essential input data for introduction and installation of the monitoring system for primary and secondary circuit systems in the nuclear power plant. The entire philosophy of the system is described and the purpose of its implementation is specified. Finally, the main purpose is the screening of the primary and secondary systems for locations where additional measurement locations shall be applied for fatigue monitoring. This is a very essential step because, during this step, if not all relevant locations are identified, then the whole fatigue monitoring system would be put into question. The main task of the work here requires, first of all, to find out the locations of the safety-related parts of the pressure-retaining boundary, which are thermally highest loaded, and therefore the fatigue load is representative for the neighboring components. These locations need to be monitored with the fatigue monitoring system.

This engineering evaluation gives the reasons as to why these locations have been selected for monitoring. Finally, a measurement plan has to be prepared where the monitored locations are summarized. The task of this manual therefore is to explain *where*,

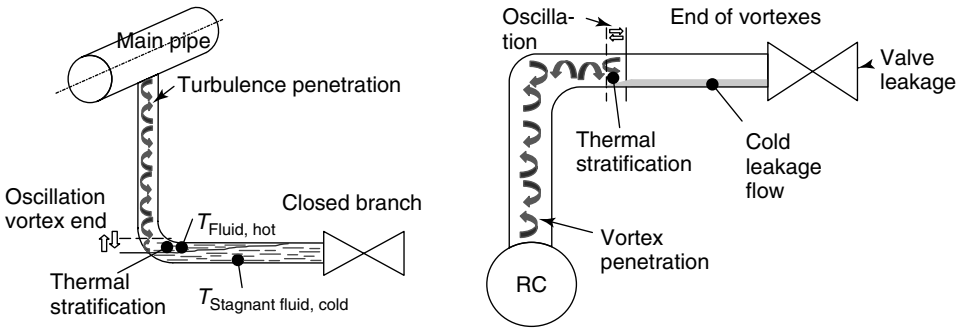


Figure 4. Turbulent mixing in a branch line containing stagnant fluid.

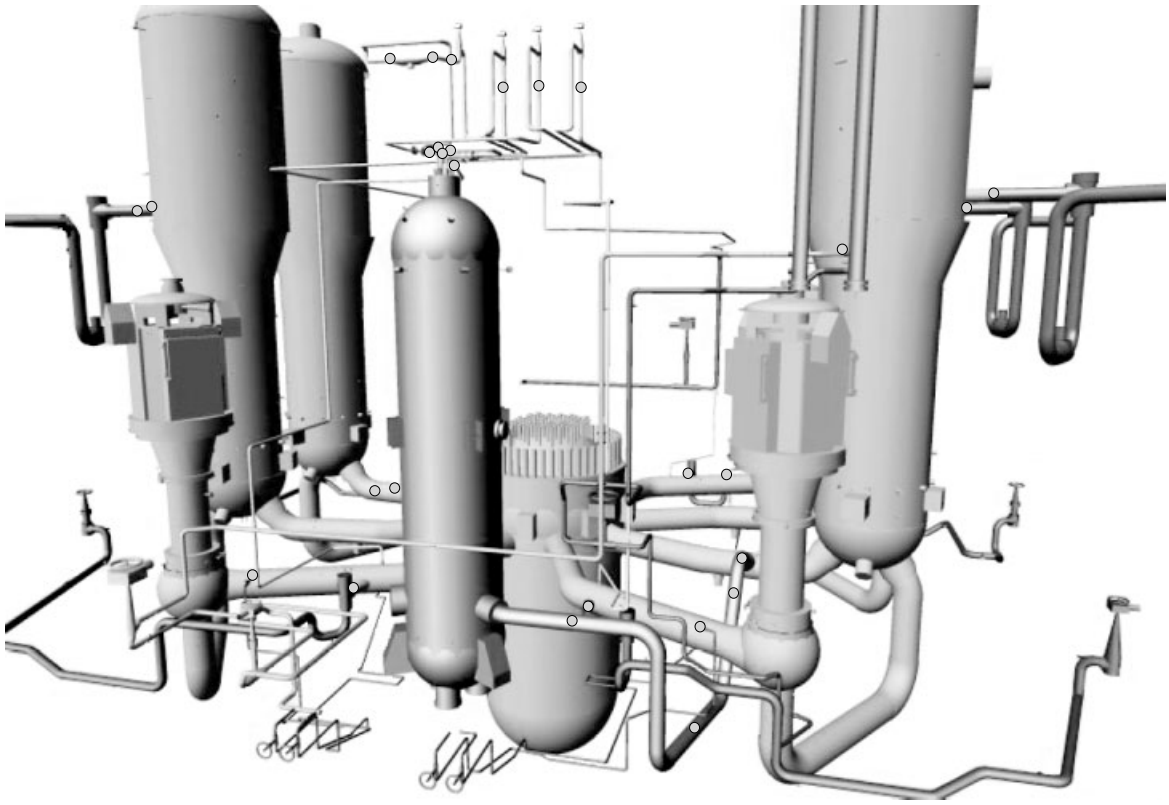


Figure 5. Selected measurement section locations of the FAMOS system.

why and *how* to implement measurement sections for fatigue monitoring, which are representative for the whole plant and cover all relevant components with respect to fatigue. An example of locations where measurement sections are required is given in Figure 5.

8 FAMOS CONCEPT

Within the FAMOS concept, the "Fatigue Monitoring Manual" provides the foundation for further work—it is the basic step. It is named "stage 0" because of the necessary information for the preparation of the measurement equipment within "stage 1". However, it is also the basis for additional FAMOS functional units for the so-called quick evaluation "stage 2" and the fatigue analysis within "stage 3". These four steps are shown in Figure 6.

9 IMPLEMENTATION OF A MEASUREMENT SYSTEM

The implementation of a fatigue monitoring system consists of two steps:

- mechanical components like measurement sections and thermocouple sensors;
- electrical components like measurement electronic system for data acquisition.

For both parts, several solutions were developed during the past years.

9.1 Mechanical components

For the measurement sections with the thermocouples, the main focus was on the improvement of the dynamic response of the temperature signal of the thermocouple. Different boundary conditions need to

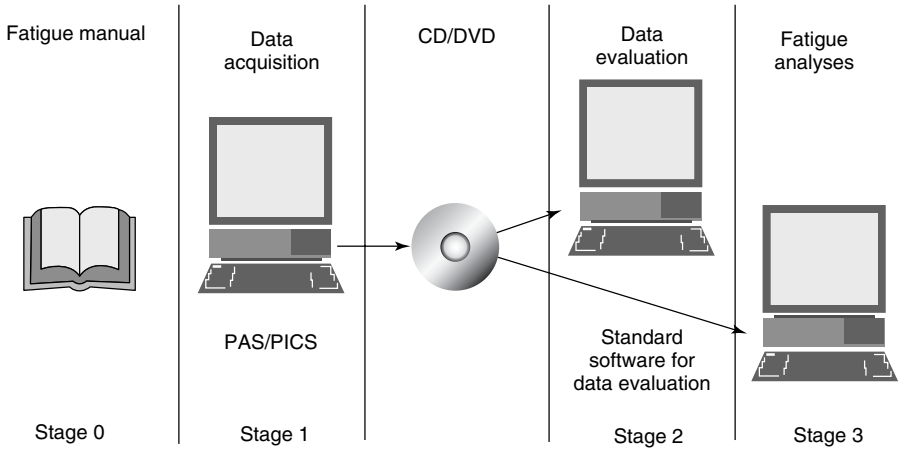


Figure 6. The four stages of the FAMOS system.

be considered for that development. On one side, the design must be very robust because the measurement section sometimes needs to be removed for in-service inspections. In addition, a very quick handling for the mounting activities is necessary since at the measurement locations high radiation exists and hence the dose rates need to be kept low for the personnel. Independent tests showed that the current design of the FAMOS measurement sections is able to reach 95% of the real surface temperature of the outer surface of the pipe [7]. With additional calculations, this deviation can be corrected to get the true outer surface temperature of the pipe.

During operation, the monitored components are exposed to different kinds of temperature loads like thermal shock or thermal stratification. For this reason, two different types of measurement sections with either two or seven thermocouples have been developed (Figure 7). The principal arrangement of a measurement section at the pipe including cover shell and insulation is shown in Figure 8. To get a robust sensor arrangement, it is important to protect the measurement section with a cover shell as the sensitive thermocouples need to be protected against damage when the metal insulation cassettes are installed. The thermocouple wires are passed through the insulation via a fixed protection pipe and then passed further within flexible or fixed protection tubes to the intermediate terminal box.

The sensors used at the measurement section are thermocouples of type K (NiCr–Ni). They are point welded at the measurement section with a thin metal foil. Thermocouples of type K are specified for a temperature up to 1200 °C (2192 °F). During the manufacture, the sensors undergo an accelerated aging process. This has the advantage that no additional aging will occur during application for fatigue monitoring in a nuclear power plant as only temperatures of up to 350 °C (662 °F) occur. This leads to a very robust behavior, which does not require additional recalibration during operation. If the sensors are damaged mechanically, a drift will be noticed and the sensor needs to be exchanged during the next outage.

9.2 Electrical components

For the data acquisition of the measured signal, different alternatives are also available. The whole measurement chain (Figure 9) consists of the following components:

- measurement sections with the thermocouples;
- terminal boxes with the thermal compensation (PT100);
- connecting cables between terminal boxes and the electronic cabinet.

The NiCr–Ni thermocouple wires as well as the corresponding thermal compensation lines are expensive compared to normal copper wires. For this reason, the transition should be made as soon as possible to save costs. Within nuclear power plants,

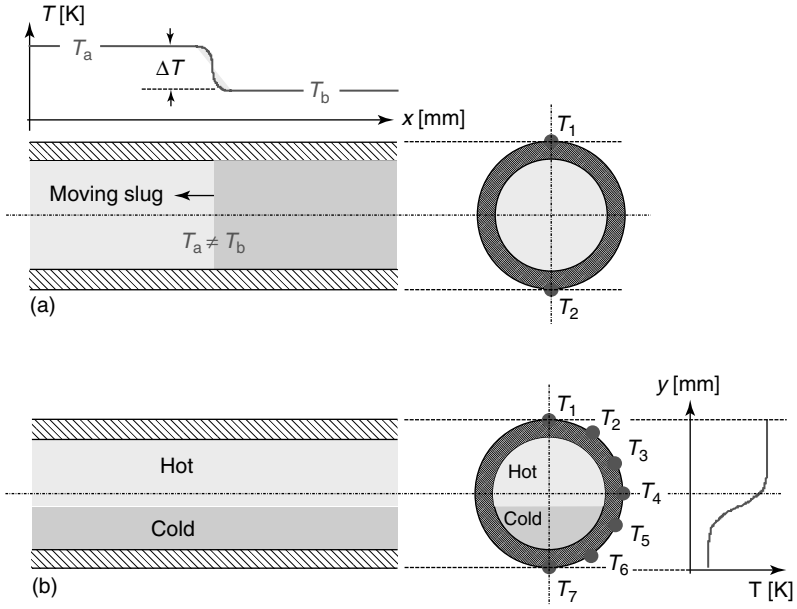


Figure 7. Schemes of idealized thermal loading of a pipeline and location of thermocouples: (a) slug flow (thermal shock) and (b) thermal stratification.

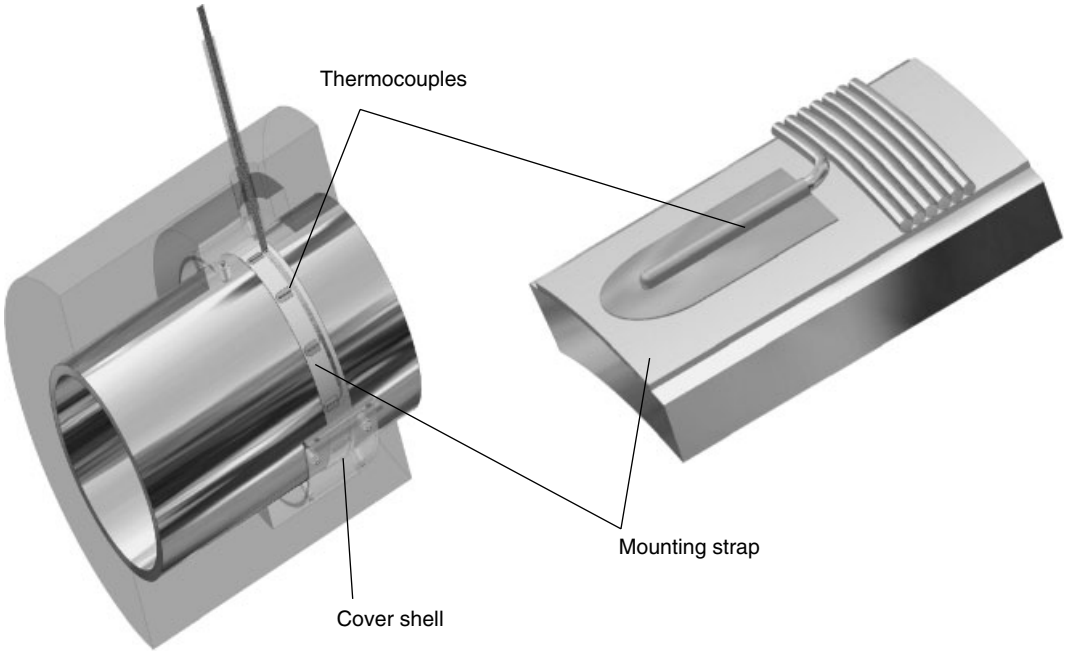


Figure 8. Measurement section with thermocouples, cover shell, and insulation.

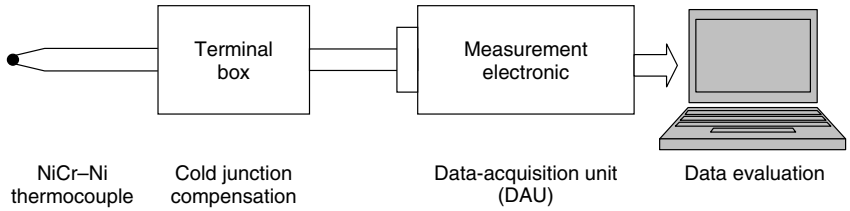


Figure 9. FAMOS measurement chain.

the radiation conditions need to be considered additionally. So it is good practice, nowadays, to have the terminal boxes arranged at a location where the radiation is lower than that directly at the pipeline. So the sojourn time for mounting the measurement sections at the pipelines is short and the dose rates for the personnel is held low (ALARA principle: as low as reasonably achievable). To make transition from the thermal compensation lines to normal copper lines, it is necessary to implement temperature compensation (cold junction compensation). It is not necessary today to have ice water available at 0 °C because the temperature compensation can be realized electronically. For this reason, the terminal of the NiCr–Ni and the copper clamps are held at the same temperature with the aid of a large copper block. The temperature of the copper block is measured with a PT100 resistance thermometer, which gives a high accuracy. This temperature is now used at the measurement electronic system for temperature compensation.

The data-acquisition unit (DAU) can also be realized in many different ways:

- local data loggers for temporary measurements;
- fixed installed cabinets with amplifiers and acquisition PC;
- data acquisition integrated within I&C and the plant process computer system.

Locally arranged data loggers can be used for temporary measurements only. The acquired data are read out after one cycle of operation for data evaluation. For installation of a complete system, either a cabinet with amplifiers and data-acquisition PC is installed or the system can be directly integrated within the I&C and the plant process computer system.

In a stand-alone measurement system with amplifiers and data-acquisition PC, it is possible to implement the most recent measurement electronic system available on the market. The disadvantage is that such systems are quickly out of date after some years of operation. This concerns the measurement electronic system, the PC, as well as the operating system and the data-acquisition software.

A longer lifetime can be expected from the components used in I&C systems. Therefore, the newest generation of integrated fatigue monitoring systems are carried out with that technology (Figure 10). The example is based on the Teleperm XP system from Siemens. The Teleperm XP cabinet is normally located outside the reactor containment in the measurement rooms, but it was also already realized to locate it inside the containment. This has the advantage that only one wall penetration with two optical wires is needed for transferring all signals to the plant process computer. In the other case, many wall penetrations are needed (e.g., 300 wall penetrations for 150 thermocouple signals, additionally approximately 50 or more signals for the PT100 depending on the number of temperature compensations).

It is important to note that data acquisition does not intervene with I&C and the operating system. The existing infrastructure of the I&C system is only used for data acquisition, storage, and visualization of the measured data. So it behaves in a manner similar to a stand-alone system on the basis of data loggers and a data-acquisition system.

10 DATA STORAGE AND HANDLING

The measurement results and archived patterns of temperatures, pressures, and other monitored parameters are stored either in the archive system of the plant process computer or at the stand-alone system on a hard disc and CD-ROM. A raid system ensures quick access and a high level of security for the database. The acquired values are displayed in real time on the

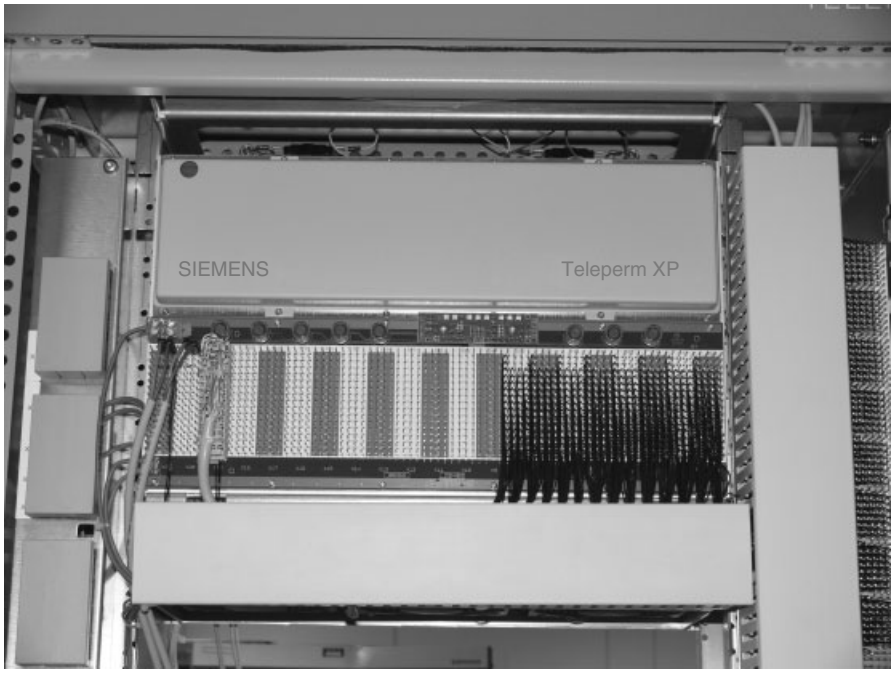


Figure 10. FAMOS integration in Teleperm XP.

plant computer screen or the local PC in the form of graphical representation. Additional printouts of these graphics are possible for direct discussion of special events. Optionally, selected segments of the load history can also be displayed via access to the archive system.

The database includes the measured parameters—patterns of temperature (recalculated and corrected to the temperature of the cold junctions), pressure, and other parameters. The database contains the complete time interval of the measuring period, but for disk capacity reason only such load vectors are stored, which exceed a given bandwidth (e.g., ± 2 K) of the individual measuring point. Standard software can be used for offline evaluation in the next step. However, before the standard procedures for evaluation are applied, it is necessary to check whether the data are correct. This is done in a plausibility check where the data are compared with operational signals and also with the individual signals within the measurement sections. In this step, it is possible to identify strikes that may occur in the case of switching on a large electrical load or if a thermocouple is damaged and causes a drift. If this data-correction step is not made, the results of the following fatigue evaluation would be faulty.

11 EVALUATION METHODOLOGIES

In general, there are three possible methods to develop accompanying fatigue statements during operation. They are listed as follows, sorted by the increasing accuracy of the fatigue damage determination.

11.1 Event-based fatigue (EBF)

In the design phase, the calculated fatigue effects are related to certain presumed events. Such an event, to which was assigned a defined set of fatigue load cycles, may be, for instance, the unit start-up, the unit shut down, power increase or decrease, exchange of water, emergency shut down, etc. The fatigue calculation assigned a fraction of fatigue damage to each of these events. The fatigue fraction was then multiplied by the presumed number of occurrences

of the event during the component's planned service life. The cumulated fatigue was obtained by the summation of all fatigue fractions of all the events considered.

Event-based fatigue (EBF) denotes the comparison of the numbers of occurrences of the events, which actually happened during operation, with those from the design phase that are the presumed ones. The evaluation does not consider the occurred loads. Thus, it is not investigated as to whether the course of events during operation complies with those from the design phase.

11.2 Cycle-based methods

For cycle-based methods, a monitoring system has to be installed, which can record the real temperature history at specific locations.

11.2.1 Cycle-based load counting (CBLC)

Cycle-based load counting (CBLC) allows to count the real number of temperature cycles in different temperature ranges, which occurred during operation. Such a cycle-counting procedure based on the rain-flow methodology is implemented in the FAMOS stage 2 software for evaluation. Additionally, also pressure cycles are counted, which are based on normal operation instrumentation.

This kind of evaluation is based solely on loads (pressures and temperatures). Whether or not the cycles are fatigue relevant can only be evaluated by comparison with allowable number of cycles during the design phase. In practice, a good alternative, which is accepted from German authorities, is the comparison of the measured cycles for the consecutive operating periods. This quickly shows whether or not significant cycles occurred, which need to be analyzed in more detail.

11.2.2 Simplified stress-based fatigue (SSBF)

This method of fatigue evaluation is more accurate than EBF, because it depends on realistic load histories, which occurred during plant operation. Simplified stress-based fatigue (SSBF) also requires a special instrumentation (such as FAMOS), which delivers reliable load histories, close to the observed components. The concept of SSBF can be done on the basis of CBLC from the previous step. Only thermal loads, which are one of the root causes of fatigue damages, are taken into consideration. Pressure transients and member forces are neglected in this step.

The first analysis step is the classification of thermal transients depending on temperature ranges. Usually, the rain-flow method, a cycle-counting algorithm, is used to sort load cycles into classes. After that, the stress intensity ranges for each load class are calculated. The allowable number of occurrences depends on the value of the stress ranges, which are extracted from design fatigue curves. The comparison of existing and allowable number of occurrences delivers a fraction of fatigue damage. The cumulated fatigue is obtained by the summation of all fatigue fractions of all the load classes.

The SSBF estimation has to be seen as a method for a qualitative evaluation of one or more operating cycles and their effects on the fatigue usage factors of the highest loaded components. SSBF has been proved as a method to estimate fatigue damages conservatively. The results serve as a basis for determining whether or not a more detailed fatigue calculation (see Section 11.3) is necessary.

11.3 Stress-based fatigue (SBF)

This level of fatigue evaluation is the most accurate one, which is usually performed at longer time intervals. It requires measured thermal and pressure loads at the observed component. Process engineering specialists analyze the load histories and compile a set of loads, which covers the entire loads in the evaluation period. The assessment criteria of thermal transients are not only the temperature ranges (see rainflow cycle counting method) but also the thermal gradients. Stress histories of the selected components are calculated by using temperature load sets. Corresponding pressure transients and member forces are taken into consideration. The usage factors are calculated by applying the rain-flow algorithm on the histories of stress intensities, by deducing fatigue fractions of stress cycles from design fatigue curves and then by accumulating these partial fatigue usage factors linearly according to Miner's law. All available information is employed including individual material characteristics; data about defects detected

during manufacturing and during operation, witness samples, etc. The results are used to review the entire facility status and to demonstrate the margins for future operation as well.

12 USING FATIGUE MONITORING SYSTEM RESULTS IN THE NPP OPERATION

Experience has shown that operation behavior of a plant could differ from design assumptions and predictions. With the help of the FAMOS instrumentation, it will be possible in the future to check if the design assumptions are met. It may be seen that load transients assumed during design did not occur or behaved differently and also it may be seen that load transients not expected during design have occurred. For this reason, FAMOS will also be used as a tool to obtain data on state and operating behavior of the plant that are as realistic as possible. So, after some years of operation, the real thermal loads will be known, and above all they will be lower and can then be used for an update of fatigue evaluations if deemed necessary.

During the years of operation with FAMOS, the system will also help the plant operators to understand how certain load cases went off and, if possible, to take measures to change unfavorable behaviors. So this information will help the shift personnel to better understand how the plant operates and thus they will be able to prevent the plant from (fatigue) damage. This knowledge will lead to a high availability and long-term operation of the plant.

Besides the purpose of long-term fatigue monitoring of individual components, the results provided by the fatigue monitoring system are also used for the following purposes:

- It enables to choose the strategy for a further safe operation of the component and to set the time intervals of nondestructive material testing. The regular exploitation of the information supplied by the monitoring system provides for enhanced component availability and safety.
- The acquired data enable the mode of operation to be analyzed. For modes of operation prone to high fatigue rates, alternative operation procedures can be proposed, thus reducing the fatigue impact in good time.
- Since the measured real operating data are permanently stored, they are available at any time to follow up on questions of service lifetime, inspection frequencies, existence of loading effects, future upgrades of evaluation procedures, new material properties, etc.

12.1 Alternative fatigue monitoring approaches

This section is a brief outline of some alternative measurement systems, which are being used in different engineering companies and manufacturers, monitoring the components' fatigue status alongside FAMOS.

First, the product WESTEMS [8] shall be roughly explained. The philosophy of this program is an approach of processing plant computer data through system algorithms to predict local component transient loads. In a following necessary step, these loads need to be applied to component models to calculate the respective stresses and fatigue usage factors. This final step has been explained in detail in Section 11.3.

Another product concerning fatigue monitoring in plants is called *Fatigue Pro* [9]. Basically, Fatigue Pro offers two possibilities to handle structural fatigue determination. One is the SBF method (see also Section 11.3), which requires a finite element model for each monitored location. Within Fatigue Pro, Green's function, a mathematical–physical approach transforms a given (measured) thermal field into thermal component stresses. Predictions of stress components at other location without a known thermal field can also be done via special transfer functions.

The automated cycle-counting module is the other option given by Fatigue Pro. More detailed information about this method is given in Section 11.2.1.

13 DIRECT SCANNING OF FATIGUE DAMAGE

Although FAMOS is called a *fatigue monitoring system*, it, nevertheless, is not able to directly monitor the fatigue damage occurred, which was already stated in the first section. The evaluation of a fatigue

usage factor today is solely based on measured pressure and temperature data, which are fatigue relevant. However, now a long-term R&D project has been launched by German research institutes with the aim of being able to detect fatigue damage directly on the components based on non-destructive evaluation (NDE) methods. Basic research results give evidence that the different stages of fatigue damage in power plant steels can be successfully related to NDE data [10]. A new practical approach to the evaluation of fatigue damage based on the fractal dimension of deformation structures is promising [11] in this context. It was found that the fractal dimension of the deformation structure derived from the surface topography increases during fatigue load. Simultaneously, the fractal dimension of the changing magnetic domain structure shows the same behavior. Thus, magnetic noise measurements collect information on small sample regions. A scaling parameter "fractal dimension" is derived and correlated to fatigue life [11]. AREVA aims at the application and verification of this method together with the research institutes.

14 SUMMARY

The AREVA integrated and sustainable concept of fatigue design, monitoring, and reassessment is an expression of the significance of design against fatigue of nuclear power plant components. On the basis of the experience during erection of the German Convoy plants, new NPPs with scheduled operating periods of 60 years, the lifetime extension of old plants, the modification of the code-based approaches, and the improvement of operational availability are driving forces in this process. This is an expression of the sense of responsibility as well as an economic requirement.

The main modules are the design analyses before operation, the advanced fatigue monitoring system, the simplified fatigue assessment, the detailed data processing combined with the specification of occurring thermohydraulic loads, and the code-based fatigue and ratcheting analysis. The fatigue monitoring process should be considered from the very beginning (initial start-up) until the end of life (e.g., 60 years). As all modules are closely connected, it is reasonable to apply the approach as a whole with an additional cost reduction effect compared to separate solutions. Thus, the integrated fatigue approach makes a significant contribution to the operational availability and the protection of investment.

ACKNOWLEDGMENTS

The implementation of the AREVA fatigue monitoring system FAMOS and the integrated fatigue concept depends on the coordination of specialists of various disciplines. The used procedures and methodologies are summarized in this article. The authors wish to express special thanks to the following colleagues for their valuable contributions, helpful discussions, and amendments to the manuscript: K. Degmayr, M. Franz, J. Rudolph, S. Krüger, S. Bergholz, and K. Wirtz.

RELATED ARTICLES

Damage Evolution Phenomena and Models

Fatigue Life Assessment of Structures

REFERENCES

[1] ASME Code: American Society of Mechanical Engineers (A.S.M.E), http://www.asme.org/.

[2] RCC-M Code: **A**ssociation **F**rançaise pour les règles de conception, de construction et de surveillance en exploitation des matériels des **C**haudières **E**lectro-**N**ucléaires (AFCEN).

[3] Nuclear Safety Standards Commission (Kerntechnischer Ausschuss—KTA), http://www.kta-gs.de/.

[4] Kleinöder W, Pöckl C. Developing and Implementation of a fatigue monitoring system for the new European pressurized water reactor EPR. *Proceedings of the International Conference "Nuclear Energy for New Europe 2007"*. Portoroz, 10–13 September, 2007, http://www.nss.si/port2007/abstracts.htm.

[5] Thermal Stratification in Piping Systems Structural Integrity Associates Inc. http://www.structint.com/tekbrefs/sib96159/sib96159.html.

[6] Pöckl C. *Erstellung eines Softwaremoduls zur Identifizierung von Temperaturschichtungen in Verzweigungen von Rohrleitungen*, Thesis Georg-Simon-Ohm-Fachhochschule Nürnberg, 10 September, 2002.

[7] Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Schriftenreihe Reaktorsicherheit und Strahlenschutz: "Einsatz von Thermoelementen zur Erfassung der Temperatur von Rohrleitungswandungen im Rahmen der Ermüdungsüberwachung", BMU—2003-632, http://www.bmu.de/files/pdfs/allgemein/application/pdf/schriftenreihe_rs632.pdf.

[8] Strauch PL, Gray MA. Consideration of fatigue issues in Westinghouse AP1000 design and operation. *Nuclear Plant Fatigue Applications Workshop*. Centennial, CO, 19–21 February, 2007.

[9] Fatigue Pro Structural Integrity Associates Inc. http://www.structint.com/products/fatiguepro/how.html.

[10] Palm S, Gertkemper H, Knoch P, Maile K. A study of the early stages of fatigue damage in power plant steels. *Proceedings of the 31st MPA-Seminar in Conjunction with the Symposium "Materials and Components Behaviour in Energy and Plant Technology"*. Stuttgart, 13–14 October, 2005.

[11] Schreiber J, Kröning M. Evaluation of fatigue damage of steel for nuclear power stations by the fractal features of deformation structures and their noise behaviour (in German language). *Proceedings of the 31st MPA-Seminar in Conjunction with the Symposium "Materials and Components Behaviour in Energy and Plant Technology"*. Stuttgart, 13–14 October, 2005.

FURTHER READING

Assessment and Management of Ageing of Major Nuclear Power Plant Components Important to Safety, http://www-pub.iaea.org/MTCD/publications/PDF/te_1361_web.pdf (accessed July 2003).