INDUSTRIAL CONTROL

In recent years control system technology has evolved at an impressive rate due in part to the growing demand for more reliable, accurate, and compact industrial products. Indeed, this field has experienced a major transition from semimanual to almost fully automatic over the past forty years thanks to major advances made in control systems technologies and control algorithms. Today, industrial control systems are designed to satisfy the highest performance requirement while operating under increasingly difficult conditions.

The importance of control systems in peoples' everyday life has been recognized since the early civilizations. The Greeks made use of float regulators as early as 300 B.C. (1). In relatively more recent history and by the end of the seventeenth century, steam boilers have been controlled by pressure regulator mechanisms invented by Dennis Papin in 1681. Despite the importance of this invention, it is generally agreed that the era of industrial control systems began when James Watt designed the fly-ball governor in 1769. This device was used for solving the problem of speed regulation of the steam engine, a system recognized by many as being the stepping stone for many discoveries that led to the industrial revolution in Europe in the eighteenth and nineteenth centuries. But it was only during and following World War II that the field of control system design was approached in a more systematic and analytical way following the discovery of the feedback amplifier and the frequency based design techniques that ensued. A full-fledged branch of what has since become known as systems theory was developed. It encompasses many areas of mathematics and different engineering disciplines. It is worth noting that since that time, and due to other revolutionary discoveries in the areas of industrial electronics (sensors, actuators, system interfaces) and computer systems, the growth rate of industrial control systems has been astonishing. But what is an industrial control system? What are its main components? And what are the most commonly used control methodologies and control architectures in the industry today?

In very broad terms, a control system is any set of interconnected elements (components, function of units), possibly involving several engineering disciplines, arranged in such a way as to provide a desired output. The output is commonly known as the controlled or manipulated variable. It may represent the rudder angle of an airplane (Fig. 1), the pressure level in a gas pipe, or the voltage of a power line. Some very well known and sometimes least understood control systems exist within the human body. The body temperature regulation system is a case in point. Some well-understood control functions within the human body have inspired control engineers to design highly reliable and efficient control systems. For instance, and in recent years, a growing number of control systems in the field of automation, consumer products, and industrial processes have been designed by approximately emulating the reasoning features of humans that are used for decision-making purposes.

Control systems exist in open or closed loop configurations. An open loop system such as the one shown in Fig. 2 receives its command signals based on a given performance goal and on prior knowledge of the system's dynamics. This input signal is provided independently of the current output of the system. Open loop control systems are usually used for processes that are not subjected to major disturbances and in which the accuracy of the output is not of prime importance. They are usually low cost and simple to construct. They are, however, very sensitive to external disturbances and cannot effectively compensate for them. The toaster is a familiar example of an open loop system. The operating mode of the device is independent of the actual color of the toast (which is in this case the controlled variable). The system behaves in exactly the same manner for toast of different types and of different thickness. This results in different outcomes for toast of different characteristics. While more expensive and more complex to design, closed loop systems, typically similar in structure to Fig. 3, are generally more reliable control systems with much greater performance than their open loop counterparts. They are also better suited for dealing with unexpected disturbances and unknown dynamics. This is mainly due to the feedback property characterizing every closed loop system. Feedback is essentially a corrective signal quantifying the difference between the desired output and the actual output of the controlled system. The signal is transmitted to the controller's input and used in such a way as to drive the actual system output to the desired result. An industrial furnace control system is an example of a closed loop system. Through the feedback signal, the controller provides the necessary commands for a central burner to decrease or increase the temperature according to well defined specifications.

COMPONENTS OF AN INDUSTRIAL CONTROL SYSTEM

A typical industrial control system in closed loop configuration is composed of four main modules: the sensing module, the actuator module, the plant module, and the controller module (Fig. 3). In the last several years major advances have been made in sensor and actuator technologies. These important devices are becoming less bulky, more precise, and are more tailored toward computer controlled systems. Similar advances have also been made in the design of more robust and more efficient control algorithms. As a result, controllers are now better adapted to handling multivariable systems, possibly involving nonlinearities and uncertainties in the dynamics of the process. A discussion of the main modules of a typical industrial control system is provided next.

Sensors

Sensors represent one of the major components of any industrial control system. This stems from their importance in quantifying the controlled output variables of the closed loop structure. The more precise and the more reliable the output measurement(s), the more efficient and the more adequate is the controller output. The focus here is on the family of sensors that provide information in terms of an analog electrical signal (most often a voltage). For digital control purposes, this signal could be discretized through appropriate devices

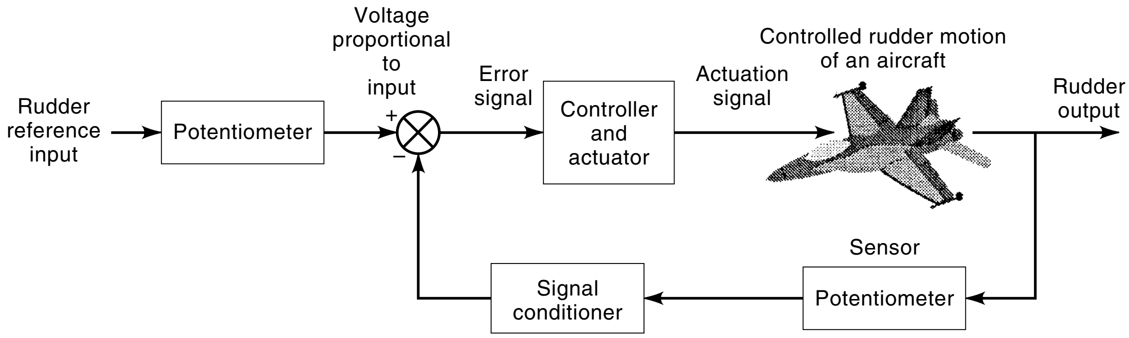

Figure 1. Aircraft rudder control system.

known as analog to digital converters (ADC). Sensors, which are also known as measuring transducers, are often composed of two main elements: the primary sensing component and the signal converter or the transducer. Very often, such elements are used in industry to convert the measured signal which could be a displacement, a temperature, a force, or a flow into an electrical signal, subsequently used by the plant controller to activate the actuator or the final control elements of the control system. Most sensors used in industry today, belong to either one of two categories: motion/force (torque) sensors or process variable sensors. While the first category is used for mechanical/electromechanical systems involving moving parts, the second is used for industrial processes characterized by output variables of the following nature: flow rate, temperature, pressure, or liquid level. A brief description of the most commonly used sensors is provided next. Further details can be found in Refs. 2 and 3.

Sensors for Mechanical/Electromechanical Systems. Motion, force, and tactile sensors provide valuable sources of feedback signals for most dynamic systems characterized by moving parts and usually found in such fields as automotive, robotics, aerospace, or manufacturing, to name a few. This is due to the nature of these systems composed mainly of a number of interconnected electromechanical devices with rotating and displacement elements. The feedback information for this class of systems could be one or more of the following variables: displacement, velocity, acceleration, contact force, and/or torque. Force and torque sensors are used in a multitude of applications including robotic systems, process testing, and diagnosis. They are particularly useful for those systems for which the motion feedback is not sufficient enough to provide high performance control.

Displacement Sensors. Displacement sensors are used to provide signals corresponding to linear or angular motion of the system or part of the system being controlled. The most common sensor used for this purpose is the potentiometer. It is a sort of resistance with sliding contact that moves from one end to the other hence providing a uniform drop of voltage proportional to the displacement (angular or linear). Other types of sensors have also been specifically designed for the measurement of linear or angular motions. The linear variable differential transformer (LVDT) for instance was developed for linear motion measurement, while the rotary variable differential transformer (RVDT) along with synchro and resolver systems were designed for the measurement of angular motion.

Velocity Sensors. Velocity sensors measure the rate of change in displacement (linear or angular). They are very useful in providing feedback for a large family of electromechanical systems. Permanent magnet sensors have been used for both the linear and the angular velocity measurements. They are based on the principle of electromagnetic induction between a permanent magnet and a conducting coil. The well-known dc tachometer represents the angular version of the permanent magnet sensor. This device is commonly used for angular velocity measurement. Other well known devices used for digital velocity measurement are based on optical encoders (2).

Acceleration Sensors. These sensors provide information on the rate of change in velocity (linear or angular). Linear acceleration is measured indirectly by assessing the force needed to accelerate a given mass. This is basically a translation of Newton's law of motion. Angular acceleration measurement is most often obtained through differentiating the signal obtained from an angular velocity sensor.

Force Sensors. A force sensor provides information on the force applied at a given location of the system being controlled. It is based on the characteristics of a piezoelectric material, which generates an electric charge every time it is subjected to a mechanical stress. Other force sensors are based

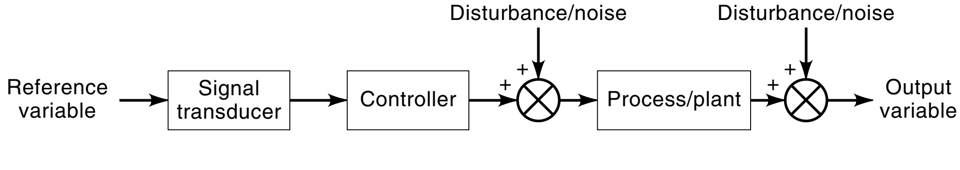

Figure 2. Typical diagram of an open loop system.

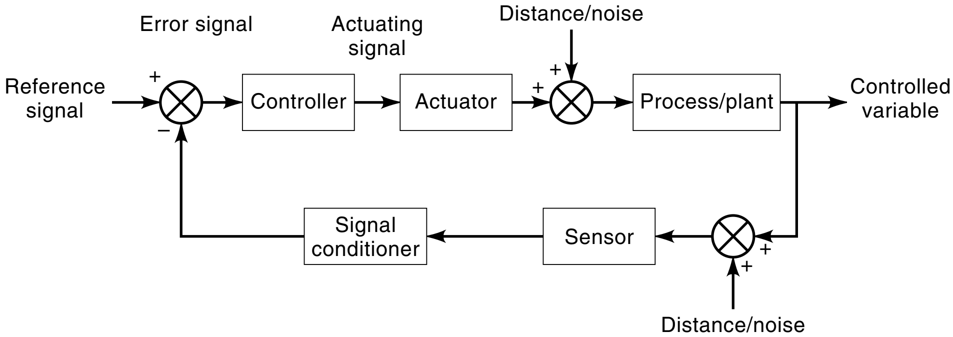

Figure 3. Typical diagram of a closed loop system.

on strain gauges which are small electromechanical devices attached to the object's surface. They convert the strain applied to them into corresponding changes in electrical resistance. This stems from the fact that the resistance of a fine wire varies as the wire is strained.

Torque Sensors. Measuring the strain produced between the drive element and the driven load through a strain gauge bridge is a common way for torque sensing. The same information could as well be obtained through measuring the angular acceleration in a known inertia element of a given rotating device. De Silva (3) describes other techniques for torque measurement.

Sensors for Industrial Processes Variables. Most often, the manipulated variables of a given industrial process (chemical plant, nuclear plant, water treatment plant, . . .) are different in nature from those of the mechanical/electromechanical systems described previously. These so-called process variables could include one or more of the following: temperature, flow rate, pressure, or liquid level. The measurement of such variables requires another family of sensors described with details by Bateson (2), and on which we give a brief overview here. Thermocouples, resistance temperature detectors, thermistors, and radiation pyrometers are among the well-known sensing devices often used in the process industry for temperature measurement. Turbine flow meters, differential pressure flow meters, and vortex shedding flow meters are commonly used for measuring flow rates. For liquid level measurement, designers have used displacement float level sensors, capacitance probe level sensors, and static pressure level sensors among other types of sensors. For pressure measurement, the strain gauge pressure sensors and the deflection type pressure sensors have been among the standard measurement devices used in many industry applications.

Actuators

Actuators represent another important module in any industrial control system. Actuators get their input or driving signals from the controller output. They usually provide the power necessary to drive a given load (in case of mechanical/electromechanical systems) or are used to operate controller components such as switches, control valves and heaters (for process control applications). For the first family of applications, they are identified as process actuators, while for the second family they are known as control actuators (3).

Process Actuators. This family of actuators is frequently used for the purpose of operating directly on the controlled system through electromechanical devices such as stepper motors and servomotors. They are used in a number of industrial applications such as in manufacturing plants, meat processing and grading plants, and gantry cranes to name a few applications. Process actuators include alternate current (ac) motors (synchronous, servo, induction), direct current (dc) motors (permanent magnet, pancake, brushless), and stepper motors (micro-stepping, half stepping, and full stepping). Designers would choose among these actuators depending on the type of application in hand. For instance, ac motors (single phase and polyphase) are mainly used for systems requiring single operating speed, dc motors are used for those applications requiring multiple speed operation, and stepper motors are used for digital control applications including robotic systems, machine tools, and x–y plotters. For a full description of these systems and their main features, one may refer to either one of the two Refs. 2 and 3.

Control Actuators. This type of actuator has the essential feature of activating or deactivating the so-called final control elements of a given control system. Final control elements are the interface between the process and the actuator. Depending on the type of the process, they could be any of the following devices: switches, contactors, mechanical relays, control valves, or heaters. Switches are among the familiar types of final control element used in the process industry to make or break a given electric circuit. They could be of mechanical, electrical, or solid, depending on the type of application and are often used for controlling the process variables in terms of flow level, pressure level, or temperature. Hydraulic and pneumatic actuators represent a major component of control actuators. They provide the necessary power for activating the final control element of a number of processes requiring hydraulic or pneumatic control valves. They are very common in process control systems in which the controlled variable could be the flow rate of a fluid or a compressible gas. It is worth noting here that while pneumatic actuators are used for moving relatively light loads with faster motion, hydraulic actuators are used for a slow and accurate positioning of heavy loads.

Controllers

The controller module represents the central component (or the brain) of any control system. Through well-defined control algorithms or modes, controllers provide the necessary signals for activating the actuator, which in turn delivers the drive signal for operating the plant or the industrial process. Since the advent of feedback, different kinds of controllers have been designed for a variety of applications and have been implemented with large classes of industrial control systems. These controllers range from the simplest type such as proportional (P), integral (I) and/or derivative (D) controllers, to the most sophisticated ones, including the new generation of robust and expert controllers. But despite the emergence in recent years of a number of advanced and sophisticated control algorithms, the conventional types (P, I, and/or D) designed on the basis of linear mathematical models of the systems remain the most widely used in many industrial applications. Their popularity stems from their relatively simple design, lower cost, and satisfactory performance. Besides, they can be easily implemented in digital form hence providing the immense advantages of using computer-controlled systems. This being said, it is expected that within the next few years, conventional controllers will give way to a newer generation of controllers capable of dealing with the ever-increasing complexity of a large number of control processes being developed nowadays. In the next few sections, the basics of control systems analysis and design tools are highlighted and some well-known conventional controllers are described and their features outlined. These include integral control or reset control, derivative or rate control, proportional-plus integral-plus derivative (PID) controllers along with phase lead-lag networks. Techniques for identification and design approaches for more advanced controllers are outlined in subsequent sections.

Analysis Objectives and Methods. One of the main objectives for analyzing feedback control systems is to quantify their degree of stability and study their transient performance and their steady state error in response to given inputs. To achieve this, a representative model of the control system has to be developed using well-established rules of physics. Once a model has been defined (in the time or frequency domain), analysis methods can then be applied. Many advanced time domain techniques have already been developed to analyze a wide range of dynamic systems for which the models are represented in terms of ordinary differential equations. But these have been proven difficult to tackle by design engineers, particularly in the case where the system is of higher order (more than 2) or is nonlinear. Other graphical analysis methods based on the frequency domain representation of the system have been developed since the late thirties and have been proven to be very successful for tackling a wide range of linear systems (or linearized systems around the operating points). These include the root-locus method, the Bode plot representation, Nyquist diagrams, and Nichols charts. Detailed information on these analysis techniques along with worked examples can be found in Ref. 4.

Control Design Objectives and Techniques. The main objective for designing a control system is to alter the dynamics of the system's plant as to achieve some preimposed design performance specifications. These are in essence design objectives required by the system user to satisfy some system constraints. The performance specifications can be expressed in either the time or frequency domain. These concepts are developed next.

Time Domain Specifications. To illustrate the concepts involved here, we use a feedback control system for which the overall transfer function (that is, of the closed loop system) is expressed in the usual standard form:

$$T(s) = \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)}$$

where ω_n is the undamped natural frequency of the system and ζ is its damping ratio. $s^2 + 2\zeta\omega_n s + \omega_n^2$ is known as the characteristic equation of the closed loop system and its roots $s_{1,2}$ are given in terms of f the natural frequency ω_n and the damping coefficient ζ of the system as:

$$s_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

It is worth noting here that for the case where $\zeta > 1$, the system has real roots and is said to be overdamped. When $\zeta = 1$, the system has repeated roots and is said to be critically damped. When $\zeta < 1$, the system has two complex conjugates roots and is said to be underdamped. Table 1 summarizes the behavior of the system as function of ζ, and Fig. 4 shows the response of the system when excited by a unit step input for different values of the parameter ζ. In the case where the control system has more than two poles, the exponent corresponding to the root closest to the origin has the largest time constant and takes longest to decay. Moreover, if the residue of such pole(s) is comparable in absolute value to

Table 1. Dependence of the System Behavior on the Value of ζ for a Second-Order System

Value of ζ	Corresponding Roots	System Behavior
$\zeta > 1$	$s_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$	Overdamped
$\zeta = 1$	$s_{1,2} = -\zeta\omega_n$	Critically damped
$\zeta < 1$	$s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$	Underdamped

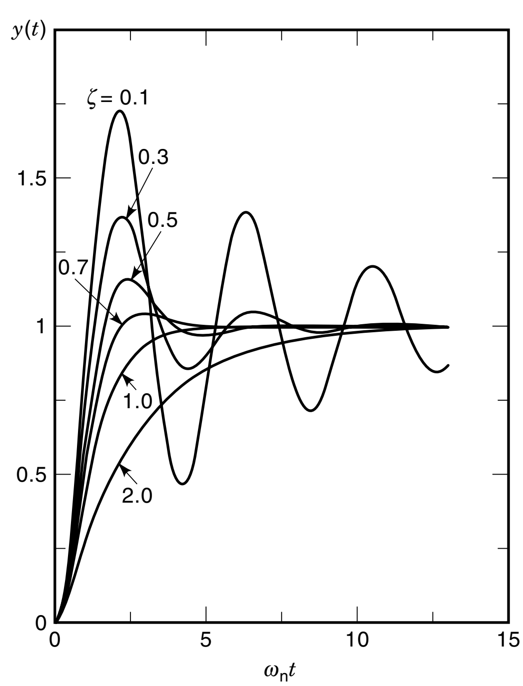

Figure 4. Transient response of a typical second-order system to a unity step input.

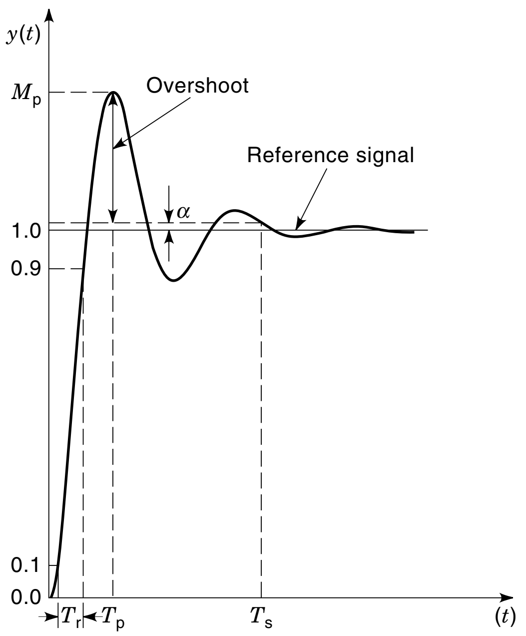

Figure 5. Transient response of a second-order system for varying values of ζ.

all the other residues of the remaining poles, the corresponding pole(s) becomes then dominant(s) and may by itself (themselves) control the behavior of the overall system. Let us again presume that the dominant behavior of the overall system is controlled by a pair of poles given by $s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$, for which we assume here that $0 < \zeta < 1$. The most commonly used time domain specifications for such type of systems are those related to the steady state behavior which itself is controlled by the error constants of the system, and the transient perfornance usually represented by a number of indices. For a given control system subjected to unit step input, these indices are illustrated in Fig. 5. In the figure, the rise time T_r represents the time taken by the system to reach (90%) of the new set point. The settling time T_s is the time taken by the system to remain within a specified region of width 2α of the reference step input (α could vary between 3% to 5% of the reference signal). The overshoot M_p is the maximum value reached by the system relative to the reference value of input. It is usually expressed in terms of percentage value of the final value of the response output. The peak time T_p is the time taken by the system to attain the overshoot value.

Frequency Response Specifications. In some applications, it may occur that the system performance specifications are expressed in terms of frequency domain terms. These are very useful when dealing with the analysis and design aspects using frequency response-based techniques. Gain margin (GM) is an indicator of the relative stability of a given control system. Suppose the overall open loop transfer function of a control system (product of all transfer function in the loop) is given by $G(s)$, the gain margin is then expressed as the inverse of the magnitude of $G(s)$ at the frequency ω_{180° corresponding to a phase angle of $G(s)$ of $-180°$:

$$GM = \frac{1}{|G(j\omega_{180^\circ})|}$$

Another measure of the relative stability of the system is provided by the phase margin (PM) index. It is basically the angle 180° compounded with the phase angle corresponding to the frequency where the open loop transfer function $G(s)$ has the unit magnitude value, also known as the gain crossover frequency. In other words, the phase margin could be expressed as:

$$PM = 180 + \arg G(j\omega_{|G|=1})$$

Two other important frequency domain specifications for a feedback control system are the resonant peak and the bandwidth. The resonant peak T_r is defined as the maximum value attained by the magnitude of the closed loop transfer function of the system $|T(j\omega)|$:

$$T_r = \max |T(j\omega)|$$

As for the bandwidth (BW) of the control system, it provides in essence the range of the frequencies over which the system responds satisfactorily. A satisfactory performance of a control system depends largely on the problem at hand and the type of application. Generally speaking, the bandwidth of a system, for which the closed loop transfer function is given by $T(s)$, is defined as the frequency at which the magnitude $|T(j\omega)|$ drops to 70.7% (or 3 dB down from) below its zero frequency value. Detailed descriptions of these frequency domain specifications can be found in Ref. 4.

Design Techniques for Industrial Control Systems. These techniques use the same tools developed for the analysis methods mentioned earlier. These include most importantly, root locus methods, Bode plot representation, Nyquist diagrams, and Nichols charts. Again the reader may wish to consult the well-detailed description of these design techniques in Ref. 4.

Integral Control. The integral control mode (I) provides a signal at the controller output, u, proportional to the integral of the error signal. But while this mode provides a better steady state error, it comes nevertheless at the expenses of a degraded transient response. This is the main reason why integral controls are most often used in conjunction with proportional control modes (P). This combination is meant to provide more flexibility for the designer to improve the steady state error e while maintaining an acceptable performance of the transient response. This is done through adjusting the controller's proportional and integral gains K_p and K_I.

$$u = K_p e + K_i \int e$$

It is worth noting here that the integral component provides a reset action to compensate for the offset created by the proportional mode.

Rate Control. Rate control or derivative feedback control provides a signal proportional to the change in the error signal rate at the controller output. While the integral mode integrates past values of the error, the derivative control mode, amplified by the derivative gain K_d, anticipates the error behavior, hence providing a correction signal to reduce the predicted error:

$$u = K_d \dot{e}$$

In practice, pure rate control is never used due to the negative effect of noise amplification and to the effect of the controller's zero output in the case of constant error signal. For this, proportional, or proportional and integral terms are always needed in conjunction with the derivative mode to drive the error signal to zero and to satisfy other performance requirement.

Proportional Integral Derivative Control. PID type controllers represent the most widely used dynamic compensators in the industry today. This applies to most industrial systems regardless of whether the control algorithms involve hierarchy, decentralization, and/or distribution in their overall architecture. PID controllers are always implemented at lower levels of the control structure. As such, they interact directly with the actuators and the final control elements of the plant/process. Depending on the type of application and the design specification requirement, PID controllers come in different configurations and different setups (pneumatic, electronic, and digital control). The output u of the PID controller is expressed in terms of the error e between the reference signal and the controlled variable as:

$$u = K_\mathrm{p} e + K_\mathrm{i} \int e + K_\mathrm{d} \dot{e}$$

where K_p, K_i, and K_d are the proportional, integral, and derivative gains of the controller, respectively. As can be seen from the expression of the output, the PID controller combines the three elementary control modes: proportional, integral, and derivative. This is used in cases where two or more control modes are not sufficient to maintain the controlled variable close to the reference signal. This often applies to processes involving large load changes. As such, the integral action, I, provides a reset action limiting the offset caused by the proportional mode, P, while the derivative mode, D, produces an anticipatory signal reducing the possible large error caused by sudden load change and limiting possible oscillatory behavior. For a second order system, this translates into creating a controller having the best properties of both the PD and the PI control modes. In fact, while the PD controller adds damping to a system and reduces the overshoot, the PI control mode improves the steady state error and contributes to damping as well.

Reset Windup. Reset or integral windup occurs in virtually every dynamic system involving integral control. This is very common for systems subjected to large change of input or large disturbances. In these circumstances, the integrator accumulates a large output, which may be beyond the activation limits of the actuator, causing it to saturate rapidly and may remain so for a long period of time, hence preventing an adequate control action. As soon as an actuator is saturated, integral action of the controller must quickly come to an end, otherwise the actuator will not provide the appropriate subsequent signal in time, hence causing large overshoot that may even lead to instability. For dealing with integral windup, designers have developed mechanisms in the feedback loop to disable the integral action of the controller as soon as the actuator saturates. This could be done for a PI system by creating a feedback loop around the integral mode component (I) of the controller. This loop is composed of an antiwindup circuit (dead zone nonlinearity with an appropriate slope). When the actuator saturates, the feedback loop around the integrator causes the input of the integrator to go automatically to zero. The integrator becomes a fast first-order lag type. Franklin (4) provides detailed descriptions of two common antiwindup circuits used for dealing with integral windup.

Phase Lead–Lag Networks. Like the PID controllers, phase lead–lag networks provide a combined effect for the proportional, derivative, and integral control modes. This is used whenever neither the lead nor the lag networks separately provides a satisfactory performance. Indeed, it may occur in some applications where the design requirement is such that it improves the steady state response (lag network effect) while providing a faster response of the system (lead network effect). This obviously cannot be achieved through a single network and instead a combination of both networks is required. This is very similar to the PID action, but here the designer has to deal with the adjustment of four controller parameters (instead of three for the PID) as can be seen from the phase lead–lag network expression $G_\mathrm{c}(s)$:

$$G_\mathrm{c}(s) = \left(\frac{1 + a\tau_1 s}{1 + \tau_1 s}\right)\left(\frac{1 + b\tau_2 s}{1 + \tau_2 s}\right)$$

where $a > 1$, $0 < b < 1$, $\tau_1 > 0$, $\tau_2 > 0$ represent the parameters of the phase lead–lag network. The first component in this combined network expression represents the phase lead contribution while the second one represents the phase lag contribution. In most applications, the phase lead portion of the network is used to provide a short rise time while the phase lag portion provides improvement for the damping behavior.

Ziegler Nichols Tuning. Among the many factors that have largely contributed to the success and the wide popularity of PID controllers in large sectors of the industry is that the parameter tuning task depends less on the knowledge of the process dynamics itself and more on experience. This experience can be acquired in a relatively short time by a technician or an engineer who is dealing regularly with the process. No prior knowledge of analytical control design rules are hence required of the process operator(s). There are, of course, well-established analytical tools to handle the tuning problem of PID. But this cannot be achieved without a complete knowledge of the system dynamics. This, unfortunately, is not always the case given the complexities of the dynamics involved in most industrial processes.

Different techniques have been proposed over the years to tackle the tuning of PID controllers without having prior knowledge of the system's dynamics. The most popular and widely used approaches are the ones developed by J. G. Ziegler and N. B. Nichols (1942, 1943). Their main assumption is that most process control systems are of type zero with no dominating complex poles and their step response tends to behave in S-shaped curves (Fig. 6). The transfer function of such systems can be expressed as:

$$G(s) = \frac{Ke^{-Ls}}{Ts + 1}$$

where K is the dc gain, T the time constant, and L the time delay of the process. The first procedure suggests that for a PID controller with a transfer function $G_\mathrm{c}(s) = K_\mathrm{p} + K_\mathrm{i}/s +$

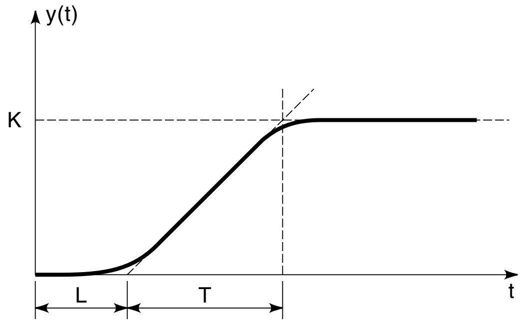

Figure 6. The S-shaped curve illustrating the step response profile of a typical process system with time delay L and time constant T.

K_ds designed for a 25% overshoot, the values of the gains are as follows:

$$\begin{array}{lll} \text{P}: & K_p = T/L & \\ \text{PI}: & K_p = 0.9T/L & K_i = 0.27(T/L^2) \\ \text{PID}: & K_p = 1.2T/L & K_i = 0.6(T/L^2) \quad K_d = 0.6T \end{array}$$

As can be seen, this method is based on the experimental assessment of the unit step response of the system and the parameters of the response.

The second procedure also known as the ultimate cycle method (UCM), proceeds as follows. With only the proportional gain present in the forward path of a simple feedback loop involving the plant, the gain K_p is increased until it reaches a critical limit K_u that would provide at the output a sustained series of oscillations with period T_u (the system reaches the critical or marginal stability stage). The suggested settings for the PID parameters are given as follows:

$$\begin{array}{lll} \text{P}: & K_p = 0.5K_u & \\ \text{PI}: & K_p = 0.45K_u & K_i = 0.54(K_u/T_u) \\ \text{PID}: & K_p = 0.6K_u & K_i = 1.2(K_u/T_u) \quad K_d = 4.8(K_u/T_u) \end{array}$$

Model Identification and Advanced Control Techniques

Conventional control techniques (frequency domain based or state space based) are usually implemented under the assumption of a good understanding of the process dynamics and its operating environment. These techniques fail, however, to provide satisfactory results when applied to poorly modeled processes, and/or processes functioning in an ill-defined environment. Even when a suitable analytical model is available, model parameters might be incompletely known. This is the case when dealing with complex systems for which the physical processes are not fully understood, hence preventing the derivation of an adequate model for the system.

System identification techniques, based on experimentally determined input-output data, are known to provide powerful alternatives for dealing with modeling difficulties and can be used as valuable tools for adequate description of the system dynamics. This can be very useful for purposes of control system adaptation. In this respect, adaptive control techniques can be regarded among the powerful and effective control schemes that make the best use of the system identification tools. At first, a model is derived experimentally through system identification. The identified model is then used as a basis for designing a controller with adjustable gains, hence capable of sustaining large changes in the system parameters, while providing desired responses. Adaptive control schemes are generally intended for plants with partially unknown dynamics and/or slowly time-varying parameters. Other advanced control techniques that are not discussed here include: robust control (such as H-infinity, variable structure control), feedback linearization, optimal control and soft computing based tools (involving fuzzy logic, neural networks, genetic algorithms).

System Identification

In broad terms, a system can be defined as a set of interconnected objects, in which a variety of variables interact to produce observable outputs (Fig. 7). A system is also termed a dynamic system when the rates of changes of the system-response variables are not negligible. In this article, we are primarily concerned with such systems. A system may be subjected to external signals, which can be classified into two categories: those that are manipulated by the observer or environment (known as inputs) and those that are not (known as disturbances).

To explain the behavior of a given system or to synthesize solutions for potential problems related to it, it is necessary to provide a working model that enables the scientist or the designer to analyze or even react to system outputs through adequate control laws. System models can be either analytical or experimental. The analytical models are based on the application of well known, basic physical laws, continuity relations and compatibility conditions (4). Experimental models, on the other hand, are derived from the application of system identification tools (5).

In some cases, particularly when the system structure is so complex that it prevents derivation of a straightforward and tractable model from analytical laws, experimental modeling techniques are sought to learn about the dynamic behavior of the system. This can also serve as a platform for designing adaptive controllers that are capable of providing the desired behavior, even under structural parametric variations of the system. This area of control is known as adaptive control (7–8). Experimental modeling aspects and adaptive control procedures are discussed next.

The main goal of system identification is to experimentally establish an adequate dynamic model of the system such that the parameters which provide the best fit between the measured input-output data and the model-generated values are estimated. Model structure as well as the model parameters may be identified. The two well-known techniques for system identification are the nonparametric and the parametric estimation methods.

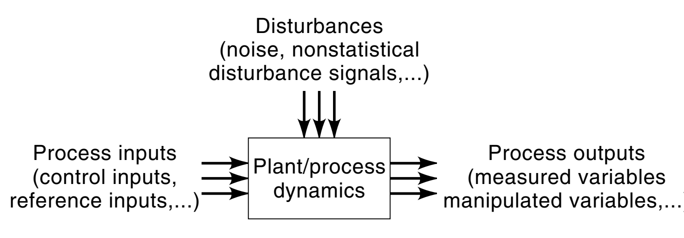

Figure 7. Block diagram of a dynamic system subject to input/output signals and to disturbances.

Nonparametric Methods. Nonparametric system identification is often used to derive an input–output description of a system model, for which the dynamics are a priori known to be linear and time invariant. The method is called nonparametric since it does not involve a parameter vector for use in the search for the best description of system dynamics. A linear-time-invariant model can be described either by its impulse response (time domain method) or by its frequency response (frequency domain method).

The nonparametric, time-domain method is useful if the designer has access to the step or impulse response of the system, known also as the transient response data. These data are usually quick and relatively easy to obtain. The main drawback of the method is that the transient response should be highly noticeable, so as to have a high level of signal-to-noise ratio. This is difficult to achieve in real world applications, and further special tests may have to be conducted as part of the overall identification process. The other aspect of nonparametric system identification pertains to the frequency-domain method. This approach is based on the collection of frequency response data when the system is excited by input signals of different frequencies. The method has the advantage of being simple in terms of data collection. However, it requires more time to apply than the time domain method. This is particularly true for systems with long time constants, which is the case, for example, in chemical processes.

Parametric Methods. Another approach for system identification pertains to the parametric estimation method. This method is particularly effective in providing on-line information on process parameters, which is very useful in designing controllers that can adapt to process-dynamic changes while providing a desired response. The very well known technique of least-squares parameter estimation represents the basic approach here. It is based on the least squares principle, which states that the unknown parameters of a mathematical model should be chosen to minimize the sum of the squares of the differences between the actually observed and the computed values. Let us presume that the estimated output $\hat{y}(T)$ of a given system at a given discrete time T, is given as a linear function of n unknown parameters a_i, each multiplied by a known coefficient γ_i, called regressor variable, which is provided through measurements and depends on input-output data; that is,

$$\hat{y}(T) = \sum_{i=1}^{n} \gamma_i(T) a_i$$

Suppose that several pairs of data $(y(t_k), \gamma(t_k))$ are collected through a series of experiments, with $\gamma(t_k)$ being the regressor vector given by

$$\gamma(t_k) = [\gamma_1(t_k), \gamma_2(t_k), \ldots, \gamma_n(t_k)]^{\mathrm{T}}$$

and $k = 1, 2, \ldots, m$ with $t_m = T$. Let us also denote by $\underline{a}$ as the unknown vector given by

$$\underline{a} = [a_1, a_2, \ldots, a_n]^{\mathrm{T}}$$

and $\underline{y}$ as being the measurement vector given by

$$\underline{y} = [y(t_1), y(t_2), \ldots, y(T)]^{\mathrm{T}}$$

It is now required to determine the parameters a_i that minimize the error δ given by

$$\delta = (1/2) \sum_{k=1}^{k=m} (y(t_k) - \hat{y}(t_k))^2$$

This is a standard optimization problem, for which there exists a unique solution given by

$$a = [\underline{a}_1, a_2, \ldots, a_n]^{\mathrm{T}} = (\Gamma^{\mathrm{T}}\Gamma)^{-1}\Gamma^{\mathrm{T}} y$$

provided that the matrix inverse $(\Gamma^{\mathrm{T}}\Gamma)^{-1}$ exists, with Γ defined as:

$$\Gamma = [\gamma(t_1), \gamma(t_2), \ldots, \gamma(T)]^{\mathrm{T}}$$

For more details on these and some other identification techniques, the reader may consult Ref. 6, where the theory is extensively covered and Refs. 7–8, where some applications are described.

Adaptive Control

An adaptive control system uses a control scheme that is capable of modifying its behavior in response to changes in process dynamics. Adaptive controllers have been extensively used in several industries including the chemical, aerospace, automotive, and pulp and paper industries. The rapid growth in the design of integrated and powerful information processors has made the use of adaptive controllers even more versatile. There are three well-known adaptive control schemes: gain scheduling, model-referenced adaptive control, and self-tuning regulators. A description of the main features of these techniques follows.

Gain Scheduling. This type of adaptive control system is based on the adjustment of controller parameters in response to the operating conditions of a process (7). This control scheme is particularly useful when the variations of the process dynamics are predictable. In fact, for a class of dynamic systems, it is possible that an explicit model of the system can be accurately described every time the operating conditions of the system take new values. Gain scheduling can be regarded as a mapping from the process parameters to the controller parameters. In practice, a gain scheduler can be implemented as a look-up table. A block diagram of this adaptive scheme is shown in Fig. 8. The two main drawbacks of this method

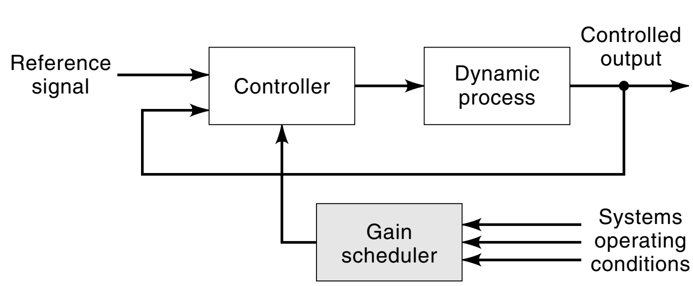

Figure 8. Block diagram of the gain scheduling based controller.

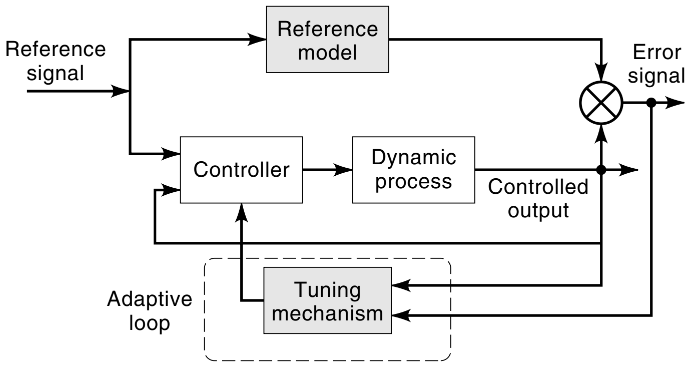

Figure 9. Block diagram of the model-referenced adaptive controller.

are related to its open loop behavior and to the discrete assignment of controller gains according to look-up table data. Indeed, for intermediate operating conditions, no explicit control gains are assigned to the system, and the designer must apply interpolation techniques to avoid instabilities.

Model-Referenced Adaptive Control. The model-referenced adaptive control (MRAC) is an adaptive control scheme capable of handling processes with unpredictable changes (7). This control procedure is based on the design of an adaptive scheme whose objective is to drive the error signal between the response of the process and that of the reference model to zero. The overall scheme, shown in Fig. 9, consists of two main loops with different time constants. The inner loop, which is the faster one, is used for the regulation of the process, while the outer loop is designed for adjustment of the parameters of the inner loop regulator in order to drive the error signal to zero. Algorithms for designing the adaptation scheme for the adjustment mechanism of the outer loop are discussed in Refs. 7 and 8. Some instability problems for applying the MRAC procedure have been observed, and remedies have been proposed in Ref. 7.

Self-Tuning Regulators. Self-tuning regulation (STR) is another adaptive control scheme characterized by its ability to handle dynamic processes that may be subjected to unpredictable changes in system parameters (7). A self-tuning regulator uses the outputs of a recursive identification scheme for plant parameters (outer loop) to adjust, through a suitable adaptation algorithm, the parameters of a controller located in the regulation loop of the system (inner loop), as shown in Fig. 10. One can easily notice some similarity in terms of the inner and the outer loop structuring for STR and MRAC. However, the main difference between the two schemes is that, while the STR design is based on an explicit separation between identification and control, the MRAC design uses a direct update of controller parameters to achieve asymptotic decay of the error signal to zero. In view of this fundamental difference in design, MRAC is referred to as a direct adaptive control scheme, while STR is known as an indirect adaptive control scheme.

INDUSTRIAL CONTROL SYSTEMS ARCHITECTURES

Industrial Large Scale Systems

A few decades ago, it became evident that standard implementation of the conventional control techniques and the standard modeling procedures were no longer capable of providing satisfactory results for the new class of industrial process systems that had been created in the chemical, petroleum, nuclear, and mining industries, to name a few. To satisfy the ever increasing demand of society manufactured goods and consumer products, large scale industrial systems were constructed. They often involved hundreds and even thousands of interconnected subsystems of different nature and sizes operating simultaneously for mass production and fast delivery of high quality finished products. The management and control of these industrial behemoths were made possible though major advances achieved in the areas of sensor and actuator technologies along with major developments made in terms of new control systems architectures and computer controlled systems. Indeed, the extraordinary advances witnessed in the computer industry in the last few years, both in software and hardware, have allowed the design of sophisticated computer controlled systems capable of handling very complex calculations while interfacing with a very large number of devices in a relatively short time.

Industrial Control Architectures

The design aspects of control systems for such large-scale systems have gone through different stages through the years using different architectures. Computer systems have proven very helpful in developing relatively powerful architectures in the 1950s and 1960s. Among the well-known architectures that have been developed over the years to tackle complex design problems were the hybrid architecture, the centralized architecture, and more recently the hierarchical distributed architecture. A description of these architectures and their main features follows.

The Hybrid Control Architecture. The first major control architecture developed as a result of improvement in computational facilities is the hybrid architecture. This architecture combines different control subsystems using different control methodologies through appropriate interfacing with a supervisory computer. In this architecture, the low level or local control of several subsystems is implemented through a set of sequential and discrete analog controllers. A supervisory controller receiving and sending appropriate information through interfacing hardware and data acquisition systems

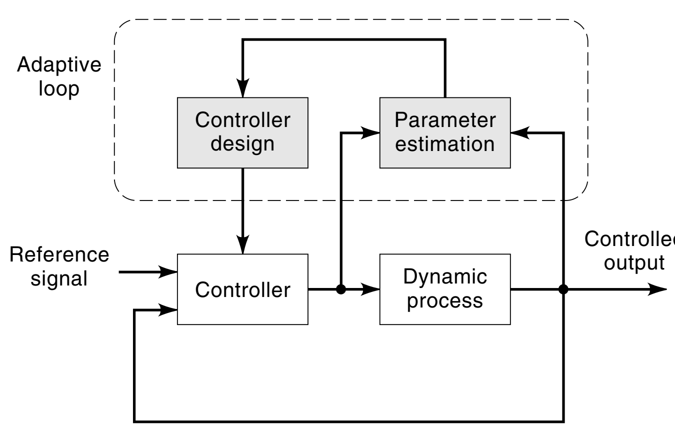

Figure 10. Block diagram of the self tuning regulator.

provides the necessary signals for overall plant management and data storage and retrieval. While this architecture has been successful in a number of applications, it suffers nevertheless from its reliance on a sometimes very large number of subsystems generating different type of signals (hybrid) and hence providing several difficulties in terms of reliable and efficient interfacing with the supervisory controller or operators. Due to its structure and the exclusive presence of discrete analog and sequential devices, this architecture cannot take advantage of the powerful computer controlled algorithms and their relatively easy implementation.

The Centralized Control Architecture. The main property of a centralized control structure is that a high performance computer located in a central location of the plant provides all functions necessary for the overall operation of the plant. Control algorithms related to the operation of the different subsystems within the plant, database updating, signal input scanning, and man/machine interfacing, are all made at the level of the main computer. The input-output signals to and from the different parts of the plants are transmitted through a large network of wiring that may exceed a few kilometers in length in certain applications. To overcome the problem of a main computer failure and to provide continuous output of the plant, a secondary backup computer is used whenever there is a malfunction of the central computer. Clearly this structure is not the best of architectures and was only widely used among power generation systems. The main problem is its sole reliance on a single computing device. Indeed, the complex management of large scale systems such industrial processes demand has not proven successful in a variety of cases involving chemical and petroleum industries using the centralized architecture due mainly to the inflexible character of such structures. This stems from the inability of the centralized architecture in coping well with module expansion and algorithms modification without having recourse to major upgrading, causing higher costs and less efficiency. This is discussed in more detail in Ref. 9.

Hierarchical Distributed Architecture. This architecture was developed as an alternative to previous architectures which were proven to be either expensive to implement and maintain, or unreliable and inflexible when it came to expansion of the plant or some of its modules. Furthermore, the two architectures described so far were not able to profit from the major features offered by the newly developed computer controlled systems and their potential for taking advantages of the new advances made in the field of communications. Hierarchical distributed architectures, on the other hand, make full use of state-of-the-art sensors and actuators and the new generation of computer-controlled systems. They have also a multi-level structure that is composed of one or more hierarchical layers depending on the complexity of the plants and the design requirements. A web of communication networks provides reliable and real time interfacing among all layers of the hierarchy as seen in Fig. 11. By design, this structure offers more autonomy to low level controllers and permits a high degree of supervision at upper levels of the structure. Low level and high level layers interface only through the shared communication networks. At the low level of the hierarchy, a large set of distributed devices operates in almost full autonomy and communicates directly with the machinery and the plant's processes. This layer of distributed systems makes full use of digital computational devices such as embedded microcontrollers. These devices can have their parameters and operational set points easily modified through low level human and computer interfaces. They can also communicate through shared communication facilities with the supervisory level of the architecture for purposes of achieving a global performance requirement. For more details on this architecture and others developed in recent years, the reader may refer to Refs. 9–10.

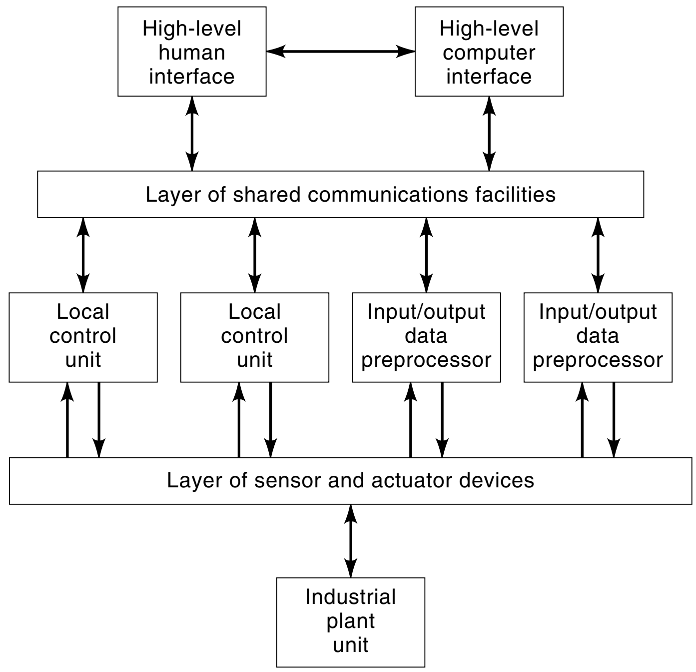

Figure 11. Typical hierarchical distributed architecture in an industrial plant.

SUMMARY

As the need for building systems characterized by high level of component integration and quality precision control continues, industrial plant designers are confronted with new and challenging modeling and control problems. Indeed, the growing trend for the design of sophisticated and highly reliable industrial processes has led to the emergence of new large scale systems that standard control techniques are no longer able to handle. This is mainly due to an increase in the complexities of the dynamics of the systems being controlled and the wide range of devices of different types today's industrial plant requires. Such complex systems are usually characterized by highly nonlinear and time varying behavior, extensive dynamic coupling, hierarchies, multiple time scales, and high dimensional decision space. Furthermore, more and more of these systems are operating in ill-defined environments and are being subjected to large parameter variations and external disturbances. Fortunately, the field of control systems continues to advance and grow at an impressive pace and with it a new generation of associated technologies including computer systems, sensors, actuators, and communication devices is emerging. At this rate, and despite the encountered design problems dictated by increases in system complexities, one should expect in the near future an overall improvement in terms of industrial plant safety, efficiency, and quality products.

BIBLIOGRAPHY

1. O. Mayr, *The Origins of Feedback Control,* Cambridge, MA: MIT Press, 1970.
2. R. Bateson, *Control Systems Technology,* Don Mills, Ontario, 4th ed., Canada: Maxwell Macmillan Canada, Inc., 1993.
3. C. De Silva, *Control Sensors and Actuators,* Englewood Cliffs, NJ: Prentice-Hall, 1989.
4. G. Franklin, D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems,* 3rd ed., Reading, MA: Addison-Wesley, 1995.
5. J. L. Shearer, A. T. Murphy, and H. H. Richardson, *Introduction to System Dynamics,* Reading, MA: Addison-Wesley, 1971.
6. L. Ljung, *System Identification-Theory for the User,* Englewood Cliffs, NJ: Prentice-Hall, 1988.
7. K. Astrom and B. Wittenmark, *Adaptive Control,* 2nd ed., Reading, MA: Addison-Wesley, 1995.
8. S. Sastry and M. Bodson, *Adaptive Control: Stability Convergence and Robustness,* Englewood Cliffs, NJ: Prentice-Hall, 1989.
9. M. Lucas, *Distributed Control Systems,* New York: Van Nostrand Reinhold Company, 1986.
10. M. Jamshidi, *Large Scale Systems: Modeling, Control, and Fuzzy Logic,* Upper Saddle River, NJ: Prentice-Hall, 1997.

FAKHREDDINE O. KARRAY
University of Waterloo

INDUSTRIAL CONTROL ARCHITECTURES. See INDUSTRIAL CONTROL.